The debiased Keyl's algorithm: a new unbiased estimator for full state tomography

Angelos Pelecanos* Jack Spilecki* John Wright*

Abstract

In the problem of quantum state tomography, one is given n copies of an unknown rank-r mixed state $\rho \in \mathbb{C}^{d\times d}$ and asked to produce an estimator $\widehat{\boldsymbol{\rho}}$ of ρ which is close to it. One of the most basic and useful properties a statistical estimator can have is that of being *unbiased*, meaning that $\widehat{\boldsymbol{\rho}}$ is equal to the true state ρ in expectation. Unfortunately, of the three estimators for tomography which are known to be either sample-optimal or almost sample-optimal, namely Keyl's algorithm [Key06] and the two tomography algorithms from [HHJ+16], none of them produce unbiased estimators. This situation has recently been highlighted as a bottleneck in several important tomographic applications [CLL24b, CLL24a].

In this work, we present the *debiased Keyl's algorithm*, the first estimator for full state tomography which is both unbiased and sample-optimal. Our algorithm is the result of taking Keyl's algorithm and carefully modifying it in order to correct for its bias. In addition, we derive an explicit formula for the second moment of our estimator which is simple and easy to use. Using it, we show the following applications.

- We show that $n = O(rd/\varepsilon^2)$ copies are sufficient to learn a rank-r mixed state $\rho \in \mathbb{C}^{d\times d}$ to trace distance error ε. This gives a new proof of the main result of [OW16] and matches the lower bound of [SSW25].
- We then improve this result and show that $n = O(rd/\varepsilon^2)$ copies are sufficient to learn to error ε in the more challenging Bures distance. This is optimal due to the lower bound of [Yue23] and improves on the best known bound of $n = \widetilde{O}(rd/\varepsilon^2)$ from prior work [HHJ+16, OW17, FO24].
- We consider the scenario of full state tomography with limited entanglement, in which one is given n copies of ρ but only allowed to perform entangled measurements across k of them at a time. We show that
$$n = O\left(\max\left(\frac{d^3}{\sqrt{k}\varepsilon^2}, \frac{d^2}{\varepsilon^2}\right)\right)$$
copies of ρ suffice to learn ρ to trace distance error ε in this scenario. This improves on the prior work of [CLL24b] and matches their lower bound which holds for $k \le 1/\varepsilon^c$, where c is a constant.
- In the problem of shadow tomography, one is given m observables $O_1, \ldots, O_m$, and the goal is to estimate each quantity $\mathrm{tr}(O_i \cdot \rho)$ up to ε error. If $\|O_i\|_\infty \le 1$ for all i, we show that $O(\log(m)/\varepsilon^2)$ copies are sufficient to solve this task in the "high accuracy regime" when $\varepsilon = O(1/d)$. This improves a result of [CLL24a], who showed this bound holds when $\varepsilon = O(1/d^{12})$. More generally, we show that if $\mathrm{tr}(O_i^2) \le F$ for all i, then
$$n = O\Big(\log(m) \cdot \Big(\min\Big\{\frac{\sqrt{rF}}{\varepsilon}, \frac{F^{2/3}}{\varepsilon^{4/3}}\Big\} + \frac{1}{\varepsilon^2}\Big)\Big)$$
copies suffice to solve this task. This improves on a result of [GLM24], who showed a bound of $n = O(\log(m) \cdot \sqrt{rF}/\varepsilon^2)$, and disproves a conjecture of [CLL24a].
- Our final application is to the field of quantum metrology. In quantum metrology, one is provided copies of a mixed state ρ_θ parameterized by a vector $\theta \in \mathbb{R}^n$, and the goal is to estimate the parameter θ. An algorithm is *locally unbiased* at a parameter θ^* if, roughly, it is unbiased given ρ_θ when θ is in a small neighborhood around θ^*. The quantum Cramér–Rao bound states that the mean squared error matrix (MSEM) of any locally unbiased algorithm is lower bounded by the inverse of the Quantum Fisher Information (QFI) matrix. We give a locally unbiased algorithm whose MSEM is *upper bounded* by twice the inverse of the QFI matrix in the asymptotic limit of large n, which is optimal.

*UC Berkeley. {apelecan,jspilecki,jswright}@berkeley.edu

Contents

Part I
Introduction

1 Introduction

A statistical estimator $\widehat{\boldsymbol{X}}$ for a quantity X is said to be *unbiased* if $\mathbf{E}[\widehat{\boldsymbol{X}}] = X$. Unbiased estimators have many desirable properties, and for many quantities of interest in statistics, the gold standard estimator is unbiased. As Voinov and Nikulin put it in *Unbiased Estimators and their Applications* [VN12, VN96], "Unbiasedness is one of the most important properties of statistical estimators". Illustrating the abundance of unbiased estimators in statistics, they collect roughly a thousand (!) such estimators for different statistical parameters.

This work is about the statistical problem of estimating a mixed quantum state $\rho \in \mathbb{C}^{d \times d}$ given n identical copies of ρ. This task, known as *quantum tomography*, is of fundamental importance to both theory and practice, and it has received significant attention by the research community in recent years. A variety of different tomography algorithms have been studied in the literature, and many of these do actually give unbiased estimators for ρ. Two well-known examples are the "uniform POVM tomography algorithm" from [Wri16, GKKT20] and the "textbook" Pauli tomography algorithm, which is covered in [Wri24a]. In addition to these, we also know of several estimators which achieve the optimal sample complexity of $n = O(d^2)$ copies for full state tomography [HHJ$^+$16, OW16]. However, we do not know of any estimators which are simultaneously unbiased *and* sample-optimal. The reason is that most existing unbiased estimators for full state tomography measure each copy of ρ independently, but it is known that any sample-optimal tomography algorithm must use entangled measurements across many copies of ρ [CHL$^+$23]. On the flip side, measurements which are entangled across many copies of ρ are often mathematically complex and difficult to understand, and even computing the bias of known sample-optimal tomography algorithms appears to be quite challenging.

It is important to understand whether sample-optimal estimators can also be unbiased if we want to understand the information theoretic limits of quantum state learning. This issue was highlighted in two recent works of Chen, Li, and Liu [CLL24b, CLL24a], who studied two problems in quantum tomography for which having an unbiased estimator would be extremely useful: full state tomography using measurements with limited entanglement and shadow tomography. However, since we do not know of any sample-optimal estimators which are unbiased, they had to develop sophisticated mathematical techniques for analyzing the known biased, sample-optimal tomography algorithms as if they were unbiased. In the course of this work, they suggested that perhaps there may not even *exist* any unbiased, sample-optimal tomography algorithms, saying that "unbiasedness seems specific to incoherent measurements".

In this work, we counter this intutition by developing an estimator for full state which is both unbiased and sample-optimal. Our estimator is based on an estimator for full state tomography introduced by Keyl in [Key06] and shown to be sample-optimal in [OW16]. Keyl's algorithm produces a biased estimator for ρ, but we show that a simple modification to this algorithm can correct for the bias and give an unbiased estimator. With this "debiased Keyl's algorithm" in hand, we show a number of interesting applications to full state tomography. For some of these applications, we are able to derive conceptually cleaner proofs of existing results. For other applications, we are able to use the debiased Keyl's algorithm to give new results which were not known before, such as the first sample-optimal bound for learning mixed states with fidelity error. Prior to our work, there were three other sample-optimal (or approximately sample-optimal) algorithms to choose from when performing full state tomography [Key06, HHJ$^+$16], but we believe that our results suggest that the debiased Keyl's algorithm may be the "right" fully entangled tomography algorithm to study in the future, and we expect that it will be useful in resolving several other remaining open problems in the study of full state tomography. Below, we describe several known unbiased estimators for full state tomography in order to motivate our construction of the debiased Keyl's algorithm. Following that, we survey the applications of this new algorithm.

1.1 Debiasing tomography algorithms

1.1.1 The single copy case

Suppose we only have a single copy of ρ, and we want to perform full state tomography on it. Exactly which measurement is the best to perform is perhaps not clear, but a natural thing to do is to measure ρ in *some* orthonormal basis, as in the following algorithm.

1. Pick an orthonormal basis $|u_1\rangle, \ldots, |u_d\rangle \in \mathbb{C}^d$.
2. Measure ρ in this basis. Let $\boldsymbol{i} \in [d]$ be the outcome.
3. Output $\widehat{\boldsymbol{\rho}}_0 = |u_{\boldsymbol{i}}\rangle\langle u_{\boldsymbol{i}}|$.

How well this algorithm performs depends on which basis we decide to measure in. For simplicity, let us assume that ρ is diagonal in the standard basis, meaning that we can write $\rho = \sum_{i=1}^{d} \alpha_i \cdot |i\rangle\langle i|$. If the basis we measure in happens to also be the standard basis, it is easy to see that

$$\mathbf{E}[\widehat{\boldsymbol{\rho}}_0] = \alpha_1 \cdot |1\rangle\langle 1| + \cdots + \alpha_d \cdot |d\rangle\langle d| = \rho,$$

and so $\widehat{\boldsymbol{\rho}}_0$ is an unbiased estimator for ρ. On the other hand, suppose the basis we measure in forms a mutually unbiased basis with the standard basis, meaning that $|\langle i|u_j\rangle|^2 = 1/d$ for all i and j. This is the case if, for example, $\{|u_i\rangle\}_{i\in[d]}$ is the Fourier basis. Then we will observe each $i \in [d]$ with equal probability, and so

$$\mathbf{E}[\widehat{\boldsymbol{\rho}}_0] = \tfrac{1}{d} \cdot |u_1\rangle\langle u_1| + \cdots + \tfrac{1}{d} \cdot |u_d\rangle\langle u_d| = \frac{I}{d},$$

giving the maximally mixed state in expectation, regardless of what the state ρ is. We can view the first case as giving us "all signal" and the second case as giving us "all noise".

It is natural to randomize the choice of basis in order to avoid accidentally falling into the "all noise" case all of the time. Doing so involves measuring according to a random basis $|\boldsymbol{u}_1\rangle, \ldots, |\boldsymbol{u}_d\rangle \in \mathbb{C}^d$ sampled from the Haar measure. Such a basis will still be *almost* mutually unbiased with the standard basis, in the sense that we expect $|\langle i|\boldsymbol{u}_j\rangle|^2$ to be close to $1/d$ for most values of i and j, but the fact that it is not *exactly* mutually unbiased means that we can still hope for a small amount of signal to leak through. In this case, our measurement, over the randomness in the choice of basis, is equivalent to performing the *uniform POVM*, the continuous POVM given by $\{d \cdot |u\rangle\langle u| \cdot du\}$. This means that we can rewrite the algorithm as follows.

1. Measure ρ with the uniform POVM $\{d \cdot |u\rangle\langle u| \cdot du\}$. Let $|\boldsymbol{u}\rangle$ be the outcome.
2. Output $\widehat{\boldsymbol{\rho}}_0 = |\boldsymbol{u}\rangle\langle \boldsymbol{u}|$.

The expectation of this estimator can be computed using standard techniques; the calculation can be found, for example, in [Wri16, Page 115]:

$$\mathbf{E}[\widehat{\boldsymbol{\rho}}_0] = \left(\frac{1}{d+1}\right) \cdot \rho + \left(\frac{d}{d+1}\right) \cdot \frac{I}{d}.$$

We can see that the first term is a small "signal" term and the second term is a large "noise" term. Moreover, the weights on these two terms are fixed constants independent of the state ρ. It is therefore straightforward to correct for this bias, and doing so results in the following estimator:

$$\widehat{\boldsymbol{\rho}} = (d+1) \cdot \widehat{\boldsymbol{\rho}}_0 - I = (d+1) \cdot |\boldsymbol{u}\rangle\langle \boldsymbol{u}| - I.$$

Indeed, this satisfies $\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \rho$ and is therefore an unbiased estimator. This single copy estimator was independently introduced by Krishnamurthy and Wright [Wri16, Section 5.1] and Guta et al. [GKKT20] and forms the basis of an optimal unentangled measurement tomography algorithm; we will discuss this algorithm further in Section 1.4 below.

1.1.2 The pure state case

Suppose we are given n copies of a rank-one mixed state ρ, i.e. $\rho = |v\rangle\langle v|$ for some pure state $|v\rangle \in \mathbb{C}^d$. In this case, there is a well-known tomography algorithm due to Hayashi [Hay98] which is given as follows.

1. Measure $\rho^{\otimes n}$ with the POVM $\{d[n] \cdot |u\rangle\langle u|^{\otimes n} \cdot du\}$, where $d[n] = \binom{d+n+1}{n}$ is the dimension of the symmetric subspace $\vee^n \mathbb{C}^d$. Let $|\boldsymbol{u}\rangle$ be the outcome.
2. Output $\widehat{\boldsymbol{\rho}}_0 = |\boldsymbol{u}\rangle\langle\boldsymbol{u}|$.

The POVM $\{d[n] \cdot |u\rangle\langle u|^{\otimes n} \cdot du\}$ can be viewed as a natural generalization of the uniform POVM from above to the n-fold symmetric subspace. It is well-known that this is a sample-optimal algorithm for pure state tomography. In particular, when $n = O(d/\varepsilon^2)$, this estimator satisfies $\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}_0) \leq \varepsilon$ with probability 99% [Wri24b]. However, $\widehat{\boldsymbol{\rho}}_0$ is *not* an unbiased estimator of ρ. This is because ρ is a pure state, whereas $\mathbf{E}[\widehat{\boldsymbol{\rho}}_0]$ will certainly be a mixed state with rank strictly larger than 1 (unless $\widehat{\boldsymbol{\rho}}_0 = \rho$ with probability 1, which is impossible). The expectation of this estimator can be computed using standard techniques; [GPS24, Lemma 13] give the following expression:

$$\mathbf{E}[\widehat{\boldsymbol{\rho}}_0] = \left(\frac{n}{d+n}\right) \cdot \rho + \left(\frac{d}{d+n}\right) \cdot \frac{I}{d}.$$

Again, the first term corresponds to the "signal" and the second term corresponds to the "noise". Note that the coefficient on the "signal" term now improves as n grows larger, and approaches 1 as n approaches ∞. Since the coefficients are fixed constants independent of the state ρ, it is straightforward to correct for this bias, resulting in the following unbiased estimator:

$$\widehat{\boldsymbol{\rho}} = \left(\frac{d+n}{n}\right) \cdot \widehat{\boldsymbol{\rho}}_0 - \frac{1}{n} \cdot I = \left(\frac{d+n}{n}\right) \cdot |\boldsymbol{u}\rangle\langle\boldsymbol{u}| - \frac{1}{n} \cdot I.$$

This estimator was introduced by Grier, Pashayan, and Schaeffer [GPS24] with applications to the shadow tomography problem; we will discuss this algorithm further in Section 1.5.

1.1.3 Keyl's algorithm

Our goal is to perform a similar debiasing for a fully entangled tomography algorithm. There are three reasonable candidates to select from: Keyl's algorithm [Key06] and the two pretty good measurement-inspired algorithms from Haah et al. [HHJ+16]. Of these, two are known to achieve optimal sample complexities of $n = O(d^2/\varepsilon^2)$ for full state tomography: Keyl's algorithm (due to [OW16]) and the first of the two Haah et al. algorithms (due to [OW15a]). Of these two, we have decided to study Keyl's algorithm, for two reasons: first, it is the more mathematically elegant of the two, and we believe this will make its bias easier to analyze. Second, as Aram Harrow once suggested to the third author, it seems plausible that it uses the *optimal* entangled measurement for full state tomography [Har16], though it is not clear how to exactly formulate what it means for it to be optimal.

Keyl's algorithm is a sophisticated representation theoretic algorithm which exploits the symmetries in the state $\rho^{\otimes n}$. In phase one of the algorithm, it performs a measurement on $\rho^{\otimes n}$ known as *weak Schur sampling* in order to estimate the eigenvalues of ρ. Weak Schur sampling returns as a measurement outcome a *Young diagram* $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_d) \vdash n$, which is a vector of nonnegative integers satisfying $\boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_d$ and $\boldsymbol{\lambda}_1 + \cdots + \boldsymbol{\lambda}_d = n$. Young diagrams are represented pictorially by arranging boxes into rows, as in Figure 1. From this picture, the Young diagram looks like a (sorted) histogram of n samples drawn from some discrete distribution over d items, and coincidentally, ρ's eigenvalues $\alpha = (\alpha_1, \ldots, \alpha_d)$ form a probability distribution over the set of integers $\{1, \ldots, d\}$. It is therefore natural to wonder if the *normalized Young diagram* $\boldsymbol{\lambda}/n = (\boldsymbol{\lambda}_1/n, \ldots, \boldsymbol{\lambda}_d/n)$ provides an accurate approximation of the spectrum α, and this question was answered in the affirmative by independent works of Alicki, Rudnicki, and Sadowski [ARS88] and Keyl and Werner [KW01]; these works showed that as $n \to \infty$, $\boldsymbol{\lambda}/n \to \alpha$ with probability 1. This estimator is known both as the *empirical Young diagram algorithm* and also as the *Keyl-Werner algorithm*, and there exists a number of follow-up works showing stronger and stronger bounds on the number of copies it requires to estimate α [HM02, CM06, OW16, OW17]. The final two of these works show that $\boldsymbol{\lambda}/n$ is close to α once

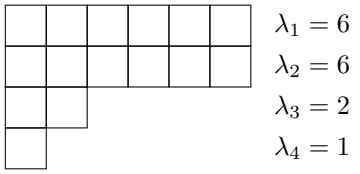


Figure 1: The Young diagram corresponding to the partition $\lambda = (6, 6, 2, 1)$. In this case, $n = 15$ and $d = 4$. This resembles the histogram, sorted from highest to lowest and then turned on its side, of $n = 15$ samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{15}$ drawn from a discrete distribution $p = (p_1, p_2, p_3, p_4)$ whose probability values, when sorted from highest to lowest, are approximately 6/15, 6/15, 2/15, and 1/15.

$n = O(d^2)$ copies are used in total variation distance, Hellinger distance, KL divergence, and chi-squared divergence; furthermore, a lower bound from [OW15b] shows that the empirical Young diagram algorithm *requires* $n = \Omega(d^2)$ copies to produce a good estimate in any of these distances. To date, it remains the best known algorithm for estimating the spectrum of a mixed state. (That said, we don't yet have matching lower bounds, suggesting the tantalizing possibility that there might exist an even better spectrum estimation algorithm.)

Having received the measurement outcome $\boldsymbol{\lambda}$, the state $\rho^{\otimes n}$ partially collapses. In phase two of Keyl's algorithm, it performs a further measurement on this collapsed state in order to learn the eigenvectors of ρ. This measurement is based on a strategy of "rotated highest weight vectors"; we will describe it in detail in Section 3.1 below once we have established the relevant representation theoretic background. It produces a unitary matrix $\boldsymbol{U} \in U(d)$ as a measurement outcome which is meant to be a good proxy for the change of basis which rotates the standard basis into ρ's eigenbasis. Setting $\widehat{\boldsymbol{\alpha}}_0 = \boldsymbol{\lambda}/n$ for our approximation of the spectrum, the estimate that Keyl's algorithm outputs is $\widehat{\boldsymbol{\rho}}_0 = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}_0 \cdot \boldsymbol{U}^\dagger$.

Keyl's algorithm produces a high quality estimate of the state ρ. It is known, for example, that $\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}_0) \leq \varepsilon$ with high probability once $n = O(d^2/\varepsilon^2)$, and furthermore that this is optimal due to a matching lower bound of Haah et al. [HHJ+16]. However, it is certainly not unbiased. For example, when $n = 1$, it turns out that Keyl's algorithm is equivalent to the biased single-copy tomography algorithm from Section 1.1.1. Furthermore, when n is allowed to be any value but ρ is restricted to be a pure state, it turns out that Keyl's algorithm is equivalent to the biased pure state tomography algorithm from Section 1.1.2 due to Hayashi. However, unlike in these special cases, the bias of Keyl's algorithm is in general quite difficult to calculate, and so there is no obvious method for debiasing it. Compounding this issue, the bias of Keyl's algorithm actually depends on the underlying state ρ, and so there cannot be just a single correction factor which applies in every case.

1.1.4 The debiased Keyl's algorithm

To debias Keyl's algorithm, we draw inspiration from the debiased algorithms in Sections 1.1.1 and 1.1.2. In both of these cases, to convert the biased estimator $\widehat{\boldsymbol{\rho}}_0$ into an unbiased estimator $\widehat{\boldsymbol{\rho}}$, we modify the eigenvalues of $\widehat{\boldsymbol{\rho}}_0$ but leave the eigenvectors intact. This suggests that to debias Keyl's algorithm, we will want an estimator of the form $\boldsymbol{\rho} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^\dagger$, where $\boldsymbol{U}$ is our original change of basis but $\widehat{\boldsymbol{\alpha}}$ is a new estimate of the spectrum which results from modifying the original estimate $\widehat{\boldsymbol{\alpha}}_0$. To define this modification, we first introduce a transformation on Young diagrams.

Definition 1.1 (Young diagram box donation)**.** Let $\lambda = (\lambda_1, \ldots, \lambda_d)$ be a Young diagram. Then $\mathrm{donate}(\lambda) \in \mathbb{Z}^d$ is the vector defined as

$$\mathrm{donate}(\lambda)_i = \lambda_i - \sum_{j=1}^{i-1} 1[\lambda_j > \lambda_i] + \sum_{j=i+1}^{d} 1[\lambda_i > \lambda_j].$$

Intuitively, each row of the Young diagram λ "donates" a box to each row which starts off longer than it and receives a box from each row which starts off shorter than it. We include a diagram of $\mathrm{donate}(\lambda)$ in Figure 2.

To our knowledge, this is a new transformation on Young diagrams, and we could find no similar transformation in the literature. Note that this transformation only shifts boxes around and does not introduce

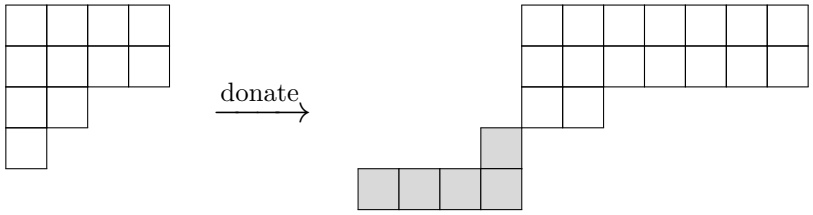


Figure 2: Let $d = 5$. On the left is the Young diagram $\lambda = (4, 4, 2, 1, 0)$. On the right is donate$(\lambda) = (7, 7, 2, -1, -4)$. The gray-colored boxes correspond to rows with a negative length. Note that donate(λ) is not necessarily a Young diagram, since it may have negative entries.

any new boxes, and so if $\lambda = (\lambda_1, \ldots, \lambda_d) \vdash n$ and $\lambda' = \text{donate}(\lambda)$, then $\lambda'_1 + \cdots + \lambda'_d = \lambda_1 + \cdots + \lambda_d = n$. With this definition in place, we can define our debiased Keyl's algorithm.

Definition 1.2 (Debiased Keyl's algorithm)**.** Suppose we are given n copies of a mixed state $\rho \in \mathbb{C}^{d \times d}$. The *debiased Keyl's algorithm* works as follows.

1. Measure $\rho^{\otimes n}$ as in Keyl's algorithm, producing a Young diagram $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_d) \vdash n$ and a unitary matrix $\boldsymbol{U} \in U(d)$.
2. Set $\boldsymbol{\lambda}^{\uparrow} = \text{donate}(\boldsymbol{\lambda})$ and $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\lambda}^{\uparrow}/n$. Output $\widehat{\boldsymbol{\rho}} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^{\dagger}$.

To gain intuition for this algorithm, let us define $|\boldsymbol{u}_i\rangle \coloneqq \boldsymbol{U} \cdot |i\rangle$ for each $1 \leq i \leq d$. Then

$$\widehat{\boldsymbol{\rho}}_0 = \sum_{i=1}^{d} (\boldsymbol{\lambda}_i/n) \cdot |\boldsymbol{u}_i\rangle\langle\boldsymbol{u}_i| \qquad \text{and} \qquad \widehat{\boldsymbol{\rho}} = \sum_{i=1}^{d} (\boldsymbol{\lambda}_i^{\uparrow}/n) \cdot |\boldsymbol{u}_i\rangle\langle\boldsymbol{u}_i|$$

For example, when $n = 1$ and $\lambda = (1, 0, \ldots, 0)$, we have

$$\begin{aligned} \widehat{\boldsymbol{\rho}}_0 &= |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| \\ \widehat{\boldsymbol{\rho}} &= d \cdot |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| - (|\boldsymbol{u}_2\rangle\langle\boldsymbol{u}_2| + \cdots + |\boldsymbol{u}_d\rangle\langle\boldsymbol{u}_d|) = (d+1) \cdot |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| - I, \end{aligned}$$

as in Section 1.1.1. Next, when ρ is a pure state, then $\lambda = (n, 0, \ldots, 0)$ always, and

$$\begin{aligned} \widehat{\boldsymbol{\rho}}_0 &= \binom{n}{n} \cdot |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| = |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| \\ \widehat{\boldsymbol{\rho}} &= \binom{n+d-1}{n} \cdot |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| - \frac{1}{n} \cdot (|\boldsymbol{u}_2\rangle\langle\boldsymbol{u}_2| + \cdots + |\boldsymbol{u}_d\rangle\langle\boldsymbol{u}_d|) = \binom{n+d}{n} \cdot |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| - \frac{1}{n} \cdot I, \end{aligned}$$

as in Section 1.1.2. More generally, Keyl's algorithm is designed to ensure that $|\boldsymbol{u}_1\rangle$, the eigenvector corresponding to the highest eigenvalue of $\widehat{\boldsymbol{\rho}}_0$, is the most accurate of all the eigenvectors, whereas $|\boldsymbol{u}_d\rangle$ is the least accurate. This means that $|\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1|$ will be the term with the highest "signal" and $|\boldsymbol{u}_d\rangle\langle\boldsymbol{u}_d|$ will be the term with the highest "noise". Thus, the noisier terms will donate boxes to each of the terms with more signal, and this "rich gets richer" approach will boost the signal while depressing the noise. The next theorem is our first main result, which shows that the Young diagram box donation transformation results in an estimator in which all of the noise is precisely cancelled out.

Theorem 1.3 (Debiased Keyl's algorithm is an unbiased estimator)**.** *Let $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $\rho^{\otimes n}$. Then $\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \rho$.*

Having computed the first moment of $\widehat{\boldsymbol{\rho}}$, we next want to compute an expression for its second moment $\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}]$. The second moment is useful because it helps us understand the extent to which $\widehat{\boldsymbol{\rho}}$ varies about its expectation. We do so in the following theorem, which is our second main result.

Theorem 1.4 (Second moment of the debiased Keyl's algorithm)**.** *Let $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $\rho^{\otimes n}$. Then*

$$\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] = \frac{n-1}{n} \cdot \rho \otimes \rho + \frac{1}{n} (\rho \otimes I + I \otimes \rho) \cdot \mathrm{SWAP} + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \cdot \mathrm{SWAP} - \mathrm{Lower}_\rho. \tag{1}$$

Here, Lower_ρ *refers to the "lower order terms" and has the following characterization: it can be written as a positive linear combination of terms of the form $(P \otimes P) \cdot \mathrm{SWAP}$, where each $P \in \mathbb{C}^{d \times d}$ is a Hermitian, positive semidefinite matrix.*

To understand this expression, the first term, which can be viewed as the "main term", is essentially equal to $\rho \otimes \rho$, which is what we would expect to see if $\widehat{\boldsymbol{\rho}}$ were always equal to its expectation. The second and third terms are the "error terms" and have the largest contribution to the variance of $\widehat{\boldsymbol{\rho}}$. As for the final term, it turns out that we can expand the second moment of $\widehat{\boldsymbol{\rho}}$ as a polynomial in the variable d^{-1}, yielding an expression of the form

$$\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] = A_\rho + B_\rho \cdot \tfrac{1}{d} + C_\rho \cdot \tfrac{1}{d^2} + \cdots,$$

where A_ρ, B_ρ, C_ρ, and so forth are matrix-valued coefficients which depend on the state ρ and the number of copies n but not on the dimension d. As we will explain in Section 15 below, the highest order term A_ρ is exactly the first three terms in Equation (1), and so the remaining low-order terms $B/d + C/d^2 + \cdots$ are given by $-\mathrm{Lower}_\rho$. In the case when $d \gg n$ these terms are small enough that they can effectively be discarded. (Indeed, as we discuss in Section 15, there are situations where it is possible to embed ρ into a larger space of dimension $D \gg d$ in order to artificially inflate its dimension and make these terms arbitrarily small.)

However, it turns out that for our applications, we can disregard these lower order terms for a simpler reason, which is that their contribution is negative in the cases we care about. In one of our applications, for example, we have to upper bound expressions of the form

$$\mathbf{E} \left|\widehat{\boldsymbol{\rho}}_{i,j}\right|^2 = \mathbf{E}[\widehat{\boldsymbol{\rho}}_{i,j} \cdot \widehat{\boldsymbol{\rho}}_{j,i}] = \mathbf{E}[\langle i| \widehat{\boldsymbol{\rho}} |j\rangle \cdot \langle j| \widehat{\boldsymbol{\rho}} |i\rangle] = \langle ij| \cdot \mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] \cdot |ji\rangle.$$

The contribution to this expression from the lower order terms is $-\langle ij| \cdot \mathrm{Lower}_\rho \cdot |ji\rangle$. From our characterization of Lower_ρ, we can expand $\langle ij| \cdot \mathrm{Lower}_\rho \cdot |ji\rangle$ as a positive linear combination of terms of the form

$$\langle ij| \cdot (P \otimes P) \cdot \mathrm{SWAP} \cdot |ji\rangle = \langle ij| \cdot (P \otimes P) \cdot |ij\rangle = P_{ii} \cdot P_{jj} \geq 0,$$

where the last step uses the fact that P is positive semidefinite. Hence, $-\langle ij| \cdot \mathrm{Lower}_\rho \cdot |ji\rangle$ is nonpositive, and we are safe to discard it. In another of our applications, we have to upper bound expressions of the form

$$\mathbf{E} \operatorname{tr}(O \cdot \widehat{\boldsymbol{\rho}})^2 = \mathbf{E} \operatorname{tr}(O \otimes O \cdot \widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}) = \operatorname{tr}(O \otimes O \cdot \mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}]),$$

where $O \in \mathbb{C}^{d \times d}$ is an observable. The contribution to this expression from the lower order terms is $-\operatorname{tr}(O \otimes O \cdot \mathrm{Lower}_\rho)$; we can expand $\operatorname{tr}(O \otimes O \cdot \mathrm{Lower}_\rho)$ as a positive linear combination of the form

$$\begin{aligned} \operatorname{tr}(O \otimes O \cdot (P \otimes P) \cdot \mathrm{SWAP}) &= \operatorname{tr}((OP) \otimes (OP) \cdot \mathrm{SWAP}) \\ &= \operatorname{tr}((OP)^2) = \operatorname{tr}(OPOP) = \operatorname{tr}((P^{1/2} O P^{1/2})^2) \geq 0, \end{aligned}$$

where we used the fact that P is positive semidefinite and therefore has a square root. Hence, we again have that $-\operatorname{tr}(O \otimes O \cdot \mathrm{Lower}_\rho)$ is nonpositive, and we are safe to discard it. The bottom line is that when we apply Theorem 1.4, we only have to worry about the first three terms, not the lower order terms, and these are easy to compute and bound.

In fact, we show that there is a whole family of unbiased estimators which one can define based on the measurements in Keyl's algorithm, of which our debiased Keyl's algorithm is just one among many. Among these, we show in Section 3.3 that our debiased Keyl's algorithm is the *minimum variance unbiased estimator*. Note that this is only among those estimators which use Keyl's measurement and does not preclude an unbiased estimator with even smaller variance which is based on a different measurement.

Discovering the estimator. Before moving on to applications of our debiased Keyl's algorithm, let us first describe how we discovered the debiased Keyl's algorithm and proved Theorems 1.3 and 1.4. Recall that when ρ is a pure state, the observed $\boldsymbol{\lambda}$ is always equal to $(n, 0, \dots, 0)$, and Keyl's algorithm is equivalent to the biased pure state tomography algorithm from Section 1.1.2. As mentioned above, the correction we use to remove the bias in this case is the same as in our pure state tomography algorithm, and it is not too hard to see that this is essentially the *only* way to correct the bias in this case. But this means we are forced to correct the bias in this way whenever we receive the Young diagram $\boldsymbol{\lambda} = (n, 0, \dots, 0)$, because there is a chance that ρ is a pure state, and we have to ensure that our estimator is unbiased in this case. When $n = 2$ and $d = 2$, say, there are only two possible Young diagrams $(2, 0)$ and $(1, 1)$, and the fact that we have already determined what to do in the first case severely limits the possible corrections that we can apply in the second case, which makes finding the right correction much easier. Applying similar logic to other small values of n and d, we solved enough special cases that we were eventually able to spot a pattern that generalized to all values of our parameters. Perhaps the biggest difficulty in this process was deciding on the best way to compute the bias of Keyl's algorithm; after trying various approaches, we settled on an approach which uses Clebsch-Gordan coefficients. Getting into what exactly Clebsch-Gordan coefficients are is perhaps a bit too technical for this introduction, and we will defer introducing them until Section 10. Suffice to say, they appear to be especially well suited for computing expectations which arise from Keyl's measurements, although they can be more than a little cumbersome to work with.

Proving Theorem 1.4 was significantly more difficult than discovering the debiased Keyl's algorithm and is the most technically involved part of this paper. The reason it was so difficult is that we also use Clebsch-Gordan coefficients to compute the variance of $\widehat{\boldsymbol{\rho}}$, and going from our large expressions involving the Clebsch-Gordan coefficients to the nice expression in Theorem 1.4 involves some miraculous cancellations which we admit to not having a great understanding of. Proving Theorem 1.4 then required us first having a guess for the precise form of the variance, and only once we had the right guess could we reverse engineer a proof to give us the right calculations. Developing a guess for the variance, in turn, required spending significant amounts of time and effort on numerical simulations, in which we pored over command line outputs and performed numerology in a lengthy and difficult process that we likened to the office jobs in the television show *Severance*. However, given that the main terms in Theorem 1.4 are so nice, and the remaining terms are negative and so can be ignored, we hope that there is an alternative means of computing variances in which these nice expressions fall out naturally, and we leave this for future work.

1.2 Application 1: full state tomography in trace distance

Our first application is to the problem of full state tomography in trace distance. Trace distance is perhaps the most fundamental distance measure between two quantum states, quantifying the extent to which they can be distinguished by a quantum measurement. O'Donnell and Wright showed that $n = O(rd/\varepsilon^2)$ copies suffice for full state tomography of rank-r states in trace distance [OW16]. In the case when $r = d$, Haah et al. [HHJ+16] showed a matching lower bound of $n = \Omega(d^2/\varepsilon^2)$ copies, and concurrent work of Scharnhorst, Spilecki, and Wright [SSW25] has extended this to a matching lower bound of $n = \Omega(rd/\varepsilon^2)$ copies for all values of r. We show that the upper bound of O'Donnell and Wright also holds for the debiased Keyl's algorithm.

Theorem 1.5 (Trace distance tomography)**.** *Let ρ be rank r, and let $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $\rho^{\otimes n}$. Then $\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}) \le \varepsilon$ with high probability when $n = O(rd/\varepsilon^2)$.*

In fact, this result follows immediately from the prior result of O'Donnell and Wright. They showed that if $\widehat{\boldsymbol{\rho}}_0$ is the output of Keyl's algorithm, then $\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}_0) \le \varepsilon$ with high probability. Next, it is easy to see that $\mathrm{D}_{\mathrm{tr}}(\widehat{\boldsymbol{\rho}}_0, \widehat{\boldsymbol{\rho}}) \le d^2/n$ because $\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\uparrow}\|_1 \le d^2$ always. Thus, by the triangle inequality,

$$\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}_0) \le \mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}_0) + \mathrm{D}_{\mathrm{tr}}(\widehat{\boldsymbol{\rho}}_0, \widehat{\boldsymbol{\rho}}) \le \varepsilon + d^2/n \le \varepsilon + \varepsilon^2 = O(\varepsilon),$$

where the second-to-last inequality uses the fact that $n = O(d^2/\varepsilon^2)$. However, we are able to give an entirely new proof of this fact which does not rely on this prior result of O'Donnell and Wright. Our proof begins by bounding the expected trace distance between ρ and $\widehat{\boldsymbol{\rho}}$ as follows:

$$\mathbf{E}[\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}})] = \tfrac{1}{2}\,\mathbf{E}\,\|\rho - \widehat{\boldsymbol{\rho}}\|_1 \le \tfrac{1}{2}\,\mathbf{E}\,\sqrt{d} \cdot \|\rho - \widehat{\boldsymbol{\rho}}\|_2 \le \tfrac{1}{2}\sqrt{d} \cdot \sqrt{\mathbf{E}\,\|\rho - \widehat{\boldsymbol{\rho}}\|_2^2},$$

where the last step uses Jensen's inequality and concavity of the square root function. Now, because $\widehat{\boldsymbol{\rho}}$ is an unbiased estimator for ρ, we have $\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \rho$, and so $\mathbf{E}\,\|\rho - \widehat{\boldsymbol{\rho}}\|_2^2$ is just the variance of $\widehat{\boldsymbol{\rho}}$. This means

$$\begin{aligned}\mathbf{E}\,\|\rho - \widehat{\boldsymbol{\rho}}\|_2^2 = \mathbf{Var}[\widehat{\boldsymbol{\rho}}] = \mathbf{E}[\mathrm{tr}(\widehat{\boldsymbol{\rho}}^2)] - \mathrm{tr}(\mathbf{E}[\widehat{\boldsymbol{\rho}}]^2) = \mathbf{E}[\mathrm{tr}(\widehat{\boldsymbol{\rho}}^2)] - \mathrm{tr}(\rho^2) &= \mathbf{E}[\mathrm{tr}(\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}} \cdot \mathrm{SWAP})] - \mathrm{tr}(\rho^2) \\ &= \mathrm{tr}(\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] \cdot \mathrm{SWAP})] - \mathrm{tr}(\rho^2).\end{aligned}$$

Now we simply substitute our expression for the second moment of $\widehat{\boldsymbol{\rho}}$ from Theorem 1.4, and from there bounding the variance is relatively straightforward. In comparison to the tomography proof of [OW16], every step in this proof feels "canonical", whereas the proof of [OW16] requires a few lucky tricks (for example, the "$r \geq 2 - \frac{1}{r}$" step always seemed like a stroke of luck to the third author) to work.

1.3 Application 2: full state tomography in Bures distance

We also give an application to *fidelity* or *Bures distance* tomography. Bures distance is a more challenging distance metric than trace distance which requires learning the small eigenspace of ρ to high accuracy, and as a result we still lack truly optimal bounds for Bures distance tomography. Haah et al. [HHJ+16] showed that $\widetilde{O}(rd/\varepsilon^2)$ samples suffice to learn a rank r mixed state ρ to ε error in Bures distance (equivalently, $\widetilde{O}(rd/\delta)$ samples suffice to learn to δ infidelity error). This result has subsequently been reproved in [OW17, FO24], the former of which showed that Keyl's algorithm itself attains this bound. We improve on all of these results and show that Bures distance tomography can be achieved "without the tilde".

Theorem 1.6 (Bures distance tomography)**.** *Given a rank r mixed state $\rho \in \mathbb{C}^{d \times d}$, $n = O(rd/\varepsilon^2)$ samples suffice to output an estimate $\widehat{\boldsymbol{\rho}}$ such that $\mathrm{D_B}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability.*

This is optimal due to a matching lower bound of $n = \Omega(rd/\varepsilon^2)$ due to Yuen [Yue23]. Indeed, this lower bound also follows from the $n = \Omega(rd/\varepsilon^2)$ lower bound for trace distance tomography from above [SSW25] combined with the fact that Bures distance tomography is at least as hard as trace distance tomography. It is the first time an optimal Bures distance bound without any logs has been shown for either entangled or unentangled measurements.

For intuition, let us sketch the proof of the general $n = O(d^2/\varepsilon^2)$ copy upper bound from Theorem 1.6, which holds when the rank $r = d$. We actually rely on a slight variant of the debiased Keyl's algorithm in order to handle the case when the state ρ is poorly conditioned (i.e. when it has extremely small eigenvalues). To begin, we bound

$$\mathbf{E}[\mathrm{D_B}(\rho, \widehat{\boldsymbol{\rho}})] \leq \mathbf{E}[\sqrt{\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \| \rho)}] \leq \sqrt{\mathbf{E}[\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \| \rho)]},$$

where $\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \| \rho)$ is the *Bures χ^2-divergence* between $\widehat{\boldsymbol{\rho}}$ and ρ. Assuming without loss of generality that ρ is diagonal in the standard basis, i.e. $\rho = \sum_{i=1}^d \alpha_i \cdot |i\rangle\langle i|$, we can express the Bures χ^2-divergence as

$$\mathbf{E}[\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \| \rho)] = \mathbf{E}\left[\sum_{i,j=1}^d \frac{2}{\alpha_i + \alpha_j} \cdot |\widehat{\boldsymbol{\rho}}_{ij} - \rho_{ij}|^2\right] = \sum_{i,j=1}^d \frac{2}{\alpha_i + \alpha_j} \cdot \mathbf{E}[|\widehat{\boldsymbol{\rho}}_{ij} - \rho_{ij}|^2].$$

Now, if $\widehat{\boldsymbol{\rho}}$ is the output of the debiased Keyl's algorithm, then it is an unbiased estimator for ρ, and so $\mathbf{E}[|\widehat{\boldsymbol{\rho}}_{ij} - \rho_{ij}|^2]$ is just the variance $\mathbf{Var}[\widehat{\boldsymbol{\rho}}_{ij}]$. To calculate this, we can use our bound on the second moment of $\widehat{\boldsymbol{\rho}}$ from Theorem 1.4, which allows us to compute the bound $\mathbf{Var}[\widehat{\boldsymbol{\rho}}_{ij}] \leq (\alpha_i + \alpha_j)/n + d/n^2$. Plugging this in above, we have

$$\mathbf{E}[\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \| \rho)] \leq \sum_{i,j=1}^d \frac{2}{\alpha_i + \alpha_j} \cdot \left(\frac{\alpha_i + \alpha_j}{n} + \frac{d}{n^2}\right) = \sum_{i,j=1}^d \left(\frac{2}{n} + \frac{2d}{(\alpha_i + \alpha_j)n^2}\right) \leq \frac{2d^2}{n} + \frac{d^3}{\min_i\{\alpha_i\} \cdot n^2}.$$

If we know that ρ is well-conditioned so that each $\alpha_i \geq d/n$, then we can bound this final expression by $3d^2/n$, which is $O(\varepsilon^2)$ when $n = O(d^2/\varepsilon^2)$. This implies a bound of $O(\varepsilon)$ on $\mathbf{E}[\mathrm{D_B}(\rho, \widehat{\boldsymbol{\rho}})]$, as desired. To ensure that ρ is sufficiently well-conditioned, we can always apply the depolarizing channel to each copy of ρ with noise rate d^2/n prior to running the debiased Keyl's algorithm in order to prepare identical copies of the state

$$\rho' = \left(1 - \frac{d^2}{n}\right) \cdot \rho + \frac{d^2}{n} \cdot \frac{I}{d},$$

which has all eigenvalues at least d/n. We show that this ρ' is still close enough to ρ that it suffices to learn ρ' instead of ρ. One final wrinkle in this proof is that $\mathrm{D_B}(\rho, \widehat{\boldsymbol{\rho}})$ is only well-defined with $\widehat{\boldsymbol{\rho}}$ is positive semi-definite, but the debiased Keyl's algorithm in general can output states with negative eigenvalues. However, we show that because ρ' is sufficiently well-conditioned and we are using $n = O(d^2/\varepsilon^2)$ copies, the estimator $\widehat{\boldsymbol{\rho}}$ is an actual density matrix (and is therefore positive semidefinite) with high probability. This proof handles the case when $r = d$; to handle the case of small r, we will require a different argument that treats the small eigenvalues of ρ with more care.

As an application of our Bures distance result, we show how to efficient learn a bipartite pure state given a bound on its Schmidt rank. We thank Benjamin Lovitz for suggesting this application.

Theorem 1.7 (Learning bipartite pure states)**.** *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert spaces of dimension d_A and d_B, respectively. Then $n = O(r(d_A + d_B)/\varepsilon^2)$ copies of a pure state $|\psi\rangle_{AB}$ with Schmidt rank r are sufficient to learn it to Bures distance ε with high probability.*

1.4 Application 3: tomography with limited entanglement

Suppose we want to perform tomography on ρ, but we can only perform entangled measurements across k copies of ρ at a time. This question is motivated by the fact that implementing large, entangled measurements can be challenging in practice; indeed, one often simply does not have enough qubits to store all n copies of ρ at once! In this case, the natural strategy is to do the following.

1. Assuming for simplicity that n is a multiple of k, divide the n copies of ρ into $n' \coloneqq n/k$ disjoint blocks.
2. Run a k-copy tomography algorithm on each block. Let $\widehat{\boldsymbol{\rho}}_1, \ldots, \widehat{\boldsymbol{\rho}}_{n'}$ be the resulting estimators.
3. Output $\widehat{\boldsymbol{\rho}} = \frac{1}{n'} \cdot (\widehat{\boldsymbol{\rho}}_1 + \cdots + \widehat{\boldsymbol{\rho}}_{n'})$.

The estimators $\widehat{\boldsymbol{\rho}}_1, \ldots, \widehat{\boldsymbol{\rho}}_{n'}$ are independent and identically distributed, which in particular means that they have the same expectation $\mathbf{E}[\widehat{\boldsymbol{\rho}}_1] = \cdots = \mathbf{E}[\widehat{\boldsymbol{\rho}}_{n'}]$. By linearity of expectation, the averaged estimator $\widehat{\boldsymbol{\rho}}$ has the same expectation as well. However, even though it has the same expectation, it will concentrate around this expectation much more strongly than the individual estimators do. This can be seen by considering its variance and recalling that the variance of independent random variables is equal to the sum of their variances:

$$\mathbf{Var}[\widehat{\boldsymbol{\rho}}] = \frac{1}{(n')^2} \cdot \mathbf{Var}[\widehat{\boldsymbol{\rho}}_1 + \cdots + \widehat{\boldsymbol{\rho}}_{n'}] = \frac{1}{(n')^2} \cdot (\mathbf{Var}[\widehat{\boldsymbol{\rho}}_1] + \cdots + \mathbf{Var}[\widehat{\boldsymbol{\rho}}_1]) = \frac{1}{(n')} \cdot \mathbf{Var}[\widehat{\boldsymbol{\rho}}_{n'}].$$

If the tomography algorithm that is used produces an unbiased estimator, then this means that the estimator $\widehat{\boldsymbol{\rho}}$ will concentrate well around ρ when n' is large. As an example, in the case when $k = 1$, the measurements must be unentangled across all n copies of ρ, and the natural tomography algorithm to use is the single-copy unbiased estimator from Section 1.1.1. In this case, the averaged estimator $\widehat{\boldsymbol{\rho}}$ is known as the "uniform POVM tomography algorithm" and satisfies $\mathrm{D_{tr}}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability once $n = O(d^3/\varepsilon^2)$, which was shown independently by Krishnamurthy and Wright [Wri16, Section 5.1] and Guta et al. [GKKT20].

However, if the tomography algorithm that is used does *not* produce an unbiased estimator, then although $\widehat{\boldsymbol{\rho}}$ will still concentrate well around the expectation of the $\widehat{\boldsymbol{\rho}}_i$'s, this will no longer be useful if this expectation is far from ρ. This was the issue faced by Chen, Li, and Liu in [CLL24b], who studied the case when the algorithm is allowed to make entangled measurements across $k > 1$ copies at a time. In spite of the fact that they did not have an unbiased, entangled tomography algorithm at their disposal, they were still able to prove nearly matching upper and lower bounds for this problem. In particular, when $k \leq \min\{d^2, (\sqrt{d}/\varepsilon)^c\}$, where c is some small constant, they showed that $n = \widetilde{O}(d^3/(\sqrt{k}\varepsilon^2))$ copies of ρ suffice to learn ρ to error ε in trace distance. Interestingly, this suggests that increasing the amount of entanglement k helps only until $k = d^2$, and further increases in k do not improve the copy complexity. They also showed that $n = \Omega(d^3/(\sqrt{k}\varepsilon^2))$ copies are necessary to learn ρ to error ε in trace distance if one performs even adaptively chosen measurements across k copies of ρ at a time, at least when $k \leq 1/\varepsilon^c$, where c is some small constant. Finally, they suggested that these restrictions on the parameter k in both their upper and lower bounds are artifacts of their proof techniques, and they conjectured that the optimal number of copies for this task is $n = \widetilde{\Theta}(d^3/(\sqrt{k}\varepsilon^2))$ when $k \leq d^2$.

For our third application of the debiased Keyl's algorithm, we show that their conjectured upper bound is attainable even "without the tilde".

Theorem 1.8 (Tomography with limited entanglement)**.** *Let $n = n' \cdot k$. Let $\widehat{\boldsymbol{\rho}}_1, \ldots, \widehat{\boldsymbol{\rho}}_{n'}$ be the outputs of the debiased Keyl's algorithm when run in n' independent copies of $\rho^{\otimes k}$. Set $\widehat{\rho} = (\widehat{\boldsymbol{\rho}}_1 + \cdots + \widehat{\boldsymbol{\rho}}_{n'})/n'$. Then $\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability when*

$$n = O\left(\max\left(\frac{d^3}{\sqrt{k}\varepsilon^2}, \frac{d^2}{\varepsilon^2}\right)\right).$$

By the lower bound of Chen, Li, and Liu [CLL24b], this bound is optimal, at least for all $k \leq 1/\varepsilon^c$. As they conjecture, we believe that our bound is optimal for all k, and we leave this question for future work.

1.5 Application 4: shadow tomography

Given m observables $O_1, \ldots, O_m \in \mathbb{C}^{d \times d}$, *shadow tomography* refers to the task of producing estimates of the observable values $\mathrm{tr}(O_1 \cdot \rho), \ldots, \mathrm{tr}(O_m \cdot \rho)$ given access to copies of ρ. First introduced by Aaronson in [Aar18], this problem occupies a central place in the field of quantum learning theory. To date, the best algorithm for general bounded observables (satisfying $\|O_i\|_\infty \leq 1$) is due to Bădescu and O'Donnell [BO24], who showed that $n = \widetilde{O}(\log^2(m)\log(d)/\varepsilon^4)$ copies suffice (see also the work of Bene Watts and Bostanci [BWB24], who reproved this bound with a different algorithm). On the flip side, the only known lower bound is due to Aaronson, who showed that $n = \Omega(\log(m)/\varepsilon^2)$ copies are necessary.

One natural approach for the shadow tomography problem is the "plug-in estimator": simply run a full state tomography algorithm to produce an estimate $\widehat{\boldsymbol{\rho}}$ of ρ and "plug it in" to compute the observable values $\widehat{\boldsymbol{o}}_1 \coloneqq \mathrm{tr}(O_1 \cdot \widehat{\boldsymbol{\rho}}), \ldots, \widehat{\boldsymbol{o}}_m \coloneqq \mathrm{tr}(O_m \cdot \widehat{\boldsymbol{\rho}})$. For example, if $\widehat{\boldsymbol{\rho}}$ is the output of Keyl's algorithm on n copies of ρ, then each $\widehat{\boldsymbol{o}}_i$ will be within ε of the true observable value $\mathrm{tr}(O_i \cdot \rho)$ once $n = O(d^2/\varepsilon^2)$, no matter how many observables m there are. This is because with this many samples, Keyl's algorithm produces an estimate $\widehat{\rho}$ with $\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability, and being ε close in trace distance is equivalent to every observable value being correct within ε. However, if we use a smaller number of copies $n \ll d^2/\varepsilon^2$, then each $\widehat{\boldsymbol{o}}_i$ might be far from the corresponding $\mathrm{tr}(O_i \cdot \rho)$, as Keyl's algorithm does not give guarantees in this regime. Keyl's algorithm fails in the "sub-d^2/ε^2 regime" for a more fundamental reason, which is that the estimator $\widehat{\boldsymbol{\rho}}$ it produces is heavily biased away from ρ in this regime, which means that $\mathrm{tr}(O_i \cdot \widehat{\boldsymbol{\rho}})$ will be far from $\mathrm{tr}(O_i \cdot \rho)$ for some observables O_i, at least in expectation. This suggests using an unbiased estimator for ρ in place of Keyl's algorithm; in this case, $\widehat{\boldsymbol{o}}_1, \ldots, \widehat{\boldsymbol{o}}_m$ will be unbiased estimators for $\mathrm{tr}(O_1 \cdot \rho), \ldots, \mathrm{tr}(O_m \cdot \rho)$, respectively, and they may approximate these values with fewer copies of ρ than it takes for $\widehat{\boldsymbol{\rho}}$ to approximate all of ρ. We note that this "plug-in estimator" approach is one method for solving the more general *oblivious shadow tomography* problem, in which one has to perform all of one's measurements on $\rho^{\otimes n}$ prior to learning the observables $O_1, \ldots, O_m$ to be estimated. (The algorithm of Bădescu and O'Donnell, for example, is not oblivious, as the measurements it performs depend on the observables $O_1, \ldots, O_m$.)

Huang, Kueng, and Preskill [HKP20] studied this "plug-in estimator" approach for the case when $\widehat{\boldsymbol{\rho}}$ is the output of the uniform POVM tomography algorithm on n' copies of ρ. In this case, they showed that each $\mathrm{tr}(O_i \cdot \widehat{\boldsymbol{\rho}})$ is an unbiased estimator for $\mathrm{tr}(O_i \cdot \rho)$ with variance $O(\mathrm{tr}(O_i^2)/n') \leq O(F/n')$, where $F = \max\{\mathrm{tr}(O_1^2), \ldots, \mathrm{tr}(O_m^2)\}$. By setting $n' = O(F/\varepsilon^2)$, each $\mathrm{tr}(O_i \cdot \widehat{\boldsymbol{\rho}})$ will have variance $O(\varepsilon^2)$ and will therefore be equal to $\mathrm{tr}(O_i \cdot \rho) \pm O(\varepsilon)$ with high probability. Suppose we then repeat this process k times, generating observable estimates $\widehat{\boldsymbol{o}}_1^1, \ldots, \widehat{\boldsymbol{o}}_m^1$ through $\widehat{\boldsymbol{o}}_1^k, \ldots, \widehat{\boldsymbol{o}}_m^k$. If we then set $\widehat{\boldsymbol{o}}_i = \mathrm{median}\{\widehat{\boldsymbol{o}}_i^1, \ldots, \widehat{\boldsymbol{o}}_i^k\}$ for all $1 \leq i \leq m$, then $\widehat{\boldsymbol{o}}_i$ will be equal to $\mathrm{tr}(O_i \cdot \rho) \pm \varepsilon$ for all $1 \leq i \leq m$ with high probability, once $k = O(\log(m))$. In total, this uses $n = n' \cdot k = O(\log(m)F/\varepsilon^2)$ copies of ρ.[1] Note that if each O_i is bounded, with $\|O_i\|_\infty \leq 1$, then F can be as large as d, in which case this approach requires $n = O(\log(m)d/\varepsilon^2)$ copies of ρ, which is better than the $O(d^2/\varepsilon^2)$ required by Keyl's algorithm but far from the copy complexities of the best algorithms. If, on the other hand, the observables are all rank-1, then $F \leq 1$, and so this algorithm

[1] We note that the step of taking the median of multiple observable values is necessitated because Huang, Kueng, and Preskill also studied a variant of the uniform POVM tomography algorithm where the uniform POVM is replaced by a more computationally efficient measurement based on Clifford unitaries. Indeed, Helsen and Walter [HW23] showed that if $\widehat{\boldsymbol{\rho}}$ is the output of the uniform POVM tomography algorithm on $n = O(\log(m)F/\varepsilon^2)$ copies of ρ, then $\mathrm{tr}(O_i \cdot \widehat{\boldsymbol{\rho}}) = \mathrm{tr}(O_i \cdot \rho) \pm \varepsilon$ for all $1 \leq i \leq m$, no medians required.

only needs $n = O(\log(m)/\varepsilon^2)$ copies of ρ. This is the same number of copies as Aaronson's lower bound from above, although strictly speaking Aaronson's lower bound holds against general and not just rank-1 observables and therefore does not apply. Huang, Kueng, and Preskill called their algorithm the "classical shadows algorithm", and as a result the more broad problem of oblivious shadow tomography is sometimes called the "classical shadows problem".

Grier, Pashayan, and Schaeffer [GPS24] studied this "plug-in estimator" approach for the special case when ρ is a pure state and $\widehat{\boldsymbol{\rho}}$ is the output of the bias-corrected pure state algorithm from Section 1.1.2. Following a similar analysis as [HKP20], they showed that

$$n = O\Big(\log(m) \cdot \Big(\frac{\sqrt{F}}{\varepsilon} + \frac{1}{\varepsilon^2}\Big)\Big)$$

copies of ρ suffice for shadow tomography. This sample complexity bound splits into two regimes, based on whether ε is large or small: when $\varepsilon \geq 1/\sqrt{F}$, their sample complexity scales as $O(\log(m)\sqrt{F}/\varepsilon)$, and when $\varepsilon \leq 1/\sqrt{F}$, their sample complexity scales as $O(\log(m)/\varepsilon^2)$. Interestingly, in the latter regime of small ε, sometimes known as the "high accuracy regime" [CLL24a], their bound matches the lower bound of $n = \Omega(\log(m)/\varepsilon^2)$ copies due to Aaronson, although again here the Aaronson bound does not necessarily apply since his lower bound is shown for general states ρ and not just for pure states. However, their sample complexity *is* optimal for the classical shadows problem, as Huang, Kueng, and Preskill [HKP20] showed that

$$n = \Omega\Big(\log(m) \cdot \Big(\frac{\sqrt{d}}{\varepsilon} + \frac{1}{\varepsilon^2}\Big)\Big)$$

copies are necessary for the classical shadows problem, even in the special case when ρ is pure.

Subsequent work has attempted to generalize the Grier, Pashayan, and Schaeffer result to the case when ρ is a general mixed state. For example, the work of Grier, Liu, and Mahajan [GLM24] gave an algorithm for the classical shadows problem which uses $n = O(\log(m) \cdot \sqrt{rF}/\varepsilon^2)$ copies in the case when ρ is rank r. For small values of r, this improves on the sample complexity of the result of [HKP20] by a factor of $\sqrt{F}$; however, it does not achieve a sample complexity of $O(\log(m)/\varepsilon^2)$ in the high accuracy regime of small ε. On the flip side, Chen, Li, and Liu [CLL24a] focused exclusively on the high accuracy regime and showed that in the case when ρ is allowed to have arbitrary rank, $n = O(\log(m)/\varepsilon^2)$ copies are sufficient for the classical shadows problem, at least in the high accuracy regime of $\varepsilon = O(1/d^{12})$. As their result holds for general states and observables, the $n = \Omega(\log(m)/\varepsilon^2)$ lower bound of Aaronson finally applies, meaning that theirs is the first provably optimal algorithm for the general shadow tomography problem, for any regime of ε. This implies, perhaps surprisingly, that in the high accuracy regime, knowing the observables in advance does not actually help when performing shadow tomography. More broadly, they conjectured that

$$n = \Theta\Big(\log(m) \cdot \Big(\frac{d}{\varepsilon} + \frac{1}{\varepsilon^2}\Big)\Big)$$

should be the optimal sample complexity for the classical shadows problem on general states and observables.

We study the "plug-in estimator" approach in the case when $\widehat{\boldsymbol{\rho}}$ is the output of the debiased Keyl's algorithm. Our main result improves on the sample complexity of [GLM24] in the low accuracy regime, improves the threshold for the high accuracy regime in [CLL24a] from $\varepsilon = O(1/d^{12})$ to $\varepsilon = O(1/d)$, and even refutes the conjecture of [CLL24a] by improving on their conjectured optimal bound.

Theorem 1.9 (Classical shadows)**.** *The classical shadows task for states of rank r and observables $O_1, \ldots, O_m$ with $\operatorname{tr}(O_i^2) \leq F$ can be solved using*

$$n = O\Big(\log(m) \cdot \Big(\min\Big\{\frac{\sqrt{rF}}{\varepsilon}, \frac{F^{2/3}}{\varepsilon^{4/3}}\Big\} + \frac{1}{\varepsilon^2}\Big)\Big)$$

copies of ρ.

To understand our sample complexity bound, let us consider two special cases. In the $r = 1$ case of pure states, $r = 1$, we have that $F^{1/2}/\varepsilon$ is less than $F^{2/3}/\varepsilon^{4/3} = (F^{1/2}/\varepsilon)^{4/3}$ unless $F^{1/2}/\varepsilon < 1$, in which case the

first term is trivial and can be discarded. This means the "min" term is just $\sqrt{F}/\varepsilon$ and the overall sample complexity is

$$n = O\Big(\log(m) \cdot \Big(\frac{\sqrt{F}}{\varepsilon} + \frac{1}{\varepsilon^2}\Big)\Big),$$

matching the bound of Grier, Pashayan, and Schaeffer [GPS24]. On the flip side, in the fully general case of $r = F = d$, the bound becomes

$$n = O\Big(\log(m) \cdot \Big(\min\Big\{\frac{d}{\varepsilon}, \frac{d^{2/3}}{\varepsilon^{4/3}}\Big\} + \frac{1}{\varepsilon^2}\Big)\Big).$$

Here, the "min" term is equal to $d^{2/3}/\varepsilon^{4/3}$ for $\varepsilon \geq 1/d$ and d/ε for $\varepsilon \leq 1/d$. But in the latter case, the $1/\varepsilon^2$ term dominates both of these terms, meaning that we can simplify the sample complexity to be

$$n = O\Big(\log(m) \cdot \Big(\frac{d^{2/3}}{\varepsilon^{4/3}} + \frac{1}{\varepsilon^2}\Big)\Big),$$

which improves on the conjectured upper bound of [CLL24a] in the low accuracy regime. We conjecture that our upper bound is optimal, and leave the task of proving matching lower bounds to future work.

1.6 Application 5: quantum metrology

Quantum metrology is the study of optimal measurements for estimating parameters of quantum systems. A parameterized quantum state is a density matrix ρ_θ which depends smoothly on some parameters $\theta = (\theta_1, \ldots, \theta_m) \in \Theta \subseteq \mathbb{R}^m$. There is a fixed state ρ_{θ^*} whose parameters $\theta^* \in \Theta$ one is assumed to already know. Now, one is given a state ρ_θ, where θ is "in the neighborhood" of θ^*, and the goal is to estimate the parameters θ. This models, for example, a common strategy for learning an unknown parameterized quantum channel Φ_θ. To do this, one might initialize a starting "probe state" ρ_{probe}, send it through the channel Φ_θ, and measure the resulting state ρ_θ to learn θ. The better a measurement one uses, the better an estimate of Φ_θ one can produce. Furthermore, one is assumed to already have prior knowledge that θ is close to some parameter vector θ^*, and one is allowed to use this knowledge to design one's measurements.

To formalize this question, we focus on *locally unbiased estimators*. As we have already seen, an estimator is an algorithm which takes as input a copy of ρ_θ and produces a (randomly distributed) estimate $\widehat{\boldsymbol{\theta}}$ of θ, and it is unbiased if $\mathbf{E}[\widehat{\boldsymbol{\theta}}] = \theta$. Being locally unbiased is a weaker condition: it requires only that

$$\text{(i)}\ \ \mathbf{E}[\widehat{\boldsymbol{\theta}} \mid \rho_{\theta^*}] = \theta^*, \quad \text{and} \quad \text{(ii)}\ \ \frac{\partial}{\partial \theta_i} \mathbf{E}[\widehat{\boldsymbol{\theta}} \mid \rho_\theta]\Big|_{\theta=\theta^*} = 0.$$

The first of these conditions states that $\widehat{\boldsymbol{\theta}}$ is indeed an unbiased estimator, but just on the state ρ_{θ^*}. The second of these conditions states that $\widehat{\boldsymbol{\theta}}$ remains unbiased, at least up to first order, in a small neighborhood around θ^*. Given a locally unbiased estimator, the figure of merit we associate with it is its *mean squared error matrix (MSEM)* $V \in \mathbb{R}^{m \times m}$, defined as

$$V_{ij} = \mathbf{E}[(\widehat{\boldsymbol{\theta}}_i - \theta_i^*) \cdot (\widehat{\boldsymbol{\theta}}_j - \theta_j^*) \mid \rho_{\theta^*}],$$

where $\widehat{\boldsymbol{\theta}}$ is the output of the estimator on the state ρ_{θ^*}.

The most well-known lower bound on V is in terms of the *Quantum Fisher Information (QFI)* matrix $\mathcal{F}$. To define $\mathcal{F}$, we first define the *symmetric logarithm derivative* matrices $L_1, \ldots, L_m$ implicitly via the equations

$$\frac{\partial}{\partial \theta_i} \rho_\theta|_{\theta=\theta^*} = \frac{1}{2} L_i \rho_{\theta^*} + \frac{1}{2} \rho_{\theta^*} L_i, \qquad \text{for all } 1 \leq i \leq m.$$

These equations do not actually uniquely specify $L_1, \ldots, L_m$, but any matrices which satisfy these equations will lead to the same QFI matrix $\mathcal{F}$. Given these, we define

$$\mathcal{F}_{ij} = \frac{1}{2} \operatorname{tr}(\rho_{\theta^*} L_i L_j) + \frac{1}{2} \operatorname{tr}(\rho_{\theta^*} L_j L_i).$$

The *quantum Cramér–Rao bound (QCRB)*, shown by Helstrom in [Hel67], states that

$$V \succeq \mathcal{F}^{-1}.$$

This places a lower bound on the variance of any locally unbiased estimator for θ.

Is the QCRB aqually achievable? The answer is yes, when $m = 1$, given asymptotically many copies of ρ_θ. To explain this, suppose we are given n copies of ρ_θ rather than one. This can be viewed as being given one copy of the parameterized state $\rho_\theta^n \coloneqq \rho_\theta^{\otimes n}$. If $\mathcal{F}$ is the QFI matrix of the ρ_θ's and $\mathcal{F}_n$ is the QFI matrix of the ρ_θ^n's, they are related via the equality $\mathcal{F}_n = n \cdot \mathcal{F}$. Hence, if V_n is the MSEM of a locally unbiased estimator for the ρ_θ^n's, then the QCRB implies that

$$V_n \succeq \mathcal{F}_n^{-1} = \frac{1}{n} \cdot \mathcal{F}^{-1}, \tag{2}$$

or, equivalently, $n \cdot V_n \succeq \mathcal{F}^{-1}$. We say that the QCRB is achievable asymptotically if there is a sequence of such locally unbiased estimators, one for each value n, such that $n \cdot V_n \to \mathcal{F}^{-1}$ as $n \to \infty$. In this case, V_n is equal to $(1/n) \cdot \mathcal{F}^{-1}$, plus other terms of order $o(1/n)$ which, albeit possibly large for small n, disappear as n increases. It was shown by Braunstein and Caves [BC94] that the QCRB is indeed achievable asymptotially in the $m = 1$ case of a single unknown parameter.

When $m > 1$, the problem is more challenging, and there are examples of multiparameter estimation problems for which the QCRB is not achievable, even asymptotically, the key issue being the noncommutativity of different quantum measurements. To simplify the algorithm's task, let us introduce a positive semidefinite cost matrix $C \in \mathbb{R}^{m \times m}$ and assign to the estimator the cost

$$\operatorname{tr}(C \cdot V) = \sum_{i,j=1}^{m} C_{ij} \cdot \mathbf{E}[(\widehat{\boldsymbol{\theta}}_i - \theta_i^*) \cdot (\widehat{\boldsymbol{\theta}}_j - \theta_j^*) \mid \rho_{\theta^*}].$$

Given this cost matrix, the QCRB implies that

$$\operatorname{tr}(C \cdot V) \geq \operatorname{tr}(C \cdot \mathcal{F}^{-1}) \tag{3}$$

for any locally unbiased estimator. Achieving the QCRB in Equation (2) is equivalent to achieving the bound in Equation (3) for all cost matrices C simultaneously; hence, achieving the bound in Equation (3) for just a single cost matrix C is potentially an easier problem, as perhaps one can hand-tailor the locally unbiased estimator for this particular cost matrix. As it turns out, the QCRB with a specific cost matrix C is still not achievable, even asymptotically.

However, there is a slight strengthening of Equation (3) which is achievable asymptotically. In [Hol11], Holevo introduced a quantity $\mathsf{Hol}(C)$ and showed that $\operatorname{tr}(C \cdot V) \geq \mathsf{Hol}(C)$, an inequality known as the *Holevo Cramér–Rao bound (HCRB)*. For our purposes, it is not important what $\mathsf{Hol}(C)$ is exactly, only that it satisfies two properties. The first property is that

$$2 \cdot \operatorname{tr}(C \cdot \mathcal{F}^{-1}) \geq \mathsf{Hol}(C) \geq \operatorname{tr}(C \cdot \mathcal{F}^{-1}),$$

and so the HCRB is only stronger than the QCRB by at most a factor of 2. The first inequality was shown by [ATD19, CSDV19] and is known to be tight in some cases; see [DDGG20, Section 3.1.1] for a simple example. The second property we will need is that if $\mathsf{Hol}_n(C)$ is the Holevo bound corresponding to the n copy states $\rho_\theta^n = \rho_\theta^{\otimes n}$, then, similarly to the QFI matrices, we have $\mathsf{Hol}(C) = n \cdot \mathsf{Hol}_n(C)$. Thus, the HCRB implies that

$$\operatorname{tr}(C \cdot V_n) \geq \mathsf{Hol}_n(C) = \frac{1}{n} \cdot \mathsf{Hol}(n).$$

We say that the HCRB is achievable asymptotically if $n \cdot \operatorname{tr}(C \cdot V_n) \to \mathsf{Hol}(n)$ as $n \to \infty$. A line of work studying quantum local asymptotic normality [KG09, YFG13, YCH19] has developed estimators which do achieve the HCRB asymptotically. This means that in the setting of asymptotically many copies, the HCRB is tight.

This paper is about designing unbiased estimators for full state tomography, and naively it seems like having good unbiased estimators for quantum tomography would help in creating good locally unbiased

estimators for quantum metrology. Motivated by this reasoning, a recent work of Zhou and Chen [ZC25] studied a natural locally unbiased estimator for quantum metrology which first computes the $n = 1$ copy unbiased estimator $\widehat{\boldsymbol{\rho}}$ from Section 1.1.1 and then performs classical postprocessing on the outcome. They showed that in the case when the parameterized states ρ_θ are all rank-one, the MSEM V of their estimator satisfies

$$V = \frac{4(d+1)}{d+2} \cdot \mathcal{F}^{-1} \preceq 4 \cdot \mathcal{F}^{-1}. \tag{4}$$

As $V \succeq \mathcal{F}^{-1}$, this shows that V is within a factor of 4 of being optimal; they referred to such estimators whose MSEM is within a constant factor of $\mathcal{F}^{-1}$ as *near optimal*. Now, it was already known that the HCRB was attainable for rank one states ρ_θ due to the work of Matsumoto [Mat02], even in the non-asymptotic regime where only one copy of ρ_θ is provided. However, Zhou and Chen's estimator has the advantage that because it provides an approximation to $\mathcal{F}^{-1}$ rather than just the $\mathsf{Hol}(C)$, it is oblivious to the cost matrix C, in the sense that the measurement is independent of the cost matrix C (in fact, it is independent of the parameterized state family ρ_θ altogether), and so one can measure first and be provided C later. (This is similar to the sense in which classical shadows algorithms are oblivious.) They also generalized their results to hold in the case when the state family ρ_θ is low rank rather than just rank one, but their guarantees degrade as ρ_{θ^*} becomes more and more poorly conditioned.

What about when more than one copy of ρ_θ is available? As we have better unbiased estimators for larger n, one might hope that we can use these to design better locally unbiased estimators for quantum metrology. To study this, we consider the locally unbiased estimator we get when we replace the unbiased estimator Zhou and Chen use with the n-copy debiased Keyl's algorithm, but keep the classical post-processing the same. We show the following result.

Theorem 1.10 (Quantum metrology)**.** *Let ρ_θ be a parameterized state family with probe state*

$$\rho_{\theta^*} = \sum_{i=1}^{d} \alpha_i \cdot |v_i\rangle\langle v_i| \,,$$

where we assume without loss of generality that the eigenvalues are sorted so that $\alpha_1 \geq \cdots \geq \alpha_r > 0$ and $\alpha_{r+1} = \cdots = \alpha_d = 0$ (so that ρ_{θ^*} is rank r). Then for every n, there is a locally unbiased estimator which takes as input n copies of ρ_θ and whose MSEM V_n has the following property. When $n = r/(\varepsilon \cdot \alpha_r)$,

$$V_n \preceq (1+\varepsilon) \cdot \frac{2}{n} \cdot \mathcal{F}^{-1}.$$

Hence, as $n \to \infty$, we have $n \cdot V_n \to 2 \cdot \mathcal{F}^{-1}$, and so this estimator achieves twice the QCRB asymptotically.

As a result, for asymptotically large n and any cost matrix C, we have that $n \cdot \mathrm{tr}(C \cdot V_n) \to 2 \cdot \mathrm{tr}(C \cdot \mathcal{F}^{-1}) \leq 2 \cdot \mathsf{Hol}(C)$. This implies the surprising fact that, in the asymptotic limit, knowing the cost matrix C ahead of time only buys you at most a factor of 2 in the final cost of the estimator. This is the optimal result of this form that one could hope for. To see this, recall that there are examples in which $\mathsf{Hol}(C) = 2 \cdot \mathrm{tr}(C \cdot \mathcal{F}^{-1})$. In these cases, we have

$$\mathrm{tr}(C \cdot V_n) \geq \mathsf{Hol}_n(C) = \frac{1}{n} \cdot \mathsf{Hol}(C) = \frac{1}{n} \cdot 2 \cdot \mathrm{tr}(C \cdot \mathcal{F}^{-1}).$$

Hence, $n \cdot \mathrm{tr}(C \cdot V_n) \geq 2 \cdot \mathrm{tr}(C \cdot \mathcal{F}^{-1})$ for all n in such cases, and so one cannot have $n \cdot V_n \succeq \beta \cdot \mathcal{F}^{-1}$ for any constant $\beta < 2$. Finally, we note that when ρ_{θ^*} is rank $r = 1$, then $\alpha_1 = 1$, and so if we apply Theorem 1.10 with $n = 1$, and therefore with $\varepsilon = 1$, we get $V_1 \preceq 4 \cdot \mathcal{F}^{-1}$, which recovers Zhou and Chen's bound from Equation (4).

1.7 Technical overview

The key technical challenge that we overcome in this work is understanding how to compute exact expressions for the first and second moments $\mathbf{E}[\widehat{\boldsymbol{\rho}}]$ and $\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}]$ of our representation-theoretic estimators. What makes this difficult is that to do so, one must understand the probability distributions on the Young diagram $\boldsymbol{\lambda}$ and

the unitary $\boldsymbol{U}$ produced by Keyl's measurement, but this distribution involves complicated representation-theoretic formulas which are unwieldy to manipulate. Indeed, although the two works [OW16, HHJ+16] were able to analyze three separate entangled tomography algorithms and show that they are either sample-optimal or almost sample-optimal, neither work was able to derive expressions for even the first moment of their estimators.

1.7.1 The single copy case

To illustrate our techniques, let us consider the simplest possible case, when we are given only a single copy of ρ. In this case, it turns out that Keyl's measurement always returns the Young diagram $\boldsymbol{\lambda} = \square$. In addition, the unitary $\boldsymbol{U}$ it returns has probability density function

$$\mathbf{Pr}[\boldsymbol{U} = U \mid \rho] = d \cdot \langle 1| U^\dagger \rho U |1\rangle \cdot \mathrm{d}U. \tag{5}$$

Since $\boldsymbol{\lambda}^\uparrow = (d, -1, \ldots, -1)$, our debiased Keyl's algorithm will output the estimator

$$\widehat{\boldsymbol{\rho}} = d \cdot |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| - |\boldsymbol{u}_2\rangle\langle\boldsymbol{u}_2| - \cdots - |\boldsymbol{u}_d\rangle\langle\boldsymbol{u}_d|,$$

where $|\boldsymbol{u}_i\rangle \coloneqq \boldsymbol{U} \cdot |i\rangle$. Intuitively, $|\boldsymbol{u}_1\rangle$ should be a rough approximation of the largest eigenvector of ρ, or should at least have decent overlap with ρ's larger eigenvectors. This is because the PDF in Equation (5) scales with $\langle 1| U^\dagger \rho U |1\rangle$, and is therefore maximized for any unitary U such that $U \cdot |1\rangle$ is the largest eigenvector of ρ. As we have already pointed out in Section 1.1.4, this is equivalent to the uniform POVM tomography algorithm from Section 1.1.1.

How to compute the expectation $\mathbf{E}[\widehat{\boldsymbol{\rho}}]$ of this estimator? It turns out that for the uniform POVM tomography algorithm, there is a slick, two-line proof that $\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \rho$ [Wri16, Section 5.1]. Instead, we will describe a significantly more cumbersome version of this argument, as it is this version that we were able to extend to the case of larger n. To begin, we can model a generic tomography algorithm by a POVM $M = \{M_{\widehat{\rho}}\}_{\widehat{\rho}}$, so that the algorithm measures $\rho^{\otimes n}$ with M, receives an outcome $\widehat{\boldsymbol{\rho}}$, and then outputs $\widehat{\boldsymbol{\rho}}$ as its estimator. Then the expectation can be written as

$$\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \sum_{\widehat{\rho}} \mathrm{tr}(M_{\widehat{\rho}} \cdot \rho^{\otimes n}) \cdot \widehat{\rho} = \mathrm{tr}_{[n]}\left(\left(\sum_{\widehat{\rho}} M_{\widehat{\rho}} \otimes \widehat{\rho}\right) \cdot \rho^{\otimes n} \otimes I\right) =: \mathrm{tr}_{[n]}(M_{\mathrm{avg}} \cdot \rho^{\otimes n} \otimes I).$$

It therefore suffices to compute a formula for the matrix M_{avg}. When $n = 1$ and the algorithm is the debiased Keyl's algorithm, this matrix is

$$M_{\mathrm{avg}} = \int_U \left(d \cdot U |1\rangle\langle 1| U^\dagger \cdot \mathrm{d}U\right) \otimes \sum_{i=1}^d \lambda_i^\uparrow \cdot U |i\rangle\langle i| U^\dagger = d \cdot \sum_{i=1}^d \lambda_i^\uparrow \cdot \int_U U^{\otimes 2} |1,i\rangle\langle 1,i| U^{\dagger,\otimes 2} \cdot \mathrm{d}U.$$

The symmetric and antisymmetric subspaces. As a first step, let us compute a formula for the integral

$$\int_U U^{\otimes 2} |1,i\rangle\langle 1,i| U^{\dagger,\otimes 2} \cdot \mathrm{d}U. \tag{6}$$

This is relatively simple in the case when $i = 1$, because the vector $|1,1\rangle$ is an element of the *symmetric subspace*

$$V_{\square\square} \coloneqq \{|\psi\rangle \in (\mathbb{C}^d)^{\otimes 2} \mid \mathrm{SWAP} \cdot |\psi\rangle = |\psi\rangle\} = \mathrm{span}\{|v\rangle \otimes |v\rangle \mid |v\rangle \in \mathbb{C}^d\}.$$

The symmetric subspace is an irreducible representation of the unitary group, which implies that for any vector $|\psi_{\square\square}\rangle \in V_{\square\square}$,

$$\int_U U^{\otimes 2} |\psi_{\square\square}\rangle\langle\psi_{\square\square}| U^{\dagger,\otimes 2} \cdot \mathrm{d}U = \frac{\Pi_{\square\square}}{\dim(V_{\square\square})}, \tag{7}$$

where $\Pi_{\square\square}$ is the projector onto $V_{\square\square}$. In particular, this applies for $|\psi_{\square\square}\rangle = |1,1\rangle$, which gives us a formula for Equation (6) when $i = 1$.

When $i > 1$, however, $|1,i\rangle\langle 1,i|$ is not entirely contained inside the symmetric subspace. Instead, it is split between the symmetric subspace and the *antisymmetric subspace*, defined as

$$V_{\square\!\!\square} \coloneqq \{|\psi\rangle \in (\mathbb{C}^d)^{\otimes 2} \mid \mathrm{SWAP} \cdot |\psi\rangle = -|\psi\rangle\}.$$

Like the symmetric subspace, the antisymmetric subspace is an irreducible representation of the unitary group, which implies that for any vector $|\psi_{\square\!\square}\rangle \in V_{\square\!\square}$,

$$\int_U U^{\otimes 2} |\psi_{\square\!\square}\rangle\langle\psi_{\square\!\square}| U^{\dagger,\otimes 2} \cdot \mathrm{d}U = \frac{\Pi_{\square\!\square}}{\dim(V_{\square\!\square})}, \tag{8}$$

where $\Pi_{\square\!\square}$ is the projector onto $V_{\square\!\square}$. In addition, because the antisymmetric subspace is non-isomorphic to the symmetric subspace, a result in representation theory known as Schur's lemma states that for any vectors $|\psi_{\square\square}\rangle \in V_{\square\square}$ and $|\psi_{\square\!\square}\rangle \in V_{\square\!\square}$,

$$\int_U U^{\otimes 2} |\psi_{\square\square}\rangle\langle\psi_{\square\!\square}| U^{\dagger,\otimes 2} \cdot \mathrm{d}U = 0. \tag{9}$$

Hence, if we could simply split $|1,i\rangle$ into its symmetric and antisymmetric components, we could use these formulas to evaluate Equation (6).

To this end, let us define the vectors

$$|\begin{smallmatrix}\boxed{i}\boxed{i}\end{smallmatrix}\rangle \coloneqq |i,i\rangle\,, \qquad |\begin{smallmatrix}\boxed{i}\boxed{j}\end{smallmatrix}\rangle \coloneqq \frac{1}{\sqrt{2}}|i,j\rangle + \frac{1}{\sqrt{2}}|j,i\rangle\,, \quad \text{and} \quad \left|\begin{smallmatrix}\boxed{i}\\\boxed{j}\end{smallmatrix}\right\rangle \coloneqq \frac{1}{\sqrt{2}}|i,j\rangle - \frac{1}{\sqrt{2}}|j,i\rangle\,, \text{ for } i<j.$$

Note that the $|\begin{smallmatrix}\boxed{i}\boxed{j}\end{smallmatrix}\rangle$ vectors are elements of the symmetric subspace, and the $\left|\begin{smallmatrix}\boxed{i}\\\boxed{j}\end{smallmatrix}\right\rangle$ vectors are elements of the antisymmetric subspace. In addition, they are orthogonal to each other by inspection, and in fact they actually form an orthonormal basis for all of $(\mathbb{C}^d)^{\otimes 2}$, as we can write the standard basis vectors as linear combinations of these vectors:

$$|i,i\rangle = |\begin{smallmatrix}\boxed{i}\boxed{i}\end{smallmatrix}\rangle\,, \qquad |i,j\rangle = \frac{1}{\sqrt{2}}|\begin{smallmatrix}\boxed{i}\boxed{j}\end{smallmatrix}\rangle + \frac{1}{\sqrt{2}}\left|\begin{smallmatrix}\boxed{i}\\\boxed{j}\end{smallmatrix}\right\rangle, \quad \text{and} \quad |j,i\rangle = \frac{1}{\sqrt{2}}|\begin{smallmatrix}\boxed{i}\boxed{j}\end{smallmatrix}\rangle - \frac{1}{\sqrt{2}}\left|\begin{smallmatrix}\boxed{i}\\\boxed{j}\end{smallmatrix}\right\rangle, \text{ for } i<j. \tag{10}$$

From this fact, we can derive several useful consequences.

1. Because these vectors form an orthonormal basis for all of $(\mathbb{C}^d)^{\otimes 2}$, they also form orthonormal bases for their respective subspaces. Thus, we can conclude that $\dim(V_{\square\square}) = d(d+1)/2$ and $\dim(V_{\square\!\square}) = d(d-1)/2$.

2. Similarly, this means that the symmetric and antisymmetric subspaces together span all of $(\mathbb{C}^d)^{\otimes 2}$. In other words, $(\mathbb{C}^d)^{\otimes 2} = V_{\square\square} \oplus V_{\square\!\square}$.

3. Finally, since these vectors form an orthonormal basis for all of $(\mathbb{C}^d)^{\otimes 2}$, they also form a complete eigenbasis for SWAP, with eigenvalues $+1$ or -1. More succinctly, $\mathrm{SWAP} = \Pi_{\square\square} - \Pi_{\square\!\square}$.

We can now use these vectors to compute our desired expectation. In particular, when $i > 1$,

$$\begin{aligned}\int_U U^{\otimes 2} |1,i\rangle\langle 1,i| U^{\dagger,\otimes 2} \cdot \mathrm{d}U = \frac{1}{2}\int_U U^{\otimes 2} |\begin{smallmatrix}\boxed{1}\boxed{i}\end{smallmatrix}\rangle\langle\begin{smallmatrix}\boxed{1}\boxed{i}\end{smallmatrix}| U^{\dagger,\otimes 2} \cdot \mathrm{d}U + \frac{1}{2}\int_U U^{\otimes 2} \left|\begin{smallmatrix}\boxed{1}\\\boxed{i}\end{smallmatrix}\right\rangle\left\langle\begin{smallmatrix}\boxed{1}\\\boxed{i}\end{smallmatrix}\right| U^{\dagger,\otimes 2} \cdot \mathrm{d}U \\ + \frac{1}{2}\int_U U^{\otimes 2} \left|\begin{smallmatrix}\boxed{1}\boxed{i}\end{smallmatrix}\right\rangle\left\langle\begin{smallmatrix}\boxed{1}\\\boxed{i}\end{smallmatrix}\right| U^{\dagger,\otimes 2} \cdot \mathrm{d}U + \frac{1}{2}\int_U U^{\otimes 2} \left|\begin{smallmatrix}\boxed{1}\\\boxed{i}\end{smallmatrix}\right\rangle\left\langle\begin{smallmatrix}\boxed{1}\boxed{i}\end{smallmatrix}\right| U^{\dagger,\otimes 2} \cdot \mathrm{d}U.\end{aligned}$$

This we can analyze using our three integration formulas Equations (7) to (9), which tell us that for $i > 1$,

$$\int_U U^{\otimes 2} |1,i\rangle\langle 1,i| U^{\dagger,\otimes 2} \cdot \mathrm{d}U = \frac{1}{2}\cdot\frac{\Pi_{\square\square}}{\dim(V_{\square\square})} + \frac{1}{2}\cdot\frac{\Pi_{\square\!\square}}{\dim(V_{\square\!\square})}.$$

Computing M_{avg}. Using our integration formulas, we can compute M_{avg} as

$$\begin{aligned}
M_{\text{avg}} &= d \cdot \sum_{i=1}^{d} \lambda_i^{\uparrow} \cdot \int_U U^{\otimes 2} \, |1,i\rangle\langle 1,i| \, U^{\dagger,\otimes 2} \cdot \mathrm{d}U \\
&= d \cdot \left(d \cdot \frac{\Pi_{\square\square}}{\dim(V_{\square\square})} + (d-1) \cdot (-1) \cdot \left(\frac{1}{2} \cdot \frac{\Pi_{\square\square}}{\dim(V_{\square\square})} + \frac{1}{2} \cdot \frac{\Pi_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}}{\dim(V_{\begin{smallmatrix}\square\\\square\end{smallmatrix}})} \right) \right) \\
&= \frac{d(d+1)}{2} \cdot \frac{\Pi_{\square\square}}{\dim(V_{\square\square})} - \frac{d(d-1)}{2} \cdot \frac{\Pi_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}}{\dim(V_{\begin{smallmatrix}\square\\\square\end{smallmatrix}})} \\
&= \Pi_{\square\square} - \Pi_{\begin{smallmatrix}\square\\\square\end{smallmatrix}} \\
&= \mathrm{SWAP}.
\end{aligned}$$

Now that we have computed M_{avg}, we can conclude our calculation of $\widehat{\boldsymbol{\rho}}$'s first moment as follows:

$$\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \mathrm{tr}_1(M_{\text{avg}} \cdot \rho \otimes I) = \mathrm{tr}_1(\mathrm{SWAP} \cdot \rho \otimes I) = \rho,$$

where we have here used the fact that $\mathrm{tr}_1(\mathrm{SWAP} \cdot A \otimes I) = A$ for any matrix $A \in \mathbb{C}^{d \times d}$. This shows that $\widehat{\boldsymbol{\rho}}$ is an unbiased estimator for ρ and therefore completes the proof.

1.7.2 The two copy case

Exploiting permutation symmetry. Next, let us illustrate how these ideas generalize to the case where our input is two copies $\rho \otimes \rho$. Here, there is a new feature not present in the single copy case, in that the input has "permutation symmetry", meaning that it is invariant under swapping the two registers. Mathematically, this can be expressed through the equation

$$\mathrm{SWAP} \cdot \rho \otimes \rho \cdot \mathrm{SWAP} = \rho \otimes \rho.$$

One consequence of this is that $\rho \otimes \rho$ commutes with SWAP, i.e. that $\rho \otimes \rho \cdot \mathrm{SWAP} = \mathrm{SWAP} \cdot \rho \otimes \rho$, and so $\rho \otimes \rho$ and SWAP share a common eigenbasis. As we have seen, SWAP has two eigenspaces, a $+1$ eigenspace corresponding to the subspace $V_{\square\square}$ and a -1 eigenspace corresponding to $V_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}$, and so $\rho \otimes \rho$'s eigenvectors also split into those that live in $V_{\square\square}$ and those that live in $V_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}$. This means that $\rho \otimes \rho$ can be written as $\rho \otimes \rho = \rho_{\square\square} + \rho_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}$, where $\rho_{\square\square}$ is a subnormalized mixed state entirely supported on $V_{\square\square}$ and $\rho_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}$ is a subnormalized mixed state entirely supported on $V_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}$.

Inside the symmetric subspace. In the case of two copies of ρ, the debiased Keyl's algorithm begins by measuring $\rho \otimes \rho$ with the projective measurement $\{\Pi_{\square\square}, \Pi_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}\}$. Let us first consider the case when the outcome $\boldsymbol{\lambda}$ is equal to $\boldsymbol{\lambda} = \square\square$. This case occurs with probability $\mathrm{tr}(\Pi_{\square\square} \cdot \rho^{\otimes 2}) = \mathrm{tr}(\rho_{\square\square})$, and when it occurs the state collapses to $\rho_{\square\square} / \mathrm{tr}(\rho_{\square\square})$. Recalling that $\boldsymbol{\lambda}^{\uparrow} = (d+2, -1, \ldots, -1)$, the debiased Keyl's algorithm will output an estimator $\widehat{\boldsymbol{\rho}}$ with eigenvalues $\boldsymbol{\lambda}^{\uparrow}/2$. But what are its eigenvectors? To learn these, we must perform a further measurement on the collapsed state $\rho_{\square\square} / \mathrm{tr}(\rho_{\square\square})$, which lies within $V_{\square\square}$. The debiased Keyl's algorithm uses the POVM

$$\{\dim(V_{\square\square}) \cdot U^{\otimes 2} \, |\boxed{1}\boxed{1}\rangle\langle\boxed{1}\boxed{1}| \, U^{\dagger,\otimes 2} \cdot \mathrm{d}U\}.$$

This is indeed a measurement on the space $V_{\square\square}$, as

$$\dim(V_{\square\square}) \cdot \int_U U^{\otimes 2} \, |\boxed{1}\boxed{1}\rangle\langle\boxed{1}\boxed{1}| \, U^{\dagger,\otimes 2} \cdot \mathrm{d}U = \Pi_{\square\square}$$

due to Equation (7). Letting $\boldsymbol{U}$ be the result of this measurement, the debiased Keyl's algorithm outputs the estimator

$$\widehat{\boldsymbol{\rho}} = \sum_{i=1}^{d} \boldsymbol{\lambda}_i^{\uparrow} \cdot |\boldsymbol{u}_i\rangle\langle\boldsymbol{u}_i| = \frac{(d+2)}{2} \cdot |\boldsymbol{u}_1\rangle\langle\boldsymbol{u}_1| - \frac{1}{2} \cdot |\boldsymbol{u}_2\rangle\langle\boldsymbol{u}_2| - \cdots - |\boldsymbol{u}_d\rangle\langle\boldsymbol{u}_d|,$$

where $|\boldsymbol{u}_i\rangle = \boldsymbol{U} \cdot |i\rangle$ for all $1 \le i \le d$. As in the single copy case, $|\boldsymbol{u}_1\rangle$ should be a rough approximation of the largest eigenvector of ρ, or should at least have decent overlap with ρ's larger eigenvectors, because the probability density function of $\boldsymbol{U}$ is given by

$$\begin{aligned}\mathbf{Pr}[\boldsymbol{U} = U \mid \rho, \square\square] &= \operatorname{tr}\left(\left(\dim(V_{\square\square}) \cdot U^{\otimes 2} \left|\boxed{1}\boxed{1}\right\rangle\!\left\langle\boxed{1}\boxed{1}\right| U^{\dagger,\otimes 2} \cdot \mathrm{d}U\right) \cdot \frac{\rho_{\square\square}}{\operatorname{tr}(\rho_{\square\square})}\right) \\ &= \frac{\dim(V_{\square\square})}{\operatorname{tr}(\rho_{\square\square})} \cdot \left\langle\boxed{1}\boxed{1}\right| U^{\dagger,\otimes 2} \rho_{\square\square} U^{\otimes 2} \left|\boxed{1}\boxed{1}\right\rangle \cdot \mathrm{d}U \\ &= \frac{\dim(V_{\square\square})}{\operatorname{tr}(\rho_{\square\square})} \cdot \left(\langle 1| U^\dagger \rho U |1\rangle\right)^2 \cdot \mathrm{d}U.\end{aligned}$$

In fact, this PDF scales not just with $\langle 1| U^\dagger \rho U |1\rangle$ but with its *square*, meaning that $|\boldsymbol{u}_1\rangle$ should be even better concentrated around ρ's highest eigenvectors than it was in the single copy case.

Highest weight vectors. If, instead, the measurement outcome is $\boldsymbol{\lambda} = \begin{smallmatrix}\square\\\square\end{smallmatrix}$, then the state collapses to $\rho_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}/\operatorname{tr}(\rho_{\begin{smallmatrix}\square\\\square\end{smallmatrix}})$. The debiased Keyl's algorithm then performs within the $V_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}$ space the POVM

$$\left\{\dim(V_{\begin{smallmatrix}\square\\\square\end{smallmatrix}}) \cdot U^{\otimes 2} \left|\boxed{\begin{smallmatrix}1\\2\end{smallmatrix}}\right\rangle\!\left\langle\boxed{\begin{smallmatrix}1\\2\end{smallmatrix}}\right| U^{\dagger,\otimes 2} \cdot \mathrm{d}U\right\},$$

which is indeed a measurement due to Equation (8). As in the single copy case and the $\boldsymbol{\lambda} = \square\square$ case, this measurement is biased towards U's for which $U \cdot |1\rangle$ and $U \cdot |2\rangle$ have good overlap with ρ's largest eigenvectors. In both the one- and two-copy cases, our measurements are chosen fixing a particular vector,

$$\text{either } |1\rangle,\ |\boxed{1}\boxed{1}\rangle, \text{ or } \left|\boxed{\begin{smallmatrix}1\\2\end{smallmatrix}}\right\rangle,$$

and considering all possible rotations of this vector by a unitary matrix U. (Note that one can unify the notation by observing that in the $n = 1$ case, the input space $\mathbb{C}^d$ is also known as the irreducible representation $V_\square$, and within this representation $|1\rangle$ is written as the vector $|\boxed{1}\rangle$.) These vectors are chosen to have the "most 1's" possible within their subspace, then the "most 2's" possible, and so forth, in order to bias the U's towards ρ's largest eigenvectors, as described above. Vectors of this form are known in representation theory as *highest weight vectors*; as a result, this measurement, which was first introduced by Keyl [Key06] in his tomography algorithm, is sometimes referred to as the "rotated highest weight measurement".

Computing M_{avg}. Computing M_{avg} in the two copy case splits into two expressions, one for the symmetric subspace and another for the antisymmetric subspace. Writing $\lambda = (2, 0, \ldots, 0)$, the expression for the symmetric subspace is

$$\begin{aligned}&\int_U \left(\dim(V_{\square\square}) \cdot U^{\otimes 2} \left|\boxed{1}\boxed{1}\right\rangle\!\left\langle\boxed{1}\boxed{1}\right| U^{\dagger,\otimes 2} \cdot \mathrm{d}U\right) \otimes \sum_{i=1}^d \lambda_i^\uparrow \cdot U |i\rangle\langle i| U^\dagger \\ = \dim(V_{\square\square}) &\cdot \sum_{i=1}^d \lambda_i^\uparrow \cdot \int_U U^{\otimes 3} \left|\boxed{1}\boxed{1}\right\rangle\!\left\langle\boxed{1}\boxed{1}\right| \otimes |i\rangle\langle i| U^{\dagger,\otimes 3} \cdot \mathrm{d}U.\end{aligned}$$

Computing this involves computing the integrals

$$\int_U U^{\otimes 3} \left|\boxed{1}\boxed{1}\right\rangle\!\left\langle\boxed{1}\boxed{1}\right| \otimes |i\rangle\langle i| U^{\dagger,\otimes 3} \cdot \mathrm{d}U.$$

As in the $n = 1$ case, this not easy to directly compute because $|\boxed{1}\boxed{1}\rangle \otimes |i\rangle$ is not contained inside an irreducible representation of the unitary group over $(\mathbb{C}^d)^{\otimes 3}$. Instead, we must first split this vector into a component for each irreducible representation. As it turns out, the irreducible representations of the unitary group on $(\mathbb{C}^d)^{\otimes 3}$ are labeled by Young diagrams λ with 3 boxes, and in particular

$$|\boxed{1}\boxed{1}\rangle \otimes |1\rangle = |\boxed{1}\boxed{1}\boxed{1}\rangle, \quad \text{and} \quad |\boxed{1}\boxed{1}\rangle \otimes |i\rangle = \sqrt{\frac{1}{3}} \cdot |\boxed{1}\boxed{1}\boxed{i}\rangle + \sqrt{\frac{2}{3}} \cdot \left|\boxed{\begin{smallmatrix}1&1\\i\end{smallmatrix}}\right\rangle, \text{ for } i > 1, \tag{11}$$

where the vectors in the decomposition are elements of the irreducible representations corresponding to $\square\square\square$ and $\begin{smallmatrix}\square\square\\\square\end{smallmatrix}$, respectively. This process, in which we take a vector in an irreducible representation of the unitary group corresponding, tensor on a $|i\rangle$ vector, and then decompose that into vectors which live in separate irreducible representations, is known as a *Clebsch-Gordan transform*, and the coefficients that arise in this process (for example, the coefficients in Equation (11), as well as the coefficients in Equation (10)) are known as *Clebsch-Gordan coefficients*. Computing this integral, and therefore computing M_{avg}, requires being able to understand and manipulate these Clebsch-Gordan coefficients and evaluate large expressions involving them. Eventually, we aim to show that $M_{\text{avg}} = \frac{1}{2} \cdot \text{SWAP}_{1,3} + \frac{1}{2} \cdot \text{SWAP}_{2,3}$. If this is true, then

$$\begin{aligned}
\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \text{tr}_{[2]}(M_{\text{avg}} \cdot \rho \otimes \rho \otimes I) &= \text{tr}_{[2]}\left(\left(\frac{1}{2} \cdot \text{SWAP}_{1,3} + \frac{1}{2} \cdot \text{SWAP}_{2,3}\right) \cdot \rho \otimes \rho \otimes I\right) \\
&= \frac{1}{2} \cdot \text{tr}_{[2]}(\text{SWAP}_{1,3} \cdot \rho \otimes \rho \otimes I) + \frac{1}{2} \cdot \text{tr}_{[2]}(\text{SWAP}_{2,3} \cdot \rho \otimes \rho \otimes I) \\
&= \frac{1}{2} \cdot \text{tr}_{[1]}(\text{SWAP} \cdot \rho \otimes I) + \frac{1}{2} \cdot \text{tr}_{[1]}(\text{SWAP} \cdot \rho \otimes I) \\
&= \rho,
\end{aligned}$$

proving that $\widehat{\boldsymbol{\rho}}$ is an unbiased estimator for ρ.

1.7.3 The general copy case

The case of general n closely resembles the case of $n = 2$ copies. First, our goal is to show that

$$M_{\text{avg}} = \frac{1}{n} \cdot (\text{SWAP}_{1,n+1} + \text{SWAP}_{2,n+1} + \cdots + \text{SWAP}_{n,n+1}), \tag{12}$$

which implies that the debiased Keyl's algorithm is indeed an unbiased estimator. (As an aside, we note that the right-hand side of this expression is a well-known element of the group algebra $\mathbb{C}[S_{n+1}]$ called the *Jucys-Murphy element*.) Computing M_{avg} then breaks into a separate sub-expression for each irreducible representation of the unitary group corresponding to a Young diagram λ with n boxes. For each such λ, computing this sub-expression involves taking the highest weight vector T^λ of λ and applying the Clebsch-Gordan transform to vectors of the form $\left|T^\lambda\right\rangle \otimes |i\rangle$ in order to understand how these vectors decompose into irreducible representations corresponding to Young diagrams with $n+1$ boxes. Having done this, proving Equation (12) then involves deriving exact formulas for certain complicated expressions involving Clebsch-Gordan coefficients, which is made all the more difficult since for general n even individual Clebsch-Gordan coefficients can be quite complicated themselves. We also then have to compute the *second* moment of our estimator in order to prove Theorem 1.4, and this involves computing a formula for the matrix

$$M_{\text{avg}}^{(2)} = \sum_{\widehat{\rho}} M_{\widehat{\rho}} \otimes \widehat{\rho} \otimes \widehat{\rho}.$$

This is the most technically challenging part of our work, as it involves understanding the effect that two back-to-back Clebsch-Gordan transforms have on a highest weight vector $\left|T^\lambda\right\rangle$.

In recent years, Clebsch-Gordan coefficients and the Clebsch-Gordan transform have become increasingly central to the study of quantum computing and information. Perhaps their first appearance in the quantum computing literature dates back to the efficient algorithm for the quantum Schur transform by Bacon, Chuang, and Harrow from 2005 [BCH05], but more recently they have been used to compute the dual and mixed Schur transforms [Ngu24, GBO24], perform optimal qudit purification [LFIC24], efficiently implement port-based teleportation [GBO24], take a quantum majority vote [BLM+23], and prove lower bounds on the number of queries needed to invert a unitary [CYZ25]. To our knowledge, our work is the first to apply the Clebsch-Gordan transform in the domain of quantum learning theory, though we expect that they will see further use in this domain in the future. Our primary technical contribution, then, is to develop a host of new tools for evaluating natural expressions that arise when studying Clebsch-Gordan coefficients, and we hope that our tools will find further applications in future works which make use of the Clebsch-Gordan transform.

1.8 Open problems

Our work leaves open a number of interesting follow-up questions, which we list below.

1. **(Learning with very high probability)** In this work, we have studied algorithms for quantum tomography which succeed with a high constant probability, say 99%. But what if we want to succeed with probability $1 - \delta$, where $\delta > 0$ is a parameter which is allowed to vary? As we show in Section 2.2 below, an immediate corollary of our results is that $n = O(d^2/\varepsilon^2 \cdot \log(1/\delta))$ copies suffice to estimate a mixed state $\rho \in \mathbb{C}^{d \times d}$ to error ε with success probability $1 - \delta$ in both trace and Bures distances. (In fact, in the case of trace distance, this also follows from the work of O'Donnell and Wright [OW16].) However, it is natural to conjecture that a stronger bound of

$$n = O\Big(\frac{d^2}{\varepsilon^2} + \frac{\log(1/\delta)}{\varepsilon^2}\Big)$$

copies should hold for both distance measures. A similar bound is known to hold in the classical setting of learning discrete distributions: in particular, $n = O(d/\varepsilon^2 + \log(1/\delta)/\varepsilon^2)$ samples are sufficient to learn a distribution $p = (p_1, \ldots, p_d)$ up to error ε with probability $1 - \delta$, in both trace and Hellinger distances [Can20]. We note that a bound of this form, with some additional logarithmic factors, can be derived from [HHJ+16, Equation 14], and so the challenge here is to achieve this bound precisely, with no additional logarithms. One natural route for accomplishing this is to study higher moments of our estimator, generalizing Theorem 1.4 beyond the second moment.

   Let us note that this question also appears to be open in the case of tomography algorithms which use independent measurements, even for learning in trace distance. It follows from [GKKT20, Theorem 2] that $n = O(d^3/\varepsilon^2 + d^2 \log(1/\delta)/\varepsilon^2)$ copies are sufficient to learn in trace distance with independent measurements (in fact, the uniform POVM tomography algorithm achieves this bound); on the other hand, it appears to follow from the argument of [CHL+23] that $n = \Omega(d^3/\varepsilon^2 + d \log(1/\delta)/\varepsilon^2)$ copies are necessary in this setting [CL24]. We believe that closing this gap, and determining the bound for Bures distance, is also an interesting open problem.

2. **(Spectrum estimation)** How many copies do we need to learn a mixed state ρ's eigenvalues $\alpha = (\alpha_1, \ldots, \alpha_d)$? This problem is certainly no harder than full state tomography, which requires learning ρ's eigenvalues *and* eigenvectors, and so $n = O(d^2/\varepsilon^2)$ samples, which suffice for tomography, also suffice for spectrum estimation. But can we do better? Or does one first have to learn the eigenvectors in order to then learn the eigenvalues? Currently, the best known spectrum estimation algorithm is the EYD algorithm, and although it is known to only require $n = O(d^2/\varepsilon^2)$ copies [OW16], it is also known to fail at spectrum estimation when $n = o(d^2/\varepsilon^2)$ [OW15b]. On the flip side, in the classical setting, it is known that $n = O(d/\log(d))$ samples are necessary and sufficient to learn the set of probability values $\{p_1, \ldots, p_d\}$ of a probability distribution $p = (p_1, \ldots, p_d)$ [VV11, VV17, HJW18] (which is the classical analogue of a mixed state's spectrum), which does outperform the $n = \Omega(d)$ samples needed to learn p itself.

   Whether we can do better than full state tomography was recently resolved in the setting of algorithms which use unentangled measurements by Pelecanos, Tan, Tang, and Wright [PTTW25], who showed that $n = o(d^3)$ samples suffice for spectrum estimation, which beats the $n = \Omega(d^3)$ copies which are required for full state tomography with unentangled measurements [CHL+23]. Their argument crucially relied on the fact that we have guarantees for full state tomography with unentangled measurements in the "very high probability" regime. As a result, we believe that a resolution of Item 1 above is likely to help design improved algorithms for spectrum estimation with entangled measurements.

3. **(Other distance measures)** For the first time, our work shows optimal bounds for learning a mixed state in Bures distance, which is a more stringent distance metric than trace distance. There are even more stringent divergences than Bures which are commonly studied in the quantum information literature, namely the quantum relative entropy and the Bures χ^2-divergence. How many copies are needed to learn a state ρ for either of these two divergences? This question was studied by Flammia and O'Donnell [FO24], who were motivated by the application of testing whether a bipartite quantum state has zero quantum mutual information. In fact, combining our Theorem 1.6 with their Theorem 2.34

produces a tomography algorithm on rank r states $\rho \in \mathbb{C}^{d \times d}$ which learns in quantum relative distance error ε using only $n = O(rd/\varepsilon) \cdot \log(d/\varepsilon)$ copies, improving on the bound in their Corollary 1.8 by a factor of $\log(d/\varepsilon)$ (see their Page 12 for a computation of their precise bound). But the best lower bound we know for this problem is $n = \Omega(rd/\varepsilon)$ (which follows from the Bures distance lower bound of [Yue23]), suggesting that $n = O(rd/\varepsilon)$ might be the optimal upper bound.

4. **(Efficient algorithms for Keyl's measurement)** Although the algorithms we consider in this work are sample-efficient, they are not yet known to be computationally efficient. The one exception is when $n = 1$, as it is known that the uniform POVM tomography algorithm still works if one substitutes the Haar random unitary used to specify the measurement basis with a random unitary chosen from a unitary 2-design or 3-design, depending on one's application [GKKT20, HKP20]. This yields an efficient algorithm, as there are examples of 2-designs and 3-designs, such as Clifford unitaries, which are computationally efficient to sample from. Can Keyl's measurement be made efficient for general n?

5. **(Optimality of the debiased Keyl's algorithm)** Is the debiased Keyl's algorithm *the* optimal entangled tomography algorithm? This is of course not a well-defined question, but perhaps one way of making it formal is asking whether the debiased Keyl's algorithm is the minimum variance estimator among all unbiased estimators, where the variance is defined as $\mathbf{E} \|\widehat{\boldsymbol{\rho}} - \rho\|_2^2$. We show in Theorem 3.9 below that the answer is yes, at least for a large family of estimators based on Keyl's measurement, but this doesn't rule out an even better unbiased estimator which uses a different measurement. A good starting point for proving this, if it is indeed true, is the $n = 1$ case and the general n pure state case.

6. **(Simpler proofs of our bounds)** Can our proofs be made simpler? As we mention in Section 1.7.1, there is a substantially simpler proof than the one we use that the debiased Keyl's algorithm is unbiased in the $n = 1$ case; could this proof be generalized to larger n? More broadly, is there a more conceptual reason why our algorithm gives an unbiased estimator, beyond just what we learn from working out the Clebsch-Gordan coefficients? Possibly even one that doesn't involve representation theory? There have been several instances in quantum learning in which results which were first derived using heavy representation theory were then given new proofs which are substantially cleaner and more conceptual (e.g. the new proof of the mixedness testing upper bound from O'Donnell and Wright [OW15b] due to Badescu, O'Donnell, and Wright [BOW19], and the new proof of the mixedness testing *lower* bound from [OW15b] due to O'Donnell and Wadhwa [OW25]). It would be interesting if the same could be done here.

1.9 Paper organization

This paper is organized into five parts.

- Part I contains this introduction, as well as preliminaries and a full description of the debiased Keyl's algorithm. The preliminaries contain general background material on quantum learning theory, as well as an overview of all the representation theory needed to state the debiased Keyl's algorithm.

- Part II contains all five of our applications. This part uses Theorems 1.3 and 1.4, our main results concerning the first and second moments of the debiased Keyl's algorithm, as a black box. With the exception of the tomography with limited entanglement application, which relies on some lemmas from the trace distance tomography section, the applications are meant to be self-contained and readable in any order.

- Part III contains an introduction to the Clebsch-Gordan transform and the Clebsch-Gordan coefficients.

- Part IV uses the Clebsch-Gordan technology developed in the previous part to prove Theorems 1.3 and 1.4, our first and second moment bounds.

- Part V is an appendix which contains a technical result that we make use of in our first and second moment proofs. Informally, it shows that the algorithm for the Schur transform due to Bacon, Chuang, and Harrow [BCH05] also happens to compute a specific choice of basis for the irreducible representations of the symmetric group known as *Young's orthogonal basis*. Previously, it was known only that it

computes a *Young-Yamanouchi basis*, a more general class of bases of which Young's orthogonal basis is a specific and especially convenient example.

Acknowledgments. A.P. is supported by DARPA under Agreement No. HR00112020023. J.S. and J.W. are supported by the NSF CAREER award CCF-233971.

2 Preliminaries

We use **boldface** to denote random variables, and define $[d] = \{1, \ldots, d\}$. Additionally, we will use δ_{ij} for the Kronecker delta, that is, $\delta_{ij} \coloneqq \mathbb{1}[i = j]$.

2.1 Quantum distances

We first review the quantum distances (and divergences) that we will use throughout this paper. The interested reader is encouraged to refer to [BOW19] for a systematic treatment of these quantum distances and their relationships.

Definition 2.1 (Schatten p-norm)**.** Let $M \in \mathbb{C}^{d\times d}$ be a matrix with singular values $\lambda_1, \cdots, \lambda_d$. The *Schatten p-norm*, for $p \geq 1$, is defined as

$$\|M\|_p = \left(\sum_{i=1}^{d} |\lambda_i|^p\right)^{1/p}.$$

Let $\rho, \sigma \in \mathbb{C}^{d\times d}$ be quantum states.

Definition 2.2 (Trace distance)**.** The *trace distance* between ρ and σ is

$$\mathrm{D}_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1.$$

Definition 2.3 (Fidelity)**.** The *fidelity* of ρ and σ is

$$\mathrm{F}(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \mathrm{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}.$$

Our version for the fidelity of quantum states is sometimes referred to as the "square root fidelity". Since fidelity is not a metric, we use the closely related Bures distance metric.

Definition 2.4 (Bures distance)**.** The *Bures distance* between ρ and σ is

$$\mathrm{D}_{\mathrm{B}}(\rho, \sigma) = \sqrt{2(1 - \mathrm{F}(\rho, \sigma))}.$$

Definition 2.5 (Bures χ^2-divergence)**.** The *Bures χ^2-divergence* of ρ from σ is given by the following formula when σ is a diagonal full-rank state with eigenvalues $(\alpha_1, \ldots, \alpha_d)$:

$$\mathrm{D}_{\chi^2}(\rho \,||\, \sigma) = \sum_{i,j=1}^{d} \frac{2}{\alpha_i + \alpha_j} \cdot |\rho_{ij} - \sigma_{ij}|^2.$$

In this paper, we will compute the trace distance of matrices that are not necessarily PSD, or even Hermitian, as well as the the fidelity, Bures distance, and Bures χ^2-divergence of sub-normalized PSD matrices. The definitions given above can be extended to these cases using the same formulas. We now state some useful results about the quantum distance measures above. The first set of inequalities relates trace distance and fidelity [NC10, Section 9.2].

Lemma 2.6 (Fuchs-van de Graaf inequalities)**.** $1 - \mathrm{F}(\rho, \sigma) \leq \mathrm{D}_{\mathrm{tr}}(\rho, \sigma) \leq \sqrt{1 - \mathrm{F}(\rho, \sigma)^2}$.

The Fuchs-van de Graaf inequalities can be used to show a pair of inequalities between trace distance and Bures distance.

Corollary 2.7. $\frac{1}{2}\mathrm{D_B}(\rho,\sigma)^2 \le \mathrm{D_{tr}}(\rho,\sigma) \le \mathrm{D_B}(\rho,\sigma)$.

Bures distance is also related to the Bures χ^2-divergence, via the following inequality.

Lemma 2.8 (Proposition 2.31, [FO24])**.** $\mathrm{D_B}(\rho,\sigma) \le \sqrt{\mathrm{D}_{\chi^2}(\rho \,||\, \sigma)}$.

Finally, the Gentle Measurement Lemma gives an expression for the fidelity between pre- and post-measurement versions of a state.

Lemma 2.9 (Gentle Measurement Lemma [Win02])**.** *Given a quantum state ρ and a projector Π, let $\rho|_\Pi$ be the post-measurement state if ρ is measured with the projective measurement $\{\Pi, I-\Pi\}$ and Π is obtained. Then*

$$\left(1 - \frac{1}{2}\mathrm{D_B}(\rho,\rho|_\Pi)^2\right)^2 = \mathrm{F}(\rho,\rho|_\Pi)^2 = \mathrm{tr}(\Pi\cdot\rho).$$

2.2 Learning with high probability and amplification

Our goal throughout this paper is to obtain tomography algorithms that output an estimator which is ε-close to the true state, in some distance D, with low probability of failure. Such an algorithm is said to succeed "with high probability". The exact threshold for success is not particularly important, but for concreteness, we will take it to mean with probability at least 0.99.

Sometimes our algorithms involve multiple (though, a small number of) steps, which each, individually, might only succeed with high probability themselves (possibly conditioned on the success of prior steps). We can use a union bound to lower bound the success probability of the overall algorithm. For example, an algorithm consisting of three steps, which each succeed with high probability, succeeds with probability at least 0.97. However, we have now dropped under our threshold. Fortunately, there is a standard procedure we may apply to amplify the success probability over 0.99, with only a constant factor increase in the number of samples, and a constant factor increase in error.

Let D be a metric (e.g. trace distance, or Bures distance). Consider a tomography algorithm $\mathcal{A}$ that takes as input n samples of a state ρ, and outputs an estimator $\widehat{\boldsymbol{\rho}}$ such that $\mathrm{D}(\widehat{\boldsymbol{\rho}},\rho) \le \varepsilon$ with probability at least, say, 0.51. We define a new algorithm $\mathcal{A}'$ that uses $n' = O(n\cdot\log(1/\delta))$ samples, and outputs a state $\widehat{\boldsymbol{\rho}}'$ such that $\mathrm{D}(\widehat{\boldsymbol{\rho}},\rho) \le 3\varepsilon$ with probability at least $1-\delta$.

The new algorithm divides the samples into $k = O(\log(1/\delta))$ groups of n samples, and executes $\mathcal{A}$ on each group of samples to obtain estimates $\widehat{\boldsymbol{\rho}}_1,\ldots,\widehat{\boldsymbol{\rho}}_k$. Then $\mathcal{A}'$ will output an estimate $\widehat{\boldsymbol{\rho}}_i$ maximizing the following quantity:

$$N_i \coloneqq \left|\{j\in[k] \mid \mathrm{D}(\widehat{\boldsymbol{\rho}}_j,\widehat{\boldsymbol{\rho}}_i) \le 2\varepsilon\}\right|.$$

See, e.g. [HKOT23, Proposition 2.4] for details.

For our purposes, we will set $\delta = 0.01$, and thus $\mathcal{A}'$ succeeds with high probability with $n' = O(n)$ samples, with a constant factor degradation in error. If $\mathcal{A}$ scales, say, inverse-polynomially in ε, then $\mathcal{A}'$ can be used to output estimates to the same accuracy, with high probability, with the same overall sample complexity.

2.3 On low-rank tomography algorithms

Our algorithms will often take copies of a rank-r quantum state ρ, and output an estimator $\widehat{\boldsymbol{\rho}}$ that is close to ρ in some distance. However, it is not necessarily the case that $\widehat{\boldsymbol{\rho}}$ is, itself, a rank-r state. However, when we are dealing with tomography in some metric (e.g. the trace distance, or Bures distance) there exists a generic transformation to our algorithm that makes it output an estimator $\widehat{\boldsymbol{\rho}}'$ of rank r.

Remark 2.10. Let D be a metric. Also let $\mathcal{A}$ be a tomography algorithm that takes n samples of a rank-r state ρ and outputs an estimator $\widehat{\boldsymbol{\rho}}$, not necessarily of rank r, such that $\mathrm{D}(\widehat{\boldsymbol{\rho}},\rho) \le \varepsilon$ with high probability. Then there exists a tomography algorithm $\mathcal{A}'$ that takes n samples of a rank-r state ρ and outputs a rank-r estimator $\widehat{\boldsymbol{\rho}}'$ such that $\mathrm{D}(\widehat{\boldsymbol{\rho}}',\rho) \le 2\varepsilon$ with high probability.

The new algorithm $\mathcal{A}'$ will set $\widehat{\boldsymbol{\rho}}'$ to be the rank-r quantum state that is closest to $\widehat{\boldsymbol{\rho}}$ in the metric D. Since ρ is also rank-r, this means that $\widehat{\boldsymbol{\rho}}'$ is at least as close to $\widehat{\boldsymbol{\rho}}$:

$$\mathrm{D}(\widehat{\boldsymbol{\rho}},\widehat{\boldsymbol{\rho}}') \le \mathrm{D}(\widehat{\boldsymbol{\rho}},\rho).$$

Thus $\mathrm{D}(\widehat{\boldsymbol{\rho}}, \widehat{\boldsymbol{\rho}}') \leq \varepsilon$ with high probability. From the triangle inequality, we conclude that with high probability,

$$\mathrm{D}(\widehat{\boldsymbol{\rho}}', \rho) \leq \mathrm{D}(\widehat{\boldsymbol{\rho}}', \widehat{\boldsymbol{\rho}}) + \mathrm{D}(\widehat{\boldsymbol{\rho}}, \rho) \leq 2\varepsilon.$$

2.4 Representation theory

In this section, we review the representation theory needed to describe the debiased Keyl's algorithm. For a more detailed treatment of applying representation theory to quantum learning algorithms, see [Wri16, Chapter 2].

2.4.1 Basics

The *symmetric group* S_n is the group of permutations of the set $[n]$. The *unitary group* $U(d)$ is the group of $d \times d$ unitary matrices. More generally, for V a complex, finite-dimensional Hilbert space, $U(V)$ is the group of unitary operators on V.

Definition 2.11 (Representations)**.** Let G be a group. A *complex, unitary, finite-dimensional representation* (henceforth, a *representation*) of G is a pair (μ, V), where V is a finite-dimensional, complex Hilbert space, and $\mu : G \to U(V)$ is a group homomorphism. That is, $\mu(e) = I_V$, $\mu(g)$ is unitary for all $g \in G$, and $\mu(g \cdot h) = \mu(g) \cdot \mu(h)$ for all $g, h \in G$. The *dimension* of the representation, denoted $\dim(\mu)$, is the dimension of V.

Commonly, a representation of G, (μ, V), is referred to either by μ or V, when the meaning is clear from context.

Definition 2.12 (Characters)**.** Let G be a group, and let μ be a representation. The *character* of μ is the function $\chi_\mu : G \to \mathbb{C}$ given by $\chi_\mu(g) = \mathrm{tr}(\mu(g))$.

Definition 2.13 (Intertwining operators)**.** Let G be a group, and let (μ_1, V_1) and (μ_2, V_2) be representations of G. Then a map $T : V_1 \to V_2$ is an *intertwining operator*, or *intertwiner*, if $T \cdot \mu_1(g) = \mu_2(g) \cdot T$ for all $g \in G$. We say T *intertwines* μ_1 and μ_2.

Definition 2.14 (Isomorphic representations)**.** Let G be a group, and let (μ_1, V_1) and (μ_2, V_2) be representations of G. We say μ_1 and μ_2 are *isomorphic* representations if there exists an invertible map T that intertwines μ_1 and μ_2. We write $\mu_1 \cong \mu_2$, or, if we want to emphasize the group, $\mu_1 \overset{G}{\cong} \mu_2$.

Definition 2.15 (Irreducible representations)**.** Let G be a group, and let (μ, V) be a representation of G. A subspace $W \subseteq V$ is *invariant* if $\mu(g) \cdot W \subseteq W$ for all $g \in G$. An invariant subspace is *trivial* if $W = \{0\}$ or $W = V$. The representation μ is *irreducible* if V has no nontrivial invariant subspaces. An irreducible representation is also called an *irrep* for short. The set of all isomorphism classes of irreducible representations of G is denoted $\widehat{G}$.

We will usually fix an irrep μ_i from each isomorphism class, and identify $\widehat{G}$ with the set $\{\mu_i\}$.

Lemma 2.16 (Schur's lemma)**.** *Let G be a group, and let (μ_1, V_1) and (μ_2, V_2) be two irreducible representations of G. Let $T : V_1 \to V_2$ intertwine μ_1 and μ_2. If μ_1 and μ_2 are non-isomorphic, then $T = 0$. If instead $\mu_1 = \mu_2$, then*

$$T = \frac{\mathrm{tr}(T)}{\dim(\mu_1)} \cdot I_{V_1}.$$

The following result, which we will make frequent use of, can be shown using Schur's lemma.

Lemma 2.17. *Let G be a group, and let (μ_1, V_1) and (μ_2, V_2) be unitary, isomorphic representations of G. Then there exists a unitary $U : V_1 \to V_2$ intertwining μ_1 and μ_2. That is, for all $g \in G$,*

$$U \cdot \mu_1(g) \cdot U^\dagger = \mu_2(g).$$

Definition 2.18 (Direct sum of representations)**.** Let G be a group, and let (μ_1, V_1), $\ldots$, (μ_k, V_k) be representations of G. Then the *direct sum* of $\mu_1, \ldots, \mu_k$ is the representation (μ, V), where $V \coloneqq V_1 \oplus \cdots \oplus V_k$ and $\mu(g) \coloneqq \mu_1(g) \oplus \cdots \oplus \mu_k(g)$, for all $g \in G$. Equivalently, we can write $\mu(g) = \sum_{i=1}^{k} |i\rangle\langle i| \otimes \mu_i(g)$.

Generally, each representation μ_i may appear multiple times. If μ_i appears m_i times, we also write $\mu = \bigoplus_{i=1}^{k} m_i \cdot \mu_i$, and refer to m_i as the *multiplicity* of μ_i.

Definition 2.19 (Complete reducibility)**.** Let G be a group, and let (μ, V) be a representation. Then μ is *completely reducible* if

$$\mu \cong \bigoplus_{i=1}^{k} m_i \cdot \mu_i,$$

with each μ_i irreducible.

Finite-dimensional unitary representations are completely reducible.

Definition 2.20 (Restriction of representations)**.** Let H be a subgroup of G, and let μ be a representation of G. The *restriction of μ to H*, a representation of H, is denoted $\mu{\downarrow}_H^G$, and is given by

$$\mu{\downarrow}_H^G(h) = \mu(h).$$

The restriction of an irreducible representation may itself be reducible. *Branching rules* describe how restrictions decompose into irreps.

Definition 2.21 (Branching rules)**.** Let H be a subgroup of G, and let μ be a representation of G. A *branching rule* is a statement of the form:

$$\mu{\downarrow}_H^G \overset{H}{\cong} \bigoplus_{\mu_i \in \widehat{H}} m_i \cdot \mu_i.$$

If $m_i \in \{0, 1\}$ for all i, then we say the branching rule is *multiplicity-free*.

2.4.2 Partitions and Young diagrams

Definition 2.22 (Partitions)**.** Let n be a nonnegative integer. A *partition* of n is a finite list of nonnegative integers $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$ and $\lambda_1 + \cdots + \lambda_m = n$. We write $\lambda \vdash n$. The *size* of λ is $|\lambda| = n$, and the *length* of λ, denoted $\ell(\lambda)$, is the largest index i with $\lambda_i > 0$.

Remark 2.23. Partitions which only differ by some number of trailing zeroes are considered equivalent. As we will see later, when considering the problem of learning a quantum state $\rho \in \mathbb{C}^d$, only partitions of length at most d will be relevant to us. In this case, we will often view a partition as a tuple of length exactly d.

Partitions have diagrammatic representations known as Young diagrams.

Definition 2.24 (Young diagrams)**.** Let $\lambda \vdash n$. The *Young diagram of shape* λ is a diagram consisting of n boxes, organized into $\ell(\lambda)$ left-justified rows, such that the i-th row contains λ_i boxes. We will also refer to boxes as *cells*.

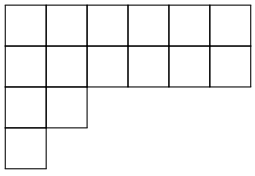


Figure 3: The Young diagram above corresponds to the partition $\lambda = (6, 6, 2, 1)$, with $|\lambda| = 15$ and $\ell(\lambda) = 4$.

See Figure 3 for an example of a Young diagram. We use the following non-standard definition to refer to a consecutive group of rows with the same length, something that will be useful for describing our family of unbiased estimators.

Definition 2.25 (Blocks in Young diagrams)**.** Let $\lambda = (\lambda_1, \ldots, \lambda_d) \vdash n$ be a Young diagram. Then $\{i, i+1, \ldots, j\}$ is a *block* of λ if

$$\lambda_i = \lambda_{i+1} = \cdots = \lambda_j,$$

and $\lambda_{i-1} > \lambda_i$ when $i > 1$, and $\lambda_j > \lambda_{j+1}$ when $j < d$.

In the Young diagram in Figure 3, $\{1, 2\}$, $\{3\}$ and $\{4\}$ are blocks. If we regard the corresponding partition as a tuple of length d, and $d \geq 5$, then $\{5, \ldots, d\}$ is also a block.

Definition 2.26 (Standard Young tableaux)**.** Let $\lambda \vdash n$. A *standard Young tableau* (SYT) of shape λ is a filling of λ with the numbers $\{1, \ldots, n\}$, such that the entries strictly increase both from left to right, and from top to bottom. The set of all SYTs of shape λ is denoted $\mathrm{SYT}(\lambda)$.

Definition 2.27 (Semistandard Young tableaux)**.** Let $\lambda \vdash n$, and let d be a positive integer. A *semistandard Young tableau* of shape λ and alphabet $[d]$ is a filling of λ with numbers from $[d]$, such that the entries weakly increase from left to right, and strictly increase from top to bottom. The set of all SSYTs of shape λ and alphabet $[d]$ is denoted $\mathrm{SSYT}(\lambda, d)$.

1	2	4	6
3	5		
7			

1	1	2	3
2	3		
3			

Figure 4: Let $\lambda = (4, 2, 1)$. On the left, a standard Young tableau of shape λ. On the right, a semistandard Young tableau of shape λ and alphabet $[3]$.

Notation 2.28. We use the following additional notation.

- We also write $\square$ to denote the Young diagram corresponding to $\lambda = (1)$.
- The box in the i-th row and j-th column is denoted by (i, j).
- The *content* of (i, j) is $\mathrm{cont}_\lambda(i, j) \coloneqq j - i$.
- The *hook* of (i, j) is the set of boxes in λ, (k, ℓ), such that $i = k$ and $j \leq \ell$, or $j = \ell$ and $i \leq k$. The *hook-length* of (i, j), denoted $\mathrm{hook}_\lambda(i, j)$, is the number of boxes in the hook of (i, j).
- If S is an SYT of shape λ, and $m \in [n]$ is the label of box (i, j) in S, then we also write $\mathrm{cont}_S(m) \coloneqq \mathrm{cont}_\lambda(i, j)$.
- Let λ and μ are Young diagrams such that $\ell(\mu) \geq \ell(\lambda)$, and suppose for each $i \in \ell(\lambda)$ we have $\lambda_i \leq \mu_i$. We can view λ as a subset of the boxes of μ, and write $\lambda \subseteq \mu$. We also write $\mu \setminus \lambda$ for the subset of boxes of μ not in λ.
- If $\lambda \subseteq \mu$ and $|\mu \setminus \lambda| = 1$, then we write $\lambda \nearrow \mu$. The Young diagram of μ can be obtained by adding a single box to some row of λ. If this is the i-th row, we also write $\mu = \lambda + e_i$, or $\lambda = \mu - e_i$.
- If $\lambda \subseteq \mu$, and μ can be obtained from λ by adding some number of boxes to λ, such that none of these additional boxes are in the same column, we write $\lambda \precsim \mu$. Equivalently, the row-lengths of λ and μ *interlace*, meaning $\mu_{i+1} \leq \lambda_i \leq \mu_i$ for all i, so that $\lambda_d \leq \mu_d \leq \lambda_{d-1} \leq \cdots \leq \mu_2 \leq \lambda_1 \leq \mu_1$.

2.4.3 The irreducible representations of S_n

We now describe the irreducible representations of the symmetric group. For more, see [Sag01] for general theory, [Har05, Chapter 7] or [VO05] for subgroup adapted bases, and [VO05] for Young's orthogonal representation and the Jucys-Murphy element.

Theorem 2.29. *The irreducible representations of S_n are indexed by partitions $\lambda \vdash n$.*

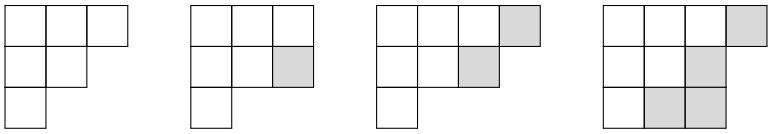


Figure 5: Four Young diagrams. On the left is $\lambda = (3, 2, 1)$. The remaining Young diagrams are μ_i, for $i = 1, 2, 3$. Each μ_i satisfies $\lambda \subseteq \mu_i$, with the boxes of $\mu \setminus \lambda$ shaded. We have $\lambda \nearrow \mu_1$, with $\mu_1 = \lambda + e_2$, since the additional box is in the second row. The shaded box in μ_1 is $(2, 3)$, and the content of this cell is $\mathrm{cont}(2, 3) = 1$. Both μ_1 and μ_2 satisfy $\lambda \preceq \mu_i$, since there is at most one added box per column. However, $\lambda \not\preceq \mu_3$, since there are two additional boxes in the third column.

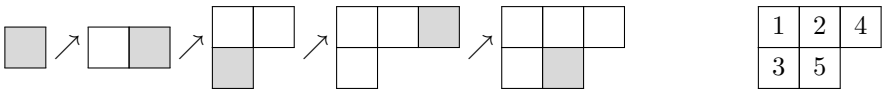


Figure 6: On the left, a sequence of Young diagrams $(\lambda^{(i)})_{i\in[5]}$, each obtained from the last by adding a single box. The new box is shaded in each diagram. On the right, the SYT of shape $\lambda^{(5)}$ corresponding to the chain. The box added at the i-th step is labeled i.

Definition 2.30. The irrep corresponding to λ is $(\kappa_\lambda, \mathrm{Sp}_\lambda)$, where Sp_λ is called the *Specht module*. We will abbreviate $\dim(\kappa_\lambda)$ by $\dim(\lambda)$.

The symmetric group has the following branching rule.

Theorem 2.31. *Let $\lambda \vdash n$ be a partition. Consider S_{n-1} as a subgroup of S_n via the embedding that maps $\pi \in S_{n-1}$ to the permutation on $[n]$ that fixes n and permutes everything else according to π. The branching rule for κ_λ, restricted to S_{n-1}, is*

$$(\kappa_\lambda)\downarrow^{S_n}_{S_{n-1}} \cong \bigoplus_{\mu \nearrow \lambda} \kappa_\mu.$$

Note that the branching rule is multiplicity-free.

A multiplicity-free branching rule can be used to construct an explicit orthonormal basis for each irreducible representation. Such a basis is generally known as a *subgroup adapted basis*. We now outline this construction for the symmetric group. From Theorem 2.31, we have the following decomposition of vector spaces.

$$\mathrm{Sp}_\lambda \cong \bigoplus_{\mu \nearrow \lambda} \mathrm{Sp}_\mu. \tag{13}$$

Taking the summands to be orthogonal reduces the problem to finding orthonormal bases for Sp_μ. We now recursively apply the branching rule to S_{n-1}, S_{n-2}, ..., S_1 to obtain

$$\mathrm{Sp}_\lambda \cong \bigoplus_{\lambda^{(1)} \nearrow \lambda^{(2)} \nearrow \cdots \nearrow \lambda^{(n-1)} \nearrow \lambda^n} \mathrm{Sp}_{\lambda^{(1)}}. \tag{14}$$

However, since S_1 is the trivial group, $\mathrm{Sp}_{\lambda^{(1)}}$ is a one-dimensional subspace, and contains a unique vector (up to phase). As a result, Sp_λ decomposes into the direct sum of one-dimensional subspaces indexed by chains of the form $\square = \lambda^{(1)} \nearrow \lambda^{(2)} \nearrow \cdots \nearrow \lambda^{(n-1)} \nearrow \lambda^n = \lambda$. Moreover, such chains are in bijection with standard Young tableaux of shape λ, where the number $i \in [n]$ is placed in the box in the unique cell in $\lambda^{(i)} \setminus \lambda^{(i-1)}$ (taking $\lambda^{(0)} = \varnothing$). See Figure 6 for an illustration.

The above construction gives us a canonical decomposition into one-dimensional subspaces. Taking these subspaces to be orthogonal, and choosing a unit-norm vector in each subspace gives us a *Young-Yamanouchi basis*.

Definition 2.32 (Young-Yamanouchi basis)**.** Given $\lambda \vdash n$, a *Young-Yamanouchi basis* for Sp_λ has a unit vector $|S\rangle$ for each standard Young tableau S of shape λ, associated to the corresponding one-dimensional subspace in Equation (14).

Corollary 2.33. $\dim(\lambda) = |\mathrm{SYT}(\lambda)|$.

Young-Yamanouchi bases are canonical up to a choice of phase in each basis vector. *Young's orthogonal basis* is a particularly nice basis, in which each permutation is represented by an orthogonal matrix.

Definition 2.34 (Young's orthogonal basis)**.** Let $\lambda \vdash n$. *Young's orthogonal basis* is a particular Young-Yamanouchi basis, which can be specified by the action of the transpositions $(i, i+1)$, for $i \in [n-1]$. If i and $i+1$ occur in the same row of S, then

$$\kappa_\lambda(i, i+1) \cdot |S\rangle = + |S\rangle .$$

If, instead, i and $i+1$ occur in the same column of S, then

$$\kappa_\lambda(i, i+1) \cdot |S\rangle = - |S\rangle .$$

Finally, if i and $i+1$ occur in neither the same row nor the same column of S, then swapping the locations of i and $i+1$ in S gives us another valid SYT S', and

$$\kappa_\lambda(i, i+1) \cdot |S\rangle = \frac{1}{\Delta_S(i)} |S\rangle + \sqrt{1 - \frac{1}{\Delta_S(i)^2}} |S'\rangle ,$$

where $\Delta_S(i) \coloneqq \mathrm{cont}_S(i+1) - \mathrm{cont}_S(i)$.

When written in this basis, the representation κ_λ is also known as *Young's orthogonal form*. Henceforth, we will take κ_λ to mean Young's orthogonal form, unless otherwise mentioned.

For a group G, the group algebra $\mathbb{C}[G]$ is the associative algebra whose elements are formal linear combinations of group elements with complex coefficients. The multiplication of basis elements is inherited from the group's operation, and can be linearly extended to multiplication of generic elements. Representations can be extended to the group algebra, as $\rho(\sum_i c_i g_i) = \sum_i c_i \rho(g_i)$.

Jucys-Murphy elements play a prominent role in our analysis.

Definition 2.35 (Jucys-Murphy element)**.** For $i \in [n]$, the *k-th Jucys-Murphy element* is the element of the group algebra $\mathbb{C}[S_n]$ given by

$$X_i \coloneqq \sum_{j<i} (j, i).$$

Theorem 2.36. *Let $\lambda \vdash n$. Consider $\kappa_\lambda(X_k)$ as an operator on Sp_λ. Then Young's orthogonal basis is an eigenbasis of $\kappa_\lambda(X_k)$, with*

$$\kappa_\lambda(X_i) \cdot |S\rangle = \mathrm{cont}_S(i) \cdot |S\rangle$$

for all standard Young tableaux S of shape λ.

We will sometimes identify X_k with its action on a representation space, and use X_k in place of $\mathcal{P}(X_k)$ or $\kappa_\lambda(X_k)$ whenever the representation is clear from context.

2.4.4 The irreducible representations of $U(d)$

In this section, we describe the irreducible representations of the unitary group. For more, see [GW09].

Definition 2.37 (Polynomial representations)**.** A finite-dimensional representation (μ, V) of a matrix group G is *polynomial*, if, expressed in some basis of V, every matrix element $\mu(g)_{ij}$ is a polynomial in the entries of $g \in G$.

A representation being polynomial is a basis-independent notion, since if the entries of $\mu(g)$ are polynomials in one basis, then they are also polynomials in any other basis.

Theorem 2.38. *The polynomial irreps of $U(d)$ are indexed by partitions λ of length at most d. The irrep corresponding to λ is written $(\nu_\lambda, V_\lambda^d)$, where V_λ^d is called the* Schur(-Weyl) module.

Since ν_λ is polynomial, we can extend the domain to be all matrices $M \in \mathbb{C}^{d\times d}$. Also, note that here we have an irrep corresponding to the empty partition, $\lambda = \varnothing$.

We now turn to the characters of these irreps.

Definition 2.39. Let $x_1, \dots, x_d$ be indeterminates. For $\lambda \vdash n$, the *Schur polynomial* $s_\lambda(x_1, \dots, x_d)$ is the degree-n homogeneous polynomial defined as $s_\lambda(x_1, \dots, x_d) \coloneqq \sum_T x^T$, where the sum is over all SSYTs T of shape λ and alphabet $[d]$, and

$$x^T = \prod_{i=1}^{d} x_i^{w_T(i)}.$$

Here, $w_T(i)$ is the number of boxes of T labeled by i.

Theorem 2.40. *Let U be a unitary with eigenvalues $\alpha_1, \dots, \alpha_d$. Then $\chi_{\nu_\lambda}(U) = s_\lambda(\alpha_1, \dots, \alpha_d)$.*

We can extend the domain of s_λ to all $M \in \mathbb{C}^{d\times d}$ by $s_\lambda(M) \coloneqq \operatorname{tr}(\nu_\lambda(M))$.

Next, we construct an explicit orthonormal basis for each polynomial irrep of the unitary group using the following result.

Theorem 2.41. *Let λ be a partition with $\ell(\lambda) \le d$, and fix a basis $\{|i\rangle\}_{i\in[d]}$. For $d \ge 2$, consider the embedding of $U(d-1)$ into $U(d)$ via the following map:*

$$U \mapsto \begin{bmatrix} U & 0 \\ 0 & 1 \end{bmatrix}.$$

Under this mapping, we can identify $U(d-1)$ with its image, a subgroup of $U(d)$. When restricted to $U(d-1)$, we have the multiplicity-free branching rule

$$(\nu_\lambda)\downarrow^{U(d)}_{U(d-1)}(U) \cong \bigoplus_{\substack{\mu \precsim \lambda \\ \ell(\mu)\le d-1,}} \nu_\mu(U).$$

We can use this branching rule to decompose V_λ^d as

$$V_\lambda^d \cong \bigoplus_{\substack{\mu \precsim \lambda \\ \ell(\mu)\le d-1,}} V_\mu^{d-1}.$$

As in the construction of Young's orthogonal basis in Section 2.4.3, we can recursively restrict, eventually decomposing

$$V_\lambda^d \cong \bigoplus_{\substack{\lambda^{(1)} \precsim \cdots \precsim \lambda^{(n)} \\ \ell(\lambda^{(i)})\le i}} V^1_{\lambda^{(1)}},$$

where $\lambda^{(n)} = \lambda$. However, since $U(1)$ is Abelian, and since $V^1_{\lambda^{(1)}}$ an irrep of $U(1)$, $V^1_{\lambda^{(1)}}$ is a one-dimensional representation, and contains a unique vector, up to phase. Hence, V_λ^d is isomorphic to a direct sum of one-dimensional subspaces labeled by chains

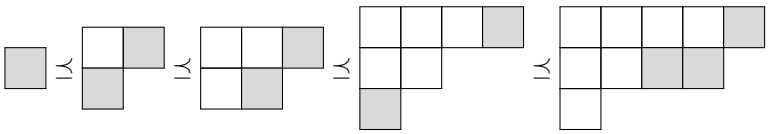
$$\lambda^{(1)} \precsim \cdots \precsim \lambda^{(n)} = \lambda,$$

where $\ell(\lambda^{(i)}) \le i$. Every such chain can be associated bijectively to a semistandard Young tableau of shape λ and alphabet $[d]$, by filling the boxes of $\lambda^{(i)} \setminus \lambda^{(i-1)}$ with the number i, for each $i \in [d]$. See Figure 7 for an illustration.

Choosing a vector in each one-dimensional subspace gives a *Gelfand-Tsetlin basis*.

Definition 2.42 (Gelfand-Tsetlin basis)**.** Given λ with $\ell(\lambda) \le d$, the *Gelfand-Tsetlin (GT) basis for V_λ^d* has a unit vector $|T\rangle$ for each semistandard Young tableau T of shape λ and alphabet $[d]$.

Corollary 2.43. $\dim(V_\lambda^d) = |\mathrm{SSYT}(\lambda, d)|$.



1	2	3	4	5
2	3	5	5	
4				

Figure 7: On the left, a sequence of Young diagrams $(\lambda^{(i)})_{i\in[5]}$, such that $\lambda^{(i)} \preceq \lambda^{(i+1)}$ and $\ell(\lambda^{(i)}) \leq i$. The new boxes added at each step are shaded in each diagram. On the right, the SSYT of shape $\lambda^{(5)}$ corresponding to the chain. The boxes added at the i-th step are labeled i.

The dimension of V_λ^d can be computed using the Weyl dimension formula [GW09, Equation (7.18)]:

$$\dim(V_\lambda^d) = \prod_{1\leq i<j\leq d} \frac{(\lambda_i - \lambda_j) + (j-i)}{j-i}. \tag{15}$$

It is also given by Stanley's hook-content formula [Sta99, Equation (7.21.2)]:

$$\dim(V_\lambda^d) = \prod_{(i,j)\in\lambda} \frac{d + \mathrm{cont}_\lambda(i,j)}{\mathrm{hook}_\lambda(i,j)}. \tag{16}$$

Certain GT basis vectors, called the *highest weight vectors*, play a prominent role in Keyl's algorithm and our debiased version.

Definition 2.44 (Highest weight SSYTs, highest weight vectors)**.** The *highest weight SSYT*, denoted T^λ, is the SSYT of shape λ for which every box in the i-th row is labelled i. The corresponding GT basis vector, $|T^\lambda\rangle$, is called the *highest weight vector*.

1	1	1	1	1
2	2	2	2	
3				

Figure 8: The highest weight SSYT of shape $\lambda = (5, 4, 1)$.

We conclude this section with an example which will be particularly important to us later on: the irrep corresponding to $\lambda = (1)$.

Example 2.45. The irrep $V_{(1)}^d$ turns out to be isomorphic to the *defining representation*, in which which U acts on $\mathbb{C}^d$ as itself. The Gelfand-Tsetlin basis is the computational basis $\{|i\rangle\}_{i\in[d]}$ (up to phase).

2.4.5 Schur-Weyl duality

In state tomography, we are given as input n copies of some unknown state ρ, i.e. we are given the state $\rho^{\otimes n}$. In this setting, there are two particularly natural representations of the groups S_n and $U(d)$, acting on $(\mathbb{C}^d)^{\otimes n}$.

Definition 2.46. The groups S_n and $U(d)$ have the following representations acting on $(\mathbb{C}^d)^{\otimes n}$. The representation $\mathcal{P}^{(n)}(\pi)$ acts by permuting the n tensor factors according to π, i.e. $\mathcal{P}^{(n)}(\pi)$ acts on a standard basis element $|i_1\rangle \otimes \cdots \otimes |i_n\rangle$ as

$$\mathcal{P}^{(n)}(\pi)\,|i_1\rangle \otimes \cdots \otimes |i_n\rangle = \left|i_{\pi^{-1}(1)}\right\rangle \otimes \cdots \otimes \left|i_{\pi^{-1}(n)}\right\rangle.$$

The representation $\mathcal{Q}^{(n,d)}(U)$ acts on each tensor factor as U, i.e. $\mathcal{Q}^{(n,d)}(U)$ acts on a standard basis element as

$$\mathcal{Q}^{(n,d)}(U)\,|i_1\rangle \otimes \cdots \otimes |i_n\rangle = U\,|i_1\rangle \otimes \cdots \otimes U\,|i_n\rangle\,.$$

We will often just write $\mathcal{P}(\pi)$ and $\mathcal{Q}(U)$, when n and d are clear from context.

Since $\mathcal{P}(\pi)$ and $\mathcal{Q}(U)$ commute for any choices of π and U, the product $\mathcal{P} \cdot \mathcal{Q}$ forms a representation of the product group $S_n \times U(d)$. Schur-Weyl duality provides us with a nice decomposition of this representation into products of irreps of S_n and $U(d)$.

Theorem 2.47 (Schur-Weyl duality)**.** *Consider the representation $\mathcal{P} \cdot \mathcal{Q}$ of $S_n \times U(d)$. The action of this representation on $(\mathbb{C}^d)^{\otimes n}$ decomposes as*

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} \mathrm{Sp}_\lambda \otimes V_\lambda^d.$$

Therefore, there exists a fixed unitary $\mathcal{U}_{\mathrm{SW}}^{(n)}$ such that for all $\pi \in S_n$ and $U \in U(d)$, we have

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \mathcal{P}(\pi)\mathcal{Q}(U) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} = \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\pi) \otimes \nu_\lambda(U). \tag{17}$$

In Section 11 we will construct the unitary $\mathcal{U}_{\mathrm{SW}}^{(n)}$, though for now, it suffices to know that such a unitary exists. After changing basis with $\mathcal{U}_{\mathrm{SW}}^{(n)}$, we will refer to the second register as the *permutation register*, and the third register as the *unitary register*, since S_n and $U(d)$ act on these registers only. The collection of vectors of the form $|\lambda\rangle \otimes |S\rangle \otimes |T\rangle$, with $\lambda \vdash n$, $S \in \mathrm{SYT}(\lambda)$, and $T \in \mathrm{SSYT}(\lambda, d)$ forms the *Schur basis*. The *λ-subspace* is the subspace spanned by Schur basis vectors with fixed λ. We write Π_λ for the projector defined by

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \Pi_\lambda \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} = |\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes I_{\dim(V_\lambda^d)}. \tag{18}$$

Recalling that the domain of $\mathcal{Q}$ can be extended to all matrices $M \in \mathbb{C}^{d \times d}$, we can apply Schur-Weyl duality to the state $\rho^{\otimes n}$, regarded as $\mathcal{P}(e) \cdot \mathcal{Q}(\rho)$, with e the identity permutation. This gives us the following result.

Corollary 2.48 (Schur-Weyl duality for quantum states)**.** *There is a fixed unitary change of basis $\mathcal{U}_{\mathrm{SW}}^{(n)}$, and therefore an allowed quantum mechanical transformation, that puts an arbitrary input state $\rho^{\otimes n}$ into the following form:*

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \rho^{\otimes n} \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} = \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \nu_\lambda(\rho).$$

In other words, there is a fixed change-of-basis unitary which simultaneously block-diagonalizes all possible input states.

As a further example, which will be useful to us later, we have

$$\begin{aligned} \mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \mathcal{P}(X_k) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} &= \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(X_k) \otimes I_{\dim(V_\lambda^d)} \\ &= \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} \sum_{S \in \mathrm{SYT}(\lambda)} \mathrm{cont}_S(k) \cdot |\lambda\rangle\langle\lambda| \otimes |S\rangle\langle S| \otimes I_{\dim(V_\lambda^d)}, \end{aligned} \tag{19}$$

using Theorems 2.36 and 2.47. Thus, in the Schur basis, the Jucys-Murphy element acts diagonally.

2.4.6 Weak Schur sampling

Definition 2.49 (Weak Schur sampling)**.** *Weak Schur sampling* (WSS) refers to performing the projective measurement $\{\Pi_\lambda\}_{\lambda\vdash n,\ell(\lambda)\leq d}$ on $\rho^{\otimes n}$, and induces a probability distribution on partitions $\lambda \vdash n$, denoted $\mathrm{WSS}_n(\rho)$.

If ρ has spectrum $\alpha = (\alpha_1, \ldots, \alpha_d)$, weak Schur sampling yields outcome λ with probability

$$\Pr_{\boldsymbol{\lambda}\sim \mathrm{WSS}_n(\rho)}[\boldsymbol{\lambda} = \lambda] = \mathrm{tr}(\Pi_\lambda \rho^{\otimes n}) = \dim(\lambda)\cdot \mathrm{tr}(\nu_\lambda(\rho)) = \dim(\lambda)\cdot s_\lambda(\alpha). \tag{20}$$

Remark 2.50. Suppose ρ has rank r. Recall that the boxes of an SSYT are strictly increasing as we move down a column. Therefore, any SSYT with more than r rows necessarily has a box containing a number larger than r. Therefore, each term of $s_\lambda(\alpha)$ contains at least one α_i for $i > r$ (see Theorem 2.40). So if $\ell(\lambda) > r$, then $s_\lambda(\alpha) = 0$. That is, we always receive a Young diagram λ with $\ell(\lambda) \leq r$.

If we perform WSS on $\rho^{\otimes n}$, and obtain outcome λ, the resulting post-measurement state is

$$\frac{\Pi_\lambda \cdot \rho^{\otimes n}\cdot \Pi_\lambda}{\mathrm{tr}(\Pi_\lambda \cdot \rho^{\otimes n})} = \mathcal{U}_{\mathrm{SW}}^{(n)\dagger}\cdot \left(|\lambda\rangle\langle\lambda| \otimes \frac{I_{\dim(\lambda)}}{\dim(\lambda)} \otimes \frac{\nu_\lambda(\rho)}{s_\lambda(\alpha)}\right)\cdot \mathcal{U}_{\mathrm{SW}}^{(n)}. \tag{21}$$

Note that, in the Schur basis, the first two registers contain no quantum information, and may be discarded as they are classically simulable.

3 The debiased Keyl's algorithm

Here, we formally introduce our unbiased entangled tomography algorithm. In Section 3.1, we describe Keyl's algorithm. In Section 3.2, we formally state our debiased version of Keyl's algorithm. Lastly, in Section 3.3 we show that in addition to our debiased Keyl's algorithm, there is an entire family of unbiased estimators for quantum state tomography. Some of these estimators have simpler descriptions than our original estimator and will be useful in some intermediate steps of our proofs. We also show that our original debiased Keyl's algorithm is the family member with the minimal variance.

3.1 Keyl's algorithm

Keyl's algorithm leverages tools from representation theory to perform quantum state tomography [Key06]. It starts by performing weak Schur sampling, and uses the Young diagram $\boldsymbol{\lambda} \vdash n$ it receives to form an estimate for the spectrum of ρ, as described in Section 2.4.6. It continues by measuring in the irrep $V_{\boldsymbol{\lambda}}^d$ to learn an estimate for the change of basis unitary that rotates the computational basis onto the eigenbasis of ρ, using the highest weight vector introduced in Theorem 2.42. Finally, it combines these ingredients to produce an estimate for the state ρ itself.

Definition 3.1 (Keyl's algorithm)**.** Given n copies of a quantum state $\rho \in \mathbb{C}^d$:

1. Perform WSS on $\rho^{\otimes n}$ to obtain a Young diagram $\boldsymbol{\lambda} \vdash n$, and the corresponding state $\rho_{\boldsymbol{\lambda}}$, as in Equation (21). Discard the first two registers, leaving only $\nu_{\boldsymbol{\lambda}}(\rho)/s_\lambda(\alpha)$. Set $\widehat{\boldsymbol{\alpha}}_0 = \boldsymbol{\lambda}/n$.
2. Measure in $V_{\boldsymbol{\lambda}}^d$ using the POVM $M_{\boldsymbol{\lambda}}$, where $M_{\boldsymbol{\lambda}} = \{M_{\boldsymbol{\lambda},U}\}_U$ is a POVM whose elements are labeled by $U \in U(d)$ and defined as follows:

$$M_{\lambda,U} = \dim(V_\lambda^d)\cdot \nu_\lambda(U)\,|T_\lambda\rangle\langle T_\lambda|\,\nu_\lambda(U)^\dagger \cdot \mathrm{d}U,$$

   where $\mathrm{d}U$ is the Haar measure on $U(d)$. Let $\boldsymbol{U}$ be the measurement outcome. We write $\boldsymbol{U} \sim K_{\boldsymbol{\lambda}}(\rho)$.
3. Output $\widehat{\boldsymbol{\rho}} \coloneqq \boldsymbol{U}\cdot \widehat{\boldsymbol{\alpha}}_0 \cdot \boldsymbol{U}^\dagger$.

Equivalently, Keyl's algorithm can be described by a single measurement, with outcomes labeled by the pair (λ, U), and

$$M_{\lambda,U} = \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} \cdot \left(\dim(V_\lambda^d) \cdot |\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \nu_\lambda(U) \, |T_\lambda\rangle\langle T_\lambda| \, \nu_\lambda(U)^\dagger \cdot \mathrm{d}U \right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)}. \tag{22}$$

To see that the POVM described in the second step of the algorithm is a valid POVM, let us define

$$N_\lambda \coloneqq \int_U M_{\lambda,U} = \dim(V_\lambda^d) \int_U \nu_\lambda(U) \, |T_\lambda\rangle\langle T_\lambda| \, \nu_\lambda(U)^\dagger \cdot \mathrm{d}U.$$

For M_λ to form a POVM, we need $N_\lambda = I_{\dim(V_\lambda^d)}$. We note that for any $V \in U(d)$,

$$\begin{aligned} \nu_\lambda(V) N_\lambda \nu_\lambda(V)^\dagger &= \dim(V_\lambda^d) \int_U \nu_\lambda(VU) \, |T_\lambda\rangle\langle T_\lambda| \, \nu_\lambda(VU)^\dagger \cdot \mathrm{d}U \\ &= \dim(V_\lambda^d) \int_W \nu_\lambda(W) \, |T_\lambda\rangle\langle T_\lambda| \, \nu_\lambda(W)^\dagger \cdot \mathrm{d}W = N_\lambda, \end{aligned}$$

so that N_λ is an intertwining operator. Here we have defined $W = VU$, and used the left-invariance of the Haar measure. Then, since ν_λ is an irreducible representation, Schur's lemma (Theorem 2.16) implies that $N_\lambda = \mathrm{tr}(N_\lambda) \cdot I_{\dim(V_\lambda^d)} / \dim(V_\lambda^d)$. However,

$$\mathrm{tr}(N_\lambda) = \dim(V_\lambda^d) \int_U \mathrm{tr}(\nu_\lambda(U) \, |T_\lambda\rangle\langle T_\lambda| \, \nu_\lambda(U)^\dagger) \cdot \mathrm{d}U = \dim(V_\lambda^d) \int_U \mathrm{d}U = \dim(V_\lambda^d),$$

so that $N_\lambda = I_{\dim(V_\lambda^d)}$.

We next provide some intuition for why the second step is able to learn a unitary that approximately diagonalizes ρ. We have

$$\mathrm{tr}(M_{\lambda,U} \cdot \nu_\lambda(\rho)) = \dim(V_\lambda^d) \cdot \langle T_\lambda| \, \nu_\lambda(U^\dagger \rho U) \, |T_\lambda\rangle \, . \tag{23}$$

As noted by Keyl [Key06, Equation 141], it turns out that the matrix elements $\langle T_\lambda| \, \nu_\lambda(U^\dagger \rho U) \, |T_\lambda\rangle$ can be re-expressed as $\Delta_\lambda(U^\dagger \rho U)$. This quantity is defined for a matrix $Z \in \mathbb{C}^{d \times d}$ as

$$\Delta_\lambda(Z) \coloneqq \prod_{i=1}^d \mathrm{pm}_i(Z)^{\lambda_i - \lambda_{i+1}},$$

where pm_i is the i-th principal minor of Z, and we take $\lambda_{d+1} = 0$. Since $U^\dagger \rho U$ has eigenvalues $\alpha_1 \geq \cdots \geq \alpha_d$, and $\mathrm{pm}_i(U^\dagger \rho U) \leq \prod_{j=1}^i \alpha_j$, we have

$$\langle T_\lambda| \, \nu_\lambda(U^\dagger \rho U) \, |T_\lambda\rangle = \prod_{i=1}^d \mathrm{pm}_i(U^\dagger \rho U)^{\lambda_i - \lambda_{i+1}} \leq \prod_{i=1}^d \prod_{j=1}^i \alpha_i^{\lambda_i - \lambda_{i+1}} = \prod_{i=1}^d \alpha_i^{\lambda_i}.$$

This maximum is attained if and only if $U^\dagger \rho U$ is diagonal in the computational basis, with decreasingly sorted eigenvalues. In other words, this maximum is attained if $\rho = U \cdot \mathrm{diag}(\alpha_1, \ldots, \alpha_d) \cdot U^\dagger$, so that U is the change of basis unitary rotating the computational basis onto the eigenbasis of ρ. Therefore, the probability density of obtaining U is maximal at the correct unitary.

3.2 The debiased Keyl's algorithm

Keyl's algorithm measures a Young diagram $\boldsymbol{\lambda} \vdash n$ and a unitary matrix $\boldsymbol{U} \in U(d)$, and outputs $\boldsymbol{U} \cdot (\boldsymbol{\lambda}/n) \cdot \boldsymbol{U}^\dagger$. In order to debias Keyl's algorithm, it turns out that we can perform a simple, deterministic transformation on $\boldsymbol{\lambda}$, called *donation*, and output $\boldsymbol{U} \cdot (\mathrm{donate}(\boldsymbol{\lambda})/n) \cdot \boldsymbol{U}^\dagger$. Donation and the debiased Keyl's algorithm were both defined in the introduction, but we restate their definitions here for convenience.

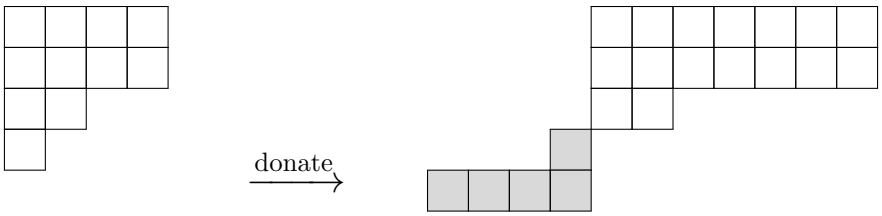


Figure 9: Take $d = 5$. On the left is the Young diagram corresponding to $\lambda = (4, 4, 2, 1, 0)$, with $d = 5$. On the right is donate$(\lambda) = (7, 7, 2, -1, -4)$. The gray-colored boxes correspond to rows with a negative length. Note that donate(λ) is not necessarily a Young diagram, since it may have negative entries, but donate(λ) is a weakly decreasing list of integers with sum $|\lambda|$, and donate(λ) is constant on the blocks of λ.

Definition 3.2 (Young diagram box donation; Theorem 1.1, restated)**.** Let $\lambda = (\lambda_1, \ldots, \lambda_d)$ be a Young diagram. Then donate$(\lambda) \in \mathbb{Z}^d$ is the vector defined as

$$\text{donate}(\lambda)_i = \lambda_i - \sum_{j=1}^{i-1} 1[\lambda_j > \lambda_i] + \sum_{j=i+1}^{d} 1[\lambda_i > \lambda_j].$$

Intuitively, each row of the Young diagram λ "donates" a box to each row which starts off longer than it and receives a box from each row which starts off shorter than it. Note that donate(λ) need not be a Young diagram, even if λ is, as it might have negative entries. We will also use the notation $\lambda^{\uparrow}$ for donate(λ).

See Figure 9 for an example of donation.

Definition 3.3 (Debiased Keyl's algorithm; Theorem 1.2, restated)**.** Given n copies of a quantum state $\rho \in \mathbb{C}^{d \times d}$, the *debiased Keyl's algorithm* works as follows.

1. Measure $\rho^{\otimes n}$ as in Keyl's algorithm (Theorem 3.1), producing a Young diagram $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_d) \vdash n$ and a unitary matrix $\boldsymbol{U} \in U(d)$.
2. Set $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\lambda}^{\uparrow}/n$. Output $\widehat{\boldsymbol{\rho}} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^{\dagger}$.

In Section 13, we prove that the debiased Keyl's algorithm is, in fact, an unbiased estimator of ρ.

3.3 A family of unbiased estimators

Note that our estimator is equivalent to performing Keyl's algorithm, and then replacing the empirical Young diagram $\boldsymbol{\lambda}/n$ with $\widehat{\boldsymbol{\alpha}}$. It turns out that in addition to the algorithm given in Theorem 3.3, there is a family of related unbiased estimators that use slightly different Young diagram transformations than $\boldsymbol{\lambda}^{\uparrow}$. In this section, we describe this family of unbiased estimators, while highlighting a few particular members that we will use in our proofs. We end this section by showing that the original estimator $\widehat{\boldsymbol{\alpha}}$ is the family member with the minimum variance.

We start by defining the set of legal Young diagram transformations, which generalizes the box donation from Theorem 1.1.

Definition 3.4 (Legal Young diagram transformation)**.** A map f from Young diagrams to $\mathbb{R}^d$ is a *legal Young diagram transformation* if for all Young diagrams $\lambda = (\lambda_1, \ldots, \lambda_d)$, $\mu = f(\lambda) \in \mathbb{R}^d$ is a vector that satisfies

$$\sum_{i \in B} \mu_i = \sum_{i \in B} \lambda_i^{\uparrow}$$

for every block of rows B in λ.

In words, any Young diagram transformation f of $\lambda \vdash n$ is legal if the total number of "boxes" in a block B after transformation f (allowing for the possibility of negative or, more generally, real-valued amounts of boxes) equals the number of boxes in that block after the box donation transformation of Theorem 1.1. An

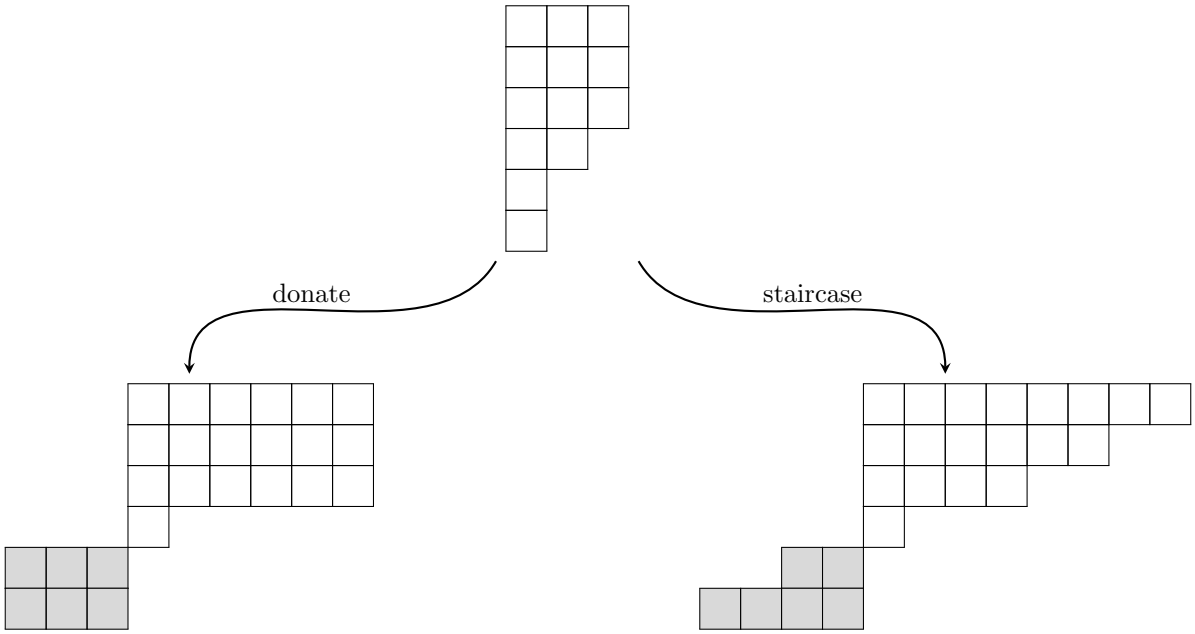


Figure 10: Two examples of legal Young diagram transformations for $\lambda = (3, 3, 3, 2, 1, 1)$. The Young diagram λ has 3 blocks that consist of the first three rows, the fourth row, and the last two rows, respectively. The diagram on the left is the one produced from the original box donation algorithm: $\lambda^{\uparrow} = (6, 6, 6, 1, -3, -3)$. The one on the right can be obtained by moving boxes within its blocks, and it coincides with the staircase box donation from Theorem 3.7: $\lambda^{\uparrow\uparrow} = (8, 6, 4, 1, -2, -4)$. The gray-colored boxes correspond to rows with negative lengths.

equivalent condition is that any legal transformation f can be obtained from $\lambda^{\uparrow}$ by arbitrarily redistributing boxes among the rows in the same block B. In addition to having real-valued lengths which may be negative or non-integral, the resulting shape is not necessarily a Young diagram, also due to the possibility of non-monotone row lengths, as illustrated in Figure 12. Given any legal Young diagram transformation, we can define an unbiased estimator for entangled state tomography.

Definition 3.5 (Family of unbiased estimators)**.** Suppose we are given n copies of a mixed state $\rho \in \mathbb{C}^{d\times d}$. For any legal Young diagram transformation f, we define an estimator as follows.

1. Measure $\rho^{\otimes n}$ as in Keyl's algorithm, producing a Young diagram $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_d) \vdash n$ and a unitary matrix $\boldsymbol{U} \in U(d)$.
2. Set $\boldsymbol{\lambda}' = f(\boldsymbol{\lambda})$ and $\widehat{\boldsymbol{\alpha}}' = \boldsymbol{\lambda}'/n$. Output $\widehat{\boldsymbol{\rho}}' = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}' \cdot \boldsymbol{U}^{\dagger}$.

We will often use the term *legal estimator* to refer to such an estimator.

We will prove in Section 13 that all legal Young diagram transformations produce unbiased estimators.

Theorem 3.6 (All legal estimators are unbiased)**.** *Let $\widehat{\boldsymbol{\rho}}'$ be the output of a legal estimator when run on $\rho^{\otimes n}$. Then $\mathbf{E}[\widehat{\boldsymbol{\rho}}'] = \rho$.*

Below, we highlight a few legal estimators that will be useful for proving the results in this paper. We start by defining the *staircase estimator*, which is perhaps the simplest one, as its defining Young diagram transformation can be described using a closed-form expression.

Definition 3.7 (Young diagram staircase box donation)**.** Let $\lambda = (\lambda_1, \ldots, \lambda_d)$ be a Young diagram. Then $\text{staircase}(\lambda) \in \mathbb{Z}^d$ is the vector defined

$$\text{staircase}(\lambda)_i = \lambda_i - \sum_{j=1}^{i-1} 1 + \sum_{j=i+1}^{d} 1 = \lambda_i - (i-1) + (d-i) = \lambda_i + d - 2i + 1.$$

Compared to the original box donation from Theorem 1.1, in the staircase donation transformation, each row of the Young diagram λ donates a box to each row above it, regardless of whether they started with the same row length. We can also model it as a two-step process, where each row first donates boxes according to the original box donation algorithm, and then the rows further donate one box to each row above them in the same block. Since the second step only involves moving boxes within the same block, it does not change the total number of boxes in each block. This confirms that the staircase donation transformation is a legal donation transformation. We include a diagram of $\text{staircase}(\lambda)$ in Figure 10. Similar to $\text{donate}(\lambda)$, $\text{staircase}(\lambda)$ need not be a Young diagram, due to potentially negative entries.

Using this Young diagram transformation, we define the *staircase estimator* and use the notation $\lambda^{\uparrow\uparrow} := \text{staircase}(\lambda)$ and $\widehat{\boldsymbol{\alpha}}^{\uparrow\uparrow} := \boldsymbol{\lambda}^{\uparrow\uparrow}/n$. The final state estimate from this algorithm is denoted by $\widehat{\boldsymbol{\rho}}^{\uparrow\uparrow} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}^{\uparrow\uparrow} \cdot \boldsymbol{U}^{\dagger}$. We chose the double arrow superscript notation for this specific estimator, since we view the staircase box donation as a more "extreme" version of the original box donation.

The advantage of the staircase estimator is the simple closed-form expression of the transformation $\lambda^{\uparrow\uparrow}$. For this reason, the staircase estimator plays a central role in proving our expressions for the first and second moments. In particular, we first prove in Section 13 that $\widehat{\boldsymbol{\rho}}^{\uparrow\uparrow}$ is an unbiased estimator, and then extend this result to all legal estimators, including our original estimator $\widehat{\boldsymbol{\rho}}$ as in Theorem 1.3. Moreover, it turns out that the staircase estimator achieves the optimal sample complexity for tomography problems over general states, when the number of copies required is $n = \Omega(d^2)$. However, it can be suboptimal in the special case when ρ is promised to be of low rank. One example is the pure state case, where weak Schur sampling always produces the tableau $\boldsymbol{\lambda} = (n, 0, \ldots, 0)$. The two estimators will have spectra

$$\widehat{\boldsymbol{\alpha}} = \left(\frac{n+d-1}{n}, -\frac{1}{n}, \ldots, -\frac{1}{n}\right), \quad \widehat{\boldsymbol{\alpha}}^{\uparrow\uparrow} = \left(\frac{n+d-1}{n}, \frac{d-3}{n}, \frac{d-5}{n}, \ldots, \frac{-d+1}{n}\right).$$

Observe that $\widehat{\boldsymbol{\rho}}^{\uparrow\uparrow}$ is far from any pure state because the absolute sum of all of its eigenvalues except the first one is going to be $O(d^2/n)$. This implies that $\widehat{\boldsymbol{\rho}}^{\uparrow\uparrow}$ is only guaranteed to be close to a pure state when

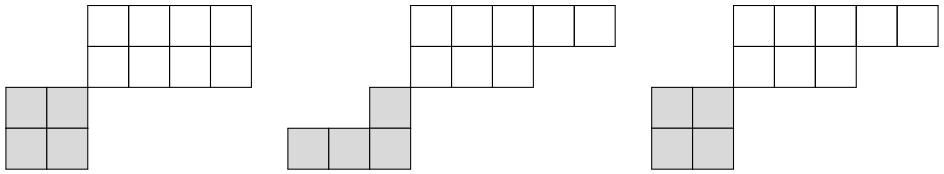


Figure 11: The original, staircase, and rank-2 Young diagram transformations for $\lambda = (2, 2, 0, 0)$.

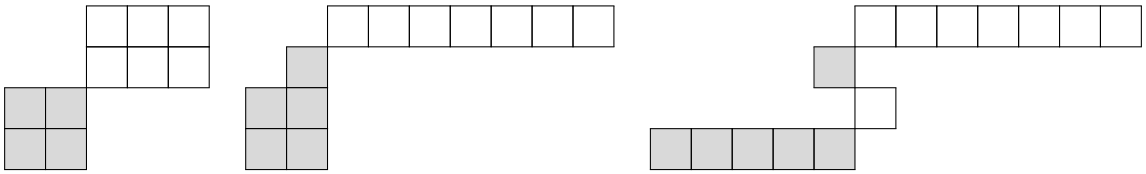


Figure 12: The transformations of $\lambda = (1, 1, 0, 0)$ represented above are also legal transformations, even though some of them seem weird. Indeed, the transformation on the right gives a shape whose row lengths are not monotonically decreasing.

the number of samples is $n \gg d^2$, which is much more than the optimal bound of $O(d/\varepsilon^2)$. The issue with the staircase transformation is that the large negative eigenvalues increase the variance of our estimator, and thus, we require more samples to approximate its true mean. As a result, when dealing with low-rank states, it will not suffice to use the staircase estimator if we want optimal bounds. However, there is a natural modification to the staircase estimator for the case of low-rank states, which reduces its variance while keeping its simplicity.

Definition 3.8 (Young diagram rank-r box donation)**.** Let $\lambda = (\lambda_1, \ldots, \lambda_d)$ be a Young diagram, with at most r nonzero rows. Then $\text{staircase}_r(\lambda) \in \mathbb{Z}^d$ is the vector defined as follows.

$$\text{staircase}_r(\lambda)_i = \begin{cases} \lambda_i + d - 2i + 1 & i \leq r \\ -r & i \in \{r+1, \ldots, d\}. \end{cases}$$

We first argue that this Young diagram transformation is a legal one. We model it as a two-step process: First, each row donates boxes according to the staircase algorithm. Since λ has at most r nonzero rows, there is a block of rows of λ that includes the last $d - r$ rows. The total number of boxes in the last $d - r$ rows of the block after the staircase donation is $n - \sum_{i=1}^{r}(\lambda_i + d - 2i + 1) = -(d - r)r$, which is a multiple of $d - r$. The second phase of the box donation will have the last $d - r$ rows donate to each other until they have the same number of boxes, for a total $-r$ each. The last step only involves donations within the same block, so the total number of boxes in each block does not change from the first phase. As the staircase transformation is legal, we conclude that the rank-r donation procedure is also legal.

First, observe that the transformation $\text{staircase}_r(\boldsymbol{\lambda})$ also has a simple closed-form expression. Using this Young diagram transformation, the final estimator will be $\widehat{\boldsymbol{\rho}}' = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}' \cdot \boldsymbol{U}^\dagger$, where $\widehat{\boldsymbol{\alpha}}' = \text{staircase}_r(\boldsymbol{\lambda})/n$. Moreover, when ρ is a rank-r state, the absolute sum of the $d - r$ smallest eigenvalues of $\widehat{\boldsymbol{\rho}}'$ will be $O(rd/n)$, compared to $O(d^2/n)$ for $\widehat{\boldsymbol{\rho}}^{\uparrow\uparrow}$. Looking forward, we will prove that this "rank-r" modification to the staircase estimator suffices to achieve the optimal bounds for rank-r states, where the number of copies required is $n = \Omega(rd)$.

Finally, we prove that the original estimator is the one with the minimum variance among all unbiased estimators from Theorem 3.5.

Lemma 3.9. *The estimator that corresponds to the Young diagram box donation of Theorem 1.1 has the minimal variance among all estimators in the family of estimators from Theorem 3.5.*

Proof. Let $\widehat{\boldsymbol{\rho}} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^\dagger$ for $\widehat{\boldsymbol{\alpha}} = \lambda^\uparrow/n$ be the original unbiased estimator, and let $\widehat{\boldsymbol{\rho}}' = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}' \cdot \boldsymbol{U}^\dagger$ be the unbiased estimator that corresponds to a legal Young diagram transformation f. Since f is a legal transformation, it holds that for any $\boldsymbol{\lambda}$ obtained from weak Schur sampling, $\widehat{\boldsymbol{\alpha}}'$ has the same total weight as

$\widehat{\boldsymbol{\alpha}}$ on each block of $\boldsymbol{\lambda}$, but not necessarily uniformly distributed among each row. The expected distance of $\widehat{\boldsymbol{\rho}}'$ from ρ in the Schatten 2-norm is

$$\begin{aligned}
\mathbf{E}\|\widehat{\boldsymbol{\rho}}' - \rho\|_2^2 &= \mathbf{E}\|(\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}}) + (\widehat{\boldsymbol{\rho}} - \rho)\|_2^2 \\
&= \mathbf{E}\|\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}}\|_2^2 + 2 \cdot \mathbf{E}\operatorname{tr}\left((\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}})(\widehat{\boldsymbol{\rho}} - \rho)\right) + \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 \\
&= \mathbf{E}\|\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}}\|_2^2 + 2 \cdot \mathbf{E}\operatorname{tr}\left((\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}})\widehat{\boldsymbol{\rho}}\right) - 2 \cdot \mathbf{E}\operatorname{tr}\left((\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}})\rho\right) + \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 \\
&= \mathbf{E}\|\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}}\|_2^2 + 2 \cdot \mathbf{E}\operatorname{tr}\left((\widehat{\boldsymbol{\rho}}' - \widehat{\boldsymbol{\rho}})\widehat{\boldsymbol{\rho}}\right) + \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 \qquad (\mathbf{E}\,\widehat{\boldsymbol{\rho}} = \mathbf{E}\,\widehat{\boldsymbol{\rho}}' = \rho) \\
&= \mathbf{E}\|\widehat{\boldsymbol{\alpha}}' - \widehat{\boldsymbol{\alpha}}\|_2^2 + 2 \cdot \mathbf{E}\operatorname{tr}\left((\widehat{\boldsymbol{\alpha}}' - \widehat{\boldsymbol{\alpha}})\widehat{\boldsymbol{\alpha}}\right) + \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2. \qquad (24)
\end{aligned}$$

In the last line, we have used the fact that $\widehat{\boldsymbol{\rho}}'$ and $\widehat{\boldsymbol{\rho}}$ are simultaneously diagonalized by $\boldsymbol{U}$. Now fix a tableau λ and consider a block of indices with λ_i constant. We will analyze the contribution of the first two terms on this block. Without loss of generality, let this block of indices be $\{1, \ldots, m\}$, and let $\{\widehat{\mu}_i'\}_{i=1}^m$ and $\{\widehat{\mu}_i\}_{i=1}^m$ be the entries of $f(\lambda)$ and $\lambda^\uparrow$ respectively on the rows in this block. Note that $\widehat{\mu}_i = \widehat{\mu} \coloneqq \frac{1}{m}\sum_{j=1}^m \widehat{\mu}_j'$ for all $i \in \{1, \ldots, m\}$. Then we have

$$\begin{aligned}
\sum_{i=1}^m (\widehat{\mu}_i' - \widehat{\mu}_i)^2 + 2\sum_{i=1}^m (\widehat{\mu}_i' - \widehat{\mu}_i)\widehat{\mu}_i &= \sum_{i=1}^m (\widehat{\mu}_i' - \widehat{\mu})^2 + 2\sum_{i=1}^m (\widehat{\mu}_i' - \widehat{\mu})\widehat{\mu} \\
&= \sum_{i=1}^m \widehat{\mu}_i'^2 - 2\widehat{\mu}\sum_{i=1}^m \widehat{\mu}_i' + m\widehat{\mu}^2 + 2\widehat{\mu}\sum_{i=1}^m \widehat{\mu}_i' - 2m\widehat{\mu}^2 \\
&= \sum_{i=1}^m \widehat{\mu}_i'^2 - m\widehat{\mu}^2 \\
&= m\left(\frac{1}{m}\sum_{i=1}^m \widehat{\mu}_i'^2 - \left(\frac{1}{m}\sum_{i=1}^m \widehat{\mu}_i'\right)^2\right) \geq 0,
\end{aligned}$$

where the last inequality follows from Cauchy-Schwarz. Since $\widehat{\boldsymbol{\alpha}}'$ and $\widehat{\boldsymbol{\alpha}}$ are diagonal matrices with entries $f(\boldsymbol{\lambda})/n$ and $\boldsymbol{\lambda}^\uparrow/n$ respectively, we conclude that for every tableau $\boldsymbol{\lambda}$ produced by weak Schur sampling, the first two terms of Equation (24) are positive, and thus

$$\mathbf{Var}(\widehat{\boldsymbol{\rho}}') = \mathbf{E}\|\widehat{\boldsymbol{\rho}}' - \rho\|_2^2 \geq \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 = \mathbf{Var}(\widehat{\boldsymbol{\rho}}). \qquad \square$$

Remark 3.10 (Unitary invariance of Keyl measurement)**.** Throughout this paper, we will consider distances or divergences that are unitarily invariant, that is, distances D that satisfy $\mathrm{D}(\rho, \sigma) = \mathrm{D}(U\rho U^\dagger, U\sigma U^\dagger)$. Any algorithm from our family of unbiased estimators, when run on a state $\rho = V \cdot \alpha \cdot V^\dagger$ with spectrum α, has the form:

$$\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha), \quad \boldsymbol{U} \sim K_{\boldsymbol{\lambda}}(\rho), \quad \widehat{\boldsymbol{\alpha}} = \operatorname{diag}(f(\boldsymbol{\lambda})/n), \quad \widehat{\boldsymbol{\rho}} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^\dagger.$$

Here, f is some legal Young diagram transformation. It holds that

$$\mathrm{D}(\widehat{\boldsymbol{\rho}}, \rho) = \mathrm{D}(\boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^\dagger, V \cdot \alpha \cdot V^\dagger) = \mathrm{D}(V^\dagger \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^\dagger V, \alpha).$$

Moreover, from Equation (23) we see that

$$\begin{aligned}
\operatorname{tr}(M_{\lambda, V^\dagger \boldsymbol{U}} \cdot \nu_\lambda(\alpha)) &= \dim(V_\lambda^d) \cdot \langle T_\lambda | \, \nu_\lambda(\boldsymbol{U}^\dagger V \alpha V^\dagger \boldsymbol{U}) \, | T_\lambda \rangle \\
&= \dim(V_\lambda^d) \cdot \langle T_\lambda | \, \nu_\lambda(\boldsymbol{U}^\dagger \rho \boldsymbol{U}) \, | T_\lambda \rangle = \operatorname{tr}(M_{\lambda, \boldsymbol{U}} \cdot \nu_\lambda(\rho)).
\end{aligned}$$

In words, the probability density of $\boldsymbol{U}$ under $K_{\boldsymbol{\lambda}}(\rho)$ is equal to the probability density of $V^\dagger \boldsymbol{U}$ under $K_{\boldsymbol{\lambda}}(\alpha)$. Therefore, when analyzing a unitarily invariant distance, we may assume, without loss of generality, that $\rho = \alpha$ is diagonal, and our algorithm operates as follows:

$$\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha), \quad \boldsymbol{U} \sim K_{\boldsymbol{\lambda}}(\alpha), \quad \widehat{\boldsymbol{\alpha}} = \operatorname{diag}(f(\boldsymbol{\lambda})/n), \quad \widehat{\boldsymbol{\rho}} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^\dagger.$$

Part II
Applications

4 Full state tomography in trace distance

In this section, we formally prove Theorem 1.5, that our debiased Keyl's estimator can learn a rank-r quantum state to trace distance ε, using $O(rd/\varepsilon^2)$ copies of ρ. Our bound matches the upper bound from Corollary 1.4 of [OW16], and the lower bound from Theorem 1.1 of [SSW25].

Theorem 4.1 (Theorem 1.5 restated)**.** *Let ρ be a rank-r quantum state, and $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $\rho^{\otimes n}$. Then $\mathrm{D}_{\mathrm{tr}}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability when $n = O(rd/\varepsilon^2)$.*

Before we prove the main theorem of this section, we prove a bound on the variance of our estimator $\widehat{\boldsymbol{\rho}}$.

Lemma 4.2 (Variance of $\widehat{\boldsymbol{\rho}}$)**.** *Let ρ be a quantum state, and $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $\rho^{\otimes n}$. Then*

$$\mathbf{Var}[\widehat{\boldsymbol{\rho}}] = \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 \leq \frac{2d}{n} + \frac{d^2 \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})}{n^2}.$$

Proof. Since $\widehat{\boldsymbol{\rho}}$ is unbiased, we have

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 = \mathbf{E}\,\mathrm{tr}\left((\widehat{\boldsymbol{\rho}} - \rho)^2\right) = \mathbf{E}\,\mathrm{tr}\big(\widehat{\boldsymbol{\rho}}^2\big) - \mathbf{E}\,\mathrm{tr}(\rho^2) = \mathbf{E}\,\mathrm{tr}\big(\widehat{\boldsymbol{\rho}}^2\big) - \mathrm{tr}(\rho^2). \tag{25}$$

We use our expression for the second moment from Theorem 1.4 to write

$$\begin{aligned}
\mathbf{E}\,\mathrm{tr}\big(\widehat{\boldsymbol{\rho}}^2\big) &= \mathbf{E}\,\mathrm{tr}\left(\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}} \cdot \mathrm{SWAP}\right) \\
&= \mathrm{tr}\left(\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] \cdot \mathrm{SWAP}\right) \\
&= \frac{n-1}{n}\,\mathrm{tr}\left(\rho \otimes \rho \cdot \mathrm{SWAP}\right) + \frac{1}{n}\,\mathrm{tr}\left(\rho \otimes I + I \otimes \rho\right) + \frac{\mathbf{E}\,\ell(\boldsymbol{\lambda})}{n^2} \cdot \mathrm{tr}(I \otimes I) - \mathrm{tr}(\mathrm{Lower}_\rho \cdot \mathrm{SWAP}) \\
&\leq \mathrm{tr}(\rho^2) + \frac{2d}{n} + \frac{d^2 \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})}{n^2}.
\end{aligned}$$

In the last step, we are able to drop $\mathrm{tr}(\mathrm{Lower}_\rho \cdot \mathrm{SWAP})$, since Lower_ρ is a positive linear combination of terms of the form $(P \otimes P) \cdot \mathrm{SWAP}$, and since $\mathrm{tr}((P \otimes P) \cdot \mathrm{SWAP} \cdot \mathrm{SWAP}) = \mathrm{tr}(P \otimes P) = \mathrm{tr}(P)^2$. Plugging this back into Equation (25) completes the proof. □

We are now ready to prove the main result of this section.

Proof of Theorem 4.1. Since ρ has rank r, the Young tableau $\boldsymbol{\lambda}$ that we obtain from weak Schur sampling will have at most r nonempty rows (see Theorem 2.50). Let us decompose $\widehat{\boldsymbol{\rho}}$ as

$$\widehat{\boldsymbol{\rho}} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}_{>0} \cdot \boldsymbol{U}^\dagger + \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}_{=0} \cdot \boldsymbol{U}^\dagger = \widehat{\boldsymbol{\rho}}_{>0} + \widehat{\boldsymbol{\rho}}_{=0},$$

where

$$\widehat{\boldsymbol{\alpha}}_{>0} \coloneqq \sum_{i=1}^{\ell(\boldsymbol{\lambda})} \widehat{\boldsymbol{\alpha}}_i\,|i\rangle\langle i| = \sum_{i,\boldsymbol{\lambda}_i > 0} \widehat{\boldsymbol{\alpha}}_i\,|i\rangle\langle i|\,, \quad \widehat{\boldsymbol{\alpha}}_{=0} \coloneqq \sum_{i=\ell(\boldsymbol{\lambda})+1}^{d} \widehat{\boldsymbol{\alpha}}_i\,|i\rangle\langle i| = \sum_{i,\boldsymbol{\lambda}_i = 0} \widehat{\boldsymbol{\alpha}}_i\,|i\rangle\langle i|\,.$$

In words, we split $\widehat{\boldsymbol{\rho}}$ into the state we obtain from the top $\ell(\boldsymbol{\lambda})$ eigenvalues (that correspond to nonempty rows of the Young diagram $\boldsymbol{\lambda}$), plus the state from the bottom $d - \ell(\boldsymbol{\lambda})$ eigenvalues (that correspond to the empty rows of $\boldsymbol{\lambda}$). Observe that $\widehat{\boldsymbol{\alpha}}_i = -\frac{\ell(\boldsymbol{\lambda})}{n}$ for all $i > \ell(\boldsymbol{\lambda})$, and thus

$$\widehat{\boldsymbol{\rho}}_{=0} = \boldsymbol{U} \cdot \Big(-\frac{\ell(\boldsymbol{\lambda})}{n} \sum_{i=\ell(\boldsymbol{\lambda})+1}^{d} |i\rangle\langle i|\Big) \cdot \boldsymbol{U}^\dagger = -\frac{\ell(\boldsymbol{\lambda})}{n} \cdot \boldsymbol{\Pi}_{=0},$$

where $\boldsymbol{\Pi}_{=0}$ is some projector of rank $d - \ell(\boldsymbol{\lambda})$. By the triangle inequality,

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_1 \leq \mathbf{E}\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_1 + \mathbf{E}\|\widehat{\boldsymbol{\rho}}_{=0}\|_1. \tag{26}$$

First, since $\widehat{\boldsymbol{\rho}}_{=0} = -\frac{\ell(\boldsymbol{\lambda})}{n} \cdot \boldsymbol{\Pi}_{=0}$ and $\ell(\boldsymbol{\lambda}) \le r$:

$$\|\widehat{\boldsymbol{\rho}}_{=0}\|_1 = \frac{\ell(\boldsymbol{\lambda})}{n} \cdot (d - \ell(\boldsymbol{\lambda})) \le \frac{rd}{n}. \tag{27}$$

Next, by Cauchy-Schwarz and Jensen's inequality:

$$\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_1 \le \sqrt{2r} \cdot \mathbf{E}\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_2 \le \sqrt{2r} \cdot \sqrt{\mathbf{E}\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_2^2}.$$

Then, we have

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_2^2 = \mathbf{E}\|(\widehat{\boldsymbol{\rho}} - \rho) - \widehat{\boldsymbol{\rho}}_{=0}\|_2^2 = \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 - 2\,\mathbf{E}\,\mathrm{tr}((\widehat{\boldsymbol{\rho}} - \rho)\widehat{\boldsymbol{\rho}}_{=0}) + \mathbf{E}\|\widehat{\boldsymbol{\rho}}_{=0}\|_2^2.$$

We expand the second term as follows

$$\begin{aligned} 2\,\mathbf{E}\,\mathrm{tr}((\widehat{\boldsymbol{\rho}} - \rho)\widehat{\boldsymbol{\rho}}_{=0}) &= 2\,\mathbf{E}\,\mathrm{tr}(\widehat{\boldsymbol{\rho}}_{>0} + \widehat{\boldsymbol{\rho}}_{=0} - \rho)\widehat{\boldsymbol{\rho}}_{=0}) \\ &= 2\,\mathbf{E}\,\mathrm{tr}(\widehat{\boldsymbol{\rho}}_{>0} \cdot \widehat{\boldsymbol{\rho}}_{=0}) + 2\,\mathbf{E}\|\widehat{\boldsymbol{\rho}}_{=0}\|_2^2 - 2\,\mathbf{E}\,\mathrm{tr}(\rho \cdot \widehat{\boldsymbol{\rho}}_{=0}) \\ &= 2\,\mathbf{E}\|\widehat{\boldsymbol{\rho}}_{=0}\|_2^2 - 2\,\mathbf{E}\,\mathrm{tr}(\rho \cdot \widehat{\boldsymbol{\rho}}_{=0}), \end{aligned}$$

where we have used the fact that $\mathrm{tr}(\widehat{\boldsymbol{\rho}}_{>0} \cdot \widehat{\boldsymbol{\rho}}_{=0}) = \mathrm{tr}(\widehat{\boldsymbol{\alpha}}_{>0} \cdot \widehat{\boldsymbol{\alpha}}_{=0}) = 0$, as $\widehat{\boldsymbol{\alpha}}_{>0}$ and $\widehat{\boldsymbol{\alpha}}_{=0}$ are supported on complementary subsets of computational basis vectors. Thus $\mathbf{E}\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_2^2$ becomes

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_2^2 = \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 + 2\,\mathbf{E}\,\mathrm{tr}(\rho \cdot \widehat{\boldsymbol{\rho}}_{=0}) - \mathbf{E}\|\widehat{\boldsymbol{\rho}}_{=0}\|_2^2 \le \mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 + 2\,\mathbf{E}\,\mathrm{tr}(\rho \cdot \widehat{\boldsymbol{\rho}}_{=0}).$$

Then, since $\ell(\boldsymbol{\lambda})$ is always at most r, Theorem 4.2 implies that

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2 \le \frac{2d}{n} + \frac{rd^2}{n^2}.$$

Moreover,

$$\mathrm{tr}(\rho \cdot \widehat{\boldsymbol{\rho}}_{=0}) = -\frac{\ell(\boldsymbol{\lambda})}{n} \cdot \mathrm{tr}(\rho \cdot \boldsymbol{\Pi}_{=0}) \le 0.$$

Thus,

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}}_{>0} - \rho\|_2^2 \le \frac{2d}{n} + \frac{rd^2}{n^2}. \tag{28}$$

Plugging Equations (27) and (28) into Equation (26), we get

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_1 \le \sqrt{\frac{4rd}{n} + \frac{2r^2d^2}{n^2}} + \frac{rd}{n},$$

which is $O(\varepsilon)$ when $n = O(rd/\varepsilon^2)$. □

Alternative proof using Young diagram statistics. There is another way of proving Theorem 4.1 without relying on our second moment calculation. Let $\widehat{\boldsymbol{\rho}}' = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}' \cdot \boldsymbol{U}^\dagger$ be the member of the family of unbiased estimators that uses the rank-r box donation transformation from Theorem 3.8. The spectrum of $\widehat{\boldsymbol{\rho}}'$ satisfies

$$\widehat{\boldsymbol{\alpha}}'_i = \begin{cases} \frac{\boldsymbol{\lambda}_i + d - 2i + 1}{n} & i \le r \\ -\frac{r}{n} & i \in \{r+1, \ldots, d\}. \end{cases}$$

We will obtain a slightly weaker bound for the variance of $\widehat{\boldsymbol{\rho}}$ as follows: $\mathbf{Var}[\widehat{\boldsymbol{\rho}}] \le \mathbf{Var}[\widehat{\boldsymbol{\rho}}'] \le \frac{2d}{n} + \frac{rd^2}{n^2}$. We proved the first inequality in Theorem 3.9, and for the second inequality, we write

$$\begin{aligned} \mathbf{Var}[\widehat{\boldsymbol{\rho}}'] &= \mathbf{E}\,\mathrm{tr}(\widehat{\boldsymbol{\rho}}'^2) - p_2(\alpha) && \text{(Equation (25))} \\ &= \mathbf{E}\left[\sum_{i=1}^{d} \widehat{\boldsymbol{\alpha}}_i'^2\right] - p_2(\alpha) \\ &= \mathbf{E}\left[\sum_{i=1}^{r} \frac{(\boldsymbol{\lambda}_i + (d - 2i + 1))^2}{n^2}\right] + \sum_{i=r+1}^{d} \frac{(-r)^2}{n^2} - p_2(\alpha) \end{aligned} \tag{29}$$

The first term is related to the second moment of the row lengths of the Young diagram random variable $\boldsymbol{\lambda}$. Fortunately, these moments can be computed using *Kerov's algebra of observables of diagrams* [Wri16]. In particular, the following member of this algebra captures the second moment of a Young diagram λ:

$$p_2^*(\lambda) = \sum_{i=1}^{d} \left(\left(\lambda_i - i + \frac{1}{2} \right)^2 - \left(-i + \frac{1}{2} \right)^2 \right) = \sum_{i=1}^{d} \left(\lambda_i^2 - \lambda_i(2i-1) \right).$$

Proposition 2.34 of [OW15b] computes the expectation of this polynomial to be

$$\mathbf{E}[p_2^*(\boldsymbol{\lambda})] = n(n-1)p_2(\alpha). \tag{30}$$

We conclude that

$$\begin{aligned}
(29) &\leq \frac{1}{n^2} \mathbf{E} \left[\sum_{i=1}^{d} \left(\boldsymbol{\lambda}_i^2 - \boldsymbol{\lambda}_i(2i-1) \right) \right] + \frac{1}{n^2} \mathbf{E} \left[\sum_{i=1}^{r} \boldsymbol{\lambda}_i(2d-2i+1) \right] + \frac{1}{n^2} \sum_{i=1}^{r} (d-2i+1)^2 + \frac{r^2 d}{n^2} - p_2(\alpha) \\
&\leq \frac{1}{n^2} \cdot \mathbf{E}[p_2^*(\boldsymbol{\lambda})] + \frac{2d}{n^2} \mathbf{E} \left[\sum_{i=1}^{r} \boldsymbol{\lambda}_i \right] + \frac{rd^2}{n^2} + \frac{r^2 d}{n^2} - p_2(\alpha) \\
&\leq \frac{2d}{n} + \frac{rd^2}{n^2}. && \text{(Equation (30), } \textstyle\sum_{i=1}^{r} \boldsymbol{\lambda}_i = n)
\end{aligned}$$

Given the variance bound, we can prove the statement of Theorem 4.1 using the same proof.

5 Full state tomography in Bures distance

In this section, we formally prove Theorem 1.6, that our debiased Keyl's estimator can learn a rank-r quantum state to Bures distance ε, using $O(rd/\varepsilon^2)$ copies of ρ. Equivalently, $O(rd/\delta)$ samples suffice to learn to infidelity error δ. This matches the lower bound proven in [Yue23].

This section is structured as follows.

1. As a warm-up, we present in Section 5.1 a simpler argument that shows how to learn a full-rank quantum state ρ to Bures distance ε using $O(d^2/\varepsilon^2)$ copies.
2. The complete argument for rank-r states appears in Section 5.2.
3. As a further application of our Bures distance tomography result, we present in Section 5.3 an algorithm for learning a bipartite pure state $|\psi\rangle_{AB}$ with Schmidt rank r using $O(r(d_A + d_B)/\varepsilon^2)$ copies of $|\psi\rangle_{AB}$.

Let us now make a couple of remarks that will be useful in multiple subsections below. First, we will use the Bures χ^2-divergence to bound the Bures distance, both of which are unitarily invariant. Following Theorem 3.10, we will therefore assume throughout this section, without loss of generality, that the state we are trying to learn ρ is diagonal with sorted eigenvalues $\alpha = (\alpha_1, \ldots, \alpha_d)$. Second, we will use the following lemma to upper bound the expectation of the square of the entries of $\widehat{\boldsymbol{\rho}}$.

Lemma 5.1. *Given a rank-r diagonal mixed state $\rho \in \mathbb{C}^{d \times d}$ with eigenvalues $\alpha_1, \ldots, \alpha_d$, let $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm on $\rho^{\otimes n}$. Then*

$$\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2 \right] \leq |\rho_{ij}|^2 + \frac{\alpha_i + \alpha_j}{n} + \frac{r}{n^2}.$$

Proof. Using the fact that $\widehat{\boldsymbol{\rho}}$ is Hermitian, we write

$$\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2 \right] = \mathbf{E}[\widehat{\boldsymbol{\rho}}_{ij} \cdot \widehat{\boldsymbol{\rho}}_{ji}] = \mathbf{E}\left[\langle i| \widehat{\boldsymbol{\rho}} |j\rangle \cdot \langle j| \widehat{\boldsymbol{\rho}} |i\rangle \right] = \langle ij| \cdot \mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] \cdot |ji\rangle.$$

We use our second moment result from Theorem 1.4 to deduce that $\mathbf{E}\,|\widehat{\boldsymbol{\rho}}_{ij}|^2$ is equal to

$$
\begin{aligned}
&= \frac{n-1}{n}\,\langle ij|\,\rho\otimes\rho\,|ji\rangle + \frac{1}{n}\,\langle ij|\,(\rho\otimes I + I\otimes\rho)\cdot \mathrm{SWAP}\,|ji\rangle + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2}\cdot\langle ij|\,\mathrm{SWAP}\,|ji\rangle - \langle ij|\,\mathrm{Lower}_\rho\,|ji\rangle \\
&= \frac{n-1}{n}\cdot|\rho_{ij}|^2 + \frac{1}{n}\cdot\langle ij|\,(\rho\otimes I + I\otimes\rho)\,|ij\rangle + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} - \langle ij|\,\mathrm{Lower}_\rho\,|ji\rangle \\
&\le |\rho_{ij}|^2 + \frac{1}{n}\,(\rho_{ii}+\rho_{jj}) + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \\
&\le |\rho_{ij}|^2 + \frac{\alpha_i+\alpha_j}{n} + \frac{r}{n^2}. && \text{(Because } \ell(\boldsymbol{\lambda})\le r \text{ when } \rho \text{ is rank } r\text{)}
\end{aligned}
$$

The first inequality follows from our characterization of Lower_ρ: we can expand $\langle ij|\cdot\mathrm{Lower}_\rho\cdot|ji\rangle$ as a positive linear combination of terms of the form

$$\langle ij|\cdot(P\otimes P)\cdot\mathrm{SWAP}\,|ji\rangle = \langle ij|\cdot(P\otimes P)\cdot|ij\rangle = P_{ii}\cdot P_{jj}\ge 0,$$

where the last step uses the fact that P is positive semidefinite. □

In contrast to learning in trace distance, the Bures distance introduces additional challenges that we have to overcome in the following subsections:

1. The first one is that our debiased Keyl's algorithm can return a matrix $\widehat{\boldsymbol{\rho}}$ with negative eigenvalues. Since the definition of the Bures distance involves the square root of matrices, it is not well-defined for an estimator that is not positive semidefinite.

2. The second challenge arises because we use Theorem 2.8 to upper bound the Bures distance by the Bures χ^2-divergence, which is sensitive to the small eigenvalues of ρ. Thus, we would like to make sure that the eigenvalues of ρ are large enough to obtain a useful bound on the Bures χ^2-divergence.

To deal with these challenges that are specific to the Bures distance, we will use two modified estimators in Sections 5.1 and 5.2 that go beyond our debiased Keyl's estimator.

5.1 Full-rank tomography

The main result of this section is an algorithm for learning a mixed state ρ to Bures distance ε using $O(d^2/\varepsilon^2)$ copies. We deal with the second challenge of learning in Bures distance, that is, when ρ has some eigenvalues which are too small, by first applying the depolarizing channel to each copy of ρ with noise rate $C\cdot\frac{d^2}{n}$, where C is a constant that we will choose later. That is, we obtain copies of the regularized state

$$\rho' \coloneqq \left(1 - C\cdot\frac{d^2}{n}\right)\cdot\rho + C\cdot\frac{d^2}{n}\cdot\frac{I}{d}.$$

Note that we will ultimately take $n=\Theta(d^2/\varepsilon^2)$ large enough so that the above procedure is well-defined (i.e., the noise rate does not exceed 1). The depolarizing noise implies that each eigenvalue of ρ' is at least $C\cdot\frac{d^2}{n}\cdot\frac{1}{d} = C\cdot\frac{d}{n}$, and as we will see, our Bures χ^2-divergence argument does not have to deal with eigenvalues that are too small.

We will show that when the number of samples is $n=\Theta(d^2/\varepsilon^2)$, the original ρ and depolarized state ρ' are $O(\varepsilon)$-close in Bures distance. Thus, we will use the debiased Keyl's algorithm to produce an estimate $\widehat{\boldsymbol{\rho}}$ that is $O(\varepsilon)$-close to ρ' in Bures distance, and it will also be a good estimate for the original state ρ.

In addition, we will show that the regularized state also deals with the first challenge of learning in the Bures distance, which is the negative eigenvalues of the estimator. In particular, we show in Theorem 5.2 that there exists a constant C such that the eigenvalues of the estimate $\widehat{\boldsymbol{\rho}}$ are all non-negative with high probability.

Lemma 5.2. *Let ρ' be a diagonal mixed state with sorted eigenvalues $\alpha'_1,\dots,\alpha'_d$, and $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $(\rho')^{\otimes n}$. There exists a constant C such that, if the eigenvalues of ρ' are all at least $C\cdot d/n$, then $\widehat{\boldsymbol{\rho}}$ is a positive semidefinite matrix with high probability.*

Proof. Recall that $\widehat{\boldsymbol{\rho}} = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}} \cdot \boldsymbol{U}^\dagger$, where $\widehat{\boldsymbol{\alpha}}$ is obtained by performing the Young diagram box donation procedure on $\boldsymbol{\lambda}$ and normalizing by n. Then the smallest eigenvalue of $\widehat{\boldsymbol{\rho}}$ is lower bounded by

$$\widehat{\boldsymbol{\alpha}}_d \geq \frac{\boldsymbol{\lambda}_d - d}{n}. \tag{31}$$

Our goal is to show that $\widehat{\boldsymbol{\alpha}}_d$ is non-negative. For this, we will need to understand the distribution of $\boldsymbol{\lambda}$ when drawn from $\mathrm{SW}^n(\alpha')$. Theorems 1.4 and 1.5 of [OW17] give bounds on the first and second moments of $\boldsymbol{\lambda}_k$ respectively. In particular, for $k \in [d]$ and $\nu_k \coloneqq \min\{1, \alpha'_k d\}$, the first moment of $\boldsymbol{\lambda}_k$ satisfies

$$\alpha'_k n - 2\sqrt{\nu_k n} \leq \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha')}[\boldsymbol{\lambda}_k] \leq \alpha'_k n + 2\sqrt{\nu_k n}, \tag{32}$$

and the second moment of $\boldsymbol{\lambda}_k$ satisfies

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha')}\left[(\boldsymbol{\lambda}_k - \alpha'_k n)^2\right] \leq O(\nu_k n). \tag{33}$$

We can bound the variance of $\boldsymbol{\lambda}_d$ using Equations (32) and (33):

$$\begin{aligned}
\mathbf{Var}[\boldsymbol{\lambda}_d] &= \mathop{\mathbf{E}}_{\boldsymbol{\lambda}}\left[(\boldsymbol{\lambda}_d - \mathop{\mathbf{E}}_{\boldsymbol{\lambda}}[\boldsymbol{\lambda}_d])^2\right] = \mathop{\mathbf{E}}_{\boldsymbol{\lambda}}\left[\left((\boldsymbol{\lambda}_d - \alpha'_d n) + (\alpha'_d n - \mathop{\mathbf{E}}_{\boldsymbol{\lambda}}[\boldsymbol{\lambda}_d])\right)^2\right] \\
&\leq \mathop{\mathbf{E}}_{\boldsymbol{\lambda}}\left[2(\boldsymbol{\lambda}_d - \alpha'_d n)^2 + 2(\alpha'_d n - \mathop{\mathbf{E}}_{\boldsymbol{\lambda}}[\boldsymbol{\lambda}_d])^2\right] \leq O(\nu_d n). && \text{(Cauchy-Schwarz)}
\end{aligned}$$

Now we will use the fact that $\nu_d \leq \alpha'_d d$ to deduce that

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha')}[\boldsymbol{\lambda}_d] \geq \alpha'_d n - 2\sqrt{\alpha'_d dn}, \qquad \mathbf{Var}[\boldsymbol{\lambda}_d] \leq O(\alpha'_d dn).$$

So, Chebyshev's inequality says that with high probability, $\boldsymbol{\lambda}_d$ is within $O(\sqrt{\mathbf{Var}[\boldsymbol{\lambda}_d]}) \leq O(\sqrt{\alpha'_d dn})$ of its expectation. In particular, there exists a constant C' such that, with high probability,

$$\boldsymbol{\lambda}_d \geq \alpha'_d n - C'\sqrt{\alpha'_d nd}.$$

Therefore, from Equation (31), it suffices to show that we can pick our C large enough so that

$$\alpha'_d n - C'\sqrt{\alpha'_d nd} - d \geq 0.$$

From the AM-GM inequality, $C'\sqrt{\alpha'_d nd} \leq \frac{1}{2}\alpha'_d n + \frac{1}{2}(C')^2 d$, and thus it suffices to show that

$$\frac{1}{2}\alpha'_d n - \left(\frac{1}{2}(C')^2 + 1\right) d \geq 0.$$

Setting C to be a constant at least $(C')^2 + 2$, and recalling that $\alpha'_d \geq C \cdot \frac{d}{n}$, we conclude that

$$\frac{1}{2}\alpha'_d n - \left(\frac{1}{2}(C')^2 + 1\right) d \geq \left(\frac{1}{2}C - \frac{1}{2}(C')^2 - 1\right) d \geq 0. \qquad \square$$

We are now ready to prove the main result of this subsection: that $\widehat{\boldsymbol{\rho}}$ is a good estimator for ρ in Bures distance.

Theorem 5.3 (Bures distance tomography)**.** *Given a mixed state $\rho \in \mathbb{C}^{d \times d}$, $n = O(d^2/\varepsilon^2)$ samples suffice to output an estimator $\widehat{\boldsymbol{\rho}}$ such that $\mathrm{D_B}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability.*

Proof. Let C be the constant from the statement of Theorem 5.2, and let ρ' be the state we obtain when we apply the depolarizing channel with noise rate $C \cdot d^2/n$ to ρ:

$$\rho' \coloneqq \left(1 - C \cdot \frac{d^2}{n}\right) \cdot \rho + C \cdot \frac{d^2}{n} \cdot \frac{I}{d}.$$

We will take $n \geq C \cdot d^2/\varepsilon^2$, and thus the noise rate does not exceed 1. We will denote the eigenvalues of ρ' by $\alpha'_1, \ldots, \alpha'_d$, and observe that the smallest eigenvalue of ρ' is at least $C \cdot d^2/n \cdot 1/d = C \cdot d/n$. On the other hand, this regularized state ρ' is still $O(\varepsilon)$-close to ρ in Bures distance, since

$$\frac{1}{2}\mathrm{D}_{\mathrm{B}}^2(\rho, \rho') \leq \mathrm{D}_{\mathrm{tr}}(\rho, \rho') = \frac{1}{2}\|\rho - \rho'\|_1 \leq \frac{1}{2}\|C \cdot d^2/n \cdot \rho\|_1 + \frac{1}{2}\|C \cdot d^2/n \cdot I/d\|_1 = O(d^2/n) = O(\varepsilon^2).$$

As a result, it suffices to learn ρ' up to $O(\varepsilon)$ error instead of ρ. From Theorem 5.2, we know that running the debiased Keyl's algorithm on $(\rho')^{\otimes n}$ will produce an estimator $\widehat{\boldsymbol{\rho}}$ that is positive semidefinite with high probability. Hence the Bures distance between $\widehat{\boldsymbol{\rho}}$ and ρ' is well defined, and can be upper bounded by

$$\mathbf{E}\left[\mathrm{D}_{\mathrm{B}}(\widehat{\boldsymbol{\rho}}, \rho')\right] \leq \mathbf{E}\left[\sqrt{\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \,\|\, \rho')}\right] \leq \sqrt{\mathbf{E}\left[\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \,\|\, \rho')\right]}. \tag{34}$$

The right-hand side of the above equation becomes:

$$\mathbf{E}\left[\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \,\|\, \rho')\right] = \sum_{i,j=1}^{d} \frac{2}{\alpha'_i + \alpha'_j} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij} - \rho'_{ij}|^2\right] = \sum_{i,j=1}^{d} \frac{2}{\alpha'_i + \alpha'_j} \cdot \left(\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2\right] - |\rho'_{ij}|^2\right). \tag{35}$$

Let us consider the term corresponding to a fixed pair (i, j):

$$\begin{aligned}
\frac{2}{\alpha'_i + \alpha'_j} \cdot \left(\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2\right] - |\rho'_{ij}|^2\right) &\leq \frac{2}{\alpha'_i + \alpha'_j} \cdot \left(\frac{\alpha'_i + \alpha'_j}{n} + \frac{d}{n^2}\right) && \text{(Theorem 5.1)} \\
&\leq \frac{2}{n} + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right). && (\alpha'_i, \alpha'_j \geq \Omega(d/n) \implies \alpha'_i + \alpha'_j \geq \Omega(d/n))
\end{aligned}$$

Equation (35) has a total of d^2 terms, each of which is at most $O\left(\frac{1}{n}\right)$ in expectation. We conclude that

$$\mathbf{E}\left[\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \,\|\, \rho')\right] = \sum_{i,j=1}^{d} \frac{2}{\alpha'_i + \alpha'_j} \cdot \left(\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2\right] - |\rho'_{ij}|^2\right) \leq O\left(\frac{d^2}{n}\right).$$

From our choice of $n \geq C \cdot d^2/\varepsilon^2 = \Theta(d^2/\varepsilon^2)$ sufficiently large, it holds that $C \cdot d^2/n = O(\varepsilon^2)$, and $\mathbf{E}\left[\mathrm{D}_{\chi^2}(\widehat{\boldsymbol{\rho}} \,\|\, \rho')\right] \leq O(\varepsilon^2)$. We conclude from Equation (34) that with high probability:

$$\mathbf{E}\left[\mathrm{D}_{\mathrm{B}}(\rho, \widehat{\boldsymbol{\rho}})\right] \leq \mathrm{D}_{\mathrm{B}}(\rho, \rho') + \mathbf{E}\left[\mathrm{D}_{\mathrm{B}}(\rho', \widehat{\boldsymbol{\rho}})\right] \leq O(\varepsilon).$$

The theorem statement follows by reducing ε by a constant factor, and applying Markov's inequality to deduce that $\mathrm{D}_{\mathrm{B}}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability. □

5.2 Low-rank tomography

The main result of this section is an algorithm for learning a mixed state of rank r to Bures distance ε using $O(rd/\varepsilon^2)$ copies.

Theorem 5.4. *Given a rank r mixed state $\rho \in \mathbb{C}^{d \times d}$, $n = O(rd/\varepsilon^2)$ samples suffice to output an estimator $\widetilde{\boldsymbol{\rho}}$ such that $\mathrm{D}_{\mathrm{B}}(\rho, \widetilde{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability.*

Recall the first challenge of learning in the Bures distance, which is the fact that $\widehat{\boldsymbol{\rho}}$ is not necessarily a positive semidefinite estimator. In the full rank case of the previous subsection, we dealt with this challenge by applying a particular depolarizing channel to each copy of our input state ρ to obtain copies of the regularized state ρ'. We then showed that the debiased Keyl's algorithm on $(\rho')^{\otimes n}$ will output a positive semidefinite estimator $\widehat{\boldsymbol{\rho}}$ with high probability, as long as the eigenvalues of ρ' are sufficiently large.

In the low rank case that we consider now, the number of samples does not suffice for this argument to succeed. Instead, we will execute our debiased Keyl's algorithm on the copies of the original state $\rho^{\otimes n}$ to produce $\widehat{\boldsymbol{\rho}}$, shift its eigenvalues by r/n, and renormalize:

$$\widetilde{\boldsymbol{\rho}} \coloneqq \frac{n}{n + rd}\left(\widehat{\boldsymbol{\rho}} + \frac{r}{n} \cdot I\right).$$

Since the smallest eigenvalue of $\widehat{\boldsymbol{\rho}}$ is at least $-\ell(\boldsymbol{\lambda})/n \geq -r/n$, the resulting state $\widetilde{\boldsymbol{\rho}}$ is positive semidefinite and will serve as our low-rank estimator for the Bures distance.

We deal with the second challenge of learning in the Bures distance, which is sensitivity of the Bures χ^2-divergence on the small eigenvalues of ρ, by following a strategy similar to the one Flammia and O'Donnell use in [FO24, Corollary 3.20] to analyze their Bures distance estimator. In particular, we will bound the Bures χ^2-divergence between $\widetilde{\boldsymbol{\rho}}$ and ρ when restricted to the large eigenvalues of ρ. We then show that the small eigenvalues of ρ do not contribute much more to the Bures distance.

To define the restriction of a state to a subset of eigenvalues, we introduce the following notation, which appears in [FO24, Notation 2.1].

Notation 5.5. Let $\sigma \in \mathbb{C}^{d\times d}$ be a matrix, and $R \subseteq [d]$. We define $\sigma[R] \in \mathbb{C}^{|R|\times|R|}$ to be the submatrix of σ formed by the rows and columns in R.

Note that when σ is a quantum state, $\sigma[R]$ may have trace less than one, but by Cauchy's interlacing theorem below, it still holds that every eigenvalue of $\sigma[R]$ is in $[0, 1]$.

Theorem 5.6 (Cauchy's interlacing theorem, [Fis05])**.** *Let $A \in \mathbb{C}^{d\times d}$ be a Hermitian matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_d$, and R be a subset of $[d]$ of size $d-1$. Then the Hermitian matrix $B = A[R] \in \mathbb{C}^{(d-1)\times(d-1)}$ has eigenvalues $\mu_1 \leq \cdots \leq \mu_{d-1}$ that satisfy*

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \cdots \leq \lambda_{d-1} \leq \mu_{d-1} \leq \lambda_d.$$

Roughly speaking, we will define $R \subseteq [d]$ to be the subset of indices corresponding to the large eigenvalues of ρ, and thus the expression

$$\mathrm{D}_{\chi^2}\left(\widetilde{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right) = \sum_{i,j\in R} \frac{2}{\alpha_i + \alpha_j} \cdot |\widetilde{\boldsymbol{\rho}}_{ij} - \rho_{ij}|^2$$

will capture the contribution of the large eigenvalues to the Bures χ^2-divergence. Note that this expression is well-defined as long as $\alpha_i > 0$ for each $i \in R$.

In addition, we use $\sigma^{\mathrm{norm}}[R]$ to denote the normalized version of $\sigma[R]$.

Notation 5.7. Let $\sigma \in \mathbb{C}^{d\times d}$ be a quantum state, and $R \subseteq [d]$. Then $\sigma^{\mathrm{norm}}[R] \in \mathbb{C}^{|R|\times|R|}$ is the $|R|$-dimensional quantum state given by

$$\sigma^{\mathrm{norm}}[R] \coloneqq \frac{\sigma[R]}{\mathrm{tr}(\sigma[R])}.$$

Notation 5.8. For a mixed state σ and projector Π, we use

$$\sigma|_{\Pi} \coloneqq \frac{\Pi \cdot \sigma \cdot \Pi}{\mathrm{tr}(\Pi \cdot \sigma)}$$

to denote the post-measurement state if σ is measured with the projective measurement $\{\Pi, I - \Pi\}$ and the outcome Π is obtained.

Observe that if Π is the projector onto the space defined by the subset R, then the matrices $\sigma^{\mathrm{norm}}[R]$ and $\sigma|_{\Pi}$ correspond to the same quantum state, with the difference that the latter matrix corresponds to the state that lives in the larger ambient space $\mathbb{C}^{d\times d}$.

Proof of Theorem 5.4. To make sure that our constants work out, we will always set $n \geq 400 \cdot rd/\varepsilon^2$. Recall that we can assume without loss of generality that ρ is diagonal with eigenvalues $\alpha_1, \ldots, \alpha_r, 0, \ldots, 0$. We define R to be the subset of indices i such that $\alpha_i \geq \frac{r}{n}$, and $L \coloneqq [d] \setminus R$ to be the remaining indices that correspond to eigenvalues less than $\frac{r}{n}$. Given these two subsets, we define the projector Π to project onto the eigenvectors of ρ that lie in R. That is,

$$\Pi = \sum_{i\in R} |i\rangle\langle i| .$$

Since L is the complement of R, the matrix $I - \Pi$ projects onto the eigenvalues that correspond to the subset L.

We will prove the theorem by using the triangle inequality to write

$$\mathrm{D_B}(\widetilde{\boldsymbol{\rho}}, \rho) \leq \mathrm{D_B}(\widetilde{\boldsymbol{\rho}}, \widetilde{\boldsymbol{\rho}}|_\Pi) + \mathrm{D_B}(\widetilde{\boldsymbol{\rho}}|_\Pi, \rho|_\Pi) + \mathrm{D_B}(\rho|_\Pi, \rho), \tag{36}$$

and bounding each of the three terms by $O(\varepsilon)$ with high probability. We bound the first and third terms using the Gentle Measurement Lemma (Theorem 2.9) on ρ and $\widetilde{\boldsymbol{\rho}}$. Recall that the Gentle Measurement Lemma relates the Bures distance between a state σ and the post-measurement state $\sigma|_\Pi$ with the following trace:

$$\left(1 - \frac{1}{2}\mathrm{D_B}(\sigma, \sigma|_\Pi)^2\right)^2 = \mathrm{tr}(\Pi \cdot \sigma).$$

For the third term, we note that L is the subset of the eigenvalues of ρ smaller than $\frac{r}{n}$, and since there are at most r nonzero eigenvalues in total, we have

$$\mathrm{tr}(\Pi \cdot \rho) = 1 - \sum_{i \notin R} \alpha_i \geq 1 - \frac{r^2}{n} \geq 1 - \frac{1}{400}\varepsilon^2.$$

For the first term, we bound $\mathrm{tr}(\Pi \cdot \widetilde{\boldsymbol{\rho}})$ by computing the expectation of the diagonal entries of $\widetilde{\boldsymbol{\rho}}$:

$$\mathbf{E}[\widetilde{\boldsymbol{\rho}}_{ii}] = \frac{n}{n + rd}\left(\mathbf{E}[\widehat{\boldsymbol{\rho}}_{ii}] + \frac{r}{n}\right) = \frac{n}{n + rd}\left(\alpha_i + \frac{r}{n}\right).$$

This means that

$$\sum_{i \notin R} \mathbf{E}[\widetilde{\boldsymbol{\rho}}_{ii}] = \sum_{i \notin R} \frac{n}{n + rd}\left(\alpha_i + \frac{r}{n}\right) \leq \sum_{i \notin R} \frac{n}{n + rd} \cdot \frac{2r}{n} \leq \frac{n}{n + rd} \cdot \frac{2rd}{n} = \frac{2rd}{n + rd} \leq \frac{2rd}{n} \leq \frac{1}{200} \cdot \varepsilon^2.$$

Since $\widetilde{\boldsymbol{\rho}}$ is positive semidefinite, the random variable $\mathrm{tr}((I - \Pi) \cdot \widetilde{\boldsymbol{\rho}}) = \sum_{i \notin R} \widetilde{\boldsymbol{\rho}}_{ii}$ is nonnegative, and Markov's inequality implies that $\mathrm{tr}((I - \Pi) \cdot \widetilde{\boldsymbol{\rho}}) \leq \frac{1}{2} \cdot \varepsilon^2$ with probability at least 0.99. This implies that $\mathrm{tr}(\Pi \cdot \widetilde{\boldsymbol{\rho}}) \geq 1 - \frac{1}{2} \cdot \varepsilon^2$ with high probability. Therefore, for the rest of this proof, we condition on the event that these inequalities hold with high probability. A straightforward application of the Gentle Measurement Lemma implies that

$$\left(1 - \frac{1}{2}\mathrm{D_B^2}(\rho|_\Pi, \rho)\right)^2 = 1 - O(\varepsilon^2) \implies \mathrm{D_B^2}(\rho|_\Pi, \rho) \leq O(\varepsilon^2), \tag{37}$$

and

$$\left(1 - \frac{1}{2}\mathrm{D_B^2}(\widetilde{\boldsymbol{\rho}}, \widetilde{\boldsymbol{\rho}}|_\Pi)\right)^2 = 1 - O(\varepsilon^2) \implies \mathrm{D_B^2}(\widetilde{\boldsymbol{\rho}}, \widetilde{\boldsymbol{\rho}}|_\Pi) \leq O(\varepsilon^2). \tag{38}$$

Our argument for bounding the second term of Equation (36) will contain several steps. First, we use the second moment expression of the debiased Keyl's estimator to bound the Bures χ^2-divergence between the unnormalized states $\widetilde{\boldsymbol{\rho}}[R]$ and $\rho[R]$. We then use this to deduce a bound on the Bures χ^2-divergence between the normalized states $\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R]$ and $\rho^{\mathrm{norm}}[R]$, which also implies a bound on their Bures distance, by Theorem 2.8. Finally, we observe that the same bound holds for the Bures distance between $\widetilde{\boldsymbol{\rho}}|_\Pi$ and $\rho|_\Pi$, since $(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R], \rho^{\mathrm{norm}}[R])$ and $(\widetilde{\boldsymbol{\rho}}|_\Pi, \rho|_\Pi)$ correspond to the same pair of states, with the difference that the first pair lives in the space spanned by R, whereas the second pair lives in the entire ambient space $\mathbb{C}^{d \times d}$.

For the first step, we write

$$\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right] = \sum_{i,j \in R} \frac{2}{\alpha_i + \alpha_j} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij} - \rho_{ij}|^2\right] = \sum_{i,j \in R} \frac{2}{\alpha_i + \alpha_j} \cdot \left(\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2\right] - |\rho_{ij}|^2\right). \tag{39}$$

Let us consider the term from Equation (39) that corresponds to a fixed pair (i, j):

$$\begin{aligned}
\frac{2}{\alpha_i + \alpha_j} \cdot \left(\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2\right] - |\rho_{ij}|^2\right) &\leq \frac{2}{\alpha_i + \alpha_j} \cdot \left(\frac{\alpha_i + \alpha_j}{n} + \frac{r}{n^2}\right) && \text{(Theorem 5.1)} \\
&\leq \frac{2}{n} + \frac{1}{n} = \frac{3}{n}. && (i, j \in R \implies \alpha_i + \alpha_j \geq \tfrac{2r}{n})
\end{aligned}$$

Equation (39) has a total of $|R|^2 \le r^2$ terms, each of which is at most $\frac{3}{n}$ in expectation. We deduce that

$$\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right] = \sum_{i,j\in R} \frac{2}{\alpha_i+\alpha_j} \cdot \left(\mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2\right] - |\rho_{ij}|^2\right) \le \frac{3r^2}{n} \le O(\varepsilon^2). \tag{40}$$

We now turn our attention to bounding $\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widetilde{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right]$. From our definition of $\widetilde{\boldsymbol{\rho}}$, it holds that

$$\widetilde{\boldsymbol{\rho}}_{ij} = \begin{cases} \frac{n}{n+rd}(\widehat{\boldsymbol{\rho}}_{ij} + \frac{r}{n}) & i = j, \\ \frac{n}{n+rd} \cdot \widehat{\boldsymbol{\rho}}_{ij} & i \neq j. \end{cases}$$

Then we write

$$\begin{aligned}
\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widetilde{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right] &= \sum_{i,j\in R} \frac{2}{\alpha_i+\alpha_j} \cdot \mathbf{E}\left[|\widetilde{\boldsymbol{\rho}}_{ij} - \rho_{ij}|^2\right] \\
&= \sum_{\substack{i,j\in R \\ i\neq j}} \frac{2}{\alpha_i+\alpha_j} \cdot \mathbf{E}\left[|\widetilde{\boldsymbol{\rho}}_{ij}|^2\right] + \sum_{i\in R} \frac{1}{\alpha_i} \cdot \mathbf{E}\left[|\widetilde{\boldsymbol{\rho}}_{ii} - \alpha_i|^2\right] \\
&= \sum_{\substack{i,j\in R \\ i\neq j}} \frac{2}{\alpha_i+\alpha_j} \mathbf{E}\left[\left|\frac{n}{n+rd} \cdot \widehat{\boldsymbol{\rho}}_{ij}\right|^2\right] + \sum_{i\in R} \frac{1}{\alpha_i} \mathbf{E}\left[\left|\frac{n}{n+rd}\left(\widehat{\boldsymbol{\rho}}_{ii} + \frac{r}{n}\right) - \alpha_i\right|^2\right].
\end{aligned} \tag{41}$$

The first term can be bounded using $\mathbf{E}\left[\left|\frac{n}{n+rd} \cdot \widehat{\boldsymbol{\rho}}_{ij}\right|^2\right] \le \mathbf{E}[|\widehat{\boldsymbol{\rho}}_{ij}|^2]$. We bound the second term as follows

$$\begin{aligned}
\mathbf{E}\left[\left|\frac{n}{n+rd}\left(\widehat{\boldsymbol{\rho}}_{ii} + \frac{r}{n}\right) - \alpha_i\right|^2\right] &= \mathbf{E}\left[\left|\frac{n}{n+rd} \cdot \widehat{\boldsymbol{\rho}}_{ii} + \frac{r}{n+rd} - \alpha_i\right|^2\right] \\
&= \mathbf{E}\left[\left|\widehat{\boldsymbol{\rho}}_{ii} - \alpha_i + \frac{r}{n+rd} - \frac{rd}{n+rd} \cdot \widehat{\boldsymbol{\rho}}_{ii}\right|^2\right] \\
&\le 3\,\mathbf{E}\left[\left|\widehat{\boldsymbol{\rho}}_{ii} - \alpha_i\right|^2\right] + 3\left(\frac{r}{n+rd}\right)^2 + 3\,\mathbf{E}\left[\left|\frac{rd}{n+rd} \cdot \widehat{\boldsymbol{\rho}}_{ii}\right|^2\right] \quad \text{(Cauchy-Schwarz)} \\
&\le 3\,\mathbf{E}\left[\left|\widehat{\boldsymbol{\rho}}_{ii} - \alpha_i\right|^2\right] + 3\left(\frac{r}{n}\right)^2 + 3\left(\frac{rd}{n}\right)^2 \cdot \mathbf{E}[|\widehat{\boldsymbol{\rho}}_{ii}|^2].
\end{aligned}$$

We plug these bounds back into Equation (41) to obtain

$$\begin{aligned}
&\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widetilde{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right] \\
&\le \sum_{\substack{i,j\in R \\ i\neq j}} \frac{2}{\alpha_i+\alpha_j} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ij}|^2\right] + 3\sum_{i\in R} \frac{1}{\alpha_i} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ii} - \alpha_i|^2\right] + 3\left(\frac{r}{n}\right)^2 \sum_{i\in R} \frac{1}{\alpha_i} + 3\left(\frac{rd}{n}\right)^2 \sum_{i\in R} \frac{1}{\alpha_i} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ii}|^2\right] \\
&\le 3\,\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right] + \frac{3r^2}{n} + 3\left(\frac{rd}{n}\right)^2 \sum_{i\in R} \frac{1}{\alpha_i} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ii}|^2\right].
\end{aligned}$$

Here, the second inequality follows because all α_i in R are at least $\frac{r}{n}$ and there are at most r of them, so that $\sum_{i\in R} \frac{1}{\alpha_i} \le n$. Now we bound the last term as follows:

$$\begin{aligned}
\sum_{i\in R} \frac{1}{\alpha_i} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ii}|^2\right] &= \sum_{i\in R} \frac{1}{\alpha_i} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ii} - \alpha_i + \alpha_i|^2\right] \\
&\le 2\sum_{i\in R} \frac{1}{\alpha_i} \cdot \mathbf{E}\left[|\widehat{\boldsymbol{\rho}}_{ii} - \alpha_i|^2\right] + 2\sum_{i\in R} \frac{1}{\alpha_i} \cdot \alpha_i^2 \qquad \text{(Cauchy-Schwarz)} \\
&\le 2\,\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right] + 2\sum_{i\in R} \alpha_i \\
&\le 2\,\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R] \,\|\, \rho[R]\right)\right] + 2.
\end{aligned}$$

Hence

$$\begin{aligned}\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widetilde{\boldsymbol{\rho}}[R]\,\|\,\rho[R]\right)\right] &\leq 3\,\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R]\,\|\,\rho[R]\right)\right] + \frac{3r^2}{n} + 3\left(\frac{rd}{n}\right)^2\left(2\,\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R]\,\|\,\rho[R]\right)\right] + 2\right)\\ &\leq O(1)\cdot\mathbf{E}\left[\mathrm{D}_{\chi^2}\left(\widehat{\boldsymbol{\rho}}[R]\,\|\,\rho[R]\right)\right] + \frac{3r^2}{n} + \frac{6r^2d^2}{n^2} \leq O(\varepsilon^2). \qquad \text{(Equation (40))}\end{aligned}$$

Therefore, Markov's inequality implies that $\mathrm{D}_{\chi^2}\left(\widetilde{\boldsymbol{\rho}}[R]\,\|\,\rho[R]\right) \leq O(\varepsilon^2)$ with high probability. For the rest of this proof, we condition on the event that the above inequality holds.

Now, we would like to translate our restricted Bures χ^2-divergence bound on the (potentially) unnormalized states $\widetilde{\boldsymbol{\rho}}[R]$ and $\rho[R]$ to the restricted Bures χ^2-divergence between their normalized versions $\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R]$ and $\rho^{\mathrm{norm}}[R]$. Recall that $\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R], \rho^{\mathrm{norm}}[R] \in \mathbb{C}^{|R|\times|R|}$. In particular, for $i, j \in R$, we will need to rescale the entries of the two states as follows

$$(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R])_{ij} = \frac{1}{\mathrm{tr}(\widetilde{\boldsymbol{\rho}}[R])}\cdot\widetilde{\boldsymbol{\rho}}_{ij} = (1+c_1)\cdot\widetilde{\boldsymbol{\rho}}_{ij}, \quad (\rho^{\mathrm{norm}}[R])_{ij} = \frac{1}{\mathrm{tr}(\rho[R])}\cdot\rho_{ij} = (1+c_2)\cdot\rho_{ij},$$

where $\mathrm{tr}(\widetilde{\boldsymbol{\rho}}[R]) = \mathrm{tr}(\Pi\cdot\widetilde{\boldsymbol{\rho}}) \geq 1 - \frac{1}{2}\cdot\varepsilon^2$ implies that $c_1 = O(\varepsilon^2)$, and similarly $c_2 = O(\varepsilon^2)$. Therefore,

$$\mathrm{D}_{\chi^2}(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R]\,\|\,\rho^{\mathrm{norm}}[R]) = \sum_{\substack{i,j\in R\\ i\neq j}}\frac{2}{\alpha_i+\alpha_j}\cdot\left|(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R])_{ij}\right|^2 + \sum_{i\in R}\frac{1}{\alpha_i}\cdot\left|(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R])_{ii} - (\rho^{\mathrm{norm}}[R])_{ii}\right|^2. \qquad (42)$$

We bound the first term using $\left|(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R])_{ij}\right|^2 \leq (1+c_1)^2\cdot\left|\widetilde{\boldsymbol{\rho}}_{ij}\right|^2 \leq O(1)\cdot\left|\widetilde{\boldsymbol{\rho}}_{ij}\right|^2$, and the second term using

$$\begin{aligned}&\left|(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R])_{ii} - (\rho^{\mathrm{norm}}[R])_{ii}\right|^2\\ &\leq 3\left|(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R])_{ii} - \widetilde{\boldsymbol{\rho}}_{ii}\right|^2 + 3\left|\widetilde{\boldsymbol{\rho}}_{ii} - \rho_{ii}\right|^2 + 3\left|\rho_{ii} - (\rho^{\mathrm{norm}}[R])_{ii}\right|^2 && \text{(Cauchy-Schwarz)}\\ &= 3c_1^2\left|\widetilde{\boldsymbol{\rho}}_{ii}\right|^2 + 3\left|\widetilde{\boldsymbol{\rho}}_{ii} - \rho_{ii}\right|^2 + 3c_2^2\left|\rho_{ii}\right|^2\\ &\leq 6c_1^2\left|\widetilde{\boldsymbol{\rho}}_{ii} - \rho_{ii}\right|^2 + 6c_1^2\left|\rho_{ii}\right|^2 + 3\left|\widetilde{\boldsymbol{\rho}}_{ii} - \rho_{ii}\right|^2 + 3c_2^2\left|\rho_{ii}\right|^2 && \text{(Cauchy-Schwarz)}\\ &= O(1)\cdot\left|\widetilde{\boldsymbol{\rho}}_{ii} - \rho_{ii}\right|^2 + O(\varepsilon^4)\cdot\left|\rho_{ii}\right|^2\\ &= O(1)\cdot\left|\widetilde{\boldsymbol{\rho}}_{ii} - \rho_{ii}\right|^2 + O(\varepsilon^4)\cdot\alpha_i^2.\end{aligned}$$

Thus $\mathrm{D}_{\chi^2}(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R]\,\|\,\rho^{\mathrm{norm}}[R])$ is at most:

$$\begin{aligned}\mathrm{D}_{\chi^2}(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R]\,\|\,\rho^{\mathrm{norm}}[R]) &\leq O(1)\cdot\sum_{\substack{i,j\in R\\ i\neq j}}\frac{2}{\alpha_i+\alpha_j}\cdot\left|\widetilde{\boldsymbol{\rho}}_{ij}\right|^2 + \sum_{i\in R}\frac{1}{\alpha_i}\cdot\left(O(1)\cdot\left|\widetilde{\boldsymbol{\rho}}_{ii} - \rho_{ii}\right|^2 + O(\varepsilon^4)\cdot\alpha_i^2\right)\\ &= O(1)\cdot\sum_{i,j\in R}\frac{2}{\alpha_i+\alpha_j}\cdot\left|\widetilde{\boldsymbol{\rho}}_{ij} - \rho_{ij}\right|^2 + O(\varepsilon^4)\cdot\sum_{i\in R}\frac{1}{\alpha_i}\cdot\alpha_i^2\\ &= O(1)\cdot\mathrm{D}_{\chi^2}(\widetilde{\boldsymbol{\rho}}[R]\,\|\,\rho[R]) + O(\varepsilon^4)\cdot\sum_{i\in R}\alpha_i\\ &\leq O(1)\cdot\mathrm{D}_{\chi^2}(\widetilde{\boldsymbol{\rho}}[R]\,\|\,\rho[R]) + O(\varepsilon^4)\\ &\leq O(\varepsilon^2).\end{aligned}$$

The inequality between the Bures distance and the Bures χ^2-divergence implies that

$$\mathrm{D}_{\mathrm{B}}(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R], \rho^{\mathrm{norm}}[R]) \leq \sqrt{\mathrm{D}_{\chi^2}(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R]\,\|\,\rho^{\mathrm{norm}}[R])} \leq O\left(\varepsilon\right).$$

Moreover, we remark that since Π is the projector onto the eigenvectors that lie in R, $\widetilde{\boldsymbol{\rho}}|_\Pi$ and $\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R]$ correspond to the same state, embedded in two ambient spaces of dimension d and $|R|$ respectively. The same holds for $\rho|_\Pi$ and $\rho^{\mathrm{norm}}[R]$. Then

$$\mathrm{F}(\widetilde{\boldsymbol{\rho}}|_\Pi, \rho|_\Pi) = \mathrm{F}(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R], \rho^{\mathrm{norm}}[R]) \implies \mathrm{D}_{\mathrm{B}}(\widetilde{\boldsymbol{\rho}}|_\Pi, \rho|_\Pi) = \mathrm{D}_{\mathrm{B}}(\widetilde{\boldsymbol{\rho}}^{\mathrm{norm}}[R], \rho^{\mathrm{norm}}[R]) \leq O(\varepsilon).$$

Recall that our goal is to use the triangle inequality $\mathrm{D_B}(\widetilde{\boldsymbol{\rho}}, \rho) \leq \mathrm{D_B}(\widetilde{\boldsymbol{\rho}}, \widetilde{\boldsymbol{\rho}}|_\Pi) + \mathrm{D_B}(\widetilde{\boldsymbol{\rho}}|_\Pi, \rho|_\Pi) + \mathrm{D_B}(\rho|_\Pi, \rho)$. We have showed that each of the terms on the right-hand side is $O(\varepsilon)$ with high probability. Therefore, we conclude that with high probability,

$$\mathrm{D_B}(\widetilde{\boldsymbol{\rho}}, \rho) \leq O(\varepsilon).$$

The theorem statement follows by reducing ε by a small constant factor. □

5.3 Learning bipartite pure states

As an application of our tomography algorithm for Bures distance, we consider the problem of learning bipartite pure states.

Suppose $|\psi\rangle_{AB}$ is a bipartite state in a Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\dim(\mathcal{H}_A) = d_A$ and $\dim(\mathcal{H}_B) = d_B$, and suppose $|\psi\rangle_{AB}$ has Schmidt rank r, i.e. there exist vectors $\{|u_i\rangle_A\}_{i\in[r]}$ in $\mathcal{H}_A$ and $\{|v_i\rangle_B\}_{i\in[r]}$ in $\mathcal{H}_B$, and nonnegative coefficients $\{\lambda_i\}_{i\in[r]}$ such that

$$|\psi\rangle_{AB} = \sum_{i=1}^{r} \lambda_i \cdot |u_i\rangle_A \otimes |v_i\rangle_B .$$

What is the sample complexity of learning $|\psi\rangle_{AB}$ to ε error in Bures distance? One strategy is to use the rank-1 special case of the Bures learning algorithm directly, Theorem 5.4, which uses $O(d_A d_B/\varepsilon^2)$ samples. In this subsection, we give an algorithm with better sample complexity when $r \ll d_A, d_B$. This algorithm was inspired by the reduction from rank-r state tomography to pure state tomography from [Yue23].

Definition 5.9 (Algorithm for learning bipartite pure states)**.** On input $|\psi\rangle_{AB}^{\otimes n}$, with $n = O(r(d_A + d_B)/\varepsilon^2)$:

1. Trace out the B register on $O(rd_A/\varepsilon^2)$ copies of $|\psi\rangle_{AB}$, obtaining copies of the rank-r mixed state ψ_A.
2. Perform rank-r tomography on the copies of ψ_A, using Theorem 5.4, producing an estimate $\widehat{\boldsymbol{\psi}}_A$ such that $\mathrm{D_B}(\widehat{\boldsymbol{\psi}}_A, \psi_A) \leq \varepsilon$ with high probability. Write $\boldsymbol{\Pi}_A$ for the projector onto the support of $\widehat{\boldsymbol{\psi}}_A$. By Theorem 2.10, we can assume $\widehat{\boldsymbol{\psi}}_A$ is a mixed state of rank at most r.
3. Measure $O(rd_B/\varepsilon^2)$ copies of $|\psi\rangle_{AB}$ with $\{\boldsymbol{\Pi}_A \otimes I_B, \overline{\boldsymbol{\Pi}}_A \otimes I_B\}$. Write
$$|\boldsymbol{\pi}\rangle_{AB} = \frac{\boldsymbol{\Pi}_A \otimes I_B \, |\psi\rangle_{AB}}{\|(\boldsymbol{\Pi}_A \otimes I_B) \, |\psi\rangle_{AB}\|}$$
for the pure state collapsed to upon obtaining the measurement outcome $\boldsymbol{\Pi}_A \otimes I_B$.
4. Since $|\boldsymbol{\pi}\rangle_{AB}$ is a pure state in $\mathrm{supp}(\boldsymbol{\Pi}_A) \otimes \mathcal{H}_B$, which has dimension at most rd_B, use the copies of $|\boldsymbol{\pi}\rangle_{AB}$ as input to the rank-1 special case of Theorem 5.4. We obtain a state $\widehat{\boldsymbol{\pi}}_{AB}$ such that $\mathrm{D_B}(\widehat{\boldsymbol{\pi}}_{AB}, \boldsymbol{\pi}_{AB}) \leq \varepsilon$ with high probability.
5. Output $\widehat{\boldsymbol{\pi}}_{AB}$.

Remark 5.10. The algorithm in Theorem 5.9 does not necessarily output a pure state. However, by Theorem 2.10, there exists an algorithm that learns a pure state $|\widehat{\boldsymbol{\psi}}\rangle_{AB}$ approximating $|\psi\rangle_{AB}$, using the same number of samples. If the base algorithm learns to Bures distance ε, then the modified algorithm learns to Bures distance 2ε.

Proposition 5.11. *The algorithm given in Theorem 5.9 learns $|\psi\rangle_{AB}$ to Bures distance $O(\varepsilon)$ using $n = O(r(d_A+d_B)/\varepsilon^2)$ copies. As a consequence, if we modify the algorithm to output pure states as in Theorem 5.10, then it outputs a pure state $|\widehat{\boldsymbol{\psi}}\rangle_{AB}$ with overlap $|\langle\psi|\widehat{\boldsymbol{\psi}}\rangle_{AB}| \geq 1 - O(\varepsilon^2)$.*

Proof. First we prove that if all of the steps succeed, then $\mathrm{D_B}(\widehat{\boldsymbol{\pi}}_{AB}, \psi_{AB}) \leq 2\varepsilon$. Since $\mathrm{D_B}(\widehat{\boldsymbol{\psi}}_A, \psi_A) \leq \varepsilon$, we have $\mathrm{F}(\widehat{\boldsymbol{\psi}}_A, \psi_A) \geq 1 - \varepsilon^2/2$. Thus, by [Wat18, Proposition 3.12, (4) and (6)],

$$1 - \frac{1}{2}\varepsilon^2 \leq F(\widehat{\boldsymbol{\psi}}_A, \psi_A) = F(\widehat{\boldsymbol{\psi}}_A, \boldsymbol{\Pi}_A \psi_A \boldsymbol{\Pi}_A) \leq \sqrt{\mathrm{tr}\left(\widehat{\boldsymbol{\psi}}_A\right) \cdot \mathrm{tr}(\boldsymbol{\Pi}_A \psi_A \boldsymbol{\Pi}_A)}$$
$$= \sqrt{\mathrm{tr}(\boldsymbol{\Pi}_A \psi_A)} = \sqrt{\mathrm{tr}((\boldsymbol{\Pi}_A \otimes I_B) \, |\psi\rangle\langle\psi|_{AB})} = \|(\boldsymbol{\Pi}_A \otimes I_B) \, |\psi\rangle_{AB}\|.$$

Therefore,

$$|\langle \boldsymbol{\pi}|\psi\rangle_{AB}| = \frac{\langle\psi|\left(\boldsymbol{\Pi}_A \otimes I_B\right)|\psi\rangle_{AB}}{\|\left(\boldsymbol{\Pi}_A \otimes I_B\right)|\psi\rangle_{AB}\|} = \|\left(\boldsymbol{\Pi}_A \otimes I_B\right)|\psi\rangle_{AB}\| \geq 1 - \frac{1}{2}\varepsilon^2.$$

This implies

$$\mathrm{D_B}(\boldsymbol{\pi}_{AB}, \psi_{AB}) = \sqrt{2(1 - |\langle \boldsymbol{\pi}|\psi\rangle_{AB}|)} \leq \varepsilon.$$

So, by the triangle inequality, we have

$$\mathrm{D_B}\big(\widehat{\boldsymbol{\pi}}_{AB}, \psi_{AB}\big) \leq \mathrm{D_B}\big(\widehat{\boldsymbol{\pi}}_{AB}, \boldsymbol{\pi}_{AB}\big) + \mathrm{D_B}\big(\boldsymbol{\pi}_{AB}, \psi_{AB}\big) \leq \varepsilon + \varepsilon = 2\varepsilon.$$

Now we consider the success probabilities. First note that the second and fourth steps succeed with high probability provided they are given the number of samples specified, by Theorem 5.4. The only other step to consider is the third step. In this stage of the algorithm, conditioned on the success of the second step, we obtain outcome $\boldsymbol{\Pi}_A \otimes I_B$ with probability

$$\|\left(\boldsymbol{\Pi}_A \otimes I_B\right)|\psi\rangle_{AB}\|^2 \geq 1 - \frac{1}{2}\varepsilon^2.$$

We can assume $\varepsilon \leq 1$, since otherwise, we can learn to this higher, still constant, accuracy instead. The probability $\|\left(\boldsymbol{\Pi}_A \otimes I_B\right)|\psi\rangle_{AB}\| \geq 1 - \frac{1}{2}\varepsilon^2$ is at least $\frac{1}{2}$. Thus, with high probability, $O(rd_B/\varepsilon^2)$ copies of $|\psi\rangle_{AB}$ suffice to produce the necessary $O(rd_B/\varepsilon^2)$ samples of $|\boldsymbol{\pi}\rangle_{AB}$ which are required for the fourth step. Altogether, since each step succeeds with high probability, the overall algorithm does too. □

6 Tomography with limited entanglement

In this section, we use our unbiased estimator to devise a tomography algorithm when we are limited to measurements spanning up to k copies of the state ρ.

Theorem 6.1. *Let $n = n' \cdot k$. Let $\widehat{\boldsymbol{\rho}}_1, \ldots, \widehat{\boldsymbol{\rho}}_{n'}$ be the outputs of the debiased Keyl's algorithm when run in n' independent copies of $\rho^{\otimes k}$. Set $\widehat{\boldsymbol{\rho}} = (\widehat{\boldsymbol{\rho}}_1 + \cdots + \widehat{\boldsymbol{\rho}}_{n'})/n'$. Then $\mathrm{D_{tr}}(\rho, \widehat{\boldsymbol{\rho}}) \leq \varepsilon$ with high probability when*

$$n = O\left(\max\left(\frac{d^3}{\sqrt{k}\varepsilon^2}, \frac{d^2}{\varepsilon^2}\right)\right).$$

Proof. From linearity of expectation and the unbiasedness of our tomography algorithm, we know that

$$\mathbf{E}\,\widehat{\boldsymbol{\rho}} = \rho.$$

We proceed to bound the variance of our estimator. Since $\widehat{\boldsymbol{\rho}}_1, \ldots, \widehat{\boldsymbol{\rho}}_{n'}$ are independent and identically distributed,

$$\mathbf{Var}[\widehat{\boldsymbol{\rho}}] = \frac{1}{(n')^2} \cdot \mathbf{Var}[\widehat{\boldsymbol{\rho}}_1 + \cdots + \widehat{\boldsymbol{\rho}}_{n'}] = \frac{1}{(n')^2} \cdot (\mathbf{Var}[\widehat{\boldsymbol{\rho}}_1] + \cdots + \mathbf{Var}[\widehat{\boldsymbol{\rho}}_1]) = \frac{1}{n'} \cdot \mathbf{Var}[\widehat{\boldsymbol{\rho}}_1].$$

Recall from Theorem 4.2 that

$$\mathbf{Var}[\widehat{\boldsymbol{\rho}}_1] \leq \frac{2d}{k} + \frac{d^2 \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})}{k^2}.$$

We show in Theorem 6.2 that $\mathbf{E}\,\ell(\boldsymbol{\lambda}) \leq 2\sqrt{k}$, which means that

$$\mathbf{Var}[\widehat{\boldsymbol{\rho}}] = \frac{1}{n'} \cdot \mathbf{Var}[\widehat{\boldsymbol{\rho}}_1] \leq \frac{2d}{n'k} + \frac{2d^2}{n'k\sqrt{k}} = \frac{2d}{n} + \frac{2d^2}{n\sqrt{k}}.$$

We upper bound the trace distance using

$$\mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_1 \leq \sqrt{d}\sqrt{\mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_2^2} \leq \sqrt{\frac{2d^2}{n} + \frac{2d^3}{n\sqrt{k}}}.$$

Therefore, taking

$$n \geq \frac{4d^2}{\varepsilon^2} + \frac{4d^3}{\sqrt{k}\varepsilon^2} = O\left(\max\left(\frac{d^3}{\sqrt{k}\varepsilon^2}, \frac{d^2}{\varepsilon^2}\right)\right)$$

many samples suffices to produce $\widehat{\boldsymbol{\rho}}$ with $\mathbf{E}\|\widehat{\boldsymbol{\rho}} - \rho\|_1 \leq \varepsilon$. Markov's inequality then implies that $\|\widehat{\boldsymbol{\rho}} - \rho\|_1 \leq O(\varepsilon)$ with high probability. The theorem statement follows by reducing ε by a small constant factor. □

Alternative proof using Young diagram statistics. Similar to Theorem 4.1, there is an alternative way of proving Theorem 6.1 without relying on our second moment calculation. Let $\widehat{\boldsymbol{\rho}}' = \boldsymbol{U} \cdot \widehat{\boldsymbol{\alpha}}' \cdot \boldsymbol{U}^\dagger$ be the member of the family of unbiased estimators that uses the rank-$\ell(\boldsymbol{\lambda})$ box donation transformation, where $\ell(\boldsymbol{\lambda}) \coloneqq \ell(\boldsymbol{\lambda})$ is the height of the Young tableau we obtain from weak Schur sampling. The spectrum of $\widehat{\boldsymbol{\rho}}'$ satisfies

$$\widehat{\boldsymbol{\alpha}}'_i = \begin{cases} \frac{\boldsymbol{\lambda}_i + d - 2i + 1}{n} & i \le \ell(\boldsymbol{\lambda}) \\ -\frac{\ell(\boldsymbol{\lambda})}{n} & i \in \{\ell(\boldsymbol{\lambda}) + 1, \dots, d\}. \end{cases}$$

We will obtain a slightly weaker bound for the variance of $\widehat{\boldsymbol{\rho}}$ as follows: $\mathbf{Var}[\widehat{\boldsymbol{\rho}}_1] \le \mathbf{Var}[\widehat{\boldsymbol{\rho}}'_1] \le \frac{23d}{k} + \frac{2d^2}{k\sqrt{k}}$. We proved the first inequality in Theorem 3.9, and for the second inequality, we write

$$\begin{aligned} \mathbf{Var}[\widehat{\boldsymbol{\rho}}'] &= \mathbf{E}\,\mathrm{tr}(\widehat{\boldsymbol{\rho}}'^2) - p_2(\alpha) && \text{(Equation (25))} \\ &= \mathbf{E}\Big[\sum_{i=1}^{d} \widehat{\boldsymbol{\alpha}}_i'^2\Big] - p_2(\alpha) \\ &= \mathbf{E}\Big[\sum_{i=1}^{\ell(\boldsymbol{\lambda})} \frac{(\boldsymbol{\lambda}_i + (d - 2i + 1))^2}{k^2}\Big] + \mathbf{E} \sum_{i=\ell(\boldsymbol{\lambda})+1}^{d} \frac{(-\ell(\boldsymbol{\lambda}))^2}{k^2} - p_2(\alpha). && (43) \end{aligned}$$

We will bound the first term again using Kerov's algebra of observables of diagrams, and in particular, the expectation of $p_2^*(\boldsymbol{\lambda})$ from Equation (30).

$$\begin{aligned} (43) &\le \frac{1}{k^2}\,\mathbf{E}\Big[\sum_{i=1}^{d} \big(\boldsymbol{\lambda}_i^2 - \boldsymbol{\lambda}_i(2i-1)\big) + \sum_{i=1}^{\ell(\boldsymbol{\lambda})} \boldsymbol{\lambda}_i(2d - 2i + 1) + \sum_{i=1}^{\ell(\boldsymbol{\lambda})} (d - 2i + 1)^2\Big] + \frac{d \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})^2}{k^2} - p_2(\alpha) \\ &\le \frac{1}{k^2}\,\mathbf{E}[p_2^*(\boldsymbol{\lambda})] + \frac{2d}{k^2}\,\mathbf{E}\Big[\sum_{i=1}^{\ell(\boldsymbol{\lambda})} \boldsymbol{\lambda}_i\Big] + \frac{d^2 \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})}{k^2} + \frac{d \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})^2}{k^2} - p_2(\alpha) \\ &\le \frac{2d}{k} + \frac{d^2 \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})}{k^2} + \frac{d \cdot \mathbf{E}\,\ell(\boldsymbol{\lambda})^2}{k^2}. && \text{(Equation (30))} \end{aligned}$$

We conclude using Theorem 6.2 that $\mathbf{Var}[\widehat{\boldsymbol{\rho}}'_1] \le \frac{23d}{k} + \frac{2d^2}{k\sqrt{k}}$. Given this variance bound, the statement of Theorem 6.1 as in the original proof.

Lemma 6.2. *Let β be a distribution over $[d]$, and let $\boldsymbol{\lambda}$ be the Young diagram produced from the RSK algorithm with k symbols drawn from the distribution β. Then the first and second moment of the height of the tableau $\boldsymbol{\lambda}$ satisfy*

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_k^d(\beta)} \ell(\boldsymbol{\lambda}) \le 2\sqrt{k}, \qquad \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_k^d(\beta)} \ell(\boldsymbol{\lambda})^2 \le 20k.$$

Proof. Theorem 1.5.4 of [Wri16] states that if β, υ are two distributions such that $\beta \succ \upsilon$, then there is a coupling $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ of $\mathrm{SW}(\beta)$ and $\mathrm{SW}(\upsilon)$ such that $\boldsymbol{\lambda} \trianglerighteq \boldsymbol{\mu}$. Here, $\beta \succ \upsilon$ means that $\sum_{i=1}^{k} \beta_i \ge \sum_{i=1}^{k} \upsilon_i$ for all $k \in [d]$, and for two partitions λ and μ, we have $\lambda \trianglerighteq \mu$ means, similarly, that $\sum_{i=1}^{k} \lambda_i \ge \sum_{i=1}^{k} \mu_i$ for all $k \in [d]$.

It is not hard to see that if $\boldsymbol{\lambda} \trianglerighteq \boldsymbol{\mu}$, then $\boldsymbol{\mu}$ necessarily has at least as many rows as $\boldsymbol{\lambda}$. In the opposite case, $\sum_{i=1}^{\ell(\mu)} \mu_i > \sum_{i=1}^{\ell(\mu)} \lambda_i$, thus contradicting the monotonicity. We will choose $\upsilon = (1/d, \dots, 1/d)$ and deduce that

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_k^d(\beta)} \ell(\boldsymbol{\lambda})^t \le \mathop{\mathbf{E}}_{\boldsymbol{\mu} \sim \mathrm{SW}_k^d(\upsilon)} \ell(\boldsymbol{\mu})^t$$

for all moments $t \in \mathbb{N}$.

What is nice is that the distribution of $\mathrm{SW}_k^d(\upsilon)$ is the limit as $d \to \infty$ is well-studied and also known as the Plancherel distribution Planch_k ([Wri16, Corollary 3.3.4]). To make use of this limiting behavior, define β_D for $D \ge d$ as the inclusion of β as a probability distribution on $[D]$, i.e. $\beta_D = (\beta_1, \dots, \beta_d, 0, \dots, 0)$. Moreover, let υ_D be the uniform distribution on $[D]$.

Then for all $D \geq d$ we have

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_k^d(\beta)} \ell(\boldsymbol{\lambda})^t = \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_k^D(\beta_D)} \ell(\boldsymbol{\lambda})^t \leq \mathop{\mathbf{E}}_{\boldsymbol{\mu} \sim \mathrm{SW}_k^D(\upsilon_D)} \ell(\boldsymbol{\mu})^t,$$

where we've used that $\mathrm{SW}_k^d(\beta)$ is identical to $\mathrm{SW}_k^D(\beta_D)$. And because this holds for arbitrary $D \geq d$, we can take the limit as $D \to \infty$ to get

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_k^d(\beta)} \ell(\boldsymbol{\lambda})^t \leq \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{Planch}_k} \ell(\boldsymbol{\lambda})^t.$$

For the Plancherel distribution, it holds that $\ell(\boldsymbol{\lambda})$ and $\boldsymbol{\lambda}_1$ are identically distributed. This is because drawing $\boldsymbol{\lambda}$ from Planch_k is equivalent to sampling a uniformly random permutation $\boldsymbol{\pi}$ of $[k]$ and performing the RSK algorithm to produce a diagram. The reverse permutation $\boldsymbol{\pi}_{\mathrm{rev}}$ will produce the transpose diagram $\boldsymbol{\lambda}'$ when we execute the RSK algorithm, which satisfies $\ell(\boldsymbol{\lambda}') = \boldsymbol{\lambda}_1$. Since the distribution of $\boldsymbol{\pi}$ and $\boldsymbol{\pi}_{\mathrm{rev}}$ are equivalent, we conclude that for all moments $t \in \mathbb{N}$

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{Planch}_k} \ell(\boldsymbol{\lambda})^t = \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{Planch}_k} \boldsymbol{\lambda}_1^t.$$

We are now ready to bound the RHS for the first and second moments. When $t = 1$, [VK85] and [Pil90] showed that (see [Wri16, Theorem 3.5.4] for an exposition of the proof)

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{Planch}_k} \boldsymbol{\lambda}_1 \leq 2\sqrt{k}. \tag{44}$$

When $t = 2$, we write

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{Planch}_k} \boldsymbol{\lambda}_1^2 = \Big(\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{Planch}_k} \boldsymbol{\lambda}_1 \Big)^2 + \mathbf{Var}[\boldsymbol{\lambda}_1] \leq 4k + 16k = 20k,$$

where *the* inequality follows from Equation (44) and [OW17, Proposition 4.8]. □

7 Shadow tomography

In this section, we prove the following theorem.

Theorem 7.1 (Theorem 1.9, restated)**.** *The classical shadows task for states of rank r and observables $O_1, \ldots, O_m$ with $\mathrm{tr}(O_i^2) \leq F$ can be solved using*

$$n = O\Big(\log(m) \cdot \Big(\min\Big\{ \frac{\sqrt{rF}}{\varepsilon}, \frac{F^{2/3}}{\varepsilon^{4/3}} \Big\} + \frac{1}{\varepsilon^2} \Big) \Big)$$

copies of ρ.

We use a "plug-in estimator" approach based on the debiased Keyl's algorithm, and show that it achieves the above sample complexity.

Definition 7.2 (Algorithm for classical shadows using the debiased Keyl's algorithm)**.** Let n' and k be parameters, and set $n = n' \cdot k$. On input $\rho^{\otimes n}$, with $n = n' \cdot k$:

1. Use n' samples of ρ to obtain an estimate $\widehat{\boldsymbol{\rho}}_1$ for ρ using the debiased Keyl's algorithm.
2. Set $\widehat{\boldsymbol{o}}_i^{(1)} \coloneqq \mathrm{tr}(O_i \cdot \widehat{\boldsymbol{\rho}}_i)$, for each $i \in \{1, \ldots, m\}$.
3. Repeat the first two steps $(k-1)$ times to obtain estimates $\widehat{\boldsymbol{o}}_i^{(j)}$, for $j \in \{2, \ldots, k\}$ and $i \in \{1, \ldots, m\}$.
4. For each i, output $\widehat{\boldsymbol{o}}_i \coloneqq \mathrm{median}(\widehat{\boldsymbol{o}}_i^{(1)}, \ldots, \widehat{\boldsymbol{o}}_i^{(k)})$.

Specifically, we will show below that this algorithm achieves the sample complexity in Theorem 7.1 with $k = O(\log m)$ and $n' = O(\min\{\frac{\sqrt{rF}}{\varepsilon}, \frac{F^{2/3}}{\varepsilon^{4/3}}\} + \frac{1}{\varepsilon^2})$. We begin our analysis with a lemma.

Lemma 7.3. *Let $O \in \mathbb{C}^{d\times d}$ be Hermitian. If $\widehat{\boldsymbol{\rho}}$ is the output of the debiased Keyl's algorithm on input $\rho^{\otimes n}$, with ρ rank-r, then*

$$\mathbf{Var}[\mathrm{tr}(O\cdot\widehat{\boldsymbol{\rho}})] \le \frac{2}{n}\|O\|_\infty^2 + \frac{\min\{r, 2\sqrt{n}\}}{n^2}\,\mathrm{tr}(O^2).$$

Proof. We have

$$\mathbf{Var}[\mathrm{tr}(O\cdot\widehat{\boldsymbol{\rho}})] = \mathbf{E}\,\mathrm{tr}(O\cdot\widehat{\boldsymbol{\rho}})^2 - (\mathbf{E}\,\mathrm{tr}(O\cdot\widehat{\boldsymbol{\rho}}))^2 = \mathrm{tr}(O\otimes O\cdot\mathbf{E}[\widehat{\boldsymbol{\rho}}\otimes\widehat{\boldsymbol{\rho}}]) - \mathrm{tr}(O\cdot\rho)^2. \tag{45}$$

By Theorem 1.4,

$$\begin{aligned}\mathrm{tr}(O\otimes O\cdot\mathbf{E}[\widehat{\boldsymbol{\rho}}\otimes\widehat{\boldsymbol{\rho}}]) &= \frac{n-1}{n}\,\mathrm{tr}(O\otimes O\cdot\rho\otimes\rho) + \frac{1}{n}\,\mathrm{tr}(O\otimes O\cdot\rho\otimes I\cdot\mathrm{SWAP}) + \frac{1}{n}\,\mathrm{tr}(O\otimes O\cdot I\otimes\rho\cdot\mathrm{SWAP})\\ &\quad + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2}\,\mathrm{tr}(O\otimes O\cdot\mathrm{SWAP}) - \mathrm{tr}(O\otimes O\cdot\mathrm{Lower}_\rho)\\ &= \frac{n-1}{n}\,\mathrm{tr}(O\cdot\rho)^2 + \frac{2}{n}\,\mathrm{tr}(O^2\cdot\rho) + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2}\,\mathrm{tr}(O^2) - \mathrm{tr}(O\otimes O\cdot\mathrm{Lower}_\rho).\end{aligned}$$

First, we note that the last term is non-negative from our characterization of Lower_ρ. In particular, we can expand $\mathrm{tr}(O\otimes O\cdot\mathrm{Lower}_\rho)$ as a positive linear combination of terms of the form

$$\mathrm{tr}(O\otimes O\cdot P\otimes P\cdot\mathrm{SWAP}) = \mathrm{tr}(OPOP) = \mathrm{tr}((\sqrt{P}O\sqrt{P})^2) \ge 0,$$

where the last equality follows because P is positive semidefinite. Next, because $O^2 \preceq \|O^2\|_\infty\cdot I_d = \|O\|_\infty^2\cdot I_d$ in the PSD order, we have $\mathrm{tr}(O^2\cdot\rho) \le \|O\|_\infty^2\cdot\mathrm{tr}(\rho) = \|O\|_\infty^2$. Finally, we note two upper bounds on $\mathbf{E}[\ell(\boldsymbol{\lambda})]$: on the one hand, we always have $\ell(\boldsymbol{\lambda}) \le r$, so that $\mathbf{E}[\ell(\boldsymbol{\lambda})] \le r$; on the other hand, $\mathbf{E}[\ell(\boldsymbol{\lambda})] \le 2\sqrt{n}$, by Theorem 6.2. Combining all of these observations gets us:

$$\mathrm{tr}(O\otimes O\cdot\mathbf{E}[\widehat{\boldsymbol{\rho}}\otimes\widehat{\boldsymbol{\rho}}]) \le \mathrm{tr}(O\cdot\rho)^2 + \frac{2}{n}\|O\|_\infty^2 + \frac{\min\{r, 2\sqrt{n}\}}{n^2}\,\mathrm{tr}(O^2).$$

Substituting this back into Equation (45) gives us the desired bound. □

Proof of Theorem 7.1. Fix any $i \in [m]$, and any $j \in [k]$. First note that the estimate $\widehat{\boldsymbol{o}}_i^{(j)}$ is unbiased, since by Theorem 1.3, we have

$$\mathbf{E}[\widehat{\boldsymbol{o}}_i^{(j)}] = \mathrm{tr}(O_i\cdot\mathbf{E}[\widehat{\boldsymbol{\rho}}]) = \mathrm{tr}(O_i\cdot\rho).$$

Moreover, Theorem 7.3 tells us

$$\mathbf{Var}[\widehat{\boldsymbol{o}}_i^{(j)}] \le \frac{2}{n'} + \frac{2}{(n')^2}\cdot\min\{r, \sqrt{n'}\}\cdot F. \tag{46}$$

Taking $n' = O(1/\varepsilon^2)$ suffices to make the first term on the right-hand side of the above equation at most $O(\varepsilon^2)$. Note that taking $n' \ge \sqrt{rF}/\varepsilon$ suffices to make $2rF/(n')^2 \le O(\varepsilon^2)$, and taking $n' \ge F^{2/3}/\varepsilon^{4/3}$ suffices to make $2F/(n')^{3/2} \le O(\varepsilon^2)$. Thus, taking $n' = O(\min\{\sqrt{rF}/\varepsilon, F^{2/3}/\varepsilon^{4/3}\}$ suffices to make at least one of the terms in the minimum $O(\varepsilon^2)$, the second term $O(\varepsilon^2)$.

So, we have that

$$n' = O\Big(\min\Big\{\frac{\sqrt{rF}}{\varepsilon}, \frac{F^{2/3}}{\varepsilon^{4/3}}\Big\} + \frac{1}{\varepsilon^2}\Big)$$

suffices to ensure $\mathbf{Var}[\widehat{\boldsymbol{o}}_i^{(j)}] \le O(\varepsilon^2)$ for all $i \in [m]$. By Chebyshev's inequality, this sample complexity is also enough to ensure that $|\widehat{\boldsymbol{o}}_i^{(j)} - \mathrm{tr}(O_i\rho)| \le \varepsilon$ with high probability, for any i, say 99%, by taking C sufficiently small.

We complete our proof with a standard median argument. Assume k is odd. Fix i and define a set $\boldsymbol{S}_i$ by

$$\boldsymbol{S}_i \coloneqq \Big|\big\{j \,:\, |\widehat{\boldsymbol{o}}_i^{(j)} - \mathrm{tr}(O_i\rho)| > \varepsilon\big\}\Big|.$$

Then the median of $\widehat{\boldsymbol{o}}_i^{(1)}, \ldots, \widehat{\boldsymbol{o}}_i^{(k)}$ is more than ε away from $\operatorname{tr}(O_i\rho)$ only when $\boldsymbol{S}_i > k/2$. However, $\mathbf{E}[\boldsymbol{S}_i] \leq k/100$, so that

$$\begin{aligned}\mathbf{Pr}\left[\left|\operatorname{median}(\boldsymbol{o}_i^{(j)})_{j\in[k]} - \operatorname{tr}(O_i\cdot\rho)\right| > \varepsilon\right] &\leq \mathbf{Pr}\left[\boldsymbol{S}_i > k/2\right] \\ &\leq \mathbf{Pr}\left[\boldsymbol{S}_i > \mathbf{E}[\boldsymbol{S}_i] + 49k/100\right] \leq \exp\left(-2k\cdot(49/100)^2\right).\end{aligned}$$

Here we have used an additive Chernoff bound, and we conclude that

$$\mathbf{Pr}\left[\left|\operatorname{median}(\boldsymbol{o}_i^{(1)}, \ldots, \boldsymbol{o}_i^{(k)}) - \operatorname{tr}(O_i\cdot\rho)\right| > \varepsilon\right] \leq \exp\left(-O(k)\right).$$

Taking $k = O(\log m)$ is enough to make this probability less than $1/(100m)$. So, by a union bound, the probability that *all* m of the $\widehat{\boldsymbol{o}}_i$ are within ε of the correct value is at most $1/100$.

We conclude that

$$n = k\cdot n' = O\Big(\log(m)\cdot\Big(\min\Big\{\frac{\sqrt{rF}}{\varepsilon}, \frac{F^{2/3}}{\varepsilon^{4/3}}\Big\} + \frac{1}{\varepsilon^2}\Big)\Big)$$

samples suffices to solve classical shadows using the algorithm described in Theorem 7.2. □

8 Quantum metrology

In this section, we construct a locally unbiased estimator achieving twice the QCRB asymptotically. Specifically, we show the following theorem, stated in the introduction.

Theorem 8.1 (Theorem 1.10, restated)**.** *Let ρ_θ be a parameterized state family with probe state*

$$\rho_{\theta^*} = \sum_{i=1}^{d} \alpha_i \cdot |v_i\rangle\langle v_i|\,,$$

where we assume without loss of generality that the eigenvalues are sorted so that $\alpha_1 \geq \cdots \geq \alpha_r > 0$ and $\alpha_{r+1} = \cdots = \alpha_d = 0$ (so that ρ_{θ^*} is rank r). Then for every n, there is a locally unbiased estimator which takes as input n copies of ρ_θ and whose MSEM V_n has the following property. When $n = 2r/(\varepsilon\cdot\alpha_r)$,

$$V_n \preceq (1+\varepsilon)\cdot\frac{2}{n\cdot\mathcal{F}}.$$

Hence, as $n\to\infty$, we have $n\cdot V_n \to 2/\mathcal{F}$, and so this estimator achieves twice the QCRB asymptotically.

In [ZC25], Zhou and Chen design a locally unbiased estimator for θ, given an unbiased estimator for ρ_θ. Since we prove the above theorem by using the debiased Keyl's algorithm as the basis for such a locally unbiased estimator, we will now give an overview of their construction. We start by reviewing some standard material. Throughout, we will write ∂_i as shorthand for $\partial/\partial\theta_i$. We will also assume without loss of generality that that ρ_{θ^*} is diagonalized in the computational basis, with eigenvalues sorted decreasingly, i.e. that $|v_i\rangle = |i\rangle$.

Recall that the symmetric logarithm derivative matrices $\{L_i\}_{i\in[m]}$ are defined implicitly as solutions to the equations

$$\partial_i\rho_\theta\big|_{\theta=\theta^*} = \frac{1}{2}L_i\rho_{\theta^*} + \frac{1}{2}\rho_{\theta^*}L_i. \tag{47}$$

These equations do not uniquely determine $\{L_i\}$, since we can add to L_i any matrix ΔL_i supported on the kernel of ρ_{θ^*} without affecting the right-hand side of Equation (47). However, this is the only source of non-uniqueness, since

$$(\partial_i\rho_\theta)_{k\ell}\big|_{\theta=\theta^*} = \langle k|\left(\frac{1}{2}L_i\rho_{\theta^*} + \frac{1}{2}\rho_{\theta^*}L_i\right)|\ell\rangle = \frac{1}{2}(\alpha_k+\alpha_\ell)(L_i)_{k\ell}$$

implies $(L_i)_{k\ell} = 2(\partial_i \rho_\theta)_{k\ell}\big|_{\theta=\theta^*}/(\alpha_k + \alpha_\ell)$ whenever $\alpha_k + \alpha_\ell > 0$, i.e. whenever $\alpha_k \neq 0$ or $\alpha_\ell \neq 0$. Taking L_i to be zero on the kernel of ρ_{θ^*} then gives rise to a canonical choice of the matrices $\{L_i\}$, with

$$L_i = 2 \sum_{\substack{k,\ell \\ \alpha_k + \alpha_\ell > 0}} \frac{(\partial_i \rho_\theta)_{k\ell}\big|_{\theta=\theta^*}}{\alpha_k + \alpha_\ell} \cdot |k\rangle\langle \ell| \,. \tag{48}$$

Note that this choice is Hermitian, since $\partial_i \rho_\theta\big|_{\theta=\theta^*}$ is Hermitian. Note also that for any choice of $\{L_i\}$,

$$\operatorname{tr}(L_i \rho_{\theta^*}) = \frac{1}{2}\operatorname{tr}(L_i \rho_{\theta^*}) + \frac{1}{2}\operatorname{tr}(\rho_{\theta^*} L_i) = \operatorname{tr}(\partial_i \rho_\theta\big|_{\theta=\theta^*}) = \partial_i \operatorname{tr}(\rho_\theta)\big|_{\theta=\theta^*} = 0, \tag{49}$$

since $\operatorname{tr}(\rho_\theta) = 1$ for any θ.

Recall that the QFI matrix is defined as

$$\mathcal{F}_{ij} \coloneqq \frac{1}{2}\operatorname{tr}(\rho_{\theta^*} L_i L_j) + \frac{1}{2}\operatorname{tr}(\rho_{\theta^*} L_j L_i) = \operatorname{tr}(L_i \cdot \partial_j \rho_\theta)\big|_{\theta=\theta^*}. \tag{50}$$

This definition is independent of the choice of $\{L_i\}$, since if $L_i' = L_i + \Delta L_i$, and ΔL_i is supported on the kernel of ρ_{θ^*}, then $\rho_{\theta^*}\Delta L_i = \Delta L_i \rho_{\theta^*} = 0$, so that $\operatorname{tr}(\rho_{\theta^*} L_i L_j) = \operatorname{tr}(\rho_{\theta^*} L_i' L_j) = \operatorname{tr}(\rho_{\theta^*} L_i' L_j')$, for all $i, j \in [m]$.

Note that $\mathcal{F}_{ij}$ is symmetric. Moreover, since we can choose Hermitian $\{L_i\}$, $\mathcal{F}_{ij}$ is also real, as

$$\mathcal{F}_{ij}^* = \frac{1}{2}\operatorname{tr}(\rho_{\theta^*} L_j^\dagger L_i^\dagger) + \frac{1}{2}\operatorname{tr}(\rho_{\theta^*} L_i^\dagger L_j^\dagger) = \mathcal{F}_{ij}.$$

We can also give an explicit expression for the entries of $\mathcal{F}$, using the matrices $\{L_i\}$ given in Equation (48), and the expression for $\mathcal{F}_{ij}$ on the right-hand side of Equation (50):

$$\mathcal{F}_{ij} = 2 \sum_{\substack{k,\ell \\ \alpha_k + \alpha_\ell > 0}} \frac{\big((\partial_i \rho_\theta)_{k\ell} \cdot (\partial_j \rho_\theta)_{\ell k}\big)\big|_{\theta=\theta^*}}{\alpha_k + \alpha_\ell}. \tag{51}$$

It is assumed that $\mathcal{F}$ is invertible. We briefly explain why. If $\mathcal{F}$ is not invertible, it must have a real eigenvector which has eigenvalue zero. Let this vector be $|\psi\rangle = \sum_{i=1}^d z_i |i\rangle$, with $z_i \in \mathbb{R}$. Then $\langle\psi| \mathcal{F} |\psi\rangle = 0 = \operatorname{tr}(\rho_{\theta^*} Q^2)$, where $Q = \sum_{i=1}^m z_i L_i$. Since Q is Hermitian, this implies that Q is supported on the kernel of ρ_{θ^*}. But since each L_i is zero on this kernel, this means that Q is also zero on the kernel, and hence $Q = 0$ as an operator. Thus, there exists a linear combination of derivatives which is zero at θ^*, by Equation (48), which means that ρ_θ is not well-parameterized at $\theta = \theta^*$. That is, the parameterization is redundant at this point. We assume this is not the case, and hence take $\mathcal{F}$ invertible.

We now develop material specific to Zhou and Chen's construction.

Definition 8.2 (Deviation observables)**.** Let $\{X_i\}_{i\in[m]}$ be a set of Hermitian operators satisfying, for all $i, j \in [m]$:

$$\text{(i)}\ \operatorname{tr}(X_i \cdot \rho_{\theta^*}) = 0, \quad \text{and} \quad \text{(ii)}\ \operatorname{tr}(X_i \cdot \partial_j \rho_\theta)\Big|_{\theta=\theta^*} = \delta_{ij}.$$

Then the X_i are called *deviation observables*.

Lemma 8.3. *Let $\{L_i\}_{i\in[m]}$ be a set of Hermitian symmetric logarithm derivative matrices. For each $i \in [m]$, define $X_i \coloneqq \sum_{j=1}^m (\mathcal{F}^{-1})_{ij} \cdot L_j$. Then $\{X_i\}_{i\in[m]}$ is a set of deviation observables.*

Proof. We verify the two conditions in Theorem 8.2. First,

$$\operatorname{tr}(X_i \cdot \rho_{\theta^*}) = \sum_{j=1}^m (\mathcal{F}^{-1})_{ij} \cdot \operatorname{tr}(L_j \cdot \rho_{\theta^*}) = 0, \tag{52}$$

since $\operatorname{tr}(L_j \cdot \rho_{\theta^*}) = 0$ using Equation (49). Second,

$$\operatorname{tr}(X_i \cdot \partial_j \rho_\theta)\Big|_{\theta=\theta^*} = \sum_{k=1}^m (\mathcal{F}^{-1})_{ik} \cdot \operatorname{tr}(L_k \cdot \partial_j \rho_\theta)\Big|_{\theta=\theta^*} = \sum_{k=1}^m (\mathcal{F}^{-1})_{ik} \cdot \mathcal{F}_{kj} = (\mathcal{F}^{-1} \cdot \mathcal{F})_{ij} = \delta_{ij},$$

since $\operatorname{tr}(L_k \cdot \partial_j \rho_\theta)\Big|_{\theta=\theta^*} = \mathcal{F}_{kj}$ by Equation (50). □

Definition 8.4 (Local shadow estimators)**.** Let $\{X_i\}_{i\in[m]}$ be a set of deviation observables, and let $\widehat{\boldsymbol{\rho}}$ be an estimator for ρ_θ. Then the corresponding *local shadow estimator* $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^m$ is given by

$$\widehat{\boldsymbol{\theta}}_i \coloneqq \theta_i^* + \operatorname{tr}(X_i \cdot \widehat{\boldsymbol{\rho}}),$$

for all $i \in [m]$.

Recall that an estimator is locally unbiased if the following conditions are met for all $i \in [m]$:

$$\text{(i)}\ \ \mathbf{E}[\widehat{\boldsymbol{\theta}} \mid \rho_{\theta^*}] = \theta^*, \quad \text{and} \quad \text{(ii)}\ \ \frac{\partial}{\partial \theta_i}\mathbf{E}[\widehat{\boldsymbol{\theta}} \mid \rho_\theta]\Big|_{\theta=\theta^*} = 0.$$

Lemma 8.5. *Let $\{X_i\}$ be a set of deviation observables, and let $\widehat{\boldsymbol{\rho}}$ be the output of an unbiased tomography algorithm, given copies of ρ_θ. Then the corresponding local shadow estimator is a locally unbiased estimator for θ.*

Proof. We verify the two conditions that define a locally unbiased estimator. Firstly, for all $i \in [m]$,

$$\mathbf{E}[\widehat{\boldsymbol{\theta}}_i \mid \rho_{\theta^*}] = \theta_i^* + \operatorname{tr}(X_i \cdot \mathbf{E}[\widehat{\boldsymbol{\rho}} \mid \rho_{\theta^*}]) = \theta_i^* + \operatorname{tr}(X_i \cdot \rho_{\theta^*}) = \theta_i^*,$$

so that $\mathbf{E}[\widehat{\boldsymbol{\theta}} \mid \rho_{\theta^*}] = \theta^*$. Secondly, for all $i, j \in [m]$,

$$\partial_i \mathbf{E}[\widehat{\boldsymbol{\theta}}_j \mid \rho_\theta]\Big|_{\theta=\theta^*} = \operatorname{tr}(X_j \cdot \partial_i \mathbf{E}[\widehat{\boldsymbol{\rho}} \mid \rho_\theta])\Big|_{\theta=\theta^*} = \operatorname{tr}(X_j \cdot \partial_i \rho_\theta)\Big|_{\theta=\theta^*} = \delta_{ij}. \qquad \square$$

Finally, Zhou and Chen's estimator is then the local shadow estimator, constructed from some fixed unbiased tomography algorithm and deviation observables $\{X_i\}_{i\in[m]}$. The $\{X_i\}$ are built from the symmetric logarithm derivative matrices $\{L_i\}_{i\in[m]}$ as in Theorem 8.3, with L_i is given explicitly by Equation (48). By Theorem 8.5, this estimator is locally unbiased. In [ZC25], Zhou and Chen apply this using the bias-corrected single copy estimator of Section 1.1.1.

We are now ready to prove the main result of the section. Our argument is largely based on the proofs of Theorem 1 and Theorem 4 of [ZC25].

Proof of Theorem 8.1. We use Zhou and Chen's construction with the debiased Keyl's algorithm. Given access to $\rho_\theta^{\otimes n}$, the debiased Keyl's algorithm produces an unbiased estimator $\widehat{\boldsymbol{\rho}}$ with

$$\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] = \frac{n-1}{n}\rho_\theta \otimes \rho_\theta + \frac{1}{n}(\rho_\theta \otimes I + I \otimes \rho_\theta) \cdot \mathrm{SWAP} + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \cdot \mathrm{SWAP} - \mathrm{Lower}_{\rho_\theta},$$

by Theorems 1.3 and 1.4. The MSEM is the matrix V_n with entries:

$$(V_n)_{ij} = \mathbf{E}[(\widehat{\boldsymbol{\theta}}_i - \theta_i^*) \cdot (\widehat{\boldsymbol{\theta}}_j - \theta_j^*) \mid \rho_{\theta^*}] = \mathbf{E}[\operatorname{tr}(X_i \cdot \widehat{\boldsymbol{\rho}}) \cdot \operatorname{tr}(X_j \cdot \widehat{\boldsymbol{\rho}}) \mid \rho_{\theta^*}] = \operatorname{tr}(X_i \otimes X_j \cdot \mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}} \mid \rho_{\theta^*}]).$$

Plugging in our variance gives:

$$\begin{aligned}(V_n)_{ij} &= \frac{n-1}{n}\operatorname{tr}(X_i\rho_{\theta^*}) \cdot \operatorname{tr}(X_j\rho_{\theta^*}) + \frac{1}{n}\big(\operatorname{tr}(\rho_{\theta^*}X_iX_j) + \operatorname{tr}(\rho_{\theta^*}X_jX_i)\big) + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2}\operatorname{tr}(X_iX_j) \\ &\quad - \operatorname{tr}(X_i \otimes X_j \cdot \mathrm{Lower}_{\rho_{\theta^*}}) \\ &= \frac{1}{n}\big(\operatorname{tr}(\rho_{\theta^*}X_iX_j) + \operatorname{tr}(\rho_{\theta^*}X_jX_i)\big) + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2}\operatorname{tr}(X_iX_j) - \operatorname{tr}(X_i \otimes X_j \cdot \mathrm{Lower}_{\rho_{\theta^*}}).\end{aligned}$$

where we have used $\operatorname{tr}(X_i\rho_{\theta^*}) = 0$, as shown in Equation (52). Since $\mathrm{Lower}_{\rho_{\theta^*}}$ is a positive combination of matrices of the form $P \otimes P \cdot \mathrm{SWAP}$, the last term is a positive combination of matrices of the form $\operatorname{tr}(X_iPX_jP)$, and hence symmetric in i and j. Moreover, recall P is positive semidefinite.

Consider the new matrix V_n', obtained from V_n by dropping the terms containing $\mathrm{Lower}_{\rho_{\theta^*}}$. That is,

$$(V_n')_{ij} = \frac{1}{n}\big(\operatorname{tr}(\rho_{\theta^*}X_iX_j) + \operatorname{tr}(\rho_{\theta^*}X_jX_i)\big) + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2}\operatorname{tr}(X_iX_j). \tag{53}$$

Both V_n and V_n' are real and symmetric matrices, and hence Hermitian. We claim $V_n \preceq V_n'$. To show this, let $|\psi\rangle = \sum_{i=1}^m z_i |i\rangle$ be an arbitrary state. We have

$$\langle\psi| (V_n' - V_n) |\psi\rangle = \sum_{i,j} \overline{z}_i z_j (V_n' - V_n)_{ij} = \sum_{i,j} \overline{z}_i z_j \operatorname{tr}(X_i \otimes X_j \cdot \operatorname{Lower}_{\rho_{\theta^*}}) = \operatorname{tr}(Q^\dagger \otimes Q \cdot \operatorname{Lower}_{\rho_{\theta^*}}),$$

where $Q \coloneqq \sum_{i=1}^m z_i X_i$. However, this trace is a positive combination of terms of the form

$$\operatorname{tr}(Q^\dagger \otimes Q \cdot P \otimes P \cdot \mathrm{SWAP}) = \operatorname{tr}(Q^\dagger P Q \cdot P).$$

Since P is positive semidefinite, so too is $Q^\dagger P Q$. Thus, $\operatorname{tr}(Q^\dagger P Q \cdot P) \geq 0$, and $\langle\psi| (V_n' - V_n) |\psi\rangle \geq 0$. Since $|\psi\rangle$ was arbitrary, $V_n' \succeq V_n$. So, it will suffice to bound V_n' in the PSD order.

Now consider the product $\mathcal{F} V_n' \mathcal{F}^T$. From Equation (53), we have

$$\begin{aligned}(\mathcal{F} V_n' \mathcal{F}^T)_{ij} &= \sum_{k,\ell} \mathcal{F}_{ik} (V_n')_{k\ell} \mathcal{F}_{j\ell} \\ &= \frac{1}{n} \operatorname{tr}(\rho_{\theta^*} \sum_{k=1}^m \mathcal{F}_{ik} X_k \sum_{\ell=1}^m \mathcal{F}_{j\ell} X_\ell) + \frac{1}{n} \operatorname{tr}(\rho_{\theta^*} \sum_{\ell=1}^m \mathcal{F}_{j\ell} X_\ell \sum_{k=1}^m \mathcal{F}_{ik} X_k) + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \operatorname{tr}(\sum_{k=1}^m \mathcal{F}_{ik} X_k \sum_{\ell=1}^m \mathcal{F}_{j\ell} X_\ell).\end{aligned}$$

However, we have

$$\sum_{k=1}^m \mathcal{F}_{ik} X_k = \sum_{k,\ell} \mathcal{F}_{ik} (\mathcal{F}^{-1})_{k\ell} L_\ell = \sum_\ell \delta_{i\ell} L_\ell = L_i,$$

where we have used $X_k \coloneqq \sum_\ell (\mathcal{F}^{-1})_{k\ell} L_\ell$. Thus

$$(\mathcal{F} V_n' \mathcal{F}^T)_{ij} = \frac{1}{n} \big(\operatorname{tr}(\rho_{\theta^*} L_i L_j) + \operatorname{tr}(\rho_{\theta^*} L_j L_i) \big) + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \operatorname{tr}(L_i L_j).$$

Now, note that the first term is equal to $2\mathcal{F}_{ij}/n$, by definition of $\mathcal{F}_{ij}$. Thus

$$\mathcal{F} V_n' \mathcal{F}^T = \frac{2}{n} \mathcal{F} + \frac{2\,\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} K, \tag{54}$$

where K is the matrix given by

$$K_{ij} \coloneqq \frac{1}{2} \operatorname{tr}(L_i L_j) = \sum_{\substack{k,\ell \\ \alpha_k + \alpha_\ell > 0}} \frac{\big((\partial_i \rho_\theta)_{k\ell} \cdot (\partial_j \rho_\theta)_{\ell k}\big)\big|_{\theta=\theta^*}}{(\alpha_k + \alpha_\ell)^2},$$

where we have used the explicit expressions for L_i in Equation (48). We now bound K in the PSD order. Let $|\psi\rangle = \sum_{i=1}^m z_i |i\rangle$ again be an arbitrary state. Define $R_{k\ell} \coloneqq \sum_{i=1}^m z_i \cdot (\partial_i \rho_\theta)_{k\ell}\big|_{\theta=\theta^*}$, and note that

$$(R^\dagger)_{k\ell} = (R)^*_{\ell k} = \sum_{i=1}^m z_i^* \cdot (\partial_i \rho_\theta)^*_{\ell k}\big|_{\theta=\theta^*} = \sum_{i=1}^m z_i^* \cdot (\partial_i \rho_\theta^\dagger)_{k\ell}\big|_{\theta=\theta^*} = \sum_{i=1}^m z_i^* \cdot (\partial_i \rho_\theta)_{k\ell}\big|_{\theta=\theta^*}.$$

So, we have

$$\langle\psi| K |\psi\rangle = \sum_{\substack{k,\ell \in [d] \\ \alpha_k + \alpha_\ell > 0}} \sum_{i,j \in [m]} z_i^* z_j \cdot \frac{\big((\partial_i \rho_\theta)_{k\ell} \cdot (\partial_j \rho_\theta)_{\ell k}\big)\big|_{\theta=\theta^*}}{(\alpha_k + \alpha_\ell)^2} = \sum_{\substack{k,\ell \\ \alpha_k + \alpha_\ell > 0}} \frac{(R^\dagger)_{k\ell} \cdot R_{\ell k}}{(\alpha_k + \alpha_\ell)^2} = \sum_{\substack{k,\ell \\ \alpha_k + \alpha_\ell > 0}} \frac{|R_{\ell k}|^2}{(\alpha_k + \alpha_\ell)^2},$$

and similarly,

$$\frac{1}{2} \langle\psi| \mathcal{F} |\psi\rangle = \sum_{\substack{k,\ell \\ \alpha_k + \alpha_\ell > 0}} \sum_{i,j} z_i^* z_j \cdot \frac{\big((\partial_i \rho_\theta)_{k\ell} \cdot (\partial_j \rho_\theta)_{\ell k}\big)\big|_{\theta=\theta^*}}{\alpha_k + \alpha_\ell} = \sum_{\substack{k,\ell \\ \alpha_k + \alpha_\ell > 0}} \frac{|R_{\ell k}|^2}{\alpha_k + \alpha_\ell}.$$

Now since

$$\langle\psi|\, K\, |\psi\rangle = \sum_{\substack{k,\ell \\ \alpha_k+\alpha_\ell>0}} \frac{|R_{\ell k}|^2}{(\alpha_k+\alpha_\ell)^2} \le \frac{1}{\alpha_r} \sum_{\substack{k,\ell \\ \alpha_k+\alpha_\ell>0}} \frac{|R_{\ell k}|^2}{\alpha_k+\alpha_\ell} = \frac{1}{2\alpha_r} \langle\psi|\, \mathcal{F}\, |\psi\rangle\,,$$

we have $K \preceq \mathcal{F}/2\alpha_r$. Hence, from Equation (54),

$$\mathcal{F} V_n' \mathcal{F}^T \preceq \left(\frac{2}{n} + \frac{4\,\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \cdot \frac{1}{2\alpha_r}\right) \cdot \mathcal{F} \preceq \frac{2}{n} \cdot \left(1 + \frac{r}{n\alpha_r}\right) \cdot \mathcal{F}.$$

Recall that $\mathcal{F}^T = \mathcal{F}$, because $\mathcal{F}$ is symmetric. So, by multiplying on the left and right by $\mathcal{F}^{-1}$, we have

$$V_n' \preceq \frac{2}{n} \cdot \left(1 + \frac{r}{n\alpha_r}\right) \cdot \mathcal{F}^{-1}.$$

Finally, since $V_n \preceq V_n'$, for $n \ge r/(\varepsilon\alpha_r)$ we conclude

$$V_n \preceq (1+\varepsilon) \cdot \frac{2}{n} \cdot \mathcal{F}^{-1}.$$

□

Part III
The Clebsch-Gordan transform

9 The Clebsch-Gordan transform

The tensor product of two irreps of $U(d)$ is generically reducible. That is, for λ and μ partitions of length at most d, we have the decomposition

$$V_\lambda^d \otimes V_\mu^d \overset{U(d)}{\cong} \bigoplus_{\ell(\tau)\leq d} m_\tau \cdot V_\tau^d,$$

for some integers $\{m_\tau\}$. It will suffice for us to consider only the special case where $\mu = (1)$. In this case, the decomposition is multiplicity-free [Har05, Chapter 7.2]:

$$V_\lambda^d \otimes V_{(1)}^d \overset{U(d)}{\cong} \bigoplus_{\substack{\lambda \nearrow \mu \\ \ell(\mu)\leq d}} V_\mu^d. \tag{55}$$

Recall that for $\mu = (1)$, our choice of V_μ^d is the defining representation on $\mathbb{C}^d$, and our choice of GT basis is the computational basis $\{|i\rangle\}_{i\in[d]}$, on which $\nu_\mu(U)$ acts as U (see Theorem 2.45). Therefore, by Theorem 2.17, for each λ there exists a unitary $\mathcal{U}_{\mathrm{CG}}^{(\lambda)} : V_\lambda^d \otimes \mathbb{C}^d \to \bigoplus_\mu V_\mu^d$ that implements the isomorphism in Equation (55). In other words, for all $U \in U(d)$,

$$\mathcal{U}_{\mathrm{CG}}^{(\lambda)} \cdot \left(\nu_\lambda(U) \otimes U\right) \cdot \mathcal{U}_{\mathrm{CG}}^{(\lambda)\dagger} = \sum_{\substack{\lambda \nearrow \mu \\ \ell(\mu)\leq d}} |\mu\rangle\langle\mu| \otimes \nu_\mu(U). \tag{56}$$

With these unitaries fixed, we can collect them into a single operator.

Definition 9.1 (The Clebsch-Gordan transform)**.** The *Clebsch-Gordan transform* is the unitary operator on $\bigoplus_\lambda \mathrm{Sp}_\lambda \otimes (V_\lambda^d \otimes \mathbb{C}^d) \cong (\mathbb{C}^d)^{\otimes(n+1)}$ given by

$$\mathcal{U}_{\mathrm{CG}}^{(n+1)} \coloneqq \sum_{\substack{\lambda\vdash n \\ \ell(\lambda)\leq d}} |\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \mathcal{U}_{\mathrm{CG}}^{(\lambda)}. \tag{57}$$

The Clebsch-Gordan transform takes the state $|\lambda\rangle \otimes |S\rangle \otimes |T\rangle \otimes |i\rangle$, to a superposition of states of the form $|\lambda\rangle \otimes |S\rangle \otimes |\mu\rangle \otimes |T'\rangle$, where $\lambda \nearrow \mu$, and $T' \in \mathrm{SSYT}(\mu, d)$.

10 Clebsch-Gordan coefficients

Definition 10.1 (Clebsch-Gordan coefficients)**.** Let $\lambda \vdash n$, $T \in \mathrm{SSYT}(\lambda, d)$, and $k \in [d]$. We can write:

$$\mathcal{U}_{\mathrm{CG}}^{(\lambda)} \cdot |T\rangle \otimes |k\rangle = \sum_{\substack{\mu\vdash(n+1,d) \\ T'\in\mathrm{SSYT}(\mu)}} \langle \mu, T'|\mathcal{U}_{\mathrm{CG}}^{(\lambda)}|T, k\rangle \cdot |\mu, T'\rangle .$$

The coefficients $\langle \mu, T'|\mathcal{U}_{\mathrm{CG}}^{(\lambda)}|T, k\rangle$ are known as the *Clebsch-Gordan (CG) coefficients*. We will typically write $\langle T'|T, k\rangle$ as shorthand for $\langle \mu, T'|\mathcal{U}_{\mathrm{CG}}^{(\lambda)}|T, k\rangle$, leaving $\mu = \mathsf{sh}(T')$ and $\mathcal{U}_{\mathrm{CG}}^{(\lambda)}$ implicit.

Our proofs will rely on explicit formulas for the CG coefficients. To state these formulas, we will first need to introduce new notation which will allow us to describe how an SSYT T changes when we restrict it to a specific subset of letters.

Notation 10.2. Let λ be a Young diagram. We define $\lambda_{\leq k} = (\lambda_1, \ldots, \lambda_k)$ to be the first k rows of λ.

Notation 10.3 (Restriction of SSYTs to alphabet $[p]$)**.** Let T be an SSYT of shape λ and alphabet $[d]$. We use $T^{[p]}$ to denote the SSYT of alphabet $[p]$ that we obtain when we delete all the boxes in T which contain letters strictly larger than p. Since the entries of T are weakly increasing along rows and strictly increasing along columns, $T^{[p]}$ is a valid SSYT of height at most p. We also define $T^{\leq p} \coloneqq \mathsf{sh}(T^{[p]})$ to be its shape.

$$T = \begin{array}{|c|c|c|c|c|}\hline 1&2&2&2&3\\\hline 2&3&3&4\\\cline{1-4} 3&4&4\\\cline{1-3} 4\\\cline{1-1}\end{array} \qquad T^{[3]} = \begin{array}{|c|c|c|c|c|}\hline 1&2&2&2&3\\\hline 2&3&3\\\cline{1-3} 3\\\cline{1-1}\end{array} \qquad T^{[2]} = \begin{array}{|c|c|c|c|}\hline 1&2&2&2\\\hline 2\\\cline{1-1}\end{array}$$

Figure 13: An example of the SSYT notation that we will use in this section. Given the SSYT T over the alphabet $[4]$, $T^{[3]}$ and $T^{[2]}$ are the SSYTs we obtain when we delete all the boxes which are bigger than 3 and 2 respectively. Thus $T^{\leq 4} = (5,4,3,1)$, $T^{\leq 3} = (5,3,1)$, and $T^{\leq 2} = (4,1)$.

Notation 10.4 (Horizontal strip)**.** Let T be an SSYT. We define

$$T^{=p} \coloneqq (T^{\leq p}, T^{\leq (p-1)}) = T^{\leq p} \setminus T^{\leq (p-1)}$$

to be the skew shape that contains all boxes with value p. Due to the properties of an SSYT, this skew shape is a *horizontal strip*, that is, a union of continuous horizontal lines of boxes that do not overlap vertically.

$$T^{[3]} = \begin{array}{|c|c|c|c|c|}\hline 1&2&2&2&3\\\hline 2&3&3\\\cline{1-3} 3\\\cline{1-1}\end{array} \qquad T^{=3} = \begin{array}{cc}\multicolumn{1}{c}{}&\multicolumn{1}{c}{}\\\cline{2-3}\multicolumn{1}{c}{}&\multicolumn{1}{|c|}{}&\multicolumn{1}{c|}{}\\\cline{2-3}\multicolumn{1}{|c|}{}&\multicolumn{1}{c}{}\\\cline{1-1}\end{array} \qquad \begin{array}{|c|}\hline\\\hline\end{array}$$

Figure 14: Let T be the SSYT from Figure 13. Above, we have $T^{[3]}$ and $T^{=3}$. The cells of $T^{=3}$ are shaded in $T^{[3]}$ – these are the cells containing the symbol 3.

We introduce two additional pieces of notation that correspond to the shapes that we get when we add boxes to the horizontal strip $T^{=p} = (T^{\leq p}, T^{\leq p-1})$.

Notation 10.5 (Adding boxes to a horizontal strip $T^{=p}$)**.** Let T be an SSYT. We define

$$T^{=p}_{+i} \coloneqq (T^{\leq p} + e_i, \quad T^{\leq p-1})$$

to be the shape of the skew diagram that we get when we add a box to the end of the i-th row of $T^{=p}$. Similarly,

$$T^{=p}_{+i,-j} \coloneqq (T^{\leq p} + e_i, \quad T^{\leq p-1} + e_j)$$

is the shape of the skew diagram we obtain when we add a box to the end of the i-th row and remove a box from the beginning of the j-th row of $T^{=p}$.

$$T^{=3}_{+1} = \begin{array}{cc}\multicolumn{1}{c}{}&\multicolumn{1}{c}{}\\\cline{2-3}\multicolumn{1}{c}{}&\multicolumn{1}{|c|}{}&\multicolumn{1}{c|}{}\\\cline{2-3}\multicolumn{1}{|c|}{}&\multicolumn{1}{c}{}\\\cline{1-1}\end{array} \qquad \begin{array}{|c|}\hline\\\hline\end{array} \qquad T^{=3}_{+1,-2} = \begin{array}{cc}\multicolumn{1}{c}{}&\multicolumn{1}{c}{}\\\cline{2-2}\multicolumn{1}{c}{}&\multicolumn{1}{|c|}{}\\\cline{2-2}\multicolumn{1}{|c|}{}&\multicolumn{1}{c}{}\\\cline{1-1}\end{array} \qquad \begin{array}{|c|}\hline\\\hline\end{array}$$

Figure 15: An example of the skew shapes that we get when we add and subtract cells from the skew shape $T^{=3}$ from Figure 14. On the left diagram, we have added a cell to the first row. On the right diagram, we have both added a cell to the first row and removed a cell from the second row.

It is well-known that the CG coefficients can be expressed as a product of numbers that only depend on the shapes of the skew diagrams $\{T^{=p}, \boxed{k}^{=p}, (T')^{=p}\}_{p\in\{1,\dots,d\}}$. In particular, the following was shown by [VK92, Section 18.2.2, Equation 3].

Theorem 10.6 (CG coefficients are products of scalar factors)**.** *When Section 9 is multiplicity-free, the Clebsch-Gordan coefficients can be written as a product:*

$$\langle T'|T,k\rangle = \prod_{p=1}^{d} \left(T^{=p}, \boxed{k}^{=p} \mid (T')^{=p}\right). \tag{58}$$

The factor $\left(T^{=p}, \boxed{k}^{=p} \mid (T')^{=p}\right)$ that appears on the right-hand side is called the $U(d-p)$-scalar factor of the Clebsch-Gordan coefficient. *These coefficients are also known as* reduced Wigner operators. *Note that each $U(d-p)$ scalar factor only depends on the shapes of the skew diagrams $T^{=p}, \boxed{k}^{=p}, (T')^{=p}$.*

Here we use $\boxed{k}$ to denote the SSYT that corresponds to the $|k\rangle$ basis vector of $V^d_{(1)}$, which has a single box filled with the number k. The shape of the restricted SSYT $\boxed{k}^{\leq p}$ has the following simple description:

$$\boxed{k}^{\leq p} = \begin{cases} (1, 0^{p-1}) & p \geq k, \\ (0, 0^{p-1}) & p < k. \end{cases}$$

Thus, the skew diagrams that can appear in the scalar factors of Equation (58) are one of three possible choices:

$$\boxed{k}^{=p} = \begin{cases} e_{11}^{=p} := ((1,0^{p-1}),(1,0^{p-2})) & p \geq k+1, \\ e_{10}^{=p} := ((1,0^{p-1}),(0,0^{p-2})) & p = k, \\ e_{00}^{=p} := ((0,0^{p-1}),(0,0^{p-2})) & p < k. \end{cases}$$

The scalar factors which appear in Equation (58) can be computed using the formulas in the following lemma.

Lemma 10.7 (Formulas for the $U(d)$-scalar factors)**.** *Let $c_s(T^{\leq k}) = T^{\leq k}_s - s$ be the content of the last box in the s-th row. Define $S(i,j) = 1$ if $i \leq j$, and $S(i,j) = -1$ if $i > j$. Then, for some choice of phases of the GT basis for ν_λ, the following three equations hold.*

(i) For any SSYT T and any p,

$$\left(T^{=p}, e_{00}^{=p} \mid T^{=p}\right) = 1, \tag{59}$$

and $\left(T^{=p}, e_{00}^{=p} \mid (T')^{=p}\right) = 0$ otherwise.

(ii) Section 18.2.10, equation 3 in [VK92]

$$\left(T^{=p}, e_{10}^{=p} \mid T^{=p}_{+i}\right) = \left| \frac{\prod_{j=1}^{p-1} \left(c_j(T^{\leq p-1}) - c_i(T^{\leq p}) - 1\right)}{\prod_{j=1, j\neq i}^{p} \left(c_j(T^{\leq p}) - c_i(T^{\leq p})\right)} \right|^{1/2}, \tag{60}$$

and $\left(T^{=p}, e_{10}^{=p} \mid (T')^{=p}\right) = 0$ otherwise.

(iii) Section 18.2.10, equation 4 in [VK92]

$$\begin{aligned} &\left(T^{=p}, e_{11}^{=p} \mid T^{=p}_{+i,-j}\right) \\ &= S(i,j) \left| \frac{\prod_{k\neq j}^{p-1} \left(c_k(T^{\leq p-1}) - c_i(T^{\leq p}) - 1\right) \prod_{k\neq i}^{p} \left(c_k(T^{\leq p}) - c_j(T^{\leq p-1})\right)}{\prod_{k\neq i}^{p} \left(c_k(T^{\leq p}) - c_i(T^{\leq p})\right) \prod_{k\neq j}^{p-1} \left(c_k(T^{\leq p-1}) - c_j(T^{\leq p-1}) - 1\right)} \right|^{1/2}, \end{aligned} \tag{61}$$

and $\left(T^{=p}, e_{11}^{=p} \mid (T')^{=p}\right) = 0$ otherwise.

Equation (58) expresses the Clebsch-Gordan coefficients as products of d scalar factors. For the tableaux that we are dealing with in this paper (that are close to highest-weight tableaux), it will be the case that most of these scalar factors are trivial. The following lemma gives a condition for a scalar factor to be trivial that will be sufficient for our purposes.

Lemma 10.8. *Let T be an SSYT, and $i \in [p]$ be a row such that T has no cells with the value p on the i-th row. Then*

$$\left(T^{=p}, e_{11}^{=p} \mid T_{+i,-i}^{=p}\right) = 1.$$

Proof. Since T has no p values in the i-th row, it holds that $c_i(T^{\leq p}) = c_i(T^{\leq p-1})$. Then Equation (61) implies that

$$\begin{aligned}\left(T^{=p}, e_{11}^{=p} \mid T_{+i,-i}^{=p}\right) &= S(i,i)\left|\frac{\prod_{k\neq i}^{p-1}\left(c_k(T^{\leq p-1}) - c_i(T^{\leq p}) - 1\right)\prod_{k\neq i}^{p}\left(c_k(T^{\leq p}) - c_i(T^{\leq p-1})\right)}{\prod_{k\neq i}^{p}\left(c_k(T^{\leq p}) - c_i(T^{\leq p})\right)\prod_{k\neq i}^{p-1}\left(c_k(T^{\leq p-1}) - c_i(T^{\leq p-1}) - 1\right)}\right|^{1/2} \\ &= \left|\frac{\prod_{k\neq i}^{p-1}\left(c_k(T^{\leq p-1}) - c_i(T^{\leq p-1}) - 1\right)\prod_{k\neq i}^{p}\left(c_k(T^{\leq p}) - c_i(T^{\leq p-1})\right)}{\prod_{k\neq i}^{p}\left(c_k(T^{\leq p}) - c_i(T^{\leq p-1})\right)\prod_{k\neq i}^{p-1}\left(c_k(T^{\leq p-1}) - c_i(T^{\leq p-1}) - 1\right)}\right|^{1/2} = 1. \qquad \square\end{aligned}$$

10.1 Clebsch-Gordan insertion

In this subsection, we describe a simple algorithm for inserting a new box containing a symbol $k \in [d]$ into an SSYT T. The output is a set of SSYTs, each with a matching coefficient, $\{(T', c_{T'})\}$. The SSYTs in this set turn out to be exactly the SSYTs with nonzero Clebsch-Gordan coefficients, and $c_{T'}$ is the Clebsch-Gordan coefficient $\langle T'|T, k\rangle$. This insertion algorithm, *Clebsch-Gordan insertion*, thus offers a more intuitive perspective on the Clebsch-Gordan transform and coefficients. Clebsch-Gordan insertion is, in fact, a special case of a continuous family of "quantum" insertion rules generalizing RSK insertion (see, for example, [Wri16, Chapter 3]), originally introduced by Sonya Berg in her thesis [Ber12].

Definition 10.9 (Clebsch-Gordan insertion)**.** Given an SSYT T with alphabet $[d]$, we insert a letter $k \in [d]$ into T to obtain a *set* of output SSYTs as follows:

1. Initialize $T' \leftarrow T$, $c_{T'} \leftarrow 1$, and $j \leftarrow$ None. Here, T' is a tentative output SSYT, $c_{T'}$ its tentative coefficient, and j stores a previous row index.
2. In all possible ways, choose a row $i \leq k$ such that $T^{\leq k} + e_i$ is a valid Young diagram.
3. Find the leftmost cell of T in row i containing a letter ℓ such that $\ell > k$, if such a cell exists. Let T'' be obtained from T' by replacing ℓ with k in this cell. Set

   $$T' \leftarrow T'', \qquad c_{T'} \leftarrow c_{T'} \cdot \mathrm{SF}(T, k; i, j), \qquad k \leftarrow \ell, \qquad j \leftarrow i,$$

   and return to Step 2. We will say ℓ has been *bumped* by k.
4. If no such cell exists, then append a new cell containing k to the end of row i to obtain T''. Set

   $$T' \leftarrow T'', \qquad c_{T'} \leftarrow c_{T'} \cdot \mathrm{SF}(T, k; i, j),$$

   append $(T', c_{T'})$ to the output, and terminate.

Here, $\mathrm{SF}(T, k; i, j)$ is a scalar factor, given by

$$\mathrm{SF}(T, k; i, j) = \begin{cases} \left(T^{=k}, e_{10}^{=k} \mid T_{+i}^{=k}\right) & j = \text{None}, \\ \left(T^{=k}, e_{11}^{=k} \mid T_{+i,-j}^{=k}\right) & \text{otherwise.} \end{cases}$$

See Figure 16 for an example of Clebsch-Gordan insertion.

Lemma 10.10. *Suppose we insert a letter $k \in [d]$ into an SSYT T with alphabet $[d]$ via Clebsch-Gordan insertion, obtaining as output a set $\{(T', c_{T'})\}$. Then the SSYTs $\{T'\}$ are exactly those with nonzero Clebsch-Gordan coefficients, and $c_{T'} = \langle T'|T, k\rangle$.*

Proof. We prove the two directions of the lemma separately.

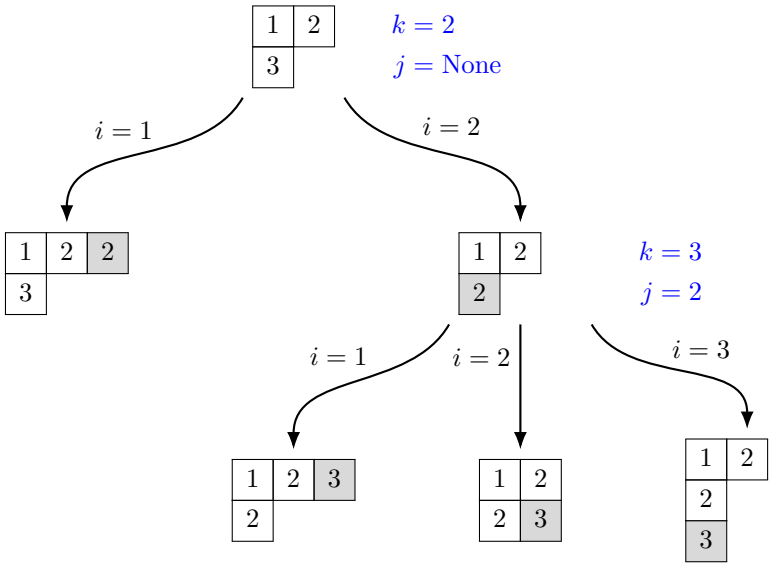


Figure 16: Here we depict Clebsch-Gordan insertion of $k = 2$ into the example SSYT

.

At the start of the insertion process, $k = 2$ is to be inserted into either of the first two rows. On the left branch, 2 is inserted at the end of the first row, and the insertion terminates. On the right branch, 2 bumps 3 from the second row, and we must now re-insert 3 into the tableau. Hence, $k \leftarrow 3$. We also update $j \leftarrow 1$ at this stage, since we had inserted the 2 into the second row. The 3 may now be inserted at the end of any of the first, second, or third rows, and the insertion terminates.

Clebsch-Gordan insertion outputs are valid SSYTs with nonzero Clebsch-Gordan coefficients: Say that the Clebsch-Gordan insertion algorithm output includes the pair $(T', c_{T'})$. Moreover, assume that T' was constructed by adding k to row j_k, and then sequentially bumping the letters $\ell_1 < \ell_2 < \cdots < \ell_t$, which were inserted into rows $j_{\ell_1}, \ldots, j_{\ell_t}$ respectively. The case of $t = 0$ corresponds to k not bumping any letter. For convenience, let us define $\ell_0 = k$ and $\ell_{t+1} = d + 1$.

We show that $c_{T'} = \langle T'|T, k\rangle$. In particular, let us define the sequence $i_k, i_{k+1}, \ldots, i_d$, such that

$$j_k = i_k = i_{k+1} = \cdots = i_{\ell_1 - 1}, \qquad j_{\ell_m} = i_{\ell_m} = \cdots = i_{\ell_{m+1}-1}, \tag{62}$$

for all $m \in [t]$. Then we claim that

$$(T')^{\leq p} = \begin{cases} T^{\leq p} & p < k, \\ T^{\leq p} + e_{i_p} & p \geq k. \end{cases}$$

To see this, note that inserting an x into a tableau can only change the locations of symbols labeled x or greater. Thus, the expression above holds for $p < k$ because k is the smallest symbol inserted into the tableau. For $p \geq k$, suppose that ℓ_m is the final letter that is inserted which satisfies $\ell_m \leq p$. Then inserting the letters $\ell_{m+1}, \ldots, \ell_t$ does not change $T^{\leq p}$ since they are all larger than p. Similarly, the letters $\ell_0 < \ell_1 < \cdots < \ell_{m-1}$ which are inserted prior to ℓ_m do not change the shape of $T^{\leq p}$ either, because they replace letters which are also $\leq p$. Thus, only ℓ_m will change the shape $T^{\leq p}$, either by adding a new box or by bumping a letter larger than p in row i_{ℓ_m}, and this will update the shape to $T^{\leq p} + e_{i_{\ell_m}} = T^{\leq p} + e_{i_p}$, as claimed. Therefore,

$$(T')^{=p} = \begin{cases} T^{=p} & p < k, \\ T^{=p}_{+i_p} & p = k, \\ T^{=p}_{+i_p, -i_{p-1}} & p > k. \end{cases} \tag{63}$$

Thus, the CG coefficient is

$$\begin{aligned} \langle T'|T, k\rangle &= \prod_{p=1}^{d} (T^{=p}, \boxed{k}^{=p} \mid (T')^{=p}) \\ &= \prod_{p \in \{\ell_0, \ell_1, \ldots, \ell_t\}} (T^{=p}, \boxed{k}^{=p} \mid (T')^{=p}) \\ &= (T^{=k}, e_{10}^{=k} \mid T^{=k}_{+j_k}) \cdot \prod_{m=1}^{t} (T^{=\ell_m}, e_{11}^{=\ell_m} \mid T^{=\ell_m}_{j_{\ell_m}, -j_{\ell_m - 1}}) \\ &= \mathrm{SF}(T, k; j_k, \mathrm{None}) \cdot \prod_{m=1}^{t} \mathrm{SF}(T, \ell_m; j_{\ell_m}, j_{\ell_{m-1}}) = c_{T'}. \end{aligned}$$

In the first step, we have used Theorem 10.6. In the second step, we have set all scalar products with $p \notin \{\ell_0, \ell_1, \ldots, \ell_t\}$ equal to 1. For $p < k = \ell_0$, this holds by item (i) of Theorem 10.7. For $p > k$ and $p \notin \{\ell_1, \ldots, \ell_t\}$, this follows by Theorem 10.8, since $i_p = i_{p-1}$ from our definition of the indices i (Equation (62)), and the fact that T contains no letter p in this row. To see this, let ℓ_m be the largest letter that was bumped which is smaller than p. Then ℓ_m was added to row $j_{\ell_m} = i_{p-1} = i_p$ and bumped the letter ℓ_{m+1}, which is larger than p. Thus, there is no letter p in the row i_p, as otherwise that would have been bumped instead.

Valid SSYTs with nonzero CG coefficients are output by the Clebsch-Gordan insertion algorithm: Let T' be a valid SSYT with nonzero CG coefficient $\langle T'|T, k\rangle$. Our first goal is to show that it must satisfy Equation (63), i.e. there exist integers $i_k, i_{k+1}, \ldots, i_d$ such that

$$(T')^{=p} = \begin{cases} T^{=p} & p < k, \\ T^{=p}_{+i_p} & p = k, \\ T^{=p}_{+i_p, -i_{p-1}} & p > k. \end{cases} \tag{64}$$

To see this, note that Equation (58) implies that for all $1 \le p \le d$,

$$\langle T' | T, k \rangle = \langle T'^{[p]} | T^{[p]}, \boxed{k}^{[p]} \rangle \cdot \prod_{i=p+1}^{d} \left(T^{=i}, \boxed{k}^{=i} \mid T'^{=i} \right). \tag{65}$$

Hence, for $\langle T' | T, k \rangle$ to be nonzero, we must have

$$\langle T'^{[p]} | T^{[p]}, \boxed{k}^{[p]} \rangle \neq 0.$$

When $p < k$, $\boxed{k}^{[p]}$ is just an empty tableau. Then we must have $T'^{[p]} = T^{[p]}$. On the other hand, when $p \ge k$, $\boxed{k}^{[p]}$ is a single box with k, and thus

$$\langle T^{[p]}, \boxed{k}^{[p]} | T'^{[p]} \rangle \neq 0 \implies \mathsf{sh}(T'^{[p]}) = \mathsf{sh}(T^{[p]}) + e_{i_p}.$$

for some integer i_p, by Equation (55). Putting these together, we have that for $p < k$,

$$T'^{=p} = (T'^{\le p}, T'^{\le p-1}) = (T^{\le p}, T^{\le p-1}) = T^{=p}.$$

Similarly, for $p = k$, we have $T'^{=k} = T^{=p}_{+i_k}$, and for $p > k$, we have $T'^{=p} = T^{=p}_{+i_p, -i_{p-1}}$. This proves Equation (64).

Next, we show that Equation (64) implies that T' can be viewed as the result of inserting a letter k into T and performing a sequence of bumps, as in the Clebsch-Gordan insertion algorithm. We do this by explicitly determining the sequence of bumped letters that produces this SSYT T'. The first letter added was k, and it was inserted at row i_k. Thus, we let $j_k = i_k$. The leftmost cell of T in row j_k that contains a letter greater than k determines the first bumped letter ℓ_1. This is because for any $p \in \{k, \dots, \ell_1 - 1\}$, $T'^{\le p} = T^{\le p} + e_{i_p}$, and since there is no letter greater than k and at most p on row i_k, it must hold that $i_p = i_k$. Thus $i_k = i_{k+1} = \cdots = i_{\ell_1 - 1}$, and we let $j_{\ell_1} = i_{\ell_1}$.

Continuing the above process, the leftmost cell of T in row $j_{\ell_{m-1}}$ that contains a letter greater than ℓ_{m-1} determines the bumped letter ℓ_m (and if no such letter exists, the Clebsch-Gordan insertion would terminate). We let $j_{\ell_m} = i_{\ell_m}$ and observe that $i_{\ell_{m-1}} = \cdots = i_{\ell_m - 1}$.

Having defined the bumped letters $\ell_1 < \cdots < \ell_t$ and their respective rows $j_k, j_{\ell_1}, \dots, j_{\ell_t}$, the argument from the first part of this proof shows that the tableau output by the Clebsch-Gordan insertion procedure is the same as T'. This is because by our choice of the j indices, the row indices $i_k, \dots, i_d$ satisfy:

$$j_k = i_k = i_{k+1} = \cdots = i_{\ell_1 - 1}, \qquad j_{\ell_m} = i_{\ell_m} = \cdots = i_{\ell_{m+1} - 1},$$

for all $m \in [t]$ (where we again assume $\ell_{t+1} = d + 1$). □

10.2 One-step coefficients

In this section, we obtain a closed-form expression for the Clebsch-Gordan coefficients $\langle T' | T^\lambda, k \rangle$, where we recall that $|T^\lambda\rangle$ is the highest weight vector of V_λ^d. We first show that the SSYTs T' for which this coefficient is nonzero have a very specific form.

Definition 10.11 (One-step semistandard Young tableaux)**.** Let λ be a Young diagram of height at most d, and $i \le k$ be positive integers at most d such that $\lambda + e_i$ is a valid Young diagram. We use $T^\lambda_{k \to i}$ to denote the *one-step semistandard Young tableau*, which is the SSYT we obtain by starting with the highest weight SSYT T^λ, and adding a new box containing the value k to the end of the i-th row.

Definition 10.12 (One-step Clebsch-Gordan coefficient)**.** Let $T^\lambda_{k \to i}$ be a one-step SSYT. We define the *one-step Clebsch-Gordan coefficient* to be

$$c^\lambda_{k \to i} := \langle T^\lambda_{k \to i} | T^\lambda, k \rangle,$$

when $T^\lambda_{k \to i}$ is well-defined (i.e. a valid SSYT), and 0 otherwise.

$$T^\lambda = \begin{array}{|c|c|c|c|c|}\hline 1&1&1&1&1\\\hline 2&2&2&2\\\cline{1-4} 3&3&3\\\cline{1-3} 4&4\\\cline{1-2}\end{array} \qquad T^\lambda_{4\to 4} = \begin{array}{|c|c|c|c|c|}\hline 1&1&1&1&1\\\hline 2&2&2&2\\\cline{1-4} 3&3&3\\\cline{1-3} 4&4&4\\\cline{1-3}\end{array} \qquad T^\lambda_{3\to 2} = \begin{array}{|c|c|c|c|c|}\hline 1&1&1&1&1\\\hline 2&2&2&2&3\\\cline{1-5} 3&3&3\\\cline{1-3} 4&4\\\cline{1-2}\end{array}$$

Figure 17: Some examples of one-step SSYTs when $\lambda = (5, 4, 3, 2)$.

Lemma 10.13. *The only basis vectors T' for which the coefficient $\langle T'|T^\lambda, k\rangle$ is nonzero correspond to the one-step SSYTs $T^\lambda_{k\to i}$, for any $i \le k$ such that $\lambda + e_i$ is a valid Young diagram. Moreover, the one-step Clebsch-Gordan coefficient satisfies the following expression:*

$$c^\lambda_{k\to i} = \left| \prod_{j=1, j\neq i}^{k-1} \left(\frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right) \cdot \frac{1}{(\lambda_i - \lambda_k) + (k-i) + \delta_{ik}} \right|^{1/2}. \tag{66}$$

Proof. Consider the Clebsch-Gordan insertion procedure from Theorem 10.9. When inserting k into row $i \le k$ during Step 2, the highest weight SSYT T^λ only contains the letter i, which is not greater than k. Thus, the insertion procedure does not reach Step 3, but always jumps to Step 4. See Figure 18. Step 4 will now append a new cell containing k to the end of row i to obtain $T^\lambda_{k\to i}$, and will set $c^\lambda_{k\to i}$ to be

$$c^\lambda_{k\to i} \leftarrow \mathrm{SF}(T^\lambda, k; i, \mathrm{None}) = \left((T^\lambda)^{=k}, e_{10}^{=k} \mid (T^\lambda)^{=k}_{+i}\right).$$

Finally, the procedure appends $(T^\lambda_{k\to i}, c^\lambda_{k\to i})$ to the output and terminates. We will now use Equation (60) to compute an expression for the one-step Clebsch-Gordan coefficient.

$$\begin{aligned} c^\lambda_{k\to i} &= \left((T^\lambda)^{=k}, e_{10}^{=k} \mid (T^\lambda)^{=k}_{+i}\right) \\ &= \left| \frac{\prod_{j=1}^{k-1}((\lambda_j - \lambda_i) + (i-j) - 1)}{\prod_{j\neq i}^{k}((\lambda_j - \lambda_i) + (i-j))} \right|^{1/2} && (T^\lambda \text{ is highest weight} \implies T^\lambda_i = \lambda_i) \\ &= \left| \prod_{j=1, j\neq i}^{k-1} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \cdot \frac{-1}{(\lambda_k - \lambda_i) + (i-k) + \delta_{ik}} \right|^{1/2}. \end{aligned}$$

The last equality holds because if $i = k$, then the denominator of the additional factor is equal to 1, and otherwise equal to $(\lambda_k - \lambda_i) + (i - k)$. □

Before we proceed, let us build some intuition about these coefficients. We first observe that all possible one-step SSYTs have a nonzero coefficient.

Observation 10.14. *Every valid one-step SSYT $T^\lambda_{k\to i}$ has a nonzero one-step Clebsch-Gordan coefficient $c^\lambda_{k\to i}$.*

$$T^\lambda = \begin{array}{|c|c|c|c|c|}\hline 1&1&1&1&1\\\hline 2&2&2&2\\\cline{1-4} 3&3&3\\\cline{1-3} 4&4\\\cline{1-2}\end{array} \xrightarrow{\text{Step 4}} T^\lambda_{k\to i} = \begin{array}{|c|c|c|c|c|}\hline 1&1&1&1&1\\\hline 2&2&2&2\\\cline{1-4} 3&3&3&4\\\cline{1-4} 4&4\\\cline{1-2}\end{array}$$

Figure 18: An example of how T' evolves during the Clebsch-Gordan insertion when $k = 4$, and $i = 3$, for $\lambda = (5, 4, 3, 2)$.

Looking at Equation (66), we can see that the only way for $c^{\lambda}_{k\to i}$ to be zero is when the expression in the numerator $(\lambda_j - \lambda_i) + (i - j) - 1$ is equal to zero for some j in $\{1, \dots, k-1\} \setminus \{i\}$. Rewriting this, it means that $\lambda_i = \lambda_j + (i - j - 1)$. We consider the following two cases:

1. If $j < i$, this means that $i - j - 1 \geq 0$ and thus $\lambda_i \geq \lambda_j$. Since the j-th row is above the i-th row, we conclude that the two row lengths must be equal. Then we cannot add another box to row i, since that would make this row longer than row j, which is above it.

2. If $j > i$, this means that $i - j - 1 \leq -2$ and thus $\lambda_i \leq \lambda_j - 2$. This is impossible, since row i is above row j, and for λ to be a valid Young diagram, it must hold that $\lambda_i \geq \lambda_j$.

We conclude that the only way for the one-step Clebsch-Gordan coefficients to vanish is by adding a box to a row i that satisfies $\lambda_{i-1} = \lambda_i$, which means that $\lambda + e_i$ is not a valid Young diagram.

10.3 Two-step coefficients

In the previous section, we introduced the one-step Clebsch-Gordan coefficients to understand the state we obtain when we tensor a qudit $|k\rangle$ to a highest weight vector. This suffices to study the first moment of our estimator $\widehat{\boldsymbol{\rho}}$. For the second moment of the estimator, we will need to understand the state we obtain when we tensor two qudits, i.e. $|k\rangle \otimes |\ell\rangle$, to a highest weight vector. We have

$$\begin{aligned} |T^{\lambda}\rangle \otimes |k\rangle \otimes |\ell\rangle &= \Big(\sum_{i \in [d]} c^{\lambda}_{k\to i} \cdot |\lambda + e_i\rangle \otimes |T^{\lambda}_{k\to i}\rangle \Big) \otimes |\ell\rangle \\ &= \sum_{i,j \in [d]} \sum_{T'} c^{\lambda}_{k\to i} \cdot \langle T' | T^{\lambda}_{k\to i}, \ell\rangle \cdot |\lambda + e_i\rangle \otimes |\lambda + e_i + e_j\rangle \otimes |T'\rangle \,, \end{aligned} \tag{67}$$

where, in the inner sum, $T' \in \mathrm{SSYT}(\lambda + e_i + e_j, d)$. Let us first try to understand which SSYTs T' appear in this superposition with nonnegative amplitude. From Theorem 10.14, the coefficient $c^{\lambda}_{k\to i}$ in Equation (67) is always nonzero if $T^{\lambda}_{k\to i}$ is a valid SSYT. Thus, it suffices to determine for which T' the coefficient $\langle T' | T^{\lambda}_{k\to i}, \ell\rangle$ is nonzero. We will need the following definition.

Definition 10.15 (Two-step semistandard Young tableaux)**.** Let λ be a Young diagram of length at most d, and $i \leq k$ and $j \leq \ell$ be positive integers at most d such that $\lambda + e_i + e_j$ is a valid Young diagram.

There is at most one valid SSYT that can be obtained from T^{λ} by adding a k to row i and an ℓ to row j. If such an SSYT exists, we denote it $T^{\lambda}_{k\ell\to ij}$ and refer to it as a *two-step semistandard Young tableau*.

Note that the ordering of the value-row pairs (k, i) and (ℓ, j) does not matter. In particular, $T^{\lambda}_{k\ell\to ij} = T^{\lambda}_{\ell k\to ji}$. Moreover, we can obtain $T^{\lambda}_{k\ell\to ij}$ from T^{λ} by first appending a k into row i, then appending an ℓ into row j. However, in the case where $k > \ell$, and $i = j$, in which case we must also swap the boxes containing k and ℓ, so that the SSYT is valid. See Figure 19 for some examples of two-step SSYTs.

$T^{\lambda} =$

1	1	1	1	1
2	2	2	2	
3	3	3		
4	4			

$T^{\lambda}_{3,4\to 2,3} =$

1	1	1	1	1
2	2	2	2	3
3	3	3	4	
4	4			

$T^{\lambda}_{2,1\to 1,1} =$

1	1	1	1	1	1	2
2	2	2	2			
3	3	3				
4	4					

Figure 19: A partition $\lambda = (5, 4, 3, 2)$ and two examples of two-step SSYTs $T^{\lambda}_{k\ell\to ij}$. The added boxes are shaded.

Lemma 10.16. *The coefficient $\langle T' | T^{\lambda}_{k\to i}, \ell\rangle$ is nonzero* only *when T' satisfies one of the following two cases:*

(i) $T' = T^{\lambda}_{k\ell\to ij}$, or

(ii) $T' = T^{\lambda}_{\ell k \to ij}$ *and* $k > \ell$,

where j *is any valid row.*

Proof. Recall the Clebsch-Gordan insertion procedure from Theorem 10.9, during which we attempt to insert ℓ into a row of $T^{\lambda}_{k\to i}$ during Step 2. We consider the following two cases:

1. If we insert ℓ into row $j \le \ell$ that is different than i, the one-step SSYT $T^{\lambda}_{k\to i}$ only contains the letter j in that row. Since j is at most ℓ, the insertion procedure does not reach Step 3, but always jumps to Step 4. Step 4 will now append a new cell containing ℓ to the end of row j to obtain $T^{\lambda}_{k\ell\to ij}$, append it (along with the respective coefficient) to the output, and terminate.

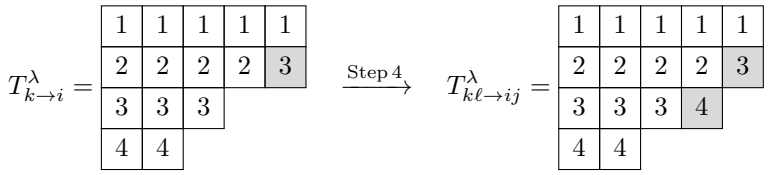


Figure 20: An example of how T' evolves during the Clebsch-Gordan insertion in the first case from the proof of Theorem 10.16 when $(k,\ell) = (3,4)$, and $(i,j) = (2,3)$, for $\lambda = (5,4,3,2)$.

2. If we insert ℓ into row i (assuming $i \le \ell$), the one-step SSYT $T^{\lambda}_{k\to i}$ contains the letter i in the first λ_i cells, and the letter k in the final cell of the row. If $k \le \ell$, then the insertion procedure does not reach Step 3, but always jumps to Step 4 and will append a new cell containing ℓ to the end of row i to obtain $T^{\lambda}_{k\ell\to ii}$.

   If, on the other hand, $k > \ell$, then ℓ will replace k at the end of row i. The current tableau becomes $T^{\lambda}_{\ell\to i}$, and the procedure will repeat Step 2 with the bumped letter k, which will be added in a valid row $j \le k$. The current SSYT will contain in row j the letters j, and possibly ℓ, which are not greater than k. Thus, we proceed to Step 4, during which a new cell with the letter k is appended at the end of row j. The final SSYT is thus $T^{\lambda}_{\ell k\to ij}$.

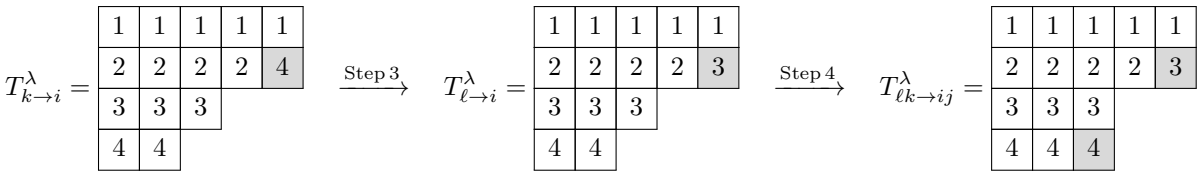


Figure 21: An example of how T' evolves during the Clebsch-Gordan insertion in the second case from the proof of Theorem 10.16 when $(k,\ell) = (4,3)$, and $(i,j) = (2,4)$, for $\lambda = (5,4,3,2)$.

We conclude that the only SSYTs T' with a nonzero coefficient are either $T^{\lambda}_{k\ell\to ij}$, or $T^{\lambda}_{\ell k\to ij}$ if $k > \ell$. □

In light of Theorem 10.16, we define the two-step Clebsch-Gordan coefficients as follows.

Definition 10.17 (Two-step Clebsch-Gordan coefficients)**.** Let λ be a Young diagram. The *two-step Clebsch-Gordan coefficients* are:

$$a^{\lambda}_{k\ell\to ij} \coloneqq \langle T^{\lambda}_{k\to i}|T^{\lambda},k\rangle \cdot \langle T^{\lambda}_{k\ell\to ij}|T^{\lambda}_{k\to i},\ell\rangle\,, \qquad b^{\lambda}_{k\ell\to ij} \coloneqq \langle T^{\lambda}_{k\to i}|T^{\lambda},k\rangle \cdot \langle T^{\lambda}_{k\ell\to ji}|T^{\lambda}_{k\to i},\ell\rangle\,.$$

Whenever a one-step or two-step SSYT is not valid, we define the corresponding coefficient to be 0.

Remark 10.18. The scalar factors that do not involve a bumped letter are always nonnegative, since they come from items (i) and (ii) of Theorem 10.7. Therefore, the two-step Clebsch-Gordan coefficients $a^{\lambda}_{k\ell\to ij}$ are always nonnegative.

11 Constructing the Schur transform

The Clebsch-Gordan transform can be used to give a recursive construction of the Schur transform. That is, we can construct a unitary $\mathcal{U}_{\mathrm{SW}}^{(n)}$ that acts as in Equation (17): for all $\pi \in S_n$ and $U \in U(d)$,

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \left(\mathcal{P}^{(n)}(\pi)\mathcal{Q}^{(n)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} = \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\pi) \otimes \nu_\lambda(U).$$

This construction is originally due to [Har05, BCH05]. Note that in these works, the Schur transform was only required to compute a Young-Yamanouchi basis in the permutation register. We will prove that the construction can be used to prepare Young's orthogonal form as well.

For intuition, we now sketch how we will do a recursive step using the Clebsch-Gordan transform. Assume we have constructed $\mathcal{U}_{\mathrm{SW}}^{(n)}$ already, so that we can rewrite vectors in $(\mathbb{C}^d)^{\otimes n}$ in the Schur basis. We want to give a construction of $\mathcal{U}_{\mathrm{SW}}^{(n+1)}$.

Start with a tensor product $|\lambda\rangle \otimes |S\rangle \otimes |T\rangle \otimes |i\rangle$, obtained by using $\mathcal{U}_{\mathrm{SW}}^{(n)}$ on the first n qudits. By applying the Clebsch-Gordan transform, $\mathcal{U}_{\mathrm{CG}}^{(n+1)}$, we take this state to a superposition of vectors of the form $|\lambda\rangle \otimes |S\rangle \otimes |\mu\rangle \otimes |T'\rangle$, where T' is of shape μ. If we now rearrange factors, we can obtain $|\mu\rangle \otimes (|\lambda\rangle \otimes |S\rangle) \otimes |T'\rangle$. In this product, $|\lambda\rangle \otimes |S\rangle$ is exactly the labeling of the Young-Yamanouchi basis vector corresponding to $|S'\rangle$, where $S' \in \mathrm{SYT}(\mu)$ is obtained from S by appending a box labeled $(n+1)$ at the location of the single box of $\mu \setminus \lambda$. Formally, the rearrangement is performed by the following unitary.

Definition 11.1 (Rearrangement unitary)**.** Let $\mathcal{U}_{\mathrm{R}}^{(n+1)} : \bigoplus_\lambda \mathrm{Sp}_\lambda \otimes \left(\bigoplus_\mu V_\mu^d\right) \to \bigoplus_\mu \left(\bigoplus_\lambda \mathrm{Sp}_\lambda \otimes V_\mu^d\right)$ mapping

$$|\lambda\rangle \otimes |S\rangle \otimes |\mu\rangle \otimes |T'\rangle \to |\mu\rangle \otimes |\lambda\rangle \otimes |S\rangle \otimes |T'\rangle,$$

for all $\lambda \vdash n$, $S \in \mathrm{SYT}(\lambda)$, μ with $\lambda \nearrow \mu$ and $T' \in \mathrm{SSYT}(\mu, d)$.

The structure of the new basis vectors suggests that the product $\widehat{\mathcal{U}}_{\mathrm{SW}}^{(n+1)} \coloneqq \mathcal{U}_{\mathrm{R}}^{(n+1)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n+1)} \cdot (\mathcal{U}_{\mathrm{SW}}^{(n)} \otimes I)$ may be performing the Schur transform on $(\mathbb{C}^d)^{\otimes(n+1)}$, i.e.

$$\widehat{\mathcal{U}}_{\mathrm{SW}}^{(n+1)} \cdot \left(\mathcal{P}^{(n+1)}(\pi)\mathcal{Q}^{(n+1)}(U)\right) \cdot \widehat{\mathcal{U}}_{\mathrm{SW}}^{(n+1)\dagger} \stackrel{?}{=} \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \kappa_\mu(\pi) \otimes \nu_\mu(U). \tag{68}$$

Indeed, loosely speaking, the "unitary register" (the register containing $|T'\rangle$) transforms according to the ν_μ representation, by the definition of the Clebsch-Gordan transform. Moreover, by the uniqueness of the decomposition into irreps by Schur-Weyl duality, the "permutation register" (the register containing $|S'\rangle = |\lambda\rangle \otimes |S\rangle$) transforms according to some representation of S_{n+1} isomorphic to κ_μ, which we recall is Young's orthogonal representation. Moreover, this representation's restriction to S_n is κ_λ (as we will show in Theorem A.4), so that the permutation register's basis vectors constitute a Young-Yamanouchi basis.

It turns out that Equation (68) *does* hold. Based on this, we make the following definition.

Definition 11.2 (Schur transform)**.** We define the *Schur transform*, $\mathcal{U}_{\mathrm{SW}}^{(n)} : (\mathbb{C}^d)^{\otimes n} \to \bigoplus_{\lambda \vdash (n,d)} \mathrm{Sp}_\lambda \otimes V_\lambda^d$, by the following recursive construction. We explicitly define $\mathcal{U}_{\mathrm{SW}}^{(1)}$ as the unitary such that

$$\mathcal{U}_{\mathrm{SW}}^{(1)} |i\rangle \coloneqq |\square\rangle \otimes |\boxed{1}\rangle \otimes |i\rangle,$$

for all $i \in [d]$. Then, for $n > 1$,

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \coloneqq \mathcal{U}_{\mathrm{R}}^{(n)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n)} \cdot (\mathcal{U}_{\mathrm{SW}}^{(n-1)} \otimes I).$$

Theorem 11.3. *The unitary in Theorem 11.2 is a Schur transform. That is, it satisfies*

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \left(\mathcal{P}^{(n)}(\pi)\mathcal{Q}^{(n)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} = \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\pi) \otimes \nu_\lambda(U), \tag{69}$$

for all $\pi \in S_n$ and $U \in U(d)$.

This theorem is useful for us, because we will eventually want to make use of explicit formulas for matrix elements of Young's orthogonal form. If the permutation registers computed a generic Young-Yamanouchi basis only, then the corresponding matrix elements may contain unknown phases, making them harder to use.

Proving Theorem 11.3 is a little involved, and we defer its proof to Part V.

12 Dual Clebsch-Gordan transform

It will be convenient for us to use the *dual Clebsch-Gordan transform* in our proofs as well. We now give a minimal introduction to what we will need.

Definition 12.1 (Rational representations)**.** A finite-dimensional representation (μ, V) of a matrix group G is *rational*, if, expressed in some basis of V, every matrix element $\mu(g)_{ij}$ is a rational function in the entries of $g \in G$.

A representation being rational is a basis-independent notion, since if the entries of $\mu(g)$ are rational in one basis, then they are also rational in any other basis. Rational representations generalize polynomial representations. The rational irreps can be classified as follows.

Theorem 12.2. *The rational irreps of $U(d)$ are indexed by tuples $\lambda \in \mathbb{Z}^d$, such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. The irrep corresponding to λ is denoted $(\nu_\lambda, V_\lambda^d)$.*

Unlike Young diagrams, the tuples indexing rational irreps do not necessarily have nonnegative row lengths. These are sometimes called *staircases* or *rational tableaux* in the literature [Ngu24]. One new rational representation is particularly important.

Definition 12.3 (Antifundamental representation)**.** The *antifundamental representation* of $U(d)$ is the representation $(\nu_{-\square}, V_{-\square}^d)$ defined by $V_{-\square}^d := \mathbb{C}^d$ and $\nu_{-\square}(U) = U^*$. This is the irrep corresponding to $\lambda = (0^{d-1}, -1)$. The computational basis elements are denoted $\{|\overline{k}\rangle\}_{k\in[d]}$ in this context.

It turns out (see for example [Ngu24, Equation (49)]) that the tensor product of a polynomial irrep V_λ^d of $U(d)$ with $V_{-\square}^d$ decomposes into rational irreps in a way similar to Equation (55).

$$V_\lambda^d \otimes V_{-\square}^d \overset{U(d)}{\cong} \bigoplus_{\substack{\mu=\lambda-\square \\ \ell(\mu)\leq d}} V_\mu^d. \tag{70}$$

Here by $\lambda - \square$ we mean a staircase tableau of the form $\lambda - e_i$, for any $i \in [d]$. So the right-hand side of the above expression iterates over all valid tuples that can be obtained from λ by removing a box. Note that with tuples, we may also remove boxes from rows with zero or fewer boxes.

Therefore, there exists a unitary $\mathcal{U}_{\mathrm{dCG}}^{(\lambda)} : V_\lambda^d \otimes \mathbb{C}^d \to \oplus_\mu V_\mu^d$ that implements the isomorphism in Equation (70). In other words, for all $U \in U(d)$,

$$\mathcal{U}_{\mathrm{dCG}}^{(\lambda)} \cdot \big(\nu_\lambda(U) \otimes U^*\big) \cdot \mathcal{U}_{\mathrm{dCG}}^{(\lambda)\dagger} = \sum_{\substack{\mu=\lambda-\square \\ \ell(\mu)\leq d}} |\mu\rangle\langle\mu| \otimes \nu_\mu(U). \tag{71}$$

We observe that the choice of $\mathcal{U}_{\mathrm{dCG}}^{(\lambda)}$ is not unique. Indeed, one can see from Schur's Lemma that any unitary of the form

$$\Big(\sum_{\mu=\lambda-\square} \alpha^\mu \cdot |\mu\rangle\langle\mu| \otimes I_{V_\mu^d}\Big) \cdot \mathcal{U}_{\mathrm{dCG}}^{(\lambda)}$$

where α^μ is a complex scalar of unit norm will also satisfy Equation (71). The above transformation defines the dual Clebsch-Gordan coefficients.

Definition 12.4 (Dual Clebsch-Gordan coefficients)**.** Let $\lambda \in \mathbb{Z}^d$, $T \in \mathrm{SSYT}(\lambda, d)$, and $k \in [d]$. We can write

$$\mathcal{U}_{\mathrm{dCG}}^{(\lambda)} \cdot |T\rangle \otimes |\overline{k}\rangle = \sum_{\mu=\lambda-\square} \sum_{T'} \langle \mu, T'| \mathcal{U}_{\mathrm{dCG}}^{(\lambda)} |T, \overline{k}\rangle \cdot |\mu, T'\rangle .$$

Here, T' indexes a basis of the irrep corresponding to μ.[2]. Then coefficients $\langle \mu, T' | \mathcal{U}_{\mathrm{dCG}}^{(\lambda)} | T, \overline{k} \rangle$ are known as the *dual Clebsch-Gordan coefficients*. We will typically write $\langle T' | T, \overline{k} \rangle$ as shorthand for $\langle \mu, T' | \mathcal{U}_{\mathrm{dCG}}^{(\lambda)} | T, \overline{k} \rangle$, leaving $\mu = \mathsf{sh}(T')$ and $\mathcal{U}_{\mathrm{dCG}}^{(\lambda)}$ implicit.

We are interested in the relationship between the Clebsch-Gordan and the dual Clebsch-Gordan coefficients when both λ and μ correspond to polynomial irreps. In particular, we use the following result from Equation (13) of Section 18.2.1 of [VK92].

Theorem 12.5 ([VK92])**.** *Let $\lambda, \mu \in \mathbb{Z}^d$ be Young diagrams with nonnegative row lengths such that $\mu = \lambda - \square$. Then there exists a unitary $\mathcal{U}_{\mathrm{dCG}}^{(\lambda)}$ that performs the dual Clebsch-Gordan transform such that:*

$$\langle \mu, T' | \mathcal{U}_{\mathrm{dCG}}^{(\lambda)} | T, \overline{k} \rangle = \sqrt{\frac{\dim(V_\mu^d)}{\dim(V_\lambda^d)}} \cdot \langle \lambda, T | \mathcal{U}_{\mathrm{CG}}^{(\mu)} | T', k \rangle \, .$$

[2]We have already seen that for polynomial representations, T' can be thought of as an SSYT. For rational representations, there is also a connection to tableaux. However, we will only make use of this in the case when λ and μ are polynomial representations, and so we will not need the more general case. See, e.g. [Ngu24], for more details.

Part IV
The moments of the debiased Keyl's estimator

13 The first moment

In this section, we formally prove Theorem 1.3 from the introduction. That is, we show that the debiased Keyl's algorithm, given in Theorem 1.2, is an unbiased estimator. We start by developing some tools related to the one-step Clebsch-Gordan coefficients.

13.1 Technical lemmas concerning one-step Clebsch-Gordan coefficients

In this subsection, we collect some identities about the one-step Clebsch-Gordan coefficients that we will use in our proofs of the first and second moments of our estimator. Recall that for a Young diagram $\lambda = (\lambda_1, \ldots, \lambda_d)$, we use $\lambda_{\leq j} = (\lambda_1, \ldots, \lambda_j)$ to denote the truncation of this tableau to its first j rows.

Notation 13.1. Let λ be a Young diagram, and let $i \in [d]$. We write $D^\lambda_{k\to i}$ as shorthand for the ratio of dimensions

$$D^\lambda_{k\to i} := \frac{\dim(V^k_{\lambda_{\leq k}+e_i})}{\dim(V^k_{\lambda_{\leq k}})},$$

when $i \leq k \leq d$, and define $D^\lambda_{k\to i} = 0$ when $0 \leq k < i$.

Notation 13.2. Let λ be a Young diagram, and let $i \in [d]$. We write C^λ_i as shorthand for the quantity $C^\lambda_i := \lambda_i - i$, which is also the content of the rightmost box in row i in λ. We will omit λ when clear from context.

The first lemma relates the one-step Clebsch-Gordan coefficients to the dimensions of truncated tableaux.

Lemma 13.3. *Let λ be a Young diagram, and $k, i \in [d]$. Then*

$$|c^\lambda_{k\to i}|^2 = D^\lambda_{k\to i} - D^\lambda_{k-1\to i}, \tag{72}$$

and

$$(C_i - C_k) \cdot |c^\lambda_{k\to i}|^2 = D^\lambda_{k-1\to i}. \tag{73}$$

These equations together immediately imply:

$$(C_i - C_k) \cdot D^\lambda_{k\to i} = (C_i - C_k + 1) \cdot D^\lambda_{k-1\to i}. \tag{74}$$

Proof. The proof follows by direct calculation. If $k < i$, each side of both Equations (72) and (73) is equal to zero, so the equations hold. When $k \geq i$, recall the Clebsch-Gordan coefficient expression from Theorem 10.13:

$$|c^\lambda_{k\to i}|^2 = \left(\prod_{j<i} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)}\right) \cdot \left(\prod_{j=i+1}^{k-1} \frac{(\lambda_i - \lambda_j) + (j-i) + 1}{(\lambda_i - \lambda_j) + (j-i)}\right) \cdot \frac{1}{(\lambda_i - \lambda_k) + (k-i) + \delta_{ik}}.$$

Now suppose $k > i$. We use the Weyl dimension formula from Equation (15) to write

$$\begin{aligned}
\frac{\dim(V^k_{\lambda_{\leq k}+e_i})}{\dim(V^k_{\lambda_{\leq k}})} &= \left(\prod_{j<i} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)}\right) \cdot \left(\prod_{j=i+1}^{k} \frac{(\lambda_i - \lambda_j) + (j-i) + 1}{(\lambda_i - \lambda_j) + (j-i)}\right) \\
&= \left(\prod_{j<i} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)}\right) \cdot \left(\prod_{j=i+1}^{k-1} \frac{(\lambda_i - \lambda_j) + (j-i) + 1}{(\lambda_i - \lambda_j) + (j-i)}\right) \cdot \left(\frac{1}{(\lambda_i - \lambda_k) + (k-i)} + 1\right) \\
&= |c^\lambda_{k\to i}|^2 + \frac{\dim(V^{k-1}_{\lambda_{\leq k-1}+e_i})}{\dim(V^{k-1}_{\lambda_{\leq k-1}})}.
\end{aligned}$$

Thus, $D^\lambda_{k\to i} = |c^\lambda_{k\to i}|^2 + D^\lambda_{k-1\to i}$. Moreover, the second equality above implies that for $k > i$,

$$D^\lambda_{k\to i} = D^\lambda_{k-1\to i} \cdot \left(\frac{1}{C_i - C_k} + 1\right) \implies (C_i - C_k) \cdot (D^\lambda_{k\to i} - D^\lambda_{k-1\to i}) = D^\lambda_{k-1\to i}.$$

Lastly, suppose $k = i$. Then

$$\begin{aligned}\frac{\dim(V^k_{\lambda_{\le k}+e_i})}{\dim(V^k_{\lambda_{\le k}})} &= \prod_{j<i} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \\ &= \left(\prod_{j<i} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)}\right) \cdot \frac{1}{(\lambda_i - \lambda_k) + (k-i) + \delta_{ik}} = |c^\lambda_{k\to i}|^2.\end{aligned}$$

Since $D^\lambda_{i-1\to i}$ is defined to be 0, we conclude that $D^\lambda_{i\to i} = |c^\lambda_{i\to i}|^2 + D^\lambda_{i-1\to i}$. Finally, both sides of Equation (73) are equal to zero when $i = k$. □

Telescoping the above sum in Equation (72) gives the following corollary.

Corollary 13.4. *The one-step Clebsch-Gordan coefficients satisfy the following relation:*

$$\sum_{k=1}^{s} |c^\lambda_{k\to i}|^2 = D^\lambda_{s\to i}.$$

The next lemma shows that the one-step CG coefficient $c^\lambda_{k\to i}$, regarded as a function of k alone (i.e. for fixed λ, d, and i), is constant on the blocks of λ.

Lemma 13.5. *Let λ be a Young diagram, and let $\{k, k+1, \ldots, \ell\}$ be a block of λ, i.e. $\lambda_k = \lambda_{k+1} = \cdots = \lambda_\ell$. Then, for each $i \le d$, the one-step Clebsch-Gordan coefficients that correspond to adding any value in $\{k, \ldots, \ell\}$ satisfy:*

$$c^\lambda_{k\to i} = c^\lambda_{k+1\to i} = \cdots = c^\lambda_{\ell\to i}. \tag{75}$$

Proof. We first observe that if $i \in \{k, \ldots, \ell\}$, then we must have $i = k$ if $\lambda + e_i$ is a valid Young diagram. If $k+1 \le i \le \ell$, then Equation (75) holds trivially as all coefficients are zero, since $\lambda + e_i$ is invalid. Furthermore, if $i > \ell$, even in the case where $\lambda + e_i$ is a valid Young diagram, $T^\lambda_{m\to i}$ is an invalid SSYT for all $m \in \{k, \ldots, \ell\}$, since $m < i$. In this case too, Equation (75) holds trivially.

It remains to show that $c^\lambda_{m+1\to i} = c^\lambda_{m\to i}$ for all $m \in \{k, \ldots, \ell - 1\}$ whenever $i \le k$. We consider the case when $i = m$ (which happens when they are both equal to k) separately.

1. If $i = m = k$, then, from Theorem 10.13, we have

$$\begin{aligned} c^\lambda_{k+1\to i} &= \left| \frac{1}{(\lambda_i - \lambda_{k+1}) + (k+1-i) + \delta_{i,k+1}} \cdot \prod_{j=1, j\neq i}^{k} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right|^{1/2} \\ &= \left| \frac{1}{(\lambda_i - \lambda_k) + (k+1-i)} \cdot \prod_{j=1, j\neq k}^{k} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right|^{1/2} && (\lambda_k = \lambda_{k+1}) \\ &= \left| \frac{1}{(\lambda_i - \lambda_k) + (k-i) + \delta_{ik}} \cdot \prod_{j=1, j\neq i}^{k-1} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right|^{1/2} = c^\lambda_{k\to i}. \end{aligned}$$

2. If $i < m$, then

$$\begin{aligned} c^{\lambda}_{m+1\to i} &= \left| \frac{1}{(\lambda_i - \lambda_{m+1}) + (m+1-i) + \delta_{i,m+1}} \cdot \prod_{j=1, j\neq i}^{m} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right|^{1/2} \\ &= \left| \frac{1}{(\lambda_i - \lambda_{m+1}) + (m+1-i)} \cdot \frac{(\lambda_m - \lambda_i) + (i-m) - 1}{(\lambda_m - \lambda_i) + (i-m)} \cdot \prod_{j=1, j\neq i}^{m-1} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right|^{1/2} \\ &= \left| \frac{1}{(\lambda_i - \lambda_m) + (m-i) + 1} \cdot \frac{(\lambda_i - \lambda_m) + (m-i) + 1}{(\lambda_i - \lambda_m) + (m-i)} \cdot \prod_{j=1, j\neq i}^{m-1} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right|^{1/2} \\ &= \left| \frac{1}{(\lambda_i - \lambda_m) + (m-i) + \delta_{im}} \cdot \prod_{j=1, j\neq i}^{m-1} \frac{(\lambda_j - \lambda_i) + (i-j) - 1}{(\lambda_j - \lambda_i) + (i-j)} \right|^{1/2} = c^{\lambda}_{m\to i}, \end{aligned}$$

where we used the fact that $\lambda_m = \lambda_{m+1}$ in the third step. □

13.2 Proof of the first moment

Theorem 13.6 (Theorem 1.3, restated)**.** *Let $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $\rho^{\otimes n}$. Then $\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \rho$.*

Let us interpret what it means for a tomography algorithm to give an unbiased estimator of a density matrix ρ. A tomography algorithm is determined by a POVM $M = \{M_i\}$ on $(\mathbb{C}^d)^{\otimes n}$ and a corresponding set of possible outputs $\{\widehat{\rho}_i\}$: the algorithm measures $\rho^{\otimes n}$ using M, and outputs $\widehat{\rho}_i$ upon obtaining outcome M_i. If the tomography algorithm is unbiased, then

$$\sum_i \widehat{\rho}_i \cdot \operatorname{tr}(M_i \cdot \rho^{\otimes n}) = \rho.$$

Some elementary manipulations of the left-hand side give

$$\sum_i \widehat{\rho}_i \cdot \operatorname{tr}(M_i \cdot \rho^{\otimes n}) = \sum_i \operatorname{tr}_{[n]}(M_i \otimes \widehat{\rho}_i \cdot \rho^{\otimes n} \otimes I) = \operatorname{tr}_{[n]}\left(\left(\sum_i M_i \otimes \widehat{\rho}_i\right) \cdot \rho^{\otimes n} \otimes I\right) = \operatorname{tr}_{[n]}(M^{(1)}_{\text{avg}} \cdot \rho^{\otimes n} \otimes I),$$

where we define

$$M^{(1)}_{\text{avg}} \coloneqq \sum_i M_i \otimes \widehat{\rho}_i.$$

So our goal is to satisfy $\operatorname{tr}_{[n]}(M^{(1)}_{\text{avg}} \cdot \rho^{\otimes n} \otimes I) = \rho$. This would happen, for example, if $M^{(1)}_{\text{avg}} = (1, n+1)$, where $(1, n+1) \coloneqq \mathcal{P}(1, n+1)$ and $\mathcal{P}(\cdot)$ is the representation of the symmetric group from Theorem 2.46, since

$$\operatorname{tr}_{[n]}((1, n+1) \cdot \rho^{\otimes n} \otimes I) = \operatorname{tr}_{[2]}((1,2) \cdot \rho \otimes I) \cdot \operatorname{tr}(\rho)^{n-1} = \rho \cdot \operatorname{tr}(\rho)^{n-1} = \rho.$$

Similarly, it would also be true if $M^{(1)}_{\text{avg}} = (k, n+1)$ for any $k \in [n]$, and also, more generally, if

$$M^{(1)}_{\text{avg}} = \sum_{k=1}^{n} \alpha_i \cdot (k, n+1),$$

for any constants with $\alpha_1 + \cdots + \alpha_n = 1$. One might then expect that any unbiased algorithm that treats its n registers symmetrically should satisfy

$$M^{(1)}_{\text{avg}} = \frac{1}{n} \cdot \Big((1, n+1) + \cdots + (n, n+1)\Big) = \frac{1}{n} \cdot X_{n+1}, \tag{76}$$

where X_{n+1} is the Jucys-Murphy element as in Theorem 2.35. We show the debiased Keyl's algorithm satisfies Equation (76) below in Theorem 13.7, from which Theorem 1.3 follows as an immediate corollary.

Lemma 13.7. *Let $M = \{M_{\lambda,U}\}$ be the POVM that corresponds to Keyl's algorithm. Consider a legal estimator, which outputs $\widehat{\rho}_{\lambda,U} \coloneqq U \cdot f(\lambda) \cdot U^\dagger/n$ when the POVM element indexed by λ and U is obtained. Let $M^{(1)}_{\mathrm{avg}} = \sum_{\lambda,U} M_{\lambda,U} \otimes \widehat{\rho}_{\lambda,U}$, where $\widehat{\rho}$. Then, for any legal estimator,*

$$M^{(1)}_{\mathrm{avg}} = \frac{1}{n} \cdot \Big((1, n+1) + \cdots + (n, n+1) \Big) = \frac{1}{n} \cdot X_{n+1}.$$

Proof. The POVM used by any legal estimator is the same as in the original Keyl's algorithm. The POVM elements are indexed by $\lambda \vdash n$ and $U \in U(d)$, and, from Equation (22), is given by:

$$M_{\lambda,U} = \mathcal{U}^{(n)\dagger}_{\mathrm{SW}} \cdot \Big(\dim(V^d_\lambda) \cdot \, |\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \nu_\lambda(U) \cdot \, |T^\lambda\rangle\langle T^\lambda| \cdot \nu_\lambda(U^\dagger) \cdot dU \Big) \cdot \mathcal{U}^{(n)}_{\mathrm{SW}}.$$

Upon receiving λ and U, the state we output is $\widehat{\rho}_{\lambda,U} = U \cdot f(\lambda) \cdot U^\dagger/n$. We can therefore rewrite $M^{(1)}_{\mathrm{avg}}$ as:

$$\begin{aligned}
&M^{(1)}_{\mathrm{avg}} \\
&= \frac{1}{n} \sum_\lambda \int_U M_{\lambda,U} \otimes \widehat{\rho}_{\lambda,U} \\
&= \frac{1}{n} \sum_\lambda \dim(V^d_\lambda) \int_U \mathcal{U}^{(n)\dagger}_{\mathrm{SW}} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \nu_\lambda(U) \cdot \, |T^\lambda\rangle\langle T^\lambda| \cdot \nu_\lambda(U^\dagger) \Big) \cdot \mathcal{U}^{(n)}_{\mathrm{SW}} \otimes U f(\lambda) U^\dagger \cdot \mathrm{d}U \\
&= \frac{1}{n} \sum_\lambda \dim(V^d_\lambda) \sum_S \sum_{k=1}^d f(\lambda)_k \int_U \mathcal{U}^{(n)\dagger}_{\mathrm{SW}} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes \, |S\rangle\langle S| \otimes \nu_\lambda(U) \cdot \, |T^\lambda\rangle\langle T^\lambda| \cdot \nu_\lambda(U^\dagger) \Big) \cdot \mathcal{U}^{(n)}_{\mathrm{SW}} \otimes U \, |k\rangle\langle k| \, U^\dagger \cdot \mathrm{d}U.
\end{aligned} \tag{77}$$

Here, $S \in \mathrm{SYT}(\lambda)$, and $k \in [d]$. We can manipulate the integrand as follows:

$$\begin{aligned}
&\mathcal{U}^{(n)\dagger}_{\mathrm{SW}} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes \, |S\rangle\langle S| \otimes \nu_\lambda(U) \cdot \, |T^\lambda\rangle\langle T^\lambda| \cdot \nu_\lambda(U^\dagger) \Big) \cdot \mathcal{U}^{(n)}_{\mathrm{SW}} \otimes U \, |k\rangle\langle k| \, U^\dagger \\
&= U^{\otimes n} \cdot \mathcal{U}^{(n)\dagger}_{\mathrm{SW}} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes \, |S\rangle\langle S| \otimes \, |T^\lambda\rangle\langle T^\lambda| \Big) \cdot \mathcal{U}^{(n)}_{\mathrm{SW}} \cdot U^{\dagger,\otimes n} \otimes U \, |k\rangle\langle k| \, U^\dagger \\
&= U^{\otimes(n+1)} \cdot \mathcal{U}^\dagger \cdot \Big(|\lambda\rangle\langle\lambda| \otimes \, |S\rangle\langle S| \otimes \, |T^\lambda\rangle\langle T^\lambda| \otimes \, |k\rangle\langle k| \Big) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+1)}.
\end{aligned}$$

Here, we are using the notation

$$\mathcal{U} \coloneqq \mathcal{U}^{(n)}_{\mathrm{SW}} \otimes I,$$

as shorthand. We have also used

$$U^{\otimes n} \cdot \mathcal{U}^{(n)\dagger}_{\mathrm{SW}} = \mathcal{U}^{(n)\dagger}_{\mathrm{SW}} \cdot \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \le d}} |\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \nu_\lambda(U).$$

We proceed by using the Clebsch-Gordan transform to rewrite the state inside parentheses using:

$$\begin{aligned}
\mathcal{U}^{(n+1)}_{\mathrm{R}} \cdot \mathcal{U}^{(n+1)}_{\mathrm{CG}} \cdot |\lambda\rangle \otimes |S\rangle \otimes \big|T^\lambda\big\rangle \otimes |k\rangle &= \mathcal{U}^{(n+1)}_{\mathrm{R}} \cdot \sum_{i=1}^d c^\lambda_{k\to i} \, |\lambda\rangle \otimes |S\rangle \otimes |\lambda + e_i\rangle \otimes \big|T^\lambda_{k\to i}\big\rangle \\
&= \sum_{i=1}^d c^\lambda_{k\to i} \, |\lambda + e_i\rangle \otimes |S_i\rangle \otimes \big|T^\lambda_{k\to i}\big\rangle \, .
\end{aligned}$$

Here, S_i is the SYT obtained from S, by adding a box labeled $n+1$ to the end of the i-th row. This is well defined for all λ such that $\lambda + e_i$ is a valid Young diagram. Recalling that the Clebsch-Gordan coefficients are real, and our definition

$$\mathcal{U}^{(n+1)}_{\mathrm{SW}} \coloneqq \mathcal{U}^{(n+1)}_{\mathrm{R}} \cdot \mathcal{U}^{(n+1)}_{\mathrm{CG}} \cdot (\mathcal{U}^{(n)}_{\mathrm{SW}} \otimes I) = \mathcal{U}^{(n+1)}_{\mathrm{R}} \cdot \mathcal{U}^{(n+1)}_{\mathrm{CG}} \cdot \mathcal{U},$$

we have

$$
\begin{aligned}
&U^{\otimes(n+1)} \cdot \mathcal{U}^\dagger \cdot \Big(|\lambda\rangle\langle\lambda| \otimes |S\rangle\langle S| \otimes |T_\lambda\rangle\langle T_\lambda| \otimes |k\rangle\langle k| \Big) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+1)} \\
&= U^{\otimes(n+1)} \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} \cdot \Big(\sum_{i,j=1}^{d} c^\lambda_{k\to i} c^\lambda_{k\to j} \, |\lambda + e_i\rangle\langle\lambda + e_j| \otimes |S_i\rangle\langle S_j| \otimes |T^\lambda_{k\to i}\rangle\langle T^\lambda_{k\to j}| \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot U^{\dagger,\otimes(n+1)} \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} \cdot \Big(\sum_{i,j=1}^{d} c^\lambda_{k\to i} c^\lambda_{k\to j} \, |\lambda + e_i\rangle\langle\lambda + e_j| \otimes |S_i\rangle\langle S_j| \otimes \nu_{\lambda+e_i}(U) \, |T^\lambda_{k\to i}\rangle\langle T^\lambda_{k\to j}| \, \nu_{\lambda+e_j}(U)^\dagger \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)}.
\end{aligned}
$$

After substituting this back into Equation (77), the integral over unitaries in $M^{(1)}_{\mathrm{avg}}$ can then be pushed under the sum and onto the unitary register. The resulting integral can be evaluated using Schur's lemma:

$$
\int_U \nu_{\lambda+e_i}(U) \, |T^\lambda_{k\to i}\rangle\langle T^\lambda_{k\to j}| \, \nu_{\lambda+e_j}(U)^\dagger \cdot \mathrm{d}U = \frac{1}{\dim(V^d_{\lambda+e_i})} \cdot I_{\dim(V^d_{\lambda+e_i})} \cdot \delta_{ij},
$$

since the integral is an intertwining operator between the irreducible representations $\nu_{\lambda+e_i}$ and $\nu_{\lambda+e_j}$, which are non-isomorphic unless $i = j$. Therefore,

$$
\begin{aligned}
&\int_U U^{\otimes(n+1)} \cdot \mathcal{U}^\dagger \cdot \Big(|\lambda\rangle\langle\lambda| \otimes |S\rangle\langle S| \otimes |T_\lambda\rangle\langle T_\lambda| \otimes |k\rangle\langle k| \Big) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+1)} \mathrm{d}U \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} \cdot \Big(\sum_{i=1}^{d} \frac{|c^\lambda_{k\to i}|^2}{\dim(V^d_{\lambda+e_i})} \cdot |\lambda + e_i\rangle\langle\lambda + e_i| \otimes |S_i\rangle\langle S_i| \otimes I_{\dim(V^d_{\lambda+e_i})} \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)}.
\end{aligned}
$$

Our expression for $M^{(1)}_{\mathrm{avg}}$ thus becomes:

$$
\begin{aligned}
M^{(1)}_{\mathrm{avg}} &= \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} \cdot \Big(\frac{1}{n} \sum_\lambda \dim(V^d_\lambda) \sum_{k=1}^{d} f(\lambda)_k \sum_S \sum_{i=1}^{d} \frac{|c^\lambda_{k\to i}|^2}{\dim(V^d_{\lambda+e_i})} \cdot |\lambda + e_i\rangle\langle\lambda + e_i| \otimes |S_i\rangle\langle S_i| \otimes I_{\dim(V^d_{\lambda+e_i})} \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)} \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} \cdot \Big(\frac{1}{n} \sum_\lambda \sum_S \sum_{i=1}^{d} \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda+e_i})} \cdot \Big(\sum_{k=1}^{d} f(\lambda)_k \cdot |c^\lambda_{k\to i}|^2 \Big) \cdot |\lambda + e_i\rangle\langle\lambda + e_i| \otimes |S_i\rangle\langle S_i| \otimes I_{\dim(V^d_{\lambda+e_i})} \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)}.
\end{aligned}
$$

We have shown so far that $M^{(1)}_{\mathrm{avg}}$ is diagonal in the $(n+1)$-copy Schur basis, and the expression below gives each diagonal entry:

$$
\frac{1}{n} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda+e_i})} \cdot \Big(\sum_{k=1}^{d} f(\lambda)_k \cdot |c^\lambda_{k\to i}|^2 \Big).
$$

We prove in Theorem 13.9 below, using our lemmas for the one-step Clebsch-Gordan coefficients, that

$$
\frac{1}{n} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda+e_i})} \cdot \Big(\sum_{k=1}^{d} f(\lambda)_k \cdot |c^\lambda_{k\to i}|^2 \Big) = \frac{1}{n} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda+e_i})} \cdot (\lambda_i + 1 - i) \cdot \frac{\dim(V^d_{\lambda+e_i})}{\dim(V^d_\lambda)} = \frac{1}{n} \cdot (\lambda_i + 1 - i). \quad (78)
$$

Finally,

$$
\begin{aligned}
M^{(1)}_{\mathrm{avg}} &= \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} \cdot \Big(\frac{1}{n} \sum_\lambda \sum_S \sum_{i=1}^{d} (\lambda_i + 1 - i) \cdot |\lambda + e_i\rangle\langle\lambda + e_i| \otimes |S_i\rangle\langle S_i| \otimes I_{\dim(V^d_{\lambda+e_i})} \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)} \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} \cdot \Big(\frac{1}{n} \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \le d}} \sum_{S' \in \mathrm{SYT}(\mu)} \mathrm{cont}_{S'}(n+1) \cdot |\mu\rangle\langle\mu| \otimes |S'\rangle\langle S'| \otimes I_{\dim(V^d_\mu)} \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)}.
\end{aligned}
$$

Comparing with Equation (19), we conclude:

$$
M^{(1)}_{\mathrm{avg}} = \frac{1}{n} \cdot X_{n+1}. \qquad \square
$$

To complete the proof of our first moment, it suffices to prove that the following identity from Equation (78) holds, for all legal Young diagram transformations f:

$$\sum_{k=1}^{d} f(\lambda)_k \cdot |c^{\lambda}_{k\to i}|^2 = (\lambda_i + 1 - i) \cdot D^{\lambda}_{d\to i}.$$

We first show in Theorem 13.8 below that this identity holds in the special case when f corresponds to the staircase transformation. Then we use the property that the one-step Clebsch-Gordan coefficients are constant on the blocks of a diagram to extend this result to all f's in Theorem 13.9.

Lemma 13.8. *Let λ be a Young diagram, and $\lambda^{\uparrow\uparrow}$ be the outcome of the staircase transformation on λ. Then*

$$\sum_{k=1}^{d} \lambda^{\uparrow\uparrow}_k \cdot |c^{\lambda}_{k\to i}|^2 = (\lambda_i + 1 - i) \cdot D^{\lambda}_{d\to i}. \tag{79}$$

Proof. Fix λ and i. We prove the lemma by induction on d. Note first though that $\lambda^{\uparrow\uparrow}_k = \lambda_k + d + 1 - 2k$ implicitly depends on d. When we would like to emphasize d, we will write $\lambda^{\uparrow\uparrow}_k(d)$. We are also taking the convention that $\lambda_k = 0$ for all $k > \ell(\lambda)$.

The base case is $d = 1$. If $i = 1$, then

$$\sum_{k=1}^{d} \lambda^{\uparrow\uparrow}_k(d=1) \cdot |c^{\lambda}_{k\to i}|^2 = \lambda^{\uparrow\uparrow}_1(d=1) \cdot |c^{\lambda}_{1\to 1}|^2 = \lambda_1 \cdot |c^{\lambda}_{1\to 1}|^2 = \lambda_1 \cdot D^{\lambda}_{1\to i} = (\lambda_i + 1 - i) \cdot D^{\lambda}_{d\to i},$$

where we have used Theorem 13.4 in the second-last step and the fact that $d = i = 1$ in the final step. If instead $i > 1$, then both sides of Equation (79) vanish.

Now assume the lemma holds for $d - 1$. We use the fact that $\lambda^{\uparrow\uparrow}_k(d) = \lambda^{\uparrow\uparrow}_k(d-1) + 1$ and the inductive hypothesis, to get:

$$\begin{aligned}\sum_{k=1}^{d} \lambda^{\uparrow\uparrow}_k(d) \cdot |c_{k\to i}|^2 &= \Big(\sum_{k=1}^{d-1} \big(\lambda^{\uparrow}_k(d-1) + 1\big) \cdot |c^{\lambda}_{k\to i}|^2\Big) + \lambda^{\uparrow\uparrow}_d(d) \cdot |c^{\lambda}_{d\to i}|^2 \\ &= (\lambda_i + 1 - i) \cdot D^{\lambda}_{d-1\to i} + \Big(\sum_{k=1}^{d-1} |c^{\lambda}_{k\to i}|^2\Big) + (\lambda_d - d + 1) \cdot |c^{\lambda}_{d\to i}|^2. \end{aligned} \tag{80}$$

We now rewrite everything in terms of C's and D's using Theorem 13.4:

$$\begin{aligned}(80) &= (C_i + 1) \cdot D^{\lambda}_{d-1\to i} + D^{\lambda}_{d-1\to i} + (C_d + 1) \cdot (D^{\lambda}_{d\to i} - D^{\lambda}_{d-1\to i}) \\ &= (C_i - C_d + 1) \cdot D^{\lambda}_{d-1\to i} + (C_d + 1) \cdot D^{\lambda}_{d\to i}. \end{aligned} \tag{81}$$

Finally, from Theorem 13.3, we have

$$D^{\lambda}_{d-1\to i} = (C_i - C_d) \cdot |c^{\lambda}_{d\to i}|^2 = (C_i - C_d) \cdot (D^{\lambda}_{d\to i} - D^{\lambda}_{d-1\to i}) \quad \Longrightarrow \quad (C_i - C_d + 1) \cdot D^{\lambda}_{d-1\to i} = (C_i - C_d) \cdot D^{\lambda}_{d\to i},$$

and substituting this gives us

$$(81) = (C_i - C_d) \cdot D^{\lambda}_{d\to i} + (C_d + 1) \cdot D^{\lambda}_{d\to i} = (C_i + 1) \cdot D^{\lambda}_{d\to i}. \qquad \square$$

We are now ready to extend the above identity to all legal transformations f.

Lemma 13.9. *Let λ be a Young diagram, and f be a legal transformation of Young diagrams. Then*

$$\sum_{k=1}^{d} f(\lambda)_k \cdot |c^{\lambda}_{k\to i}|^2 = (\lambda_i + 1 - i) \cdot D^{\lambda}_{d\to i}. \tag{82}$$

Proof. For a generic legal transformation f, we use Theorem 13.5, which states that the one-step Clebsch-Gordan coefficients $c^\lambda_{k\to i}$ are constant (as functions of k) on indices in the same blocks of λ. Write $c^\lambda_{k\to i} = c^\lambda_{B\to i}$, for all $k \in B$, for B a block of λ. Then, for any estimator, we have

$$\sum_{k=1}^{d} f(\lambda)_k \cdot |c^\lambda_{k\to i}|^2 = \sum_{\text{blocks } B} |c^\lambda_{B\to i}|^2 \sum_{k\in B} f(\lambda)_k = \sum_{\text{blocks } B} |c^\lambda_{B\to i}|^2 \sum_{k\in B} \lambda^\uparrow_k,$$

where the second step is by the definition of a legal estimator, and recall that $\lambda^\uparrow$ refers to the box donation transformation from Theorem 1.1. Thus, the left-hand side of Equation (82) is the same for *any* legal transformation, and so

$$\sum_{k=1}^{d} f(\lambda)_k \cdot |c^\lambda_{k\to i}|^2 = \sum_{k=1}^{d} \lambda^{\uparrow\uparrow}_k \cdot |c^\lambda_{k\to i}|^2 = (\lambda_i + 1 - i) \cdot D^\lambda_{d\to i}. \qquad \square$$

14 The second moment

In this section, we formally prove Theorem 1.4 from the introduction, which obtains an expression for the second moment of the output $\widehat{\boldsymbol{\rho}}$ of the debiased Keyl's algorithm.

14.1 Strategy overview

Since our proof for the second moment is more involved than the first moment, we start with a high-level overview of the main ingredients.

A central part of the first moment proof was the matrix $M^{(1)}_{\text{avg}}$ and its relation to the expected output of the debiased Keyl's algorithm on n copies of ρ:

$$\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \operatorname{tr}_{[n]}(M^{(1)}_{\text{avg}} \cdot \rho^{\otimes n} \otimes I).$$

Using Schur-Weyl duality and the Clebsch-Gordan transform, we showed that $M^{(1)}_{\text{avg}}$ is diagonal in the Schur basis, and that it is equal to the Jucys-Murphy element $\frac{1}{n} \cdot X_{n+1}$, which implies that our estimator is unbiased. That is, $\mathbf{E}[\widehat{\boldsymbol{\rho}}] = \rho$.

For the second moment, we define the matrix $M^{(2)}_{\text{avg}}$, which captures the second moment of the debiased Keyl's estimator on n copies of the state ρ. We define, for an estimator that measures with POVM $\{M_i\}$ and outputs $\widehat{\rho}_i$ upon obtaining M_i:

$$M^{(2)}_{\text{avg}} \coloneqq \sum_i M_i \otimes \widehat{\rho}_i \otimes \widehat{\rho}_i.$$

Then

$$\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] = \operatorname{tr}_{[n]}(M^{(2)}_{\text{avg}} \cdot \rho^{\otimes n} \otimes I \otimes I).$$

We show in Section 14.4 the Main Lemma (Theorem 14.15) for the second moment, which says that the matrix $M^{(2)}_{\text{avg}}$ satisfies the equation

$$M^{(2)}_{\text{avg}} + M^{(2)}_{\text{corr}} \cdot \text{SWAP} = \frac{1}{n^2} \cdot X_{n+1}X_{n+2}. \tag{83}$$

Throughout this section we use SWAP for $\text{SWAP}_{n+1,n+2}$, and $M^{(2)}_{\text{corr}}$ is a "correction" term for which we have an exact formula. Later in this subsection, we use Equation (83) and the exact form of $M^{(2)}_{\text{corr}}$ to prove Theorem 1.4, which is restated below.

The remainder of this section is dedicated to proving the Main Lemma. In Section 14.2, we use the Clebsch-Gordan transform to deduce that $M^{(2)}_{\text{avg}}$ and $M^{(2)}_{\text{corr}}$ have the same block-diagonal structure in the Schur basis as SWAP, with submatrices of size 2×2 and 1×1 on their main diagonal. The entries of these submatrices are given by expressions involving sums of the two-step Clebsch-Gordan coefficients. The Jucys-Murphy element $X_{n+1}X_{n+2}$ is a diagonal matrix in the Schur basis whose entries are related to the

contents of boxes of standard Young tableaux. Therefore, to prove Equation (83), it suffices to compute the exact value of $M_{\text{avg}}^{(2)} + M_{\text{corr}}^{(2)} \cdot \text{SWAP}$ for each submatrix in the Schur basis, and show that it matches the values of $\frac{1}{n^2} \cdot X_{n+1}X_{n+2}$. This reduces to proving an expression for the diagonal terms in each submatrix, the Diagonal Expression Lemma (Theorem 14.17), and this we do in Section 14.5.

For this purpose, we develop several tools for the two-step Clebsch-Gordan coefficients in Section 14.3. The two most important ones are Theorem 14.10, which gives a closed-form formula for the partial sums of the two-step Clebsch-Gordan coefficients, and Theorem 14.14, which relates the two-step Clebsch-Gordan coefficients and the blocks of the underlying Young diagram. Both of these technical lemmas are extensions of Theorems 13.4 and 13.5 from the one-step to the two-step coefficients, respectively.

Theorem 14.1 (Theorem 1.4, restated)**.** *Let $\widehat{\boldsymbol{\rho}}$ be the output of the debiased Keyl's algorithm when run on $\rho^{\otimes n}$. Then*

$$\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] = \frac{n-1}{n} \cdot \rho \otimes \rho + \frac{1}{n} (\rho \otimes I + I \otimes \rho) \cdot \text{SWAP} + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \cdot \text{SWAP} - \text{Lower}_\rho.$$

Here, Lower_ρ *refers to the "lower order terms' and has the following characterization: it can be written as a positive linear combination of terms of the form $(P \otimes P) \cdot \text{SWAP}$, where each $P \in \mathbb{C}^{d \times d}$ is a Hermitian, positive semidefinite matrix.*

Proof. Recall from the proof of unbiasedness that we interpreted our tomography algorithm as measuring $\rho^{\otimes n}$ using the POVM $M = \{M_i\}$. Upon obtaining outcome M_i, the algorithm outputs $\widehat{\boldsymbol{\rho}}_i$. It holds that

$$\begin{aligned}
\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] &= \sum_i \widehat{\rho}_i \otimes \widehat{\rho}_i \cdot \operatorname{tr}(M_i \cdot \rho^{\otimes n}) \\
&= \sum_i \operatorname{tr}_{[n]} \left(M_i \otimes \widehat{\rho}_i \otimes \widehat{\rho}_i \cdot \rho^{\otimes n} \otimes I \otimes I \right) \\
&= \operatorname{tr}_{[n]} \left(\sum_i M_i \otimes \widehat{\rho}_i \otimes \widehat{\rho}_i \cdot \rho^{\otimes n} \otimes I \otimes I \right) \\
&= \operatorname{tr}_{[n]} \left(M_{\text{avg}}^{(2)} \cdot \rho^{\otimes n} \otimes I \otimes I \right),
\end{aligned}$$

where we define

$$M_{\text{avg}}^{(2)} \coloneqq \sum_i M_i \otimes \widehat{\rho}_i \otimes \widehat{\rho}_i.$$

For the debiased Keyl's algorithm, the POVM elements are labeled by $\lambda \vdash n$ and $U \in U(d)$. See Equation (22). Similar calculations as the first moment section provide the following expression for $M_{\text{avg}}^{(2)}$:

$$\begin{aligned}
M_{\text{avg}}^{(2)} &= \sum_\lambda \int_U M_{\lambda,U} \otimes U\widehat{\alpha}U^\dagger \otimes U\widehat{\alpha}U^\dagger \\
&= \sum_\lambda \dim(V_\lambda^d) \int_U \mathcal{U}_{\text{SW}}^{(n)\dagger} \cdot \left(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \nu_\lambda(U) \left|T^\lambda\right\rangle\left\langle T^\lambda\right| \nu_\lambda(U)^\dagger \right) \cdot \mathcal{U}_{\text{SW}}^{(n)} \otimes U\widehat{\alpha}U^\dagger \otimes U\widehat{\alpha}U^\dagger \cdot \mathrm{d}U \\
&= \sum_\lambda \dim(V_\lambda^d) \int_U U^{\otimes n} \cdot \mathcal{U}_{\text{SW}}^{(n)\dagger} \cdot \left(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \left|T^\lambda\right\rangle\left\langle T^\lambda\right| \right) \cdot \mathcal{U}_{\text{SW}}^{(n)} \cdot U^{\dagger,\otimes n} \otimes U\widehat{\alpha}U^\dagger \otimes U\widehat{\alpha}U^\dagger \cdot \mathrm{d}U \\
&= \sum_\lambda \dim(V_\lambda^d) \int_U U^{\otimes(n+2)} \cdot \mathcal{U}^\dagger \cdot \left(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \left|T^\lambda\right\rangle\left\langle T^\lambda\right| \otimes \widehat{\alpha} \otimes \widehat{\alpha} \right) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U \\
&= \sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell=1}^d \widehat{\alpha}_k \widehat{\alpha}_\ell \int_U U^{\otimes(n+2)} \cdot \mathcal{U}^\dagger \cdot \left(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \left|T^\lambda\right\rangle\left\langle T^\lambda\right| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \right) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U \\
&= \frac{1}{n^2} \sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell} \lambda_k^\uparrow \lambda_\ell^\uparrow \int_U U^{\otimes(n+2)} \cdot \mathcal{U}^\dagger \cdot \left(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \left|T^\lambda\right\rangle\left\langle T^\lambda\right| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \right) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U,
\end{aligned}$$

where here $\mathcal{U} \coloneqq \mathcal{U}_{\text{SW}}^{(n)} \otimes I \otimes I$. As mentioned above (Equation (83)), we show in Theorem 14.15 that

$$M_{\text{avg}}^{(2)} + M_{\text{corr}}^{(2)} \cdot \text{SWAP} = \frac{1}{n^2} \cdot X_{n+1}X_{n+2},$$

where $M_{\mathrm{corr}}^{(2)}$ is a "correction" term that is given by

$$\begin{aligned} M_{\mathrm{corr}}^{(2)} &\coloneqq \frac{1}{n}\sum_{\lambda}\dim(V_\lambda^d)\sum_{k,\ell}\widehat{\alpha}_{\max(k,\ell)}\int_U U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\left(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\left|T^\lambda\rangle\langle T^\lambda\right|\otimes|k\rangle\langle k|\otimes|\ell\rangle\langle\ell|\right)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\mathrm{d}U \\ &= \frac{1}{n^2}\sum_{\lambda}\dim(V_\lambda^d)\sum_{k,\ell}\lambda^{\uparrow}_{\max(k,\ell)}\int_U U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\left(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\left|T^\lambda\rangle\langle T^\lambda\right|\otimes|k\rangle\langle k|\otimes|\ell\rangle\langle\ell|\right)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\mathrm{d}U. \end{aligned}$$

Note that this is identical to our expression for $M_{\mathrm{avg}}^{(2)}$ above except with the $\lambda_k^{\uparrow}\lambda_\ell^{\uparrow}$ part replaced by a $\lambda^{\uparrow}_{\max(k,\ell)}$.
From Theorem 14.2 below, we can write $M_{\mathrm{corr}}^{(2)}$ in the following way, with $\{m_j^\lambda\}_{j\in[d]}$ a set of nonnegative integers, and $\Pi_{\le j}\coloneqq\sum_{i\le j}|i\rangle\langle i|$ the projector onto the first j computational basis vectors:

$$\begin{aligned} M_{\mathrm{corr}}^{(2)} &= \frac{1}{n^2}\sum_{\lambda}\dim(V_\lambda^d)\sum_{j=1}^{d}m_j^\lambda\int_U U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\left(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\left|T^\lambda\rangle\langle T^\lambda\right|\otimes\Pi_{\le j}\otimes\Pi_{\le j}\right)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\mathrm{d}U \\ &\quad-\frac{1}{n^2}\sum_{\lambda}\dim(V_\lambda^d)\cdot\ell(\lambda)\int_U U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\left(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\left|T^\lambda\rangle\langle T^\lambda\right|\otimes I\otimes I\right)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\mathrm{d}U. \end{aligned}$$

We will call the first term above the negative correction $M_{\mathrm{corr},-}^{(2)}$ and the second term as the positive correction $M_{\mathrm{corr},+}^{(2)}$, from the sign of the term they contribute to $\mathbf{E}[\widehat{\boldsymbol{\rho}}\otimes\widehat{\boldsymbol{\rho}}]$. Therefore

$$\begin{aligned} \mathbf{E}[\widehat{\boldsymbol{\rho}}\otimes\widehat{\boldsymbol{\rho}}] &= \mathrm{tr}_{[n]}(M_{\mathrm{avg}}^{(2)}\cdot\rho^{\otimes n}\otimes I\otimes I) \\ &= \frac{1}{n^2}\,\mathrm{tr}_{[n]}(X_{n+1}X_{n+2}\cdot\rho^{\otimes n}\otimes I\otimes I)+\mathrm{tr}_{[n]}(M_{\mathrm{corr},+}^{(2)}\cdot\mathrm{SWAP}\cdot\rho^{\otimes n}\otimes I\otimes I) \\ &\quad-\mathrm{tr}_{[n]}(M_{\mathrm{corr},-}^{(2)}\cdot\mathrm{SWAP}\cdot\rho^{\otimes n}\otimes I\otimes I). \end{aligned}$$

Let us compute each term separately. For the first term, we expand the product of Jucys-Murphy elements as

$$\begin{aligned} X_{n+1}X_{n+2} &= \left(\sum_{i=1}^{n}(i,n+1)\right)\cdot\left(\sum_{j=1}^{n+1}(j,n+2)\right) \\ &= \sum_{i\neq j}(i,n+1)(j,n+2)+\sum_{i=1}^{n}(i,n+1,n+2)+\sum_{i=1}^{n}(i,n+2,n+1). \end{aligned}$$

Thus

$$\begin{aligned} &\mathrm{tr}_{[n]}\left(X_{n+1}X_{n+2}\cdot\rho^{\otimes n}\otimes I\otimes I\right) \\ &= \sum_{i\neq j}\mathrm{tr}_{[n]}\left((i,n+1)(j,n+2)\cdot\rho^{\otimes n}\otimes I\otimes I\right)+\sum_{i=1}^{n}\mathrm{tr}_{[n]}\left((i,n+1,n+2)\cdot\rho^{\otimes n}\otimes I\otimes I\right) \\ &\qquad+\sum_{i=1}^{n}\mathrm{tr}_{[n]}\left((i,n+2,n+1)\cdot\rho^{\otimes n}\otimes I\otimes I\right) \\ &= n(n-1)\cdot(\rho\otimes\rho)+n\cdot(\rho\otimes I+I\otimes\rho)\cdot\mathrm{SWAP}. \end{aligned}$$

Now we compute the second term. First, we simplify $M^{(2)}_{\mathrm{corr},+}$ as follows

$$\begin{aligned}
M^{(2)}_{\mathrm{corr},+} &= \frac{1}{n^2}\sum_{\lambda}\dim(V_\lambda^d)\cdot\ell(\lambda)\int_U U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\big|T^\lambda\big\rangle\big\langle T^\lambda\big|\otimes I\otimes I\Big)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\mathrm{d}U\\
&= \frac{1}{n^2}\sum_{\lambda}\dim(V_\lambda^d)\cdot\ell(\lambda)\cdot\Big(\int_U U^{\otimes n}\cdot\mathcal{U}_{\mathrm{SW}}^{(n)\dagger}\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\big|T^\lambda\big\rangle\big\langle T^\lambda\big|\Big)U^{\dagger,\otimes n}\cdot\mathcal{U}_{\mathrm{SW}}^{(n)}\cdot\mathrm{d}U\Big)\otimes I\otimes I\\
&= \frac{1}{n^2}\sum_{\lambda}\dim(V_\lambda^d)\cdot\ell(\lambda)\cdot\Big(\mathcal{U}_{\mathrm{SW}}^{(n)\dagger}\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\int_U\nu_\lambda(U)\big|T^\lambda\big\rangle\big\langle T^\lambda\big|\nu_\lambda(U)^\dagger\cdot\mathrm{d}U\Big)\cdot\mathcal{U}_{\mathrm{SW}}^{(n)}\Big)\otimes I\otimes I\\
&= \frac{1}{n^2}\sum_{\lambda}\dim(V_\lambda^d)\cdot\ell(\lambda)\cdot\Big(\mathcal{U}_{\mathrm{SW}}^{(n)\dagger}\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\frac{I_{\dim(V_\lambda^d)}}{\dim(V_\lambda^d)}\Big)\cdot\mathcal{U}_{\mathrm{SW}}^{(n)}\Big)\otimes I\otimes I\quad\text{(Schur's Lemma)}\\
&= \frac{1}{n^2}\sum_{\lambda}\ell(\lambda)\cdot\Pi_\lambda\otimes I\otimes I.\qquad\text{(Equation (18))}
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{tr}_{[n]}\left(M^{(2)}_{\mathrm{corr},+}\cdot\mathrm{SWAP}\cdot\rho^{\otimes n}\otimes I\otimes I\right) &= \frac{1}{n^2}\sum_{\lambda}\ell(\lambda)\cdot\mathrm{tr}_{[n]}\left(\Pi_\lambda\otimes I\otimes I\cdot\mathrm{SWAP}\cdot\rho^{\otimes n}\otimes I\otimes I\right)\\
&= \frac{1}{n^2}\left(\sum_{\lambda}\ell(\lambda)\cdot\mathrm{tr}\left(\Pi_\lambda\cdot\rho^{\otimes n}\right)\right)\cdot\mathrm{SWAP}\\
&= \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2}\cdot\mathrm{SWAP}.
\end{aligned}$$

Finally, we prove that the last term corresponds to Lower_ρ. In particular, $M^{(2)}_{\mathrm{corr},-}$ can be written as a positive linear combination of the following terms:

$$\int_U U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\big|T^\lambda\big\rangle\big\langle T^\lambda\big|\otimes\Pi_{\le j}\otimes\Pi_{\le j}\Big)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\mathrm{d}U.$$

Hence $\mathrm{tr}_{[n]}\left(M^{(2)}_{\mathrm{corr},-}\cdot\mathrm{SWAP}\cdot\rho^{\otimes n}\otimes I\otimes I\right)$ can also be written as a positive linear combination of the terms:

$$\begin{aligned}
&\mathrm{tr}_{[n]}\left(\int_U U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\big|T^\lambda\big\rangle\big\langle T^\lambda\big|\otimes\Pi_{\le j}\otimes\Pi_{\le j}\Big)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\mathrm{SWAP}\cdot\rho^{\otimes n}\otimes I\otimes I\cdot\mathrm{d}U\right)\\
&= \int_U\mathrm{tr}_{[n]}\left(U^{\otimes(n+2)}\cdot\mathcal{U}^\dagger\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\big|T^\lambda\big\rangle\big\langle T^\lambda\big|\otimes\Pi_{\le j}\otimes\Pi_{\le j}\Big)\cdot\mathcal{U}\cdot U^{\dagger,\otimes(n+2)}\cdot\rho^{\otimes n}\otimes I\otimes I\right)\cdot\mathrm{SWAP}\cdot\mathrm{d}U\\
&= \int_U\mathrm{tr}\left(U^{\otimes n}\cdot\mathcal{U}_{\mathrm{SW}}^{(n)\dagger}\cdot\Big(|\lambda\rangle\langle\lambda|\otimes I_{\dim(\lambda)}\otimes\big|T^\lambda\big\rangle\big\langle T^\lambda\big|\Big)\cdot\mathcal{U}_{\mathrm{SW}}^{(n)\dagger}\cdot U^{\dagger,\otimes n}\cdot\rho^{\otimes n}\right)\cdot\left(U\Pi_{\le j}U^\dagger\otimes U\Pi_{\le j}U^\dagger\right)\cdot\mathrm{SWAP}\cdot\mathrm{d}U.
\end{aligned}$$

The trace inside the integrand is nonnegative since $\mathrm{tr}(AB)\ge 0$ if A and B are positive semidefinite. We conclude that each term itself can be further written as a positive linear combination of terms $(P\otimes P)\cdot\mathrm{SWAP}$, where each P is of the form $U\Pi_{\le j}U^\dagger$, and is thus a Hermitian, positive semidefinite matrix. □

Lemma 14.2. *Let $n\ge 1$ and $\lambda\vdash n$. It holds that*

$$\sum_{k,\ell}\lambda^{\uparrow}_{\max(k,\ell)}\cdot|k\rangle\langle k|\otimes|\ell\rangle\langle\ell| = \Big(\sum_{j=1}^{d}m_j^\lambda\cdot\Pi_{\le j}\otimes\Pi_{\le j}\Big)-\ell(\lambda)\cdot I\otimes I,$$

where $\Pi_{\le j}\coloneqq\sum_{i\le j}|i\rangle\langle i|$ is the projector onto the first j computational basis vectors, and $\{m_j^\lambda\}_{j\in[d]}$ is a set of nonnegative integers defined as follows. Regard λ as a tuple of length d, and let $B_a=\{i_a,\dots,j_a\}$ be the a-th block of λ, for $a\in[A]$, where A is the total number of blocks. The coefficients $\{m_j^\lambda\}_{j\in[d]}$ are given by:

$$m_j^\lambda=\begin{cases}\lambda_{B_a}-\lambda_{B_{a+1}}+|B_a|+|B_{a+1}|, & j=j_a\text{ for some }a\in[A-1],\\ \lambda_{B_A}+|B_A|, & j=j_A\text{ and }\lambda_A>0\\ 0, & \text{otherwise.}\end{cases}$$

Proof. We have

$$\sum_{j=1}^{d} m_j^\lambda \cdot \Pi_{\leq j} \otimes \Pi_{\leq j} = \sum_{j=1}^{d} m_j^\lambda \cdot \Big(\sum_{k=1}^{j} |k\rangle\langle k|\Big) \otimes \Big(\sum_{\ell=1}^{j} |\ell\rangle\langle \ell|\Big) = \sum_{k,\ell=1}^{d} \Big(\sum_{j=\max(k,\ell)}^{d} m_j^\lambda\Big) \cdot |k\rangle\langle k| \otimes |\ell\rangle\langle \ell|\,.$$

First, we consider the case where $\lambda_A = 0$. In this case, $\ell(\lambda) = \sum_{a'=1}^{A-1} |B_{a'}|$. As a subcase, suppose $\max(k,\ell) \in B_a$, with $a < A$, then

$$\begin{aligned}
\sum_{j=\max(k,\ell)}^{d} m_j^\lambda &= \lambda_{B_a} + |B_a| + 2\Big(\sum_{a'=a+1}^{A-1} |B_{a'}|\Big) + |B_A| \\
&= \lambda_{B_a} + |B_a| + 2\Big(\sum_{a'=a+1}^{A-1} |B_{a'}|\Big) + |B_A| + \Big(\ell(\lambda) - \sum_{a'=1}^{A-1} |B_{a'}|\Big) \\
&= \Big(\lambda_{B_a} - \sum_{a'<a} |B_{a'}| + \sum_{a'>a} |B_{a'}|\Big) + \ell(\lambda) \\
&= \lambda^{\uparrow}_{B_a} + \ell(\lambda).
\end{aligned}$$

In the other subcase, with $\max(k,\ell) \in B_A$, we have $\lambda_{B_A} = 0$ and

$$\sum_{j=\max(k,\ell)}^{d} m_j^\lambda = 0 = \lambda_{B_A} + \Big(\ell(\lambda) - \sum_{a'=1}^{A-1} |B_{a'}|\Big) = \Big(\lambda_{B_A} - \sum_{a'=1}^{A-1} |B_{a'}|\Big) + \ell(\lambda) = \lambda^{\uparrow}_{B_A} + \ell(\lambda).$$

Thus, when $\lambda_A = 0$, we have

$$\begin{aligned}
\Big(\sum_{j=1}^{d} m_j^\lambda \cdot \Pi_{\leq j} \otimes \Pi_{\leq j}\Big) - \ell(\lambda) \cdot I \otimes I &= \sum_{k,\ell} \Big(\sum_{j=\max(k,\ell)}^{d} m_j^\lambda - \ell(\lambda)\Big) \cdot |k\rangle\langle k| \otimes |\ell\rangle\langle \ell| \\
&= \sum_{k,\ell} \lambda^{\uparrow}_{\max(k,\ell)} \cdot |k\rangle\langle k| \otimes |\ell\rangle\langle \ell|\,.
\end{aligned}$$

The second case, when $\lambda_A > 0$, is similar, but now $\ell(\lambda) = \sum_{a'=1}^{A} |B_{a'}|$. If we add a formal $(A+1)$-th block with $|B_{A+1}| = 0$ and $\lambda_{A+1} = 0$, then this case reduces to the first subcase of the $\lambda_A = 0$ case above. □

14.2 Block-diagonal expressions in the Schur basis

Having defined the expressions $M^{(2)}_{\text{avg}}$ and $M^{(2)}_{\text{corr}}$, in this section, we show that they can be written in the Schur basis as block-diagonal matrices, using the Clebsch-Gordan transform.

Notation 14.3. For convenience, we will adopt the following notation and terminology.

- We will abbreviate $\lambda + e_i + e_j$ as λ_{ij}, the Young diagram obtained when we add boxes to λ in rows i and j. Note that $\lambda_{ij} = \lambda_{ji}$.
- If $\lambda + e_i$ and $\lambda + e_i + e_j$ are valid Young diagrams, and S is an SYT of shape λ, we let S_{ij} denote the SYT obtained by starting with S, and adding $n+1$ to the end of the i-th row, and $n+2$ to the end of the j-th row. Note that $S_{ij} \neq S_{ji}$, unless $i = j$. It can also be the case that S_{ij} is valid, whereas S_{ji} is not, or vice versa. For example, if $\lambda = (1,1)$, then we may insert into the first row, and then the second, but not in the reverse order.
- For S an SYT of shape λ, we can use the pair $(S, \{i,j\})$ to index the subspace of $\text{Sp}_{\lambda_{ij}}$ spanned by the valid SYTs among $|S_{ij}\rangle$ and $|S_{ji}\rangle$.
- We refer to a block with both S_{ij} and S_{ji} valid as *swappable*, since the new boxes in rows i and j may be inserted in either order. Otherwise, we call a block *non-swappable*. There are two non-swappable subcases: if $j = i$, then we refer to this as the *horizontal* case; and if $j = i+1$, we refer to this as the *vertical* case. See Figure 22 for examples.

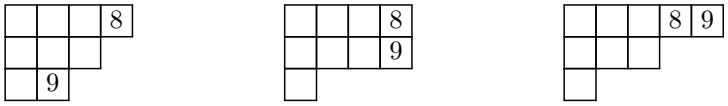


Figure 22: In the above examples, we have an SYT of shape $\lambda = (3,3,1)$ (not all of its values are visible because they are irrelevant), and we add two new boxes in rows i and j with values 8 and 9 respectively. The three resulting SYTs correspond to the cases when i, j are swappable, vertical, and horizontal, respectively.

Lemma 14.4. *Any expression M of the form*

$$M = \sum_{\lambda} \dim(V_\lambda^d) \int_U U^{\otimes(n+2)} \cdot \mathcal{U}^\dagger \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes |T^\lambda\rangle\langle T^\lambda| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \Big) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U,$$

can be written in the Schur basis as a block-diagonal matrix:

$$\mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot M \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} = \sum_{\lambda \vdash n} \sum_{i \le j} |\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes \sum_{S \in \lambda} |S, \{i,j\}\rangle\langle S, \{i,j\}| \otimes M_{\{i,j\}}^S \otimes I_{\dim(V_{\lambda_{ij}}^d)}.$$

Here, the vectors $|S, \{i,j\}\rangle$ index into the blocks (swappable, horizontal and vertical), and each $M_{\{i,j\}}^S$ is a 2×2 matrix if S_{ij} and S_{ji} are valid and distinct SYTs supported on the block $\mathrm{span}(|S_{ij}\rangle, |S_{ji}\rangle)$ (i.e. the swappable case), and is a 1×1 matrix otherwise supported on the block $\mathrm{span}(|S_{ij}\rangle)$ (i.e. the non-swappable case). The diagonal entries of $M_{\{i,j\}}^S$ are equal to

$$M_{ij,ij}^S \coloneqq \langle S_{ij}| M_{\{i,j\}}^S |S_{ij}\rangle = \frac{\dim(V_\lambda^d)}{\dim(V_{\lambda_{ij}}^d)} \cdot \big(|a_{k\ell \to ij}^\lambda|^2 + |b_{k\ell \to ij}^\lambda|^2 \big).$$

Whenever $M_{\{i,j\}}^S$ is a 2×2 matrix, its off-diagonal entries are equal to

$$M_{ij,ji}^S \coloneqq \langle S_{ij}| M_{\{i,j\}}^S |S_{ji}\rangle = \frac{\dim(V_\lambda^d)}{\dim(V_{\lambda_{ij}}^d)} \cdot \big(a_{k\ell \to ij}^\lambda b_{k\ell \to ji}^\lambda + a_{k\ell \to ji}^\lambda b_{k\ell \to ij}^\lambda + |a_{kk \to ij}^\lambda|^2 \cdot \delta_{k\ell} \big).$$

Proof. We will prove the first part of the lemma by applying the Clebsch-Gordan transform to M. For brevity, we will use $\lambda_{ij} = \lambda + e_i + e_j$ to denote the Young diagram when we add boxes in rows i and j to λ, and analogously, S_{ij} (though note $S_{ij} \neq S_{ji}$). The Clebsch-Gordan transform gives

$$\begin{aligned}
&|\lambda\rangle \otimes |S\rangle \otimes |T^\lambda\rangle \otimes |k\rangle \otimes |\ell\rangle \\
&= \Big(\mathcal{U}_{\mathrm{R}}^{(n+2)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n+2)} \cdot \mathcal{U}_{\mathrm{R}}^{(n+1)} \otimes I \cdot \mathcal{U}_{\mathrm{CG}}^{(n+1)} \otimes I \Big)^\dagger \cdot \sum_{i,j} |\lambda_{ij}\rangle \otimes |S_{ij}\rangle \otimes \Big(a_{k\ell \to ij}^\lambda |T_{k\ell \to ij}^\lambda\rangle + b_{k\ell \to ij}^\lambda |T_{k\ell \to ji}^\lambda\rangle \Big) \\
&= \Big(\mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot \mathcal{U}^\dagger \Big)^\dagger \cdot \sum_{i,j} |\lambda_{ij}\rangle \otimes |S_{ij}\rangle \otimes \Big(a_{k\ell \to ij}^\lambda |T_{k\ell \to ij}^\lambda\rangle + b_{k\ell \to ij}^\lambda |T_{k\ell \to ji}^\lambda\rangle \Big).
\end{aligned}$$

Thus

$$\begin{aligned}
&|\lambda\rangle\langle\lambda| \otimes |S\rangle\langle S| \otimes |T^\lambda\rangle\langle T^\lambda| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \\
&\cong \sum_{i,j} \sum_{i',j'} |\lambda_{ij}\rangle\langle\lambda_{i'j'}| \otimes |S_{ij}\rangle\langle S_{i'j'}| \otimes \Big(a_{k\ell \to ij}^\lambda |T_{k\ell \to ij}^\lambda\rangle + b_{k\ell \to ij}^\lambda |T_{k\ell \to ji}^\lambda\rangle \Big) \Big(a_{k\ell \to i'j'}^\lambda \langle T_{k\ell \to i'j'}^\lambda| + b_{k\ell \to i'j'}^\lambda \langle T_{k\ell \to j'i'}^\lambda| \Big),
\end{aligned} \tag{84}$$

where $\cong$ indicates equality up to conjugation by $\mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot \mathcal{U}^\dagger$. When we compute the integral of Equation (84), due to Schur-Weyl duality (Theorem 2.47), we will get a linear combination of expressions of the form:

$$\begin{aligned}
&\int_U U^{\otimes(n+2)} \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} \cdot \Big(|\lambda_{ij}\rangle\langle\lambda_{i'j'}| \otimes |S_{ij}\rangle\langle S_{i'j'}| \otimes |T_{k\ell \to pq}^\lambda\rangle\langle T_{k\ell \to p'q'}^\lambda| \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} \cdot \Big(|\lambda_{ij}\rangle\langle\lambda_{i'j'}| \otimes |S_{ij}\rangle\langle S_{i'j'}| \otimes \int_U \nu_{\lambda_{pq}}(U) |T_{k\ell \to pq}^\lambda\rangle\langle T_{k\ell \to p'q'}^\lambda| \nu_{\lambda_{p'q'}}(U)^\dagger \cdot \mathrm{d}U \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)},
\end{aligned}$$

where p, q and p', q' are such that $\{p, q\} = \{i, j\}$ and $\{p', q'\} = \{i', j'\}$. The above integral defines an intertwining operator between the irreps $\nu_{\lambda_{pq}}$ and $\nu_{\lambda_{p'q'}}$, and so can be computed using Schur's lemma. In particular, if the Young diagrams λ_{pq} and $\lambda_{p'q'}$ are different, then the integral is zero. This happens if and only if $\{p, q\} \neq \{p', q'\}$. Otherwise, when $\{p, q\} = \{p', q'\}$ (so that $\{i, j\} = \{p, q\} = \{p', q'\} = \{i', j'\}$) the integral evaluates to a multiple of the identity $c \cdot I_{\dim(V^d_{\lambda_{ij}})}$, where the constant c is given by Theorem 2.16:

$$c = \frac{1}{\dim(V^d_{\lambda_{ij}})} \cdot \operatorname{tr}\Big(\int_U \nu_{\lambda_{ij}}(U) \, |T^\lambda_{k\ell\to pq}\rangle\langle T^\lambda_{k\ell\to p'q'}| \, \nu_{\lambda_{ij}}(U)^\dagger \cdot \mathrm{d}U\Big) = \frac{\langle T^\lambda_{k\ell\to p'q'} | T^\lambda_{k\ell\to pq}\rangle}{\dim(V^d_{\lambda_{ij}})}.$$

Thus, in the Schur basis, M is a linear combination of terms of the form

$$|\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes |S_{ij}\rangle\langle S_{i'j'}| \otimes I_{\dim(V^d_{\lambda_{ij}})}.$$

Since $\{i, j\} = \{i', j'\}$, we see that each standard Young tableau $|S_{ij}\rangle$ can only possibly be paired with itself and the tableau $|S_{ji}\rangle$ that we obtain if we swap i and j, provided this is a valid tableau. Thus, we obtain the block structure claimed in the first part of the lemma.

Let us now compute the diagonal entries of $M^S_{\{i,j\}}$. A term with $|S_{ij}\rangle\langle S_{ij}|$ in the second register arises from terms as in Equation (84) with $(i, j) = (i', j')$:

$$|\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes |S_{ij}\rangle\langle S_{ij}| \otimes \Big(a^\lambda_{k\ell\to ij} \, |T^\lambda_{k\ell\to ij}\rangle + b^\lambda_{k\ell\to ij} \, |T^\lambda_{k\ell\to ji}\rangle\Big)\Big(a^\lambda_{k\ell\to ij} \, \langle T^\lambda_{k\ell\to ij}| + b^\lambda_{k\ell\to ij} \, \langle T^\lambda_{k\ell\to ji}|\Big).$$

With such a term, after integrating over the unitary group and using Schur's lemma, we are left with terms of the form

$$\frac{1}{\dim(V^d_{\lambda_{ij}})} \cdot |\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes |S_{ij}\rangle\langle S_{ij}| \otimes \Big(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2\Big) \cdot I_{\dim(V^d_{\lambda_{ij}})}.$$

This is because $T^\lambda_{k\ell\to ij}$ and $T^\lambda_{k\ell\to ji}$ are distinct SSYTs when $k \neq \ell$, and when $k = \ell$, we have that $b^\lambda_{k\ell\to ij} = 0$.

We proceed to the off-diagonal entries of $M^S_{\{i,j\}}$, which occur only in the swappable case, when necessarily $i \neq j$. A term with $|S_{ji}\rangle\langle S_{ji}|$ in the second register arises from terms as in Equation (84) with $(i, j) = (j', i')$:

$$|\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes |S_{ij}\rangle\langle S_{ji}| \otimes \Big(a^\lambda_{k\ell\to ij} \, |T^\lambda_{k\ell\to ij}\rangle + b^\lambda_{k\ell\to ij} \, |T^\lambda_{k\ell\to ji}\rangle\Big)\Big(a^\lambda_{k\ell\to ji} \, \langle T^\lambda_{k\ell\to ji}| + b^\lambda_{k\ell\to ji} \, \langle T^\lambda_{k\ell\to ij}|\Big).$$

With such a term, after integrating over the unitary group and using Schur's lemma, we are left with

$$\frac{1}{\dim(V^d_{\lambda_{ij}})} \cdot |\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes |S_{ij}\rangle\langle S_{ij}| \otimes \big(a^\lambda_{k\ell\to ij} b^\lambda_{k\ell\to ji} + a^\lambda_{k\ell\to ji} b^\lambda_{k\ell\to ij} + |a^\lambda_{kk\to ij}|^2 \cdot \delta_{k\ell}\big) \cdot I_{\dim(V^d_{\lambda_{ij}})}.$$

This is because $T^\lambda_{k\ell\to ij}$ and $T^\lambda_{k\ell\to ji}$ are distinct SSYTs when $k \neq \ell$, but not when $k = \ell$. □

In addition to the block-diagonal structure of the expressions from Theorem 14.4, we show below that each 2×2 block matrix is a linear combination of the identity and SWAP when the expression is real and symmetric in k and ℓ.

Notation 14.5. Following Theorem 2.34, SWAP is block-diagonal in the Schur basis, and preserves horizontal, vertical, and swappable blocks. In particular, we can write:

$$\begin{aligned}
\mathrm{SWAP} &= \mathcal{P}((n+1, n+2)) \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} \cdot \Big(\sum_{\lambda\vdash n} \sum_{i\le j} |\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes \kappa_{\lambda_{ij}}(n+1, n+2) \otimes I_{\dim(V^d_{\lambda_{ij}})}\Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} \cdot \Big(\sum_{\lambda\vdash n} \sum_{i\le j} |\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes \sum_{S\in \mathrm{SYT}(\lambda)} |S, \{i, j\}\rangle\langle S, \{i, j\}| \otimes \mathrm{SWAP}^S_{\{i,j\}} \otimes I_{\dim(V^d_{\lambda_{ij}})}\Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger}.
\end{aligned}$$

Moreover, $\mathrm{SWAP}^S_{\{i,j\}}$ is 1 in the horizontal case, -1 in the vertical case, and in the swappable case,

$$\mathrm{SWAP}^S_{\{i,j\}} = \begin{pmatrix} \frac{1}{\Delta_{ji}} & \sqrt{1 - \frac{1}{\Delta_{ji}^2}} \\ \sqrt{1 - \frac{1}{\Delta_{ji}^2}} & -\frac{1}{\Delta_{ji}} \end{pmatrix}.$$

Here, $\Delta_{ji} \coloneqq \Delta_S(n+1) = \mathrm{cont}_S(n+2) - \mathrm{cont}_S(n+1)$, is the difference of contents between the two new cells in rows j and i. We will also use Δ_{ji} in the horizontal and vertical cases, where $\Delta_{ji} = +1$ and $\Delta_{ji} = -1$ respectively. Note that in the swappable case, $|\Delta_{ji}| \geq 2$.

Lemma 14.6. *Any expression M of the form*

$$M = \sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell=1}^{d} f(k,\ell) \int_U U^{\otimes(n+2)} \cdot \mathcal{U}^\dagger \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \big|T^\lambda\big\rangle\big\langle T^\lambda\big| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \Big) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U, \tag{85}$$

with f is a real symmetric function in k, ℓ, can be written in the Schur basis as a block-diagonal matrix

$$\mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot M \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} = \sum_{\lambda \vdash n} |\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes \sum_{S \in \lambda} \sum_{i \leq j} |S, \{i,j\}\rangle\langle S, \{i,j\}| \otimes M^S_{\{i,j\}} \otimes I_{\dim(V^d_{\lambda + e_i + e_j})},$$

where each $M^S_{\{i,j\}}$ of dimension 2 must have the form $x^S_{\{i,j\}} \cdot I + y^S_{\{i,j\}} \cdot \mathrm{SWAP}^S_{\{i,j\}}$, where $x^S_{\{i,j\}}$ and $y^S_{\{i,j\}}$ are scalars.

Proof. Observe that if $f(k,\ell) = f(\ell,k)$, then M commutes with SWAP, since M is equal to

$$\begin{aligned}
&\sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell} f(k,\ell) \int_U U^{\otimes(n+2)} \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \big|T^\lambda\big\rangle\big\langle T^\lambda\big| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \Big) U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U \\
&= \sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell} f(k,\ell) \int_U \Big(U^{\otimes n} \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \big|T^\lambda\big\rangle\big\langle T^\lambda\big| \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)} \cdot U^{\dagger,\otimes n} \Big) \otimes \Big(U |k\rangle\langle k| U^\dagger \otimes U |\ell\rangle\langle\ell| U^\dagger \Big) \cdot \mathrm{d}U.
\end{aligned}$$

Then $\mathrm{SWAP} \cdot M \cdot \mathrm{SWAP}$ becomes

$$\begin{aligned}
&\sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell} f(k,\ell) \int_U \Big(U^{\otimes n} \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \big|T^\lambda\big\rangle\big\langle T^\lambda\big| \Big) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)} \cdot U^{\dagger,\otimes n} \Big) \otimes \Big(U |\ell\rangle\langle\ell| U^\dagger \otimes U |k\rangle\langle k| U^\dagger \Big) \cdot \mathrm{d}U \\
&= \sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell} f(k,\ell) \int_U U^{\otimes(n+2)} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \big|T^\lambda\big\rangle\big\langle T^\lambda\big| \otimes |\ell\rangle\langle\ell| \otimes |k\rangle\langle k| \Big) \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U \\
&= \sum_\lambda \dim(V_\lambda^d) \sum_{k,\ell} f(\ell,k) \int_U U^{\otimes(n+2)} \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \big|T^\lambda\big\rangle\big\langle T^\lambda\big| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \Big) \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U,
\end{aligned}$$

which is equal to M if $f(k,\ell) = f(\ell,k)$. So, if f is a real symmetric function, then M and SWAP commute. In this case, M and SWAP are simultaneously diagonalizable as they are both Hermitian. By writing both of them in the Schur basis, where they have the same block diagonal structure, we conclude that their principal submatrices that correspond to each $(S, \{i,j\})$ block must also commute.

For a swappable block $(S, \{i,j\})$, the projectors on the two eigenspaces of $\mathrm{SWAP}^S_{\{i,j\}}$ are $\frac{1}{2}\big(I \pm \mathrm{SWAP}^S_{\{i,j\}}\big)$. Thus, we conclude that any $M^S_{\{i,j\}}$ can be written as a linear combination of $\frac{1}{2}\big(I \pm \mathrm{SWAP}^S_{\{i,j\}}\big)$, or equivalently, as a linear combination of I and $\mathrm{SWAP}^S_{\{i,j\}}$. □

14.3 Technical lemmas concerning two-step Clebsch-Gordan coefficients

In this subsection, we collect some identities about the two-step Clebsch-Gordan coefficients that we will use in our proof for the second moment of our estimator. Recall from Theorem 13.1 that we use $D^\lambda_{k \to i}$ to refer to the ratio of dimensions $\dim(V^k_{\lambda \leq k + e_i}) / \dim(V^k_{\lambda \leq k})$ when $i \leq k \leq d$, and $D^\lambda_{k \to i} = 0$ otherwise.

Notation 14.7. We will often consider expressions of the form $|a^\lambda_{k\ell \to ij}|^2 + |b^\lambda_{k\ell \to ij}|^2$. To simplify notation, we introduce the following shorthand:

$$|c^\lambda_{k\ell \to ij}|^2 \coloneqq |a^\lambda_{k\ell \to ij}|^2 + |b^\lambda_{k\ell \to ij}|^2.$$

Notation 14.8. We will use the following shorthand for the partial sum of two-step CG coefficients:

$$F(s,t) \coloneqq \sum_{k=1}^{s} \sum_{\ell=1}^{t} |c^\lambda_{k\ell \to ij}|^2.$$

Notation 14.9. We will also use the shorthand

$$D^{\lambda}_{s\to ij} \coloneqq D^{\lambda}_{s\to i} \cdot D^{\lambda+e_i}_{s\to j} = \frac{\dim\left(V^s_{\lambda^{\leq s}+e_i}\right)}{\dim\left(V^s_{\lambda^{\leq s}}\right)} \cdot \frac{\dim\left(V^s_{\lambda^{\leq s}+e_i+e_j}\right)}{\dim\left(V^s_{\lambda^{\leq s}+e_i}\right)} = \frac{\dim\left(V^s_{\lambda^{\leq s}+e_i+e_j}\right)}{\dim\left(V^s_{\lambda^{\leq s}}\right)}.$$

Note that $D^{\lambda}_{s\to ij} = D^{\lambda}_{s\to ji}$.

Lemma 14.10 (Partial sums of two-step Clebsch-Gordan coefficients)**.** *Let λ be a Young diagram, and $i, j, k, \ell \in [d]$. Then*

$$F(s,t) = \sum_{k=1}^{s}\sum_{\ell=1}^{t} |c^{\lambda}_{k\ell\to ij}|^2 = \begin{cases} D^{\lambda}_{s\to i} \cdot D^{\lambda+e_i}_{t\to j} & s \leq t, \\ D^{\lambda}_{t\to j} \cdot D^{\lambda+e_j}_{s\to i} + \frac{1}{\Delta^2_{ji}}\left(D^{\lambda}_{t\to i} \cdot D^{\lambda+e_i}_{s\to j} - D^{\lambda}_{t\to j} \cdot D^{\lambda+e_j}_{s\to i}\right) & s \geq t. \end{cases}$$

Proof. We prove the lemma in three steps.

Step 1: First, we prove the result in the case when $s \leq t$, using the dual Clebsch-Gordan transform. We observe that $|c^{\lambda}_{k\ell\to ij}|^2$ can be written as:

$$\begin{aligned} |c^{\lambda}_{k\ell\to ij}|^2 &= |a^{\lambda}_{k\ell\to ij}|^2 + |b^{\lambda}_{k\ell\to ij}|^2 \\ &= |\left\langle T^{\lambda}, k|T^{\lambda}_{k\to i}\right\rangle \cdot \left\langle T^{\lambda}_{k\to i}, \ell|T^{\lambda}_{k\ell\to ij}\right\rangle|^2 + |\left\langle T^{\lambda}, k|T^{\lambda}_{k\to i}\right\rangle \cdot \left\langle T^{\lambda}_{k\to i}, \ell|T^{\lambda}_{k\ell\to ji}\right\rangle|^2 \\ &= |\left\langle T^{\lambda}, k|T^{\lambda}_{k\to i}\right\rangle|^2 \cdot \left(|\left\langle T^{\lambda}_{k\to i}, \ell|T^{\lambda}_{k\ell\to ij}\right\rangle|^2 + |\left\langle T^{\lambda}_{k\to i}, \ell|T^{\lambda}_{k\ell\to ji}\right\rangle|^2\right) \\ &= |c^{\lambda}_{k\to i}|^2 \sum_{T'} |\left\langle T^{\lambda}_{k\to i}, \ell|T'\right\rangle|^2, \end{aligned}$$

where $T' \in \mathrm{SSYT}(\lambda_{ij}, d)$. The last equality follows because the only SSYTs T' for which $\langle T^{\lambda}_{k\to i}, \ell|T'\rangle$ is nonzero are $T^{\lambda}_{k\ell\to ij}$ and $T^{\lambda}_{k\ell\to ji}$. Now $F(s,t)$ can be written as

$$\sum_{k=1}^{s}\sum_{\ell=1}^{t} |c^{\lambda}_{k\ell\to ij}|^2 = \sum_{k=1}^{s} |c^{\lambda}_{k\to i}|^2 \sum_{\ell=1}^{t}\sum_{T'} |\left\langle T^{\lambda}_{k\to i}, \ell|T'\right\rangle|^2.$$

We proceed by showing that for $k \leq t$, the value of the inner summation does not depend on k, and has a quite simple expression:

$$\sum_{\ell=1}^{t}\sum_{T'} |\left\langle T^{\lambda}_{k\to i}, \ell|T'\right\rangle|^2 = D^{\lambda+e_i}_{t\to j}. \tag{86}$$

From the description of the Clebsch-Gordan insertion algorithm in Theorem 10.9, since $k, \ell \leq t$, no letter greater than t is bumped during the insertion of the letter ℓ in $T^{\lambda}_{k\to i}$ to obtain T'. Thus, the Clebsch-Gordan coefficient $\langle T^{\lambda}_{k\to i}, \ell|T'\rangle$ is a product of scalar factors that only depend on the restrictions of $T^{\lambda}_{k\to i}$ and T' to the alphabet $[t]$. We conclude that in the case of $k, \ell \leq t$, it holds that:

$$\left\langle T^{\lambda}_{k\to i}, \ell|T'\right\rangle = \left\langle (T^{\lambda}_{k\to i})^{[t]}, \ell|(T')^{[t]}\right\rangle.$$

This gives an alternative expression for the left-hand side of Equation (86):

$$\sum_{\ell=1}^{t}\sum_{T'} |\left\langle T^{\lambda}_{k\to i}, \ell|T'\right\rangle|^2 = \sum_{\ell=1}^{t}\sum_{T''} |\left\langle (T^{\lambda}_{k\to i})^{[t]}, \ell|T''\right\rangle|^2.$$

The variable T'' on the right-hand side is iterating over the SSYTs of shape $(\lambda_{ij})_{\leq t}$, and thus has height at most t. From Theorem 12.5, we can rewrite the above coefficient using the dual Clebsch-Gordan coefficients:

$$\sum_{\ell=1}^{t}\sum_{T''} |\left\langle (T^{\lambda}_{k\to i})^{[t]}, \ell|T''\right\rangle|^2 = \frac{\dim(V^t_{\lambda_{ij}})}{\dim(V^t_{\lambda+e_i})} \sum_{\ell=1}^{t}\sum_{T''} |\left\langle (T^{\lambda}_{k\to i})^{[t]}|T'', \bar{\ell}\right\rangle|^2.$$

We observe that since ℓ iterates over $[t]$, then $\overline{\ell}$ iterates over all basis elements of the space $V^t_{-\square}$. Hence

$$\begin{aligned}
\frac{\dim(V^t_{\lambda_{ij}})}{\dim(V^t_{\lambda+e_i})} \sum_{\ell=1}^{t} \sum_{T''} |\langle (T^\lambda_{k\to i})^{[t]} | T'', \overline{\ell}\rangle|^2 &= D^{\lambda+e_i}_{t\to j} \sum_{\ell=1}^{t} \sum_{T''} \langle (T^\lambda_{k\to i})^{[t]} | T'', \overline{\ell}\rangle \cdot \langle T'', \overline{\ell} | (T^\lambda_{k\to i})^{[t]}\rangle \quad \text{(Theorem 13.1)} \\
&= D^{\lambda+e_i}_{t\to j} \langle (T^\lambda_{k\to i})^{[t]} | \left(\sum_{\ell=1}^{t} \sum_{T''} |T'', \overline{\ell}\rangle\langle T'', \overline{\ell}| \right) | (T^\lambda_{k\to i})^{[t]}\rangle \\
&= D^{\lambda+e_i}_{t\to j} \cdot \langle (T^\lambda_{k\to i})^{[t]} | \left(I_{V^t_{\lambda_{ij}}} \otimes I_{V^t_{-\square}} \right) | (T^\lambda_{k\to i})^{[t]}\rangle \\
&= D^{\lambda+e_i}_{t\to j} \cdot \langle (T^\lambda_{k\to i})^{[t]} | \left(\bigoplus_{\mu=\lambda_{ij}-\square} I_{V^t_\mu} \right) | (T^\lambda_{k\to i})^{[t]}\rangle \quad \text{(Equation (70))} \\
&= D^{\lambda+e_i}_{t\to j}.
\end{aligned}$$

The last equation follows because the space $V^t_{\lambda+e_i}$ appears with unit multiplicity in the direct sum, and $|(T^\lambda_{k\to i})^{[t]}\rangle$ is a vector of unit norm in this space. We conclude that when $s \le t$:

$$\begin{aligned}
\sum_{k=1}^{s} \sum_{\ell=1}^{t} |c^\lambda_{k\ell\to ij}|^2 &= \sum_{k=1}^{s} |c^\lambda_{k\to i}|^2 \sum_{\ell=1}^{t} \sum_{T''} |\langle T^\lambda_{k\to i}, \ell | T''\rangle|^2 \\
&= \sum_{k=1}^{s} |c^\lambda_{k\to i}|^2 \cdot D^{\lambda+e_i}_{t\to j} \quad \text{(Equation (86))} \\
&= D^\lambda_{s\to i} \cdot D^{\lambda+e_i}_{t\to j}. \quad \text{(Theorem 13.4)}
\end{aligned}$$

Step 2: Next, we note that $F(s,t)$ is related to certain entries of a particular operator in the Schur basis. Specifically, consider the operator M given by

$$M_{st} \coloneqq \sum_\lambda \dim(V^d_\lambda) \sum_{k,\ell=1}^{d} f_{st}(k,\ell) \int_U U^{\otimes(n+2)} \cdot \mathcal{U}^\dagger \cdot \left(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes |T^\lambda\rangle\langle T^\lambda| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \right) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U,$$

for f_{st} given by

$$f_{st}(k,\ell) \coloneqq \begin{cases} 1 & k \le s, \text{ and } \ell \le t, \\ 0 & \text{otherwise.} \end{cases}$$

Previous results, Theorems 14.4 and 14.6, have characterized such operators in the Schur basis. Specifically, M_{st} is block-diagonal in the blocks of SWAP. Fix S and a pair of indices $\{i,j\}$ with $i \le j$. By Theorem 14.4 and linearity, the $i \le j$ entry is given by:

$$(M_{st})^S_{ij,ij} = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k=1}^{s} \sum_{\ell=1}^{t} \left(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2 \right) = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot F(s,t). \tag{87}$$

The other diagonal entry, if it exists, is given by swapping i and j in the above expression (noting that F itself generally also has implicit dependence on i and j). The off-diagonal entries, if they exist, are given by

$$(M_{st})^S_{ij,ji} = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k=1}^{s} \sum_{\ell=1}^{t} \left(a^\lambda_{k\ell\to ij} b^\lambda_{k\ell\to ji} + a^\lambda_{k\ell\to ji} b^\lambda_{k\ell\to ij} + |a^\lambda_{kk\to ij}|^2 \cdot \delta_{k\ell} \right). \tag{88}$$

We will now restrict to the case we already understand from the previous step: $s \le t$. We have two cases:

(i) In the non-swappable case, Equation (87) is the sole entry of the 1×1 matrix, and, by the previous step, is equal to

$$(M_{st})^S_{ij,ij} = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot D^\lambda_{s\to i} \cdot D^{\lambda+e_i}_{t\to j}.$$

(ii) In the swappable case, we have a 2×2 matrix. The diagonal entries are likewise given by

$$(M_{st})^S_{ij,ij} = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot D^\lambda_{s\to i} \cdot D^{\lambda+e_i}_{t\to j}, \qquad (M_{st})^S_{ji,ji} = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot D^\lambda_{s\to j} \cdot D^{\lambda+e_j}_{t\to i}. \tag{89}$$

The off-diagonal entries, we claim, are zero.

We show this first for the case when $s = t$. In this case, f_{st} is symmetric, and hence M_{st} commutes with SWAP. Thus $(M_{st})^S_{\{i,j\}} = x \cdot I + y \cdot \mathrm{SWAP}^S_{\{i,j\}}$, for some coefficients $x, y \in \mathbb{C}$. However, when $s = t$, we also have that the two diagonal entries given in Equation (89) are equal, and hence $y = 0$. Thus, the off-diagonal entries are zero in this case.

We now extend this to the case $s < t$. Abbreviate

$$g(k,\ell) \coloneqq a^\lambda_{k\ell\to ij} b^\lambda_{k\ell\to ji} + a^\lambda_{k\ell\to ji} b^\lambda_{k\ell\to ij} + |a^\lambda_{kk\to ij}|^2 \cdot \delta_{k\ell}.$$

Note that $g(k,\ell) = 0$ for $k < \ell$, because $\delta_{k\ell} = 0$, and because no bumping can occur when CG inserting ℓ after k, so that $b^\lambda_{k\ell\to ij} = b^\lambda_{k\ell\to ji} = 0$. From the $s = t$ case, we also know that, for all s,

$$\sum_{k=1}^{s}\sum_{\ell=1}^{s} g(k,\ell) = 0,$$

since this sum is proportional to the off-diagonal entries, by Equation (88). Thus, the off-diagonal entries are

$$(M_{st})^S_{ij,ji} = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k=1}^{s}\sum_{\ell=1}^{t} g(k,\ell) = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \Big(\sum_{k=1}^{s}\sum_{\ell=1}^{s} g(k,\ell) + \sum_{k=1}^{s}\sum_{\ell=s+1}^{t} g(k,\ell) \Big) = 0,$$

since the first sum vanishes, and each entry of the second sum vanishes. Thus, the off-diagonal entries are zero for $s < t$ too.

Thus, in the case where $s \le t$ and i and j are swappable, we have

$$(M_{st})^S_{\{i,j\}} = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \begin{pmatrix} D^\lambda_{s\to i} \cdot D^{\lambda+e_i}_{t\to j} & 0 \\ 0 & D^\lambda_{s\to j} \cdot D^{\lambda+e_j}_{t\to i} \end{pmatrix}. \tag{90}$$

Step 3: In the last step, we show the result for $s \ge t$. We begin by noting that since

$$\mathrm{SWAP} \cdot \Big(\sum_{k=1}^{t}\sum_{\ell=1}^{s} |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \Big) \cdot \mathrm{SWAP} = \sum_{k=1}^{t}\sum_{\ell=1}^{s} |\ell\rangle\langle\ell| \otimes |k\rangle\langle k| = \sum_{k=1}^{s}\sum_{\ell=1}^{t} |k\rangle\langle k| \otimes |\ell\rangle\langle\ell|,$$

we have $M_{st} = \mathrm{SWAP} \cdot M_{ts} \cdot \mathrm{SWAP}$. Thus, in the block determined by S and $\{i,j\}$, we have the identity:

$$(M_{st})^S_{\{i,j\}} = \mathrm{SWAP}^S_{\{i,j\}} \cdot (M_{ts})^S_{\{i,j\}} \cdot \mathrm{SWAP}^S_{\{i,j\}}. \tag{91}$$

Now, take $t \le s$. We consider two cases:

(i) In the non-swappable case, the matrices in Equation (91) are 1×1, and hence all commute. So,

$$(M_{st})^S_{\{i,j\}} = (M_{ts})^S_{\{i,j\}} \qquad \Longrightarrow \qquad \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot F(s,t) = \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot F(t,s). \qquad \text{(Equation (87))}$$

Therefore,

$$F(s,t) = F(t,s) = D^\lambda_{t\to i} \cdot D^{\lambda+e_i}_{s\to j} = D^\lambda_{t\to j} \cdot D^{\lambda+e_j}_{s\to i} + \frac{1}{\Delta^2_{ji}} \big(D^\lambda_{t\to i} \cdot D^{\lambda+e_i}_{s\to j} - D^\lambda_{t\to j} \cdot D^{\lambda+e_j}_{s\to i} \big),$$

where the second step holds since $t \le s$, and the third step holds since $\Delta^2_{ji} = 1$.

(ii) In the swappable case, we can explicitly compute the right-hand side of Equation (91) in the $s \geq t$ case using Equation (90). We get:

$$\begin{aligned}
&(M_{st})^S_{\{i,j\}} \\
&= \frac{\dim(V_\lambda^d)}{\dim(V_{\lambda_{ij}}^d)} \begin{pmatrix} \frac{1}{\Delta_{ji}} & \sqrt{1-\frac{1}{\Delta_{ji}^2}} \\ \sqrt{1-\frac{1}{\Delta_{ji}^2}} & -\frac{1}{\Delta_{ji}} \end{pmatrix} \cdot \begin{pmatrix} D_{t\to i}^\lambda \cdot D_{s\to j}^{\lambda+e_i} & 0 \\ 0 & D_{t\to j}^\lambda \cdot D_{s\to i}^{\lambda+e_j} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\Delta_{ji}} & \sqrt{1-\frac{1}{\Delta_{ji}^2}} \\ \sqrt{1-\frac{1}{\Delta_{ji}^2}} & -\frac{1}{\Delta_{ji}} \end{pmatrix} \\
&= \frac{\dim(V_\lambda^d)}{\dim(V_{\lambda_{ij}}^d)} \begin{pmatrix} D_{t\to j}^\lambda \cdot D_{s\to i}^{\lambda+e_j} + \frac{1}{\Delta_{ji}^2}\big(D_{t\to i}^\lambda \cdot D_{s\to j}^{\lambda+e_i} - D_{t\to j}^\lambda \cdot D_{s\to i}^{\lambda+e_j}\big) & * \\ * & * \end{pmatrix},
\end{aligned}$$

where the asterisks represent entries we will not need. Finally, comparing this computation with Equation (87) gives:

$$F(s,t) = D_{t\to j}^\lambda \cdot D_{s\to i}^{\lambda+e_j} + \frac{1}{\Delta_{ji}^2}\big(D_{t\to i}^\lambda \cdot D_{s\to j}^{\lambda+e_i} - D_{t\to j}^\lambda \cdot D_{s\to i}^{\lambda+e_j}\big),$$

as claimed. □

There are many useful corollaries of Theorem 14.10. We use the following corollaries in our proof of the Diagonal Expression Lemma, Theorem 14.17 below, which itself is important for proving the Main Lemma, Theorem 14.15.

Corollary 14.11. *$M_{\text{corr}}^{(2)}$ is diagonal in the Schur basis.*

Proof. Fix a 2×2 block indexed by S and $\{i,j\}$, with $i < j$. From Theorem 14.6, it suffices to show the two diagonal entries of $M_{\text{corr}}^{(2)}$ in this block are equal. We start by considering the following sum:

$$\begin{aligned}
\sum_{k=1}^d \sum_{\ell=1}^d \lambda^\uparrow_{\max(k,\ell)} \cdot |c_{k\ell\to ij}^\lambda|^2 &= \sum_{k=1}^d \lambda_k^\uparrow \cdot \Big(\sum_{k',\ell=1}^k |c_{k'\ell\to ij}^\lambda|^2 - \sum_{k',\ell=1}^{k-1} |c_{k'\ell\to ij}^\lambda|^2 \Big) \\
&= \sum_{k=1}^d \lambda_k^\uparrow \cdot \Big(D_{k\to i}^\lambda \cdot D_{k\to j}^{\lambda+e_i} - D_{k-1\to i}^\lambda \cdot D_{k-1\to j}^{\lambda+e_i} \Big) \qquad \text{(Theorem 14.10)} \\
&= \sum_{k=1}^d \lambda_k^\uparrow \cdot \Big(D_{k\to ij}^\lambda - D_{k-1\to ij}^\lambda \Big). \qquad \text{(Theorem 14.9)}
\end{aligned}$$

Note that the remaining sum is symmetric in exchanging $i \leftrightarrow j$, by Theorem 14.9. That is,

$$\begin{aligned}
\sum_{k=1}^d \sum_{\ell=1}^d \lambda^\uparrow_{\max(k,\ell)} \cdot |c_{k\ell\to ij}^\lambda|^2 &= \sum_{k=1}^d \lambda_k^\uparrow \cdot \Big(D_{k\to ij}^\lambda - D_{k-1\to ij}^\lambda \Big) \\
&= \sum_{k=1}^d \lambda_k^\uparrow \cdot \Big(D_{k\to ji}^\lambda - D_{k-1\to ji}^\lambda \Big) = \sum_{k=1}^d \sum_{\ell=1}^d \lambda^\uparrow_{\max(k,\ell)} \cdot |c_{k\ell\to ji}^\lambda|^2.
\end{aligned}$$

Finally, from the expression for the diagonal entries given in Theorem 14.4,

$$M_{ij,ij}^S = \frac{1}{n} \cdot \frac{\dim(V_\lambda^d)}{\dim(V_{\lambda_{ij}}^d)} \sum_{k,\ell=1}^d \lambda^\uparrow_{\max(k,\ell)} \cdot |c_{k\ell\to ij}^\lambda|^2 = \frac{1}{n} \cdot \frac{\dim(V_\lambda^d)}{\dim(V_{\lambda_{ji}}^d)} \sum_{k,\ell=1}^d \lambda^\uparrow_{\max(k,\ell)} \cdot |c_{k\ell\to ji}^\lambda|^2 = M_{ji,ji}^S. \qquad \square$$

Corollary 14.12. *We have the following equations:*

(i) $\sum_{k,\ell=1}^d \lambda_k^{\uparrow\uparrow} \cdot |c_{k\ell\to ij}^\lambda|^2 = (C_i^\lambda + 1) \cdot D_{d\to ij}^\lambda$.

(ii) $\sum_{k,\ell=1}^d \big(\lambda_\ell^{\uparrow\uparrow} + \frac{1}{\Delta_{ji}}\big) \cdot |c_{k\ell\to ij}^\lambda|^2 = (C_j^{\lambda+e_i} + 1) \cdot D_{d\to ij}^\lambda$.

Proof. Our general strategy for showing both equations will be to rewrite it in terms of partial sums of two-step CG coefficients, and then to use Theorem 14.10 and Theorem 13.8. We start with the first equation. We have:

$$\begin{aligned}
\sum_{k=1}^{d}\sum_{\ell=1}^{d}\lambda_k^{\uparrow\uparrow}\cdot|c^\lambda_{k\ell\to ij}|^2 &= \sum_{k=1}^{d}\lambda_k^{\uparrow\uparrow}\cdot\Big(\sum_{k'=1}^{k}\sum_{\ell=1}^{d}|c^\lambda_{k'\ell\to ij}|^2-\sum_{k'=1}^{k-1}\sum_{\ell=1}^{d}|c^\lambda_{k'\ell\to ij}|^2\Big)\\
&= \sum_{k=1}^{d}\lambda_k^{\uparrow\uparrow}\cdot\big(D^\lambda_{k\to i}\cdot D^{\lambda+e_i}_{d\to j}-D^\lambda_{k-1\to i}\cdot D^{\lambda+e_i}_{d\to j}\big) && \text{(Theorem 14.10)}\\
&= \Big(\sum_{k=1}^{d}\lambda_k^{\uparrow\uparrow}\cdot\big(D^\lambda_{k\to i}-D^\lambda_{k-1\to i}\big)\Big)\cdot D^{\lambda+e_i}_{d\to j}\\
&= \Big(\sum_{k=1}^{d}\lambda_k^{\uparrow\uparrow}\cdot|c^\lambda_{k\to i}|^2\Big)\cdot D^{\lambda+e_i}_{d\to j} && \text{(Theorem 13.3)}\\
&= (C^\lambda_i+1)\cdot D^\lambda_{d\to i}\cdot D^{\lambda+e_i}_{d\to j} && \text{(Theorem 13.8)}\\
&= (C^\lambda_i+1)\cdot D^\lambda_{d\to ij}. && \text{(Theorem 14.9)}
\end{aligned}$$

This proves the first equation.

The second is similar, but slightly more complicated, since when we appeal to Theorem 14.10, we will be using the more complicated case for when $s\geq t$. We have:

$$\begin{aligned}
&\sum_{k=1}^{d}\sum_{\ell=1}^{d}\lambda_\ell^{\uparrow\uparrow}\cdot|c^\lambda_{k\ell\to ij}|^2\\
&= \sum_{\ell=1}^{d}\lambda_\ell^{\uparrow\uparrow}\cdot\Big(\sum_{k=1}^{d}\sum_{\ell'=1}^{\ell}|c^\lambda_{k\ell'\to ij}|^2-\sum_{k=1}^{d}\sum_{\ell'=1}^{\ell-1}|c^\lambda_{k\ell'\to ij}|^2\Big)\\
&= \sum_{\ell=1}^{d}\lambda_\ell^{\uparrow\uparrow}\cdot\Big(\big(1-\frac{1}{\Delta_{ji}^2}\big)(D^\lambda_{\ell\to j}-D^\lambda_{\ell-1\to j})\cdot D^{\lambda+e_j}_{d\to i}+\frac{1}{\Delta_{ji}^2}\cdot(D^\lambda_{\ell\to i}-D^\lambda_{\ell-1\to i})\cdot D^{\lambda+e_i}_{d\to j}\Big) && \text{(Theorem 14.10)}\\
&= \sum_{\ell=1}^{d}\lambda_\ell^{\uparrow\uparrow}\cdot\Big(\big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot|c^\lambda_{\ell\to j}|^2\cdot D^{\lambda+e_j}_{d\to i}+\frac{1}{\Delta_{ji}^2}\cdot|c^\lambda_{\ell\to i}|^2\cdot D^{\lambda+e_i}_{d\to j}\Big) && \text{(Theorem 13.3)}\\
&= (C^\lambda_j+1)\cdot\big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot D^\lambda_{d\to j}\cdot D^{\lambda+e_j}_{d\to i}+(C^\lambda_i+1)\cdot\frac{1}{\Delta_{ji}^2}\cdot D^\lambda_{d\to i}\cdot D^{\lambda+e_i}_{d\to j} && \text{(Theorem 13.8)}\\
&= (C^\lambda_j+1)\cdot\big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot D^\lambda_{d\to ij}+(C^\lambda_i+1)\cdot\frac{1}{\Delta_{ji}^2}\cdot D^\lambda_{d\to ij}. && \text{(Theorem 14.9)}
\end{aligned}$$

We also have:

$$\sum_{k=1}^{d}\sum_{\ell=1}^{d}\frac{1}{\Delta_{ji}}\cdot|c^\lambda_{k\ell\to ij}|^2=\frac{1}{\Delta_{ji}}\cdot D^\lambda_{d\to ij}=\frac{(C^{\lambda+e_i}_j+1)-(C^\lambda_i+1)}{\Delta_{ji}^2}\cdot D^\lambda_{d\to ij}.$$

Combining these gives:

$$\sum_{k=1}^{d}\sum_{\ell=1}^{d}\big(\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\big)\cdot|c^\lambda_{k\ell\to ij}|^2=\Big(\big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot(C^\lambda_j+1)+\frac{1}{\Delta_{ji}^2}\cdot(C^{\lambda+e_i}_j+1)\Big)\cdot D^\lambda_{d\to ij}.$$

In the horizontal subcase, we have $\Delta_{ji}^2=1$. In the non-horizontal case, $C^\lambda_j=C^{\lambda+e_i}_j$. In both cases, the sum resolves to the second equation:

$$\sum_{k=1}^{d}\sum_{\ell=1}^{d}\big(\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\big)\cdot|c^\lambda_{k\ell\to ij}|^2=(C^{\lambda+e_i}_j+1)\cdot D^\lambda_{d\to ij}.$$

□

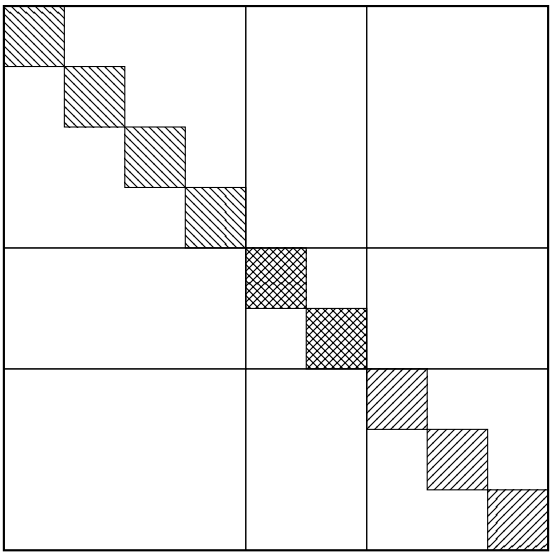


Figure 23: An example of the relation of the two-step Clebsch-Gordan coefficients $|c^{\lambda}_{k\ell\to ij}|^2$ with respect to the blocks of $\lambda = (3,3,3,2,2,1,1,1)$ for some fixed $i,j \in [d]$. The horizontal axis corresponds to the value of $k \in [9]$, and the vertical axis to the value of $\ell \in [9]$. The Young diagram has 3 blocks, consisting of the indices $\{1,2,3,4\}$, $\{5,6\}$, and $\{7,8,9\}$. The blocks define 9 regions in which the two-step coefficients are equal. The only exceptions are the diagonal entries $|c^{\lambda}_{kk\to ij}|^2$ in a block, which are equal to a distinct constant.

Lemma 14.13. *We have the following equation:*

$$\sum_{\substack{k,\ell \\ \max(k,\ell)=d}} (\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow} + \frac{1}{\Delta_{ji}}\lambda_d^{\uparrow\uparrow}) \cdot |c^{\lambda}_{k\ell\to ij}|^2 = (C_i^{\lambda}+1)(C_j^{\lambda+e_i}+1)\cdot D^{\lambda}_{d\to ij} - (C_i^{\lambda}+2)(C_j^{\lambda+e_i}+2)\cdot D^{\lambda}_{d-1\to ij}.$$

The proof of Theorem 14.13 is more involved, but requires no new ideas, and we defer it to Section 14.6.1. We also prove a two-step analog to Theorem 13.5, which shows how the two-step Clebsch-Gordan coefficients $|c^{\lambda}_{k\ell\to ij}|^2$, regarded as a function of k and ℓ, are related with respect to the blocks of λ.

Lemma 14.14. *Let λ be a Young diagram, and $i,j \in [d]$. The two-step Clebsch-Gordan coefficients $|c^{\lambda}_{k\ell\to ij}|^2$ satisfy:*

(i) Let B be a block of λ. Then for each pair of $k,\ell \in B$, with $k \neq \ell$,

$$|c^{\lambda}_{k\ell\to ij}|^2 = |c^{\lambda}_{BB\to ij}|^2.$$

In addition, when $k = \ell$,

$$|c^{\lambda}_{kk\to ij}|^2 = \big(1 + \frac{1}{\Delta_{ji}}\big)\cdot |c^{\lambda}_{BB\to ij}|^2.$$

(ii) Let B_1, B_2 be two distinct blocks of λ. Then the two-step Clebsch-Gordan coefficient $|c^{\lambda}_{k\ell\to ij}|^2$ is equal for all $k \in B_1$ and $\ell \in B_2$, that is:

$$|c^{\lambda}_{k\ell\to ij}|^2 = |c^{\lambda}_{B_1B_2\to ij}|^2.$$

The proof of this lemma is also deferred; see Section 14.6.2.

14.4 Main Lemma

Lemma 14.15 (Main Lemma)**.** *It holds that*

$$M^{(2)}_{\text{avg}} + M^{(2)}_{\text{corr}} \cdot \text{SWAP} = \frac{1}{n^2}\cdot X_{n+1}X_{n+2}, \tag{92}$$

where

$$M^{(2)}_{\mathrm{corr}} = \frac{1}{n}\sum_{\lambda} \dim(V^d_\lambda) \sum_{k,\ell} \lambda^{\uparrow}_{\max(k,\ell)} \int_U U^{\otimes(n+2)} \cdot \mathcal{U}^\dagger \cdot \Big(|\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)} \otimes \big|T^\lambda\big\rangle\big\langle T^\lambda\big| \otimes |k\rangle\langle k| \otimes |\ell\rangle\langle\ell| \Big) \cdot \mathcal{U} \cdot U^{\dagger,\otimes(n+2)} \cdot \mathrm{d}U.$$

Proof. Theorem 14.4 above shows that $M^{(2)}_{\mathrm{avg}}$ and $M^{(2)}_{\mathrm{corr}}$ have the same block structure in the Schur basis as SWAP. In particular, their blocks are indexed by an SYT S with n boxes, and a set of indices $\{i,j\} \in [d]^2$. Let S_{ij} be the SYT obtained from S by inserting a box containing $(n+1)$ into the i-th row, and then a box containing $(n+2)$ into the j-th row. However, we note that S_{ij} may not be a valid SYT. For example, if $\lambda = (1,1)$, and S is the SYT of shape λ, then S_{12} is valid, but S_{21} is not. Necessarily, however, at least one of S_{ij} and S_{ji} must be valid. Whenever both S_{ij} and S_{ji} are valid SYTs, the block corresponding to $(S,\{i,j\})$ is 2×2. Otherwise, if one of S_{ij} or S_{ji} is not a valid SYT, the corresponding block is 1×1, a scalar.

We know from Theorem 2.36 that $X_{n+1}X_{n+2}$ is a diagonal matrix in the Schur basis with the entry $(C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1)$ at $|S_{ij}\rangle\langle S_{ij}|$. Thus it suffices to prove Equation (92) for each block $(S,\{i,j\})$:

$$[M^{(2)}_{\mathrm{avg}}]^S_{\{i,j\}} + [M^{(2)}_{\mathrm{corr}} \cdot \mathrm{SWAP}]^S_{\{i,j\}} = \frac{1}{n^2} \cdot [X_{n+1}X_{n+2}]^S_{\{i,j\}}. \tag{93}$$

To show the above equality, we will use the following identity:

$$\sum_{k,\ell} (\lambda^\uparrow_k \lambda^\uparrow_\ell + 1/\Delta_{ji} \cdot \lambda^\uparrow_{\max(k,\ell)}) \cdot (|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2) = (C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1) \cdot \frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_\lambda)}, \tag{94}$$

which we prove in Theorem 14.17 below. For the rest of this proof, we fix any such block $(S,\{i,j\})$ and consider the case it corresponds to a 2×2 block, or a 1×1 block separately. Moreover, for notational simplicity, we will omit the superscript S and subscript $\{i,j\}$ for these cases.

1. If both S_{ij} and S_{ji} are distinct, valid SYTs, then $i \neq j$, and we are dealing with a 2×2 block. Theorem 14.6 below implies that on each block $(S,\{i,j\})$, our matrices have the following form:

$$M^{(2)}_{\mathrm{avg}} = x_{\mathrm{avg}} \cdot I + y_{\mathrm{avg}} \cdot \mathrm{SWAP},$$

$$M^{(2)}_{\mathrm{corr}} = x_{\mathrm{corr}} \cdot I + y_{\mathrm{corr}} \cdot \mathrm{SWAP}.$$

Hence

$$M^{(2)}_{\mathrm{avg}} + M^{(2)}_{\mathrm{corr}} \cdot \mathrm{SWAP} = \big(x_{\mathrm{avg}} + y_{\mathrm{corr}}\big) \cdot I + \big(y_{\mathrm{avg}} + x_{\mathrm{corr}}\big) \cdot \mathrm{SWAP}.$$

So we would like to show that

$$x_{\mathrm{avg}} + y_{\mathrm{corr}} = \frac{1}{n^2} \cdot (C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1), \tag{95}$$

$$y_{\mathrm{avg}} + x_{\mathrm{corr}} = 0. \tag{96}$$

From the structure of SWAP, we can compute x_{avg} and y_{avg} as follows:

$$x_{\mathrm{avg}} = \frac{1}{2}\big((M^{(2)}_{\mathrm{avg}})_{ij,ij} + (M^{(2)}_{\mathrm{avg}})_{ji,ji}\big), \qquad y_{\mathrm{avg}} = \frac{\Delta_{ji}}{2}\big((M^{(2)}_{\mathrm{avg}})_{ij,ij} - (M^{(2)}_{\mathrm{avg}})_{ji,ji}\big),$$

and similarly

$$x_{\mathrm{corr}} = \frac{1}{2}\big((M^{(2)}_{\mathrm{corr}})_{ij,ij} + (M^{(2)}_{\mathrm{corr}})_{ji,ji}\big), \qquad y_{\mathrm{corr}} = \frac{\Delta_{ji}}{2}\big((M^{(2)}_{\mathrm{corr}})_{ij,ij} - (M^{(2)}_{\mathrm{corr}})_{ji,ji}\big).$$

The left-hand side of Equation (95) becomes

$$
\begin{aligned}
&x_{\text{avg}} + y_{\text{corr}} \\
&= \frac{1}{2} \cdot (M^{(2)}_{\text{avg}})_{ij,ij} + \frac{1}{2} \cdot (M^{(2)}_{\text{avg}})_{ji,ji} + \frac{\Delta_{ji}}{2}\big((M^{(2)}_{\text{corr}})_{ij,ij} - (M^{(2)}_{\text{corr}})_{ji,ji}\big) \\
&= \frac{1}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k,\ell} \lambda^\uparrow_k \lambda^\uparrow_\ell \big(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2\big) \\
&\quad + \frac{1}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k,\ell} \lambda^\uparrow_k \lambda^\uparrow_\ell \big(|a^\lambda_{k\ell\to ji}|^2 + |b^\lambda_{k\ell\to ji}|^2\big) \\
&\quad + \frac{\Delta_{ji}}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k,\ell} \lambda^\uparrow_{\max(k,\ell)} \big(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2 - |a^\lambda_{k\ell\to ji}|^2 - |b^\lambda_{k\ell\to ji}|^2\big) \qquad \text{(Theorem 14.4)} \\
&= \frac{1}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k,\ell} \Big(\lambda^\uparrow_k \lambda^\uparrow_\ell + \frac{\lambda^\uparrow_{\max(k,\ell)}}{\Delta_{ji}}\Big) \big(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2\big) \\
&\quad + \frac{1}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k,\ell} \Big(\lambda^\uparrow_k \lambda^\uparrow_\ell + \frac{\lambda^\uparrow_{\max(k,\ell)}}{\Delta_{ij}}\Big) \big(|a^\lambda_{k\ell\to ji}|^2 + |b^\lambda_{k\ell\to ji}|^2\big) \\
&\quad + \frac{1}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \Big(\Delta_{ji} - \frac{1}{\Delta_{ji}}\Big) \sum_{k,\ell} \lambda^\uparrow_{\max(k,\ell)} \big(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2 - |a^\lambda_{k\ell\to ji}|^2 - |b^\lambda_{k\ell\to ji}|^2\big).
\end{aligned}
$$

From Equation (94), the first two summations are equal to

$$
\frac{1}{2n^2}(C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1) + \frac{1}{2n^2}(C^\lambda_j + 1)(C^{\lambda+e_j}_i + 1) = \frac{1}{n^2}(C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1),
$$

where we used that the fact that $C^{\lambda+e_i}_j = C^\lambda_j$ and $C^{\lambda+e_j}_i = C^\lambda_i$ when $i \neq j$. It remains to show that the last summation is zero. We will prove that for any $s \in [d]$,

$$
\sum_{\substack{k,\ell \\ \max(k,\ell)=s}} \big(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2 - |a^\lambda_{k\ell\to ij}|^2 - |b^\lambda_{k\ell\to ij}|^2\big) = 0.
$$

This follows from the partial sums of the two-step Clebsch-Gordan coefficients. In particular, Theorem 14.10 implies that for any $s \in [d]$:

$$
\sum_{k=1}^{s}\sum_{\ell=1}^{s}(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2) = D^\lambda_{s\to i} D^{\lambda+e_i}_{s\to j} = \frac{\dim(V^s_{\lambda_{\le s}+e_i+e_j})}{\dim(V^s_{\lambda_{\le s}})}.
$$

Exchanging i and j in the above equation does not change the right-hand side, and thus

$$
\begin{aligned}
&\sum_{k=1}^{s}\sum_{\ell=1}^{s}(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2) = \sum_{k=1}^{s}\sum_{\ell=1}^{s}(|a^\lambda_{k\ell\to ji}|^2 + |b^\lambda_{k\ell\to ji}|^2) \\
\implies &\sum_{k=1}^{s}\sum_{\ell=1}^{s}(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2 - |a^\lambda_{k\ell\to ji}|^2 - |b^\lambda_{k\ell\to ji}|^2) = 0. \qquad (97)
\end{aligned}
$$

Subtracting Equation (97) for $s-1$ from the expression above indeed gives:

$$
\sum_{\substack{k,\ell \\ \max(k,\ell)=s}} \big(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2 - |a^\lambda_{k\ell\to ij}|^2 - |b^\lambda_{k\ell\to ij}|^2\big) = 0.
$$

The left-hand side of Equation (96) becomes

$$\begin{aligned}
&y_{\text{avg}} + x_{\text{corr}} \\
&= \frac{1}{2} \cdot \left((M^{(2)}_{\text{corr}})_{ij,ij} + \Delta_{ji} \cdot (M^{(2)}_{\text{avg}})_{ij,ij}\right) + \frac{1}{2} \cdot \left((M^{(2)}_{\text{corr}})_{ji,ji} - \Delta_{ji} \cdot (M^{(2)}_{\text{avg}})_{ji,ji}\right) \\
&= \frac{1}{2} \cdot \left((M^{(2)}_{\text{corr}})_{ij,ij} + \Delta_{ji} \cdot (M^{(2)}_{\text{avg}})_{ij,ij}\right) + \frac{1}{2} \cdot \left((M^{(2)}_{\text{corr}})_{ji,ji} + \Delta_{ij} \cdot (M^{(2)}_{\text{avg}})_{ji,ji}\right) \\
&= \frac{1}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \sum_{k,\ell=1}^{d} (\lambda^{\uparrow}_{\max(k,\ell)} + \Delta_{ji} \cdot \lambda^{\uparrow}_k \lambda^{\uparrow}_\ell)(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2) \\
&\quad + \frac{1}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \sum_{k,\ell=1}^{d} (\lambda^{\uparrow}_{\max(k,\ell)} + \Delta_{ij} \cdot \lambda^{\uparrow}_k \lambda^{\uparrow}_\ell)(|a^\lambda_{k\ell\to ji}|^2 + |b^\lambda_{k\ell\to ji}|^2) \\
&= \frac{\Delta_{ji}}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \sum_{k,\ell=1}^{d} (\frac{1}{\Delta_{ji}} \cdot \lambda^{\uparrow}_{\max(k,\ell)} + \lambda^{\uparrow}_k \lambda^{\uparrow}_\ell)(|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2) \\
&\quad + \frac{\Delta_{ij}}{2n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \sum_{k,\ell=1}^{d} (\frac{1}{\Delta_{ij}} \cdot \lambda^{\uparrow}_{\max(k,\ell)} + \lambda^{\uparrow}_k \lambda^{\uparrow}_\ell)(|a^\lambda_{k\ell\to ji}|^2 + |b^\lambda_{k\ell\to ji}|^2) \\
&= \frac{\Delta_{ji}}{2n^2} \cdot (C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1) + \frac{\Delta_{ij}}{2n^2} \cdot (C^\lambda_j + 1)(C^{\lambda+e_j}_i + 1) = 0,
\end{aligned}$$

where we used the fact that $C^{\lambda+e_i}_j = C^\lambda_j$ and $C^{\lambda+e_j}_i = C^\lambda_i$ when $i \neq j$.

2. If S_{ij} is a valid SYT, but S_{ji} is either not valid, or not a distinct SYT, then we are dealing with a 1×1 block. First, we obtain the following expressions for the diagonal entries of $M^{(2)}_{\text{avg}}$ and $M^{(2)}_{\text{corr}}$ from Theorem 14.4:

$$M^{(2)}_{\text{avg}} = \frac{1}{n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \sum_{k,\ell} \lambda^{\uparrow}_k \lambda^{\uparrow}_\ell \cdot (|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2),$$

$$M^{(2)}_{\text{corr}} = \frac{1}{n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \sum_{k,\ell} \lambda^{\uparrow}_{\max(k,\ell)} \cdot (|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2).$$

The SWAP matrix has the scalar $\frac{1}{\Delta_{ji}}$ in the $(S, \{i,j\})$ block. Hence

$$\begin{aligned}
M^{(2)}_{\text{avg}} + M^{(2)}_{\text{corr}} \cdot \text{SWAP} &= \frac{1}{n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \cdot \sum_{k,\ell} (\lambda^{\uparrow}_k \lambda^{\uparrow}_\ell + \frac{1}{\Delta_{ji}} \cdot \lambda^{\uparrow}_{\max(k,\ell)}) \cdot (|a^\lambda_{k\ell\to ij}|^2 + |b^\lambda_{k\ell\to ij}|^2) \\
&= \frac{1}{n^2} \cdot (C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1) \\
&= \frac{1}{n^2} \cdot X_{n+1} X_{n+2}.
\end{aligned}$$

□

14.5 Proof of the Diagonal Expression Lemma

In this section, we prove the Diagonal Expression Lemma, which states that:

$$\sum_{k,\ell} (\lambda^{\uparrow}_k \lambda^{\uparrow}_\ell + 1/\Delta_{ji} \cdot \lambda^{\uparrow}_{\max(k,\ell)}) \cdot |c^\lambda_{k\ell\to ij}|^2 = (C^\lambda_i + 1)(C^{\lambda+e_i}_j + 1) \cdot \frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_\lambda)}.$$

This identity can be thought of as the second moment analog of the identity

$$\sum_{k,\ell} \lambda^{\uparrow}_k \cdot |c^\lambda_{k\to i}|^2 = (C^\lambda_i + 1) \cdot \frac{\dim(V^d_{\lambda+e_i})}{\dim(V^d_\lambda)}$$

from the first moment. Similar to the first moment, we will first show this identity for the staircase transformation $\lambda^{\uparrow\uparrow}$ in Equation (98), and then "average" the $\lambda^{\uparrow\uparrow}$'s over the blocks of λ to obtain the desired statement in Theorem 14.17. In contrast to the first moment, this identity does not hold for all legal estimators.

Lemma 14.16 (Diagonal Expression Lemma for the staircase estimator)**.** *Let λ be a Young diagram. Then the Diagonal Expression Lemma holds for the staircase estimator:*

$$\sum_{k=1}^{d}\sum_{\ell=1}^{d}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2=(C_i^\lambda+1)(C_j^{\lambda+e_i}+1)\cdot\frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_\lambda)}. \tag{98}$$

Proof. Fix λ, i and j. We prove the lemma by induction on d. Our base case is $d=1$. If either i or j are greater than 1, both sides of the desired equation vanish, since $|c^\lambda_{k\ell\to ij}|^2=0$ and $\dim(V^d_{\lambda_{ij}})/\dim(V^d_\lambda)=D^\lambda_{d\to ij}=0$ in cases where $d<i,j$. So, assume $i=j=1$. Note that in this case, $\Delta_{ji}=1$, $C_i^\lambda=\lambda_1-1$, and $C_j^{\lambda+e_i}=\lambda_1$. The only nonzero CG coefficient is $c^\lambda_{11\to 11}$, which must therefore have unit norm. Since $d=1$, we have $\dim(V^d_{\lambda_{ij}})=\dim(V^d_\lambda)=1$, as the only valid SSYT is filled with 1's. Lastly, we have $\lambda_1^{\uparrow\uparrow}=\lambda_1$. Thus,

$$\left((\lambda_1^{\uparrow\uparrow})^2+\lambda_1^{\uparrow\uparrow}\right)\cdot|c^\lambda_{11\to 11}|^2=\lambda_1^{\uparrow\uparrow}(\lambda_1^{\uparrow\uparrow}+1)=\lambda_1(\lambda_1+1)=(C_i^\lambda+1)(C_j^{\lambda+e_i}+1)=(C_i^\lambda+1)(C_j^{\lambda+e_i}+1)\cdot\frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_\lambda)},$$

which proves the base case.

Now assume the result holds for $d-1$. We split the left-hand side of Equation (98) up into the following terms:

$$\begin{aligned}&\sum_{k=1}^{d}\sum_{\ell=1}^{d}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2\\&=\sum_{k=1}^{d-1}\sum_{\ell=1}^{d-1}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2+\sum_{\substack{k,\ell\\\max(k,\ell)=d}}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2.\end{aligned} \tag{99}$$

We now consider the first term of Equation (99). Recall that $\lambda_k^{\uparrow\uparrow}(d)=\lambda_k^{\uparrow\uparrow}(d-1)+1$. We have

$$\begin{aligned}&\sum_{k,\ell=1}^{d-1}\left(\lambda_k^{\uparrow\uparrow}(d)\lambda_\ell^{\uparrow\uparrow}(d)+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}(d)\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2\\&=\sum_{k,\ell=1}^{d-1}\left((\lambda_k^{\uparrow\uparrow}(d-1)+1)(\lambda_\ell^{\uparrow\uparrow}(d-1)+1)+\frac{1}{\Delta_{ji}}\cdot(\lambda^{\uparrow\uparrow}_{\max(k,\ell)}(d-1)+1)\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2\\&=\sum_{k,\ell=1}^{d-1}\left(\lambda_k^{\uparrow\uparrow}(d-1)\lambda_\ell^{\uparrow\uparrow}(d-1)+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}(d-1)\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2+\sum_{k,\ell=1}^{d-1}\left(\lambda_k^{\uparrow\uparrow}(d-1)+\lambda_\ell^{\uparrow\uparrow}(d-1)+1+\frac{1}{\Delta_{ji}}\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2.\end{aligned}$$

We can use the inductive hypothesis on the first sum here, and Theorem 14.12 and Theorem 14.10 for the second. We get:

$$\begin{aligned}\sum_{k,\ell=1}^{d-1}\left(\lambda_k^{\uparrow\uparrow}(d)\lambda_\ell^{\uparrow\uparrow}(d)+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}(d)\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2&=(C_i^\lambda+1)(C_j^{\lambda+e_i}+1)\cdot D^\lambda_{d-1\to ij}+(C_i^\lambda+C_j^{\lambda+e_i}+3)\cdot D^\lambda_{d-1\to ij}\\&=(C_i^\lambda+2)(C_j^{\lambda+e_i}+2)\cdot D^\lambda_{d-1\to ij}.\end{aligned} \tag{100}$$

We now turn to the second term in Equation (99). This term was calculated in Theorem 14.13, and we have:

$$\sum_{\substack{k,\ell\\\max(k,\ell)=d}}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda^{\uparrow\uparrow}_{\max(k,\ell)}\right)\cdot|c^{\lambda}_{k\ell\to ij}|^2=(C_i^\lambda+1)(C_j^{\lambda+e_i}+1)\cdot D^\lambda_{d\to ij}-(C_i^\lambda+2)(C_j^{\lambda+e_i}+2)\cdot D^\lambda_{d-1\to ij}. \tag{101}$$

Finally, substituting Equations (100) and (101) back into Equation (99) we obtain the claimed expression:

$$\sum_{k=1}^{d}\sum_{\ell=1}^{d}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{k\ell\to ij}^{\lambda}|^2=(C_i^{\lambda}+1)(C_j^{\lambda+e_i}+1)\cdot D_{d\to ij}^{\lambda}. \qquad \square$$

We next show that the Diagonal Expression Lemma holds for the debiased Keyl's estimator as well. We remind the reader that our general approach is to "average" the analogous result for the staircase estimator.

Lemma 14.17 (Diagonal Expression Lemma for the debiased Keyl's estimator)**.** *Let λ be a Young diagram. Then the Diagonal Expression Lemma holds for the debiased Keyl's estimator:*

$$\sum_{k=1}^{d}\sum_{\ell=1}^{d}\left(\lambda_k^{\uparrow}\lambda_\ell^{\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow}\right)\cdot|c_{k\ell\to ij}^{\lambda}|^2=(C_i^{\lambda}+1)(C_j^{\lambda+e_i}+1)\cdot\frac{\dim(V_{\lambda_{ij}}^d)}{\dim(V_\lambda^d)}.$$

Proof. To prove the lemma, we will make use of the results in Theorem 14.14. Our starting point is the Diagonal Expression Lemma for the staircase estimator (Theorem 14.16), which we rewrite in terms of the blocks $\{B\}$ of λ:

$$\begin{aligned}
&\sum_{k=1}^{d}\sum_{\ell=1}^{d}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{k\ell\to ij}^{\lambda}|^2\\
&=\sum_{B\neq B'}\sum_{k\in B}\sum_{\ell\in B'}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{BB'\to ij}^{\lambda}|^2+\sum_{B}\sum_{\substack{k,\ell\in B\\k\neq\ell}}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{BB\to ij}^{\lambda}|^2\\
&\quad+\sum_{B}\sum_{k\in B}\left((\lambda_k^{\uparrow\uparrow})^2+\frac{1}{\Delta_{ji}}\cdot\lambda_k^{\uparrow\uparrow}\right)\cdot|c_{BB\to ij}^{\lambda}|^2\cdot\left(1+\frac{1}{\Delta_{ji}}\right).
\end{aligned}\tag{102}$$

We aim to show that we can replace $\lambda_k^{\uparrow\uparrow}$ to $\lambda_k^{\uparrow}$ without changing the value of the combined expression. We informally say that an expression where we can make such a replacement is *averageable*. We begin by showing the first sum by itself is averageable. Assume $B<B'$, meaning that all rows of B occur above all rows of B'. Then

$$\begin{aligned}
\sum_{k\in B}\sum_{\ell\in B'}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{BB'\to ij}^{\lambda}|^2&=\Big(\sum_{k\in B}\sum_{\ell\in B'}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_\ell^{\uparrow\uparrow}\right)\Big)\cdot|c_{BB'\to ij}^{\lambda}|^2\\
&=\Big(\sum_{k\in B}\left(\lambda_k^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\right)\Big)\cdot\Big(\sum_{\ell\in B'}\lambda_\ell^{\uparrow\uparrow}\Big)\cdot|c_{BB'\to ij}^{\lambda}|^2\\
&=\Big(\sum_{k\in B}\left(\lambda_k^{\uparrow}+\frac{1}{\Delta_{ji}}\right)\Big)\cdot\Big(\sum_{\ell\in B'}\lambda_\ell^{\uparrow}\Big)\cdot|c_{BB'\to ij}^{\lambda}|^2.
\end{aligned}$$

The last step holds since the average value on any block of a legal estimator is the same. Undoing the factorizations thus gives:

$$\sum_{k\in B}\sum_{\ell\in B'}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{BB'\to ij}^{\lambda}|^2=\sum_{k\in B}\sum_{\ell\in B'}\left(\lambda_k^{\uparrow}\lambda_\ell^{\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow}\right)\cdot|c_{BB'\to ij}^{\lambda}|^2.$$

The result also holds if $B>B'$, by swapping $k\leftrightarrow\ell$ in the above reasoning. So,

$$\sum_{B\neq B'}\sum_{k\in B}\sum_{\ell\in B'}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{BB'\to ij}^{\lambda}|^2=\sum_{B\neq B'}\sum_{k\in B}\sum_{\ell\in B'}\left(\lambda_k^{\uparrow}\lambda_\ell^{\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow}\right)\cdot|c_{BB'\to ij}^{\lambda}|^2.\tag{103}$$

We now rewrite the remaining two sums of Equation (102) in the following way:

$$\sum_{B}\sum_{k,\ell\in B}\left(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\cdot\lambda_{\max(k,\ell)}^{\uparrow\uparrow}\right)\cdot|c_{BB\to ij}^{\lambda}|^2+\sum_{B}\sum_{k\in B}\frac{1}{\Delta_{ji}}\cdot\left((\lambda_k^{\uparrow\uparrow})^2+\frac{1}{\Delta_{ji}}\cdot\lambda_k^{\uparrow\uparrow}\right)\cdot|c_{BB\to ij}^{\lambda}|^2.\tag{104}$$

The first part of the first sum in Equation (104) is individually averageable:

$$\sum_B \sum_{k,\ell\in B} \lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow} \cdot |c^\lambda_{BB\to ij}|^2 = \sum_B \Big(\sum_{k\in B} \lambda_k^{\uparrow\uparrow}\Big)\Big(\sum_{\ell\in B} \lambda_\ell^{\uparrow\uparrow}\Big) \cdot |c^\lambda_{BB\to ij}|^2$$
$$= \sum_B \Big(\sum_{k\in B} \lambda_k^{\uparrow}\Big)\Big(\sum_{\ell\in B} \lambda_\ell^{\uparrow}\Big) \cdot |c^\lambda_{BB\to ij}|^2 = \sum_B \sum_{k,\ell\in B} \lambda_k^{\uparrow}\lambda_\ell^{\uparrow} \cdot |c^\lambda_{BB\to ij}|^2.$$

Similarly the second part of the second sum:

$$\frac{1}{\Delta_{ji}^2} \sum_{k\in B} \lambda_k^{\uparrow\uparrow} \cdot |c^\lambda_{BB\to ij}|^2 = \frac{1}{\Delta_{ji}^2} \sum_{k\in B} \lambda_k^{\uparrow} \cdot |c^\lambda_{BB\to ij}|^2.$$

The remaining two pieces of Equation (104) combine to give an averageable expression. We have

$$\begin{aligned}
&\frac{1}{\Delta_{ji}} \sum_B \sum_{k,\ell\in B} \lambda^{\uparrow\uparrow}_{\max(k,\ell)} \cdot |c^\lambda_{BB\to ij}|^2 + \frac{1}{\Delta_{ji}} \sum_B \sum_{k\in B} (\lambda_k^{\uparrow\uparrow})^2 \cdot |c^\lambda_{BB\to ij}|^2 \\
&= \frac{1}{\Delta_{ji}} \sum_B \Big(\sum_{k\in B} \big(N_B(k)\cdot \lambda_k^{\uparrow\uparrow} + (\lambda_k^{\uparrow\uparrow})^2\big)\Big) \cdot |c^\lambda_{BB\to ij}|^2 \\
&= \frac{1}{\Delta_{ji}} \sum_B \Big(\sum_{k\in B} \lambda_k^{\uparrow\uparrow} \cdot \big(N_B(k) + \lambda_k^{\uparrow\uparrow}\big)\Big) \cdot |c^\lambda_{BB\to ij}|^2, \qquad (105)
\end{aligned}$$

where $N_B(k) = \big|\{(k_1,k_2)\in B\times B \,|\, \max(k_1,k_2)=k\}\big|$. That is, $N_B(k)$ is the number of times k appears in the maximum of a pair of indices in B. We have $N_B(k) = 2k - 1 - 2\sum_{B'<B} |B'|$, for all $k\in B$. Thus,

$$N_B(k) + \lambda_k^{\uparrow\uparrow} = \Big(2k - 1 - 2\sum_{B'<B} |B'|\Big) + \Big(\lambda_B + d + 1 - 2k\Big) = \lambda_B + d - 2\sum_{B'<B} |B'| =: N'_B,$$

where N'_B is independent of $k\in B$. Therefore, using the fact that $\lambda_k^{\uparrow} = \lambda_B^{\uparrow}$ for all $k\in B$, we have:

$$\begin{aligned}
(105) &= \frac{1}{\Delta_{ji}} \sum_B N'_B \cdot \Big(\sum_{k\in B} \lambda_k^{\uparrow\uparrow}\Big) \cdot |c^\lambda_{BB\to ij}|^2 \\
&= \frac{1}{\Delta_{ji}} \sum_B N'_B \cdot \Big(\sum_{k\in B} \lambda_k^{\uparrow}\Big) \cdot |c^\lambda_{BB\to ij}|^2 \\
&= \frac{1}{\Delta_{ji}} \sum_B \lambda_B^{\uparrow} \cdot \Big(\sum_{k\in B} N_B(k) + \lambda_k^{\uparrow\uparrow}\Big) \cdot |c^\lambda_{BB\to ij}|^2 \\
&= \frac{1}{\Delta_{ji}} \sum_B \lambda_B^{\uparrow} \cdot \Big(\sum_{k\in B} N_B(k) + \lambda_k^{\uparrow}\Big) \cdot |c^\lambda_{BB\to ij}|^2 \\
&= \frac{1}{\Delta_{ji}} \sum_B \Big(\sum_{k\in B} \big(N_B(k)\cdot \lambda_k^{\uparrow} + (\lambda_k^{\uparrow})^2\big)\Big) \cdot |c^\lambda_{BB\to ij}|^2
\end{aligned}$$

That is,

$$\begin{aligned}
&\frac{1}{\Delta_{ji}} \sum_B \sum_{k,\ell\in B} \lambda^{\uparrow\uparrow}_{\max(k,\ell)} \cdot |c^\lambda_{BB\to ij}|^2 + \frac{1}{\Delta_{ji}} \sum_B \sum_{k\in B} (\lambda_k^{\uparrow\uparrow})^2 \cdot |c^\lambda_{BB\to ij}|^2 \\
&= \frac{1}{\Delta_{ji}} \sum_B \sum_{k,\ell\in B} \lambda^{\uparrow}_{\max(k,\ell)} \cdot |c^\lambda_{BB\to ij}|^2 + \frac{1}{\Delta_{ji}} \sum_B \sum_{k\in B} (\lambda_k^{\uparrow})^2 \cdot |c^\lambda_{BB\to ij}|^2.
\end{aligned}$$

Thus, Equation (104) is averageable, and hence Equation (102) is averageable too. □

14.6 Deferred proofs

14.6.1 Proof of Theorem 14.13

The proofs in this section involve elementary, but lengthy calculations. We begin with a few helper lemmas.

Lemma 14.18. *Let λ be a Young diagram of length at most d, and let $i \in [d]$. We have*

$$\sum_{k=1}^{d} \left(\lambda_k^{\uparrow\uparrow} - \lambda_{k+1}^{\uparrow\uparrow}\right) \cdot D^{\lambda}_{k\to i} = (C_i^{\lambda} + 1) \cdot D^{\lambda}_{d\to i},$$

where we take $\lambda_{d+1}^{\uparrow\uparrow} = 0$.

Proof. We have

$$\sum_{k=1}^{d} \left(\lambda_k^{\uparrow\uparrow} - \lambda_{k+1}^{\uparrow\uparrow}\right) \cdot D^{\lambda}_{k\to i} = \sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow} \cdot \left(D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}\right) = \sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow} \cdot |c^{\lambda}_{k\to i}|^2 = (C_i^{\lambda} + 1) \cdot D^{\lambda}_{d\to i}.$$

In the second sum, $D^{\lambda}_{0\to i} = 0$ by convention. The second equality is by Theorem 13.3; the third is by Theorem 13.8. □

Notation 14.19. In the following lemmas and their proofs, it will be useful to abbreviate the following quantity:

$$B^{\lambda}_{ij} \coloneqq (C_i^{\lambda} - C_d^{\lambda}) \cdot D^{\lambda}_{d\to ij} - (C_i^{\lambda} - C_d^{\lambda} + 1) \cdot D^{\lambda}_{d-1\to ij}.$$

We will drop the superscript when λ is clear from context.

Lemma 14.20. *Let λ be a Young diagram of length at most d, and let $i, j \in [d]$. We have*

$$\sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow} \cdot |c^{\lambda}_{kd\to ij}|^2 = B^{\lambda}_{ij} + (C_d^{\lambda} + 1) \cdot \big(F(d,d) - F(d,d-1)\big).$$

Proof. Write $G_1(k) \coloneqq F(k,d) - F(k,d-1)$, and note that $|c^{\lambda}_{kd\to ij}|^2 = G_1(k) - G_1(k-1)$, where we take $G_1(0) = 0$. We have

$$\sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow} \cdot |c^{\lambda}_{kd\to ij}|^2 = \sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow} \cdot \big(G_1(k) - G_1(k-1)\big) = \Big(\sum_{k=1}^{d-1} (\lambda_k^{\uparrow\uparrow} - \lambda_{k+1}^{\uparrow\uparrow}) \cdot G_1(k)\Big) + \lambda_d^{\uparrow\uparrow} \cdot G_1(d). \tag{106}$$

The last term can be rewritten as

$$\lambda_d^{\uparrow\uparrow} \cdot G_1(d) = (C_d^{\lambda} + 1) \cdot \big(F(d,d) - F(d,d-1)\big). \tag{107}$$

In the remaining sum, we have $k \le d-1$, so that from Theorem 14.10, we have

$$G_1(k) = D^{\lambda}_{k\to i} \cdot \big(D^{\lambda+e_i}_{d\to j} - D^{\lambda+e_i}_{d-1\to j}\big).$$

Then, by Theorem 14.18, we have

$$\begin{aligned}
\sum_{k=1}^{d-1} (\lambda_k^{\uparrow\uparrow} - \lambda_{k+1}^{\uparrow\uparrow}) \cdot D^{\lambda}_{k\to i} &= \Big(\sum_{k=1}^{d} (\lambda_k^{\uparrow\uparrow} - \lambda_{k+1}^{\uparrow\uparrow}) \cdot D^{\lambda}_{k\to i}\Big) - \lambda_d^{\uparrow\uparrow} \cdot D^{\lambda}_{d\to i} \\
&= (C_i^{\lambda} + 1) \cdot D^{\lambda}_{d\to i} - (C_d^{\lambda} + 1) \cdot D^{\lambda}_{d\to i} \\
&= (C_i^{\lambda} - C_d^{\lambda}) \cdot D^{\lambda}_{d\to i}.
\end{aligned}$$

So,

$$\begin{aligned}\sum_{k=1}^{d-1}(\lambda_k^{\uparrow\uparrow}-\lambda_{k+1}^{\uparrow\uparrow})\cdot G_1(k) &= \Big(\sum_{k=1}^{d-1}(\lambda_k^{\uparrow\uparrow}-\lambda_{k+1}^{\uparrow\uparrow})\cdot D^{\lambda}_{k\to i}\Big)\cdot\big(D^{\lambda+e_i}_{d\to j}-D^{\lambda+e_i}_{d-1\to j}\big)\\ &= (C_i^\lambda - C_d^\lambda)\cdot D^{\lambda}_{d\to i}\cdot\big(D^{\lambda+e_i}_{d\to j}-D^{\lambda+e_i}_{d-1\to j}\big)\\ &= (C_i^\lambda - C_d^\lambda)\cdot D^{\lambda}_{d\to ij} - (C_i^\lambda - C_d^\lambda)\cdot D^{\lambda}_{d\to i}\cdot D^{\lambda+e_i}_{d-1\to j}\\ &= (C_i^\lambda - C_d^\lambda)\cdot D^{\lambda}_{d\to ij} - (C_i^\lambda - C_d^\lambda+1)\cdot D^{\lambda}_{d-1\to ij} \qquad \text{(Theorem 13.3)}\\ &= B_{ij}. \qquad (108)\end{aligned}$$

The lemma follows by substituting Equations (107) and (108) into Equation (106).

□

Lemma 14.21. *Let λ be a Young diagram of length at most d, and let $i, j \in [d]$. We have*

$$\sum_{\ell=1}^{d}\lambda_\ell^{\uparrow\uparrow}\cdot|c^\lambda_{d\ell\to ij}|^2 = \big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot B_{ji} + \frac{1}{\Delta_{ji}^2}\cdot B_{ij} + (C_d^\lambda+1)\cdot\big(F(d,d)-F(d-1,d)\big).$$

This lemma's proof is almost the same as the previous lemma's, only slightly more complicated by having to use the $s \geq t$ case of Theorem 14.10.

Proof. Write $G_2(\ell) \coloneqq F(d,\ell) - F(d-1,\ell)$, and note that $|c^\lambda_{d\ell\to ij}|^2 = G_2(\ell) - G_2(\ell-1)$, where we take $G_2(0) = 0$. We have

$$\sum_{\ell=1}^{d}\lambda_\ell^{\uparrow\uparrow}\cdot|c^\lambda_{d\ell\to ij}|^2 = \sum_{\ell=1}^{d}\lambda_\ell^{\uparrow\uparrow}\cdot\big(G_2(\ell)-G_2(\ell-1)\big) = \Big(\sum_{\ell=1}^{d-1}(\lambda_\ell^{\uparrow\uparrow}-\lambda_{\ell+1}^{\uparrow\uparrow})\cdot G_2(\ell)\Big) + \lambda_d^{\uparrow\uparrow}\cdot G_2(d). \qquad (109)$$

The last term can be rewritten as

$$\lambda_d^{\uparrow\uparrow}\cdot G_2(d) = (C_d^\lambda+1)\cdot\big(F(d,d)-F(d-1,d)\big). \qquad (110)$$

In the remaining sum, we have $\ell \leq d-1$, so that from Theorem 14.10, we have

$$G_2(\ell) = \big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot D^\lambda_{\ell\to j}\cdot\big(D^{\lambda+e_j}_{d\to i}-D^{\lambda+e_j}_{d-1\to i}\big) + \frac{1}{\Delta_{ji}^2}\cdot D^\lambda_{\ell\to i}\cdot\big(D^{\lambda+e_i}_{d\to j}-D^{\lambda+e_i}_{d-1\to j}\big).$$

Then, by calculations similar to those in Equation (108) in the previous lemma's proof, we have

$$\begin{aligned}\sum_{\ell=1}^{d-1}(\lambda_\ell^{\uparrow\uparrow}-\lambda_{\ell+1}^{\uparrow\uparrow})\cdot G_2(\ell) &= \big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot\Big((C_j^\lambda-C_d^\lambda)\cdot D^\lambda_{d\to ij} - (C_j^\lambda-C_d^\lambda+1)\cdot D^\lambda_{d-1\to ij}\Big)\\ &\quad + \frac{1}{\Delta_{ji}^2}\cdot\Big((C_i^\lambda-C_d^\lambda)\cdot D^\lambda_{d\to ij} - (C_i^\lambda-C_d^\lambda+1)\cdot D^\lambda_{d-1\to ij}\Big)\\ &= \big(1-\frac{1}{\Delta_{ji}^2}\big)\cdot B_{ji} + \frac{1}{\Delta_{ji}^2}\cdot B_{ij}. \qquad (111)\end{aligned}$$

The lemma follows by substituting Equations (110) and (111) into Equation (109). □

We are now ready to prove Theorem 14.13.

Lemma 14.22 (Theorem 14.13, restated)**.** *We have the following equation:*

$$\sum_{\substack{k,\ell\\ \max(k,\ell)=d}}(\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow}+\frac{1}{\Delta_{ji}}\lambda_d^{\uparrow\uparrow})\cdot|c^\lambda_{k\ell\to ij}|^2 = (C_i^\lambda+1)(C_j^{\lambda+e_i}+1)\cdot D^\lambda_{d\to ij} - (C_i^\lambda+2)(C_j^{\lambda+e_i}+2)\cdot D^\lambda_{d-1\to ij}.$$

Proof. We start by splitting the left-hand side into four terms:

$$\sum_{\substack{k,\ell \\ \max(k,\ell)=d}} (\lambda_k^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow} + \frac{1}{\Delta_{ji}}\lambda_d^{\uparrow\uparrow}) \cdot |c^\lambda_{k\ell\to ij}|^2 = \sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow}\lambda_d^{\uparrow\uparrow} \cdot |c^\lambda_{kd\to ij}|^2 + \sum_{\ell=1}^{d} \lambda_d^{\uparrow\uparrow}\lambda_\ell^{\uparrow\uparrow} \cdot |c^\lambda_{d\ell\to ij}|^2$$
$$- (\lambda_d^{\uparrow\uparrow})^2 \cdot |c^\lambda_{dd\to ij}|^2 + \sum_{\substack{k,\ell \\ \max(k,\ell)=d}} \frac{1}{\Delta_{ji}}\lambda_d^{\uparrow\uparrow} \cdot |c^\lambda_{k\ell\to ij}|^2. \tag{112}$$

Note that every term on the right is proportional to $\lambda_d^{\uparrow\uparrow} = C_d^\lambda + 1$. Thus, we start by considering instead the right-hand side of this equation, divided by $\lambda_d^{\uparrow\uparrow}$:

$$\sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow} \cdot |c^\lambda_{kd\to ij}|^2 + \sum_{\ell=1}^{d} \lambda_\ell^{\uparrow\uparrow} \cdot |c^\lambda_{d\ell\to ij}|^2 - \lambda_d^{\uparrow\uparrow} \cdot |c^\lambda_{dd\to ij}|^2 + \sum_{\substack{k,\ell \\ \max(k,\ell)=d}} \frac{1}{\Delta_{ji}} \cdot |c^\lambda_{k\ell\to ij}|^2. \tag{113}$$

The first term of Equation (113) is given by Theorem 14.20 to be:

$$\sum_{k=1}^{d} \lambda_k^{\uparrow\uparrow} \cdot |c^\lambda_{kd\to ij}|^2 = B_{ij} + (C_d^\lambda + 1) \cdot \big(F(d,d) - F(d,d-1)\big).$$

The second term of Equation (113), by Theorem 14.21, is:

$$\begin{aligned}\sum_{\ell=1}^{d} \lambda_\ell^{\uparrow\uparrow} \cdot |c^\lambda_{d\ell\to ij}|^2 &= \big(1 - \frac{1}{\Delta_{ji}^2}\big) \cdot B_{ji} + \frac{1}{\Delta_{ji}^2} \cdot B_{ij} + (C_d^\lambda + 1) \cdot \big(F(d,d) - F(d-1,d)\big) \\ &= B_{ji} + \frac{1}{\Delta_{ji}^2} \cdot (B_{ij} - B_{ji}) + (C_d^\lambda + 1) \cdot \big(F(d,d) - F(d-1,d)\big).\end{aligned}$$

Before pressing on, let us rewrite the difference $B_{ij} - B_{ji}$ as:

$$\begin{aligned}B_{ij} - B_{ji} &= (C_i^\lambda - C_j^\lambda) \cdot \big(D^\lambda_{d\to ij} - D^\lambda_{d-1\to ij}\big) \\ &= (\delta_{ij} - \Delta_{ij}) \cdot \big(D^\lambda_{d\to ij} - D^\lambda_{d-1\to ij}\big),\end{aligned}$$

where we have used $\Delta_{ji} = C_j^{\lambda+e_i} - C_i^\lambda = C_j^\lambda - C_i^\lambda + \delta_{ij}$. Lastly, notice that $\delta_{ij} \neq 0$ implies $i = j$, which implies $\Delta_{ji} = 1$. Thus, $\delta_{ij}/\Delta_{ji}^2 = \delta_{ij}$, and

$$\sum_{\ell=1}^{d} \lambda_\ell^{\uparrow\uparrow} \cdot |c^\lambda_{d\ell\to ij}|^2 = B_{ji} + \big(\delta_{ij} - \frac{1}{\Delta_{ji}}\big) \cdot \big(D^\lambda_{d\to ij} - D^\lambda_{d-1\to ij}\big) + (C_d^\lambda + 1) \cdot \big(F(d,d) - F(d-1,d)\big).$$

The third term of Equation (113), by Theorem 14.10, is:

$$-\lambda_d^{\uparrow\uparrow} \cdot |c^\lambda_{dd\to ij}|^2 = -(C_d^\lambda + 1) \cdot \big(F(d,d) - F(d,d-1) - F(d-1,d) + F(d-1,d-1)\big).$$

The fourth term of Equation (113), by Theorem 14.10, is:

$$\sum_{\substack{k,\ell \\ \max(k,\ell)=d}} \frac{1}{\Delta_{ji}} \cdot |c^\lambda_{k\ell\to ij}|^2 = \frac{1}{\Delta_{ji}} \cdot \Big(\sum_{k,\ell=1}^{d} |c^\lambda_{k\ell\to ij}|^2 - \sum_{k,\ell=1}^{d-1} |c^\lambda_{k\ell\to ij}|^2 \Big) = \frac{1}{\Delta_{ji}} \cdot \big(D^\lambda_{d\to ij} - D^\lambda_{d-1\to ij}\big).$$

Recombining the terms gives:

$$\begin{aligned}(113) &= (B_{ij} + B_{ji}) + \delta_{ij} \cdot \big(D^\lambda_{d\to ij} - D^\lambda_{d-1\to ij}\big) + (C_d^\lambda + 1) \cdot \big(F(d,d) - F(d-1,d-1)\big) \\ &= (B_{ij} + B_{ji}) + \big(C_d^\lambda + 1 + \delta_{ij}\big) \cdot \big(D^\lambda_{d\to ij} - D^\lambda_{d-1\to ij}\big),\end{aligned} \tag{114}$$

here using Theorem 14.10 to substitute $F(s,s) = D^{\lambda}_{s\to ij}$. Now, we also have

$$B_{ij} + B_{ji} = (C_i^\lambda + C_j^\lambda - 2C_d^\lambda) \cdot D^\lambda_{d\to ij} - (C_i^\lambda + C_j^\lambda - 2C_d^\lambda + 2) \cdot D^\lambda_{d-1\to ij},$$

so that

$$\begin{aligned}(114) &= (C_i^\lambda + C_j^\lambda - C_d^\lambda + 1 + \delta_{ij}) \cdot D^\lambda_{d\to ij} - (C_i^\lambda + C_j^\lambda - C_d^\lambda + 3 + \delta_{ij}) \cdot D^\lambda_{d-1\to ij} \\ &= (C_i^\lambda + C_j^{\lambda+e_i} - C_d^\lambda + 1) \cdot D^\lambda_{d\to ij} - (C_i^\lambda + C_j^{\lambda+e_i} - C_d^\lambda + 3) \cdot D^\lambda_{d-1\to ij}.\end{aligned}$$

Thus,

$$(112) = (C_d^\lambda + 1) \cdot (C_i^\lambda + C_j^{\lambda+e_i} - C_d^\lambda + 1) \cdot D^\lambda_{d\to ij} - (C_d^\lambda + 1) \cdot (C_i^\lambda + C_j^{\lambda+e_i} - C_d^\lambda + 3) \cdot D^\lambda_{d-1\to ij}. \quad (115)$$

Our final ingredient comes from the following identity:

$$\begin{aligned}(C_i^\lambda - C_d^\lambda) \cdot (C_j^{\lambda+e_i} - C_d^{\lambda+e_i}) \cdot D^\lambda_{d\to ij} &= \left((C_i^\lambda - C_d^\lambda) \cdot D^\lambda_{d\to i}\right) \cdot \left((C_j^{\lambda+e_i} - C_d^{\lambda+e_i}) \cdot D^{\lambda+e_i}_{d\to j}\right) \\ &= \left((C_i^\lambda - C_d^\lambda + 1) \cdot D^\lambda_{d-1\to i}\right) \cdot \left((C_j^{\lambda+e_i} - C_d^{\lambda+e_i} + 1) \cdot D^{\lambda+e_i}_{d-1\to j}\right) \\ &= (C_i^\lambda - C_d^\lambda + 1) \cdot (C_j^{\lambda+e_i} - C_d^{\lambda+e_i} + 1) \cdot D^\lambda_{d-1\to ij},\end{aligned}$$

using Theorem 13.3. In particular, we will add zero to Equation (115) as:

$$0 = (C_i^\lambda - C_d^\lambda) \cdot (C_j^{\lambda+e_i} - C_d^{\lambda+e_i}) \cdot D^\lambda_{d\to ij} - (C_i^\lambda - C_d^\lambda + 1) \cdot (C_j^{\lambda+e_i} - C_d^{\lambda+e_i} + 1) \cdot D^\lambda_{d-1\to ij}.$$

However, the coefficient of $D^\lambda_{d\to ij}$ then becomes:

$$\begin{aligned}&(C_d^\lambda + 1) \cdot (C_i^\lambda + C_j^{\lambda+e_i} - C_d^\lambda + 1) + (C_i^\lambda - C_d^\lambda) \cdot (C_j^{\lambda+e_i} - C_d^{\lambda+e_i}) \\ &= (C_i^\lambda + 1)(C_j^{\lambda+e_i} + 1) + (C_d^\lambda - C_d^{\lambda+e_i}) \cdot (C_i^\lambda - C_d^\lambda) \\ &= (C_i^\lambda + 1)(C_j^{\lambda+e_i} + 1),\end{aligned}$$

where the last step holds since the term $C_d^\lambda - C_d^{\lambda+e_i} = -\delta_{id}$ is nonzero if and only if $C_i^\lambda = C_d^\lambda$. Similarly, the coefficient of $D^\lambda_{d-1\to ij}$ becomes:

$$\begin{aligned}&(C_d^\lambda + 1) \cdot (C_i^\lambda + C_j^{\lambda+e_i} - C_d^\lambda + 3) + (C_i^\lambda - C_d^\lambda + 1) \cdot (C_j^{\lambda+e_i} - C_d^{\lambda+e_i} + 1) \\ &= (C_i^\lambda + 2)(C_j^{\lambda+e_i} + 2) + (C_d^\lambda - C_d^{\lambda+e_i}) \cdot (C_i^\lambda - C_d^\lambda) \\ &= (C_i^\lambda + 2)(C_j^{\lambda+e_i} + 2).\end{aligned}$$

Thus,

$$(115) = (C_i^\lambda + 1)(C_j^{\lambda+e_i} + 1) \cdot D^\lambda_{d\to ij} - (C_i^\lambda + 2)(C_j^{\lambda+e_i} + 2) \cdot D^\lambda_{d-1\to ij},$$

which completes the proof of the lemma. □

14.6.2 Proof of Theorem 14.14

In this subsection, we prove the relations between the two-step Clebsch-Gordan coefficients with respect to the blocks of the Young diagram λ.

Lemma 14.23 (Theorem 14.14, restated)**.** *Let λ be a Young diagram, and $i, j \in [d]$. There exist constants $\{c^\lambda_{B_1 B_2 \to ij}\}_{B_1,B_2}$, where B_1, B_2 are blocks of λ, such that the two-step Clebsch-Gordan coefficients $|c^\lambda_{k\ell\to ij}|^2$ satisfy:*

(i) Let B be a block of λ. Then the two-step Clebsch-Gordan coefficients $|c^\lambda_{k\ell\to ij}|^2$ are equal for all pairs of distinct $k, \ell \in B$, that is:

$$|c^\lambda_{k\ell\to ij}|^2 = |c^\lambda_{BB\to ij}|^2.$$

In addition, when $k = \ell$,

$$|c^\lambda_{kk\to ij}|^2 = \left(1 + \frac{1}{\Delta_{ji}}\right) \cdot |c^\lambda_{BB\to ij}|^2.$$

(ii) Let B_1, B_2 be two distinct blocks of λ. Then the two-step Clebsch-Gordan coefficients $|c^\lambda_{k\ell\to ij}|^2$ are equal for all $k \in B_1$ and $\ell \in B_2$, that is:

$$|c^\lambda_{k\ell\to ij}|^2 = |c^\lambda_{B_1B_2\to ij}|^2.$$

Proof. The proof follows by combining Theorems 14.25 to 14.27 and 14.29. These lemmas are proven below.

We first prove item (ii). Since blocks contain consecutive indices, and B_1, B_2 are distinct blocks of λ, one of the blocks has strictly smaller row indices than the other block. Without loss of generality, let B_1 be the block with the smaller indices. We would like to show that for all $k, k' \in B_1$ and $\ell, \ell' \in B_2$:

$$|c^\lambda_{k\ell\to ij}|^2 = |c^\lambda_{k'\ell'\to ij}|^2.$$

Since $k, k' < \ell, \ell'$ Theorem 14.25 implies the equality above for all such k, k' and ℓ, ℓ'.

Let us now prove item (i). For a block B of λ, let k, k', ℓ, ℓ' be elements of B. If $|B| = 1$, then the statement trivially holds; thus, we assume that $|B| \geq 2$, and use $m, m+1$ to denote the two smallest indices of the block.

We first consider the case of distinct $k \neq \ell$. If $k < \ell$ and $k' < \ell'$, or $k > \ell$ and $k' > \ell'$, then Theorem 14.25 again implies that $|c^\lambda_{k\ell\to ij}|^2 = |c^\lambda_{k'\ell'\to ij}|^2$. Otherwise, we assume without loss of generality that $k < \ell$ and $k' > \ell'$ (the proof of the $k > \ell$ and $k' < \ell'$ case is identical to this one). Theorem 14.25 implies that

$$|c^\lambda_{k\ell\to ij}|^2 = |c^\lambda_{m,m+1\to ij}|^2, \qquad |c^\lambda_{k'\ell'\to ij}|^2 = |c^\lambda_{m+1,m\to ij}|^2.$$

We show in Theorem 14.26 that the right-hand sides of the two equations above are equal, which completes the $k \neq \ell$ case of item (i).

The case of $k = \ell$ remains. If k is not the first row in the block B, then Theorem 14.27 implies that $|c^\lambda_{kk\to ij}|^2 = (1+1/\Delta_{ji}) \cdot |c^\lambda_{k-1,k\to ij}|^2$. Since $k-1, k \in B$, we have argued earlier that $|c^\lambda_{k-1,k\to ij}|^2 = |c^\lambda_{BB\to ij}|^2$, and the desired statement holds. Note that the above argument does not work if k is the first row of block B, since $k-1$ is not in block B in that case. This case is proved separately in Theorem 14.29. □

Before we proceed with the proofs of the remaining lemmas, we introduce the following piece of notation that will simplify our calculations.

Notation 14.24. Let $F_1(s,t)$ and $F_2(s,t)$ denote the two expressions that arise in the partial sums of the two-step Clebsch-Gordan coefficients:

$$F_1(s,t) \coloneqq D^\lambda_{s\to i} \cdot D^{\lambda+e_i}_{t\to j}, \qquad F_2(s,t) \coloneqq \Big(1 - \frac{1}{\Delta^2_{ji}}\Big) \cdot D^\lambda_{t\to j} \cdot D^{\lambda+e_j}_{s\to i} + \frac{1}{\Delta^2_{ji}} \cdot D^\lambda_{t\to i} \cdot D^{\lambda+e_i}_{s\to j}.$$

We extend the definition of $F_1(s,t)$ and $F_2(s,t)$ to all values of s, t, even though they have a more restricted domain in Theorem 14.10.

In particular, the statement of Theorem 14.10 simplifies to

$$F(s,t) = \sum_{k=1}^{s} \sum_{\ell=1}^{t} |c^\lambda_{k\ell\to ij}|^2 = \begin{cases} F_1(s,t) & s \leq t, \\ F_2(s,t) & s > t. \end{cases}$$

Lemma 14.25. *Let λ be a Young diagram, and B_1, B_2 be two blocks of λ (not necessarily distinct), such that $k, k' \in B_1$, and $\ell, \ell' \in B_2$. If either*

(i) $k < \ell$ and $k' < \ell'$, or

(ii) $k > \ell$ and $k' > \ell'$,

then

$$|c^\lambda_{k\ell\to ij}|^2 = |c^\lambda_{k'\ell'\to ij}|^2.$$

Proof. Let us consider the case when $k < \ell$ and $k' < \ell'$ first. We write

$$\begin{aligned}|c^{\lambda}_{k\ell\to ij}|^2 &= F(k,\ell) - F(k-1,\ell) - F(k,\ell-1) + F(k-1,\ell-1)\\ &= D^{\lambda}_{k\to i}D^{\lambda+e_i}_{\ell\to j} - D^{\lambda}_{k-1\to i}D^{\lambda+e_i}_{\ell\to j} - D^{\lambda}_{k\to i}D^{\lambda+e_i}_{\ell-1\to j} + D^{\lambda}_{k-1\to i}D^{\lambda+e_i}_{\ell-1\to j} && \text{(Theorem 14.10)}\\ &= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i})\cdot(D^{\lambda+e_i}_{\ell\to j} - D^{\lambda+e_i}_{\ell-1\to j}). && (116)\end{aligned}$$

Similarly,

$$|c^{\lambda}_{k'\ell'\to ij}|^2 = (D^{\lambda}_{k'\to i} - D^{\lambda}_{k'-1\to i})\cdot(D^{\lambda+e_i}_{\ell'\to j} - D^{\lambda+e_i}_{\ell'-1\to j}). \tag{117}$$

Since k and k' are in the same block of λ, it holds from Theorem 13.5 that

$$D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i} = |c^{\lambda}_{k\to i}|^2 = |c^{\lambda}_{k'\to i}|^2 = D^{\lambda}_{k'\to i} - D^{\lambda}_{k'-1\to i}.$$

Moreover, if ℓ and ℓ' are in the same block of $\lambda + e_i$, then it also follows that

$$D^{\lambda+e_i}_{\ell\to j} - D^{\lambda+e_i}_{\ell-1\to j} = |c^{\lambda+e_i}_{\ell\to j}|^2 = |c^{\lambda+e_i}_{\ell'\to j}|^2 = D^{\lambda+e_i}_{\ell'\to j} - D^{\lambda+e_i}_{\ell'-1\to j},$$

which implies that the right-hand sides of Equation (116) and Equation (117) are equal.

If ℓ and ℓ' are not in the same block of $\lambda + e_i$, then, because ℓ and ℓ' are in the same block of λ, we must have $i = \ell$ or $i = \ell'$. Assume $i = \ell$; the $i = \ell'$ case is similar. Since $k < \ell$, this means that $k < i$, and thus the respective one-step Clebsch-Gordan coefficient is equal to zero:

$$0 = |c^{\lambda}_{k\to i}|^2 = D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i} = D^{\lambda}_{k'\to i} - D^{\lambda}_{k'-1\to i}.$$

We conclude that in all cases the right-hand sides of Equation (116) and Equation (117) are equal, and so are the two-step Clebsch-Gordan coefficients.

We now turn our attention to the case when $k > \ell$ and $k' > \ell'$. Using Theorem 14.10 we write:

$$\begin{aligned}|c^{\lambda}_{k\ell\to ij}|^2 &= F(k,\ell) - F(k-1,\ell) - F(k,\ell-1) + F(k-1,\ell-1)\\ &= \big(1 - \frac{1}{\Delta^2_{ji}}\big)\cdot(D^{\lambda}_{\ell\to j} - D^{\lambda}_{\ell-1\to j})\cdot(D^{\lambda+e_j}_{k\to i} - D^{\lambda+e_j}_{k-1\to i}) + \frac{1}{\Delta^2_{ji}}\cdot(D^{\lambda}_{\ell\to i} - D^{\lambda}_{\ell-1\to i})\cdot(D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j})\\ &= \big(1 - \frac{1}{\Delta^2_{ji}}\big)\cdot|c^{\lambda}_{\ell k\to ji}|^2 + \frac{1}{\Delta^2_{ji}}\cdot|c^{\lambda}_{\ell k\to ij}|^2, && (118)\end{aligned}$$

where the last equality follows from Equation (116). Similarly,

$$|c^{\lambda}_{k'\ell'\to ij}|^2 = \big(1 - \frac{1}{\Delta^2_{ji}}\big)\cdot|c^{\lambda}_{\ell' k'\to ji}|^2 + \frac{1}{\Delta^2_{ji}}\cdot|c^{\lambda}_{\ell' k'\to ij}|^2. \tag{119}$$

Since $\ell < k$ and $\ell' < k'$, the first part of this proof implies that

$$|c^{\lambda}_{\ell k\to ji}|^2 = |c^{\lambda}_{\ell' k'\to ji}|^2, \qquad |c^{\lambda}_{\ell k\to ij}|^2 = |c^{\lambda}_{\ell' k'\to ij}|^2,$$

and we conclude that the right-hand sides of Equation (118) and Equation (119) are equal, and so are the corresponding two-step Clebsch-Gordan coefficients. □

Lemma 14.26. *Let λ be a Young diagram, and B be a block of λ, such that $k, k+1 \in B$. Then*

$$|c^{\lambda}_{k,k+1\to ij}|^2 = |c^{\lambda}_{k+1,k\to ij}|^2.$$

Proof. First, we observe that since $k, k+1$ are in the same block, when $i = k+1$, both two-step coefficients are equal to zero, since $\lambda + e_{k+1}$ is an invalid Young diagram. Similarly, both coefficients are zero when $j = k+1$ and $i \neq k$. Thus, for the remainder of this proof, we will consider the case when $i = k, j = k+1$, or when both $i, j \neq k+1$.

Using Theorem 14.10 we write:

$$\begin{aligned}|c^{\lambda}_{k,k+1\to ij}|^2 &= F(k,k+1) - F(k-1,k+1) - F(k,k) + F(k-1,k)\\ &= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i})\cdot(D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j})\\ &= (D^{\lambda}_{k+1\to i} - D^{\lambda}_{k\to i})\cdot(D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j}), && (k, k+1 \in B)\end{aligned}$$

and similarly

$$\begin{aligned}|c^{\lambda}_{k+1,k\to ij}|^2 &= F(k+1,k) - F(k,k) - F(k+1,k-1) + F(k,k-1)\\ &= \big(1-\frac{1}{\Delta^2_{ji}}\big)\cdot(D^{\lambda}_{k\to j} - D^{\lambda}_{k-1\to j})\cdot(D^{\lambda+e_j}_{k+1\to i} - D^{\lambda+e_j}_{k\to i}) + \frac{1}{\Delta^2_{ji}}\cdot(D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i})\cdot(D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j})\\ &= \big(1-\frac{1}{\Delta^2_{ji}}\big)\cdot(D^{\lambda}_{k+1\to j} - D^{\lambda}_{k\to j})\cdot(D^{\lambda+e_j}_{k+1\to i} - D^{\lambda+e_j}_{k\to i}) + \frac{1}{\Delta^2_{ji}}\cdot(D^{\lambda}_{k+1\to i} - D^{\lambda}_{k\to i})\cdot(D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j}),\end{aligned}$$

where the last equation follows because k and $k+1$ are in the same block of λ. Equating the two previous expressions, it suffices to show that:

$$\big(1-\frac{1}{\Delta^2_{ji}}\big)\cdot(D^{\lambda}_{k+1\to j} - D^{\lambda}_{k\to j})\cdot(D^{\lambda+e_j}_{k+1\to i} - D^{\lambda+e_j}_{k\to i}) = \big(1-\frac{1}{\Delta^2_{ji}}\big)\cdot(D^{\lambda}_{k+1\to i} - D^{\lambda}_{k\to i})\cdot(D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j}).$$

When $i=k$ and $j=k+1$, then $\Delta^2_{ji}=1$, and thus the equation above is trivially satisfied. Let us now restrict our attention to the case when $\Delta^2_{ji}\neq 1$ and both i,j are not equal to $k+1$. We observe that

$$D^{\lambda}_{k\to i}D^{\lambda+e_i}_{k\to j} = \frac{\dim(V^k_{\lambda\leq k+e_i+e_j})}{\dim(V^k_{\lambda\leq k})} = D^{\lambda}_{k\to j}D^{\lambda+e_j}_{k\to i} = D^{\lambda}_{k\to ij}, \tag{120}$$

hence, what we want to show above simplifies to

$$D^{\lambda}_{k+1\to j}D^{\lambda+e_j}_{k\to i} + D^{\lambda}_{k\to j}D^{\lambda+e_j}_{k+1\to i} = D^{\lambda}_{k+1\to i}D^{\lambda+e_i}_{k\to j} + D^{\lambda}_{k\to i}D^{\lambda+e_i}_{k+1\to j}. \tag{121}$$

Towards using Equation (120) again, we would like to use Theorem 13.3 to change the "$D_{k+1\to\cdot}$" factors to "$D_{k\to\cdot}$" factors. Since $i,j\neq k+1$, we can write:

$$\begin{aligned}D^{\lambda}_{k+1\to j} &= \Big(1+\frac{1}{C^{\lambda}_j - C^{\lambda}_{k+1}}\Big)\cdot D^{\lambda}_{k\to j}, \qquad D^{\lambda}_{k+1\to i} = \Big(1+\frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}}\Big)\cdot D^{\lambda}_{k\to j},\\ D^{\lambda+e_j}_{k+1\to i} &= \Big(1+\frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}}\Big)\cdot D^{\lambda+e_j}_{k\to i}, \qquad D^{\lambda+e_i}_{k+1\to j} = \Big(1+\frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big)\cdot D^{\lambda+e_i}_{k\to j}.\end{aligned}$$

To show Equation (121), it suffices to show that

$$\begin{aligned}&\iff \Big(1+\frac{1}{C^{\lambda}_j - C^{\lambda}_{k+1}} + 1 + \frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}}\Big)D^{\lambda}_{k\to ij} = \Big(1+\frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}} + 1 + \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big)D^{\lambda}_{k\to ij}\\ &\iff \Big(\frac{1}{C^{\lambda}_j - C^{\lambda}_{k+1}} + \frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}}\Big)D^{\lambda}_{k\to ij} = \Big(\frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}} + \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big)D^{\lambda}_{k\to ij}.\end{aligned} \tag{122}$$

Theorem 13.3 also implies the following:

$$\begin{aligned}D^{\lambda}_{k+1\to ij} &= D^{\lambda}_{k+1\to i}D^{\lambda+e_i}_{k+1\to j} = \Big(1+\frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}}\Big)D^{\lambda}_{k\to i}\Big(1+\frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big)D^{\lambda+e_i}_{k+1\to j}\\ &= \Big(1+\frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}}\Big)\cdot\Big(1+\frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big)\cdot D^{\lambda}_{k\to ij},\end{aligned}$$

and

$$\begin{aligned}D^{\lambda}_{k+1\to ij} &= D^{\lambda}_{k+1\to j}D^{\lambda+e_j}_{k+1\to i} = \Big(1+\frac{1}{C^{\lambda}_j - C^{\lambda}_{k+1}}\Big)D^{\lambda}_{k\to j}\Big(1+\frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}}\Big)D^{\lambda+e_j}_{k+1\to i}\\ &= \Big(1+\frac{1}{C^{\lambda}_j - C^{\lambda}_{k+1}}\Big)\cdot\Big(1+\frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}}\Big)\cdot D^{\lambda}_{k\to ij}.\end{aligned}$$

Hence

$$\left(1+\frac{1}{C_i^\lambda - C_{k+1}^\lambda}\right)\cdot\left(1+\frac{1}{C_j^{\lambda+e_i} - C_{k+1}^{\lambda+e_i}}\right)\cdot D_{k\to ij}^\lambda = \left(1+\frac{1}{C_j^\lambda - C_{k+1}^\lambda}\right)\cdot\left(1+\frac{1}{C_i^{\lambda+e_j} - C_{k+1}^{\lambda+e_j}}\right)\cdot D_{k\to ij}^\lambda. \quad (123)$$

Subtracting Equation (123) from Equation (122), we only need to show that

$$\frac{1}{C_i^\lambda - C_{k+1}^\lambda}\cdot\frac{1}{C_j^{\lambda+e_i} - C_{k+1}^{\lambda+e_i}}\cdot D_{k\to ij}^\lambda = \frac{1}{C_j^\lambda - C_{k+1}^\lambda}\cdot\frac{1}{C_i^{\lambda+e_j} - C_{k+1}^{\lambda+e_j}}\cdot D_{k\to ij}^\lambda.$$

We remark that $C_y^{\lambda+e_x} = C_y^\lambda + \delta_{xy}$, and thus we rewrite the above expression as

$$(C_i^\lambda - C_{k+1}^\lambda)\cdot(C_j^\lambda - C_{k+1}^\lambda + \delta_{ij} - \delta_{i(k+1)})\cdot D_{k\to ij}^\lambda = (C_j^\lambda - C_{k+1}^\lambda)\cdot(C_i^\lambda - C_{k+1}^\lambda - \delta_{ij} - \delta_{j(k+1)})\cdot D_{k\to ij}^\lambda.$$

Collecting terms, it suffices to show that

$$\begin{aligned}&(C_j^\lambda - C_{k+1}^\lambda)\cdot(\delta_{ji} - \delta_{j(k+1)}) = (C_i^\lambda - C_{k+1}^\lambda)\cdot(\delta_{ij} - \delta_{i(k+1)})\\ \iff\ &(C_i^\lambda - C_{k+1}^\lambda)\cdot\delta_{i(k+1)} - (C_j^\lambda - C_{k+1}^\lambda)\cdot\delta_{j(k+1)} + (C_j^\lambda - C_i^\lambda)\cdot\delta_{ij} = 0.\end{aligned}$$

The last equation holds because the expression $(C_x^\lambda - C_y^\lambda)\cdot\delta_{xy}$ is zero for all x, y. □

Lemma 14.27. *Let λ be a Young diagram, and B be a block of λ, such that $k, k+1 \in B$. Then*

$$|c_{k+1,k+1\to ij}^\lambda|^2 = \left(1+\frac{1}{\Delta_{ji}}\right)\cdot|c_{k,k+1\to ij}^\lambda|^2.$$

Proof. We first observe that if $i > k+1$, then the Clebsch-Gordan coefficients on both sides are equal to zero. Moreover, both coefficients vanish when $i = k+1$. This is because the first insertion of the left coefficient corresponds to the Young diagram of shape $\lambda + e_{k+1}$, and this is not a valid shape since $k, k+1$ are in the same block. Similarly, the first insertion of the right coefficient corresponds to adding k to row $i > k$, which results in an invalid SSYT. Hence, we can restrict our attention to when $i \le k$.

Similarly, j has to be at most $k+1$ since we add a letter that is at most $k+1$ to row j. And since $k, k+1$ are in the same block, we can only have $j = k+1$ when $i = k$, in which case the new boxes are added in a vertical strip. Then $\Delta_{ji} = -1$ and the right-hand side vanishes. In that "vertical" case, the left-hand side also vanishes, because the coefficient corresponds to the SSYT $T_{k+1,k+1\to ij}^\lambda$, which is an invalid tableau, since it contains the letter $k+1$ in the same column as the same letter. For the rest of this proof, we assume that $i, j \le k$.

On the left-hand side, we have

$$\begin{aligned}&|c_{k+1,k+1\to ij}^\lambda|^2\\ &= F_1(k+1,k+1) - F_1(k,k+1) - F_2(k+1,k) + F_1(k,k)\\ &= \big(F_1(k+1,k+1) - F_1(k,k+1) - F_1(k+1,k) + F_1(k,k)\big) + \big(F_1(k+1,k) - F_2(k+1,k)\big)\\ &= (D_{k+1\to i}^\lambda - D_{k\to i}^\lambda)\cdot(D_{k+1\to j}^{\lambda+e_i} - D_{k\to j}^{\lambda+e_i}) + \big(F_1(k+1,k) - F_2(k+1,k)\big).\end{aligned}$$

On the other hand,

$$\begin{aligned}|c_{k,k+1\to ij}^\lambda|^2 &= F_1(k,k+1) - F_1(k-1,k+1) - F_1(k,k) + F_1(k-1,k)\\ &= (D_{k\to i}^\lambda - D_{k-1\to i}^\lambda)\cdot(D_{k+1\to j}^{\lambda+e_i} - D_{k\to j}^{\lambda+e_i})\\ &= (D_{k+1\to i}^\lambda - D_{k\to i}^\lambda)\cdot(D_{k+1\to j}^{\lambda+e_i} - D_{k\to j}^{\lambda+e_i}) \qquad (k, k+1 \in B)\end{aligned}$$

Comparing the two expressions, it suffices to show that

$$F_1(k+1,k) - F_2(k+1,k) = \frac{1}{\Delta_{ji}}\cdot(D_{k+1\to i}^\lambda - D_{k\to i}^\lambda)(D_{k+1\to j}^{\lambda+e_i} - D_{k\to j}^{\lambda+e_i}). \quad (124)$$

We prove this equation holds in Theorem 14.28 below. This proof is deferred to a separate lemma because it is also used in the proof of Theorem 14.29, and because proving Equation (124) does not require that k and $k+1$ are in the same block. □

Lemma 14.28. *Let λ be a Young diagram, and let $i, j, k \in [d]$ such that $k+1 > i, j$. Then*

$$F_1(k+1,k) - F_2(k+1,k) = \frac{1}{\Delta_{ji}} \cdot (D^{\lambda}_{k+1\to i} - D^{\lambda}_{k\to i}) \cdot (D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j}).$$

Proof. We use Theorem 13.3 to rewrite the right-hand side of the desired statement:

$$\begin{aligned}\frac{1}{\Delta_{ji}} \cdot (D^{\lambda}_{k+1\to i} - D^{\lambda}_{k\to i}) \cdot (D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j}) &= \frac{1}{\Delta_{ji}} \cdot \frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}} \cdot D^{\lambda}_{k\to i} \cdot \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}} \cdot D^{\lambda+e_i}_{k\to j} \\ &= \frac{1}{\Delta_{ji}} \cdot \frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}} \cdot \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}} \cdot D^{\lambda}_{k\to ij}. \qquad (125)\end{aligned}$$

The fact that $i, j < k+1$ ensures that the denominators of the above expressions are never zero. Similarly, we write the left-hand side of the lemma statement as:

$$\begin{aligned}&F_1(k+1,k) - F_2(k+1,k) \\ &= D^{\lambda}_{k+1\to i} \cdot D^{\lambda+e_i}_{k\to j} - \Big(1 - \frac{1}{\Delta^2_{ji}}\Big) \cdot D^{\lambda}_{k\to j} \cdot D^{\lambda+e_j}_{k+1\to i} - \frac{1}{\Delta^2_{ji}} \cdot D^{\lambda}_{k\to i} \cdot D^{\lambda+e_i}_{k+1\to j} \\ &= \Big(1 + \frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}}\Big) \cdot D^{\lambda}_{k\to i} \cdot D^{\lambda+e_i}_{k\to j} - \Big(1 - \frac{1}{\Delta^2_{ji}}\Big) \cdot D^{\lambda}_{k\to j} \cdot \Big(1 + \frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}}\Big) \cdot D^{\lambda+e_j}_{k\to i} \\ &\quad - \frac{1}{\Delta^2_{ji}} \cdot D^{\lambda}_{k\to i} \cdot \Big(1 + \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big) \cdot D^{\lambda+e_i}_{k\to j} \\ &= \Big[\Big(1 + \frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}}\Big) - \Big(1 - \frac{1}{\Delta^2_{ji}}\Big)\Big(1 + \frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}}\Big) - \frac{1}{\Delta^2_{ji}}\Big(1 + \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big)\Big] \cdot D^{\lambda}_{k\to ij} \\ &= \Big[\frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}} - \Big(1 - \frac{1}{\Delta^2_{ji}}\Big) \cdot \frac{1}{C^{\lambda+e_j}_i - C^{\lambda+e_j}_{k+1}} - \frac{1}{\Delta^2_{ji}} \cdot \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big] \cdot D^{\lambda}_{k\to ij}. \qquad (126)\end{aligned}$$

We show that the two sides are equal by considering the following two cases:

Case 1: $i = j$. In this case, $\Delta_{ji} = C^{\lambda+e_i}_j - C^{\lambda}_i = 1$, so that we have:

$$\begin{aligned}(126) &= \Big[\frac{1}{C^{\lambda}_i - C^{\lambda}_{k+1}} - \frac{1}{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1}}\Big] \cdot D^{\lambda}_{k\to ij} \\ &= \frac{C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1} - C^{\lambda}_i + C^{\lambda}_{k+1}}{(C^{\lambda}_i - C^{\lambda}_{k+1}) \cdot (C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1})} \cdot D^{\lambda}_{k\to ij} \\ &= \frac{1}{(C^{\lambda}_i - C^{\lambda}_{k+1}) \cdot (C^{\lambda+e_i}_j - C^{\lambda+e_i}_{k+1})} \cdot D^{\lambda}_{k\to ij} = (125) \qquad (i \neq k+1 \implies C^{\lambda+e_i}_{k+1} = C^{\lambda}_{k+1})\end{aligned}$$

Case 2: $i \neq j$. In this case, $C^{\lambda+e_j}_i = C^{\lambda}_i$ and $C^{\lambda+e_i}_j = C^{\lambda}_j$. Let us use C_i, C_j to refer to these quantities. Moreover, $\Delta_{ji} = (C^{\lambda+e_i}_j + 1) - (C^{\lambda}_i + 1) = C_j - C_i$. Then, we get:

$$\begin{aligned}(126) &= \Big[\frac{1}{C_i - C_{k+1}} - \Big(1 - \frac{1}{\Delta^2_{ji}}\Big) \cdot \frac{1}{C_i - C_{k+1}} - \frac{1}{\Delta^2_{ji}} \cdot \frac{1}{C_j - C_{k+1}}\Big] \cdot D^{\lambda}_{k\to ij} \\ &= \frac{1}{\Delta^2_{ji}} \cdot \frac{C_j - C_i}{(C_i - C_{k+1})(C_j - C_{k+1})} \cdot D^{\lambda}_{k\to ij} \\ &= \frac{1}{\Delta^2_{ji}} \cdot \frac{\Delta_{ji}}{(C_i - C_{k+1})(C_j - C_{k+1})} \cdot D^{\lambda}_{k\to ij} = (125).\end{aligned}$$

□

Lemma 14.29. *Let λ be a Young diagram, and B be a block of λ, such that $k, k+1 \in B$. Then*

$$|c^{\lambda}_{kk\to ij}|^2 = \Big(1 + \frac{1}{\Delta_{ji}}\Big) \cdot |c^{\lambda}_{k,k+1\to ij}|^2.$$

Proof. We first observe that if $i > k$, then the Clebsch-Gordan coefficients on both sides are equal to zero. Hence, we can restrict our attention to when $i \leq k$.

Similarly, if $j > k+1$, then both coefficients are zero. The same happens when $i < k, j = k+1$, since the left coefficient has an invalid SSYT, and the right coefficient corresponds to the invalid shape $\lambda + e_i + e_{k+1}$. When $i = k, j = k+1$, then the left-hand coefficient is zero, and $\Delta_{ji} = -1$ (since $k, k+1$ are in the same block), thus the right-hand side also vanishes. We conclude that we can further restrict our attention to when $i, j \leq k$.

On the left-hand side, we have

$$\begin{aligned}
|c^{\lambda}_{kk\to ij}|^2 &= F_1(k,k) - F_1(k-1,k) - F_2(k,k-1) + F_1(k-1,k-1) \\
&= \big(F_1(k,k) - F_1(k-1,k) - F_1(k,k-1) + F_1(k-1,k-1)\big) + \big(F_1(k,k-1) - F_2(k,k-1)\big) \\
&= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) + \big(F_1(k,k-1) - F_2(k,k-1)\big). \qquad (127)
\end{aligned}$$

The right-hand side is equal to

$$\begin{aligned}
\big(1 + \frac{1}{\Delta_{ji}}\big) \cdot |c^{\lambda}_{k,k+1\to ij}|^2 &= \big(1 + \frac{1}{\Delta_{ji}}\big) \cdot \big(F_1(k,k+1) - F_1(k-1,k+1) - F_1(k,k) + F_1(k-1,k)\big) \\
&= \big(1 + \frac{1}{\Delta_{ji}}\big) \cdot (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j}). \qquad (128)
\end{aligned}$$

Let us now consider the following four cases, depending on whether i, j are equal to k or strictly less than k.

Case 1: $i, j < k$**.** Since $i < k$, k and $k+1$ are in the same block of $\lambda + e_i$, which allows us to write:

$$(128) = \big(1 + \frac{1}{\Delta_{ji}}\big) \cdot (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}).$$

Comparing the expressions for the left and right-hand sides, it suffices to show that

$$F_1(k,k-1) - F_2(k,k-1) = \frac{1}{\Delta_{ji}} \cdot (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}). \qquad (129)$$

The above equation follows directly from Theorem 14.28 when we use $k-1$ in the place of k, and since $i, j < k$.

Case 2: $i = j = k$**.** In that case, $\Delta_{ji} = 1$, and recall that the term $D^{\lambda}_{k-1\to i}$ is zero when $i > k-1$. The left-hand side can be simplified to

$$(127) = D^{\lambda}_{k\to i} D^{\lambda+e_i}_{k\to j} + \big(F_1(k,k-1) - F_2(k,k-1)\big) = D^{\lambda}_{k\to i} D^{\lambda+e_i}_{k\to j},$$

where the last equality holds because both $F_1(k,k-1)$ and $F_2(k,k-1)$ have terms of the form $D^{\mu}_{k-1\to k}$, which are zero for any diagram μ. The right-hand side can be simplified to

$$\begin{aligned}
(128) &= \big(1 + \frac{1}{\Delta_{ji}}\big) \cdot D^{\lambda}_{k\to i} \cdot (D^{\lambda+e_i}_{k+1\to j} - D^{\lambda+e_i}_{k\to j}) \\
&= 2 \cdot D^{\lambda}_{k\to i} \cdot \frac{1}{C^{\lambda+e_i}_{j} - C^{\lambda+e_i}_{k+1}} \cdot D^{\lambda+e_i}_{k\to j} && \text{(Theorem 13.3)} \\
&= 2 \cdot D^{\lambda}_{k\to i} \cdot \frac{1}{(\lambda_k + 1 - k) - (\lambda_{k+1} - k - 1)} \cdot D^{\lambda+e_i}_{k\to j} = D^{\lambda}_{k\to i} D^{\lambda+e_i}_{k\to j}. && (i = j = k \text{ and } \lambda_k = \lambda_{k+1})
\end{aligned}$$

Hence, the two sides are equal.

Case 3: $i = k, j < k$. In that case, $\Delta_{ji} = C_j^{\lambda+e_i} - C_i^{\lambda} = C_j^{\lambda} - C_k^{\lambda}$. The left-hand side can be simplified to

$$\begin{aligned}
(127) &= D^{\lambda}_{k\to i} \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) + \big(F_1(k,k-1) - F_2(k,k-1)\big) \\
&= D^{\lambda}_{k\to i} \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) + D^{\lambda}_{k\to i} \cdot D^{\lambda+e_i}_{k-1\to j} - \Big(1 - \frac{1}{\Delta_{ji}^2}\Big) \cdot D^{\lambda}_{k-1\to j} \cdot D^{\lambda+e_j}_{k\to i} \\
&= D^{\lambda}_{k\to i} \cdot D^{\lambda+e_i}_{k\to j} - \Big(1 - \frac{1}{\Delta_{ji}^2}\Big) \cdot \Big(1 - \frac{1}{C_j^{\lambda} - C_k^{\lambda} + 1}\Big) \cdot D^{\lambda}_{k\to j} \cdot D^{\lambda+e_j}_{k\to i} && \text{(Theorem 13.3)} \\
&= D^{\lambda}_{k\to ij} - \frac{(\Delta_{ji}-1)(\Delta_{ji}+1)}{\Delta_{ji}^2} \cdot \frac{\Delta_{ji}}{\Delta_{ji}+1} \cdot D^{\lambda}_{k\to ij} = \frac{1}{\Delta_{ji}} \cdot D^{\lambda}_{k\to ij}.
\end{aligned}$$

The right-hand side is equal to

$$\begin{aligned}
(128) &= \Big(1 + \frac{1}{\Delta_{ji}}\Big) \cdot D^{\lambda}_{k\to i} \cdot \frac{1}{C_j^{\lambda+e_i} - C_{k+1}^{\lambda+e_i}} \cdot D^{\lambda+e_i}_{k\to j} && \text{(Theorem 13.3)} \\
&= \frac{\Delta_{ji}+1}{\Delta_{ji}} \cdot \frac{1}{C_j^{\lambda} - C_k^{\lambda} + 1} \cdot D^{\lambda}_{k\to ij} && (C_{k+1}^{\lambda+e_i} = \lambda_{k+1} - k - 1 = \lambda_k - k - 1 = C_k^{\lambda} - 1) \\
&= \frac{\Delta_{ji}+1}{\Delta_{ji}} \cdot \frac{1}{\Delta_{ji}+1} D^{\lambda}_{k\to ij} = \frac{1}{\Delta_{ji}} \cdot D^{\lambda}_{k\to ij}.
\end{aligned}$$

Hence, the two sides are equal.

Case 4: $i < k, j = k$. In that case, $\Delta_{ji} = C_j^{\lambda+e_i} - C_i^{\lambda} = C_k^{\lambda} - C_i^{\lambda}$. The left-hand side can be simplified to

$$\begin{aligned}
(127) &= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) - \frac{1}{\Delta_{ji}^2} \cdot D^{\lambda}_{k-1\to i} \cdot D^{\lambda+e_i}_{k\to j} \\
&= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) - \frac{1}{\Delta_{ji}^2} \cdot \Big(1 - \frac{1}{C_i^{\lambda} - C_k^{\lambda} + 1}\Big) \cdot D^{\lambda}_{k\to i} D^{\lambda+e_i}_{k\to j} && \text{(Theorem 13.3)} \\
&= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) + \frac{1}{1-\Delta_{ji}} \cdot \frac{\Delta_{ji}}{\Delta_{ji}^2} \cdot D^{\lambda}_{k\to ij}.
\end{aligned}$$

Since $i < k$, k and $k+1$ are in the same block of $\lambda + e_i$, which simplifies the right-hand side to

$$\begin{aligned}
(128) &= \big(1 + \frac{1}{\Delta_{ji}}\big) \cdot (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) \\
&= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) + \frac{1}{\Delta_{ji}} \cdot (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot D^{\lambda+e_i}_{k\to j} \\
&= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) + \frac{1}{\Delta_{ji}} \cdot \frac{1}{C_i^{\lambda} - C_k^{\lambda} + 1} \cdot D^{\lambda}_{k\to i} D^{\lambda+e_i}_{k\to j} && \text{(Theorem 13.3)} \\
&= (D^{\lambda}_{k\to i} - D^{\lambda}_{k-1\to i}) \cdot (D^{\lambda+e_i}_{k\to j} - D^{\lambda+e_i}_{k-1\to j}) + \frac{1}{\Delta_{ji}} \cdot \frac{1}{1-\Delta_{ji}} \cdot D^{\lambda}_{k\to ij}.
\end{aligned}$$

Hence, the two sides are equal. □

15 Large-d expansion of the second moment

Recall our expression for the second moment from Theorem 1.4:

$$\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] = \frac{n-1}{n} \cdot \rho \otimes \rho + \frac{1}{n} \cdot (I \otimes \rho + \rho \otimes I) \cdot \mathrm{SWAP} + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \cdot \mathrm{SWAP} - \mathrm{Lower}_{\rho}.$$

We have previously argued that Lower_{ρ} can be written as a positive linear combination of terms of the form $(P \otimes P) \cdot \mathrm{SWAP}$, for some PSD matrices P. This nice structure allowed us to drop the contribution of Lower_{ρ} in our applications.

In this section, we argue that Lower_ρ consists of terms which are *lower-order* in d. That is, for fixed n and $\lambda \vdash n$, we have

$$\lim_{d\to\infty} \mathrm{Lower}_\rho = 0.$$

In particular, this allows one to obtain a clean expression for $\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}]$, by embedding the input state copies $\rho^{\otimes n}$ from a d-dimensional Hilbert space to one of dimension $D \gg d$, such that the lower-order terms vanish:

$$\mathbf{E}[\widehat{\boldsymbol{\rho}} \otimes \widehat{\boldsymbol{\rho}}] \approx \frac{n-1}{n} \cdot \rho \otimes \rho + \frac{1}{n} \cdot (I \otimes \rho + \rho \otimes I) \cdot \mathrm{SWAP} + \frac{\mathbf{E}[\ell(\boldsymbol{\lambda})]}{n^2} \cdot \mathrm{SWAP}. \tag{130}$$

We show that Lower_ρ only consists of lower-order terms in d by computing the constant-in-d terms of $M^{(2)}_{\mathrm{avg}}$, and demonstrating that they correspond exactly to the terms in Equation (130). The remaining terms are subleading, and must be Lower_ρ.

Following the proof of Theorem 14.4, we can write $M^{(2)}_{\mathrm{avg}}$ as a block-diagonal matrix in the Schur basis, with each block $[M^{(2)}_{\mathrm{avg}}]^S_{\{i,j\}}$ corresponding to an SYT and an unordered pair of indices $S, \{i,j\}$,

$$M^{(2)}_{\mathrm{avg}} = \sum_{\lambda\vdash n} \sum_{i\le j} |\lambda_{ij}\rangle\langle\lambda_{ij}| \otimes \sum_{S\in\lambda} |S,\{i,j\}\rangle\langle S,\{i,j\}| \otimes (M^{(2)}_{\mathrm{avg}})^S_{\{i,j\}} \otimes I_{\dim(V^d_{\lambda+e_i+e_j})},$$

with diagonal entry

$$[M^{(2)}_{\mathrm{avg}}]^S_{ij,ij} = \frac{1}{n^2} \cdot \frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} \sum_{k,\ell} \lambda^\uparrow_k \lambda^\uparrow_\ell \cdot |c^\lambda_{k\ell\to ij}|^2.$$

We use the following formula for the dimension of the space V^d_λ (Equation (16)):

$$\dim(V^d_\lambda) = \prod_{\square\in\lambda} \frac{d + \mathrm{cont}_\lambda(\square)}{h_\lambda(\square)},$$

hence

$$\begin{aligned}
\frac{\dim(V^d_\lambda)}{\dim(V^d_{\lambda_{ij}})} &= \prod_{\square\in\lambda} \frac{h_{\lambda_{ij}}(\square)}{h_\lambda(\square)} \cdot \frac{d+\mathrm{cont}_\lambda(\square)}{d+\mathrm{cont}_{\lambda_{ij}}(\square)} \cdot \prod_{\square\in\lambda_{ij}\setminus\lambda} \frac{h_{\lambda_{ij}}(\square)}{d+\mathrm{cont}_{\lambda_{ij}}(\square)} \\
&= \prod_{\square\in\lambda} \frac{h_{\lambda_{ij}}(\square)}{h_\lambda(\square)} \cdot \prod_{\square\in\lambda_{ij}\setminus\lambda} \frac{h_{\lambda_{ij}}(\square)}{d+\mathrm{cont}_{\lambda_{ij}}(\square)} \\
&= O_n(1) \cdot \frac{O(1)}{\prod_{\square\in\lambda_{ij}\setminus\lambda}(d+\mathrm{cont}_{\lambda_{ij}}(\square))} \\
&= \frac{O_n(1)}{\prod_{\square\in\lambda_{ij}\setminus\lambda}(d+\mathrm{cont}_{\lambda_{ij}}(\square))}.
\end{aligned}$$

where $O_n(1)$ is a factor that only depends on n, and not on d. Moreover, we observe that the contents of the two new boxes on rows i and j, $\{\mathrm{cont}_{\lambda_{ij}}(\square)\}_{\square\in\lambda_{ij}\setminus\lambda}$ have magnitude at most n, and thus the denominator is $d^2 + O(d)$. Thus, the only constant-in-d terms that survive in each $(S,\{i,j\})$ block are the $\Theta(d^2)$ terms in the numerator

$$\begin{aligned}
&\sum_{k,\ell} \lambda^\uparrow_k \lambda^\uparrow_\ell \cdot |c^\lambda_{k\ell\to ij}|^2 \\
&= \sum_{k,\ell=1}^{\ell(\lambda)} \lambda^\uparrow_k \lambda^\uparrow_\ell \cdot |c^\lambda_{k\ell\to ij}|^2 + \sum_{k=1}^{\ell(\lambda)} \sum_{\ell=\ell(\lambda)+1}^{d} \lambda^\uparrow_k \lambda^\uparrow_\ell \cdot |c^\lambda_{k\ell\to ij}|^2 + \sum_{k=\ell(\lambda)+1}^{d} \sum_{\ell=1}^{\ell(\lambda)} \lambda^\uparrow_k \lambda^\uparrow_\ell \cdot |c^\lambda_{k\ell\to ij}|^2 + \sum_{k,\ell=\ell(\lambda)+1}^{d} \lambda^\uparrow_k \lambda^\uparrow_\ell \cdot |c^\lambda_{k\ell\to ij}|^2.
\end{aligned}$$

In the above equation, we broke down the summation into the non-zero and zero rows of λ, since $\lambda_i = 0$ if and only if $i > \ell(\lambda)$. Since the Clebsch-Gordan coefficients capture overlaps between unit vectors, they have magnitude at most 1. On the other hand, we can write

$$\lambda^\uparrow_i = \begin{cases} \lambda_i - \sum_{\lambda_j>\lambda_i} 1 + \sum_{\lambda_i>\lambda_j>0} 1 + d - \ell(\lambda), & \lambda_i > 0 \\ -\ell(\lambda), & \lambda_i = 0 \end{cases}.$$

In the first case, only the fourth term depends on d, whereas the remaining terms have magnitude at most n. Moreover, there are at most n rows with $\lambda_i > 0$. On the other hand, there are $d - \ell(\lambda)$ rows with zero length. Hence, the $\Theta(d^2)$ terms of the numerator are included in the following expression:

$$d^2 \cdot \Big(\sum_{k,\ell=1}^{\ell(\lambda)} |c^{\lambda}_{k\ell\to ij}|^2 \Big) - d\ell(\lambda) \cdot \Big(\sum_{k=1}^{\ell(\lambda)} \sum_{\ell=\ell(\lambda)+1}^{d} |c^{\lambda}_{k\ell\to ij}|^2 + \sum_{k=\ell(\lambda)+1}^{d} \sum_{\ell=1}^{\ell(\lambda)} |c^{\lambda}_{k\ell\to ij}|^2 \Big) + \ell(\lambda)^2 \cdot \Big(\sum_{k,\ell=\ell(\lambda)+1}^{d} |c^{\lambda}_{k\ell\to ij}|^2 \Big).$$

Using the formulas for the partial sums in Theorem 14.10, we can verify that the expression above is equal to $d^2(X_{n+1}X_{n+2} + \ell(\lambda)/\Delta_{ji}) + O(d)$, which gives rise to the terms in Equation (130).

References

[Aar18] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing*, pages 325–338, 2018.

[ARS88] Robert Alicki, Sławomir Rudnicki, and Sławomir Sadowski. Symmetry properties of product states for the system of N n-level atoms. *Journal of mathematical physics*, 29(5):1158–1162, 1988.

[ATD19] Francesco Albarelli, Mankei Tsang, and Animesh Datta. Upper bounds on the holevo cram\'er-rao bound for multiparameter quantum parametric and semiparametric estimation. Technical report, arXiv:1911.11036, 2019.

[BC94] Samuel Braunstein and Carlton Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439, 1994.

[BCH05] Dave Bacon, Isaac Chuang, and Aram Harrow. The quantum Schur transform: I. efficient qudit circuits. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2005.

[Ber12] Sonya Berg. *A quantum algorithm for the quantum Schur-Weyl transform*. PhD thesis, University of California, Davis, 2012.

[BLM+23] Harry Buhrman, Noah Linden, Laura Mančinska, Ashley Montanaro, and Maris Ozols. Quantum majority vote. In *Proceedings of the 14th Innovations in Theoretical Computer Science*, 2023.

[BO24] Costin Bădescu and Ryan O'Donnell. Improved quantum data analysis. *TheoretiCS*, 3, 2024.

[BOW19] Costin Bădescu, Ryan O'Donnell, and John Wright. Quantum state certification. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing*, pages 503–514, 2019.

[BWB24] Adam Bene Watts and John Bostanci. Quantum event learning and gentle random measurements. In *Proceedings of the 15th Innovations in Theoretical Computer Science*, pages 97–1, 2024.

[Can20] Clément Canonne. A short note on learning discrete distributions, 2020.

[CHL+23] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning? In *Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science*, pages 391–404, 2023.

[CL24] Sitan Chen and Jerry Li. Personal communication, 2024.

[CLL24a] Sitan Chen, Jerry Li, and Allen Liu. Optimal high-precision shadow estimation. In *27th Conference on Quantum Information Processing*, 2024.

[CLL24b] Sitan Chen, Jerry Li, and Allen Liu. An optimal tradeoff between entanglement and copy complexity for state tomography. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1331–1342, 2024.

[CM06] Matthias Christandl and Graeme Mitchison. The spectra of quantum states and the Kronecker coefficients of the symmetric group. *Communications in mathematical physics*, 261(3):789–797, 2006.

[CSDV19] Angelo Carollo, Bernardo Spagnolo, Alexander Dubkov, and Davide Valenti. On quantumness in multi-parameter quantum estimation. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(9):094010, 2019.

[CYZ25] Kean Chen, Nengkun Yu, and Zhicheng Zhang. Tight bound for quantum unitary time-reversal. Technical report, arXiv:2507.05736, 2025.

[DDGG20] Rafał Demkowicz-Dobrzański, Wojciech Górecki, and Mădălin Guţă. Multi-parameter estimation beyond quantum fisher information. *Journal of Physics A: Mathematical and Theoretical*, 53(36):363001, 2020.

[Fis05] Steve Fisk. A very short proof of Cauchy's interlace theorem for eigenvalues of Hermitian matrices. *Technical report, arXiv:0502408*, 2005.

[FO24] Steven Flammia and Ryan O'Donnell. Quantum chi-squared tomography and mutual information testing. *Quantum*, 8:1381, 2024.

[GBO24] Dmitry Grinko, Adam Burchardt, and Maris Ozols. Gelfand-Tsetlin basis for partially transposed permutations, with applications to quantum information. *27th Conference on Quantum Information Processing*, 2024.

[GKKT20] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020.

[GLM24] Daniel Grier, Sihan Liu, and Gaurav Mahajan. Improved classical shadows from local symmetries in the Schur basis. Technical report, arXiv:2405.09525, 2024.

[GPS24] Daniel Grier, Hakop Pashayan, and Luke Schaeffer. Sample-optimal classical shadows for pure states. *Quantum*, 8:1373, 2024.

[GW09] Roe Goodman and Nolan Wallach. *Symmetry, representations, and invariants*. Springer, 2009.

[Har05] Aram Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, Massachusetts Institute of Technology, 2005.

[Har16] Aram Harrow. Personal communication, 2016.

[Hay98] Masahito Hayashi. Asymptotic estimation theory for a finite-dimensional pure state model. *Journal of Physics A: Mathematical and General*, 31(20):4633, 1998.

[Hel67] Carl Helstrom. Minimum mean-squared error of estimates in quantum statistics. *Physics letters A*, 25(2):101–102, 1967.

[HHJ+16] Jeongwan Haah, Aram Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, August 2016. Preprint.

[HJW18] Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Local moment matching: a unified methodology for symmetric functional estimation and distribution estimation under Wasserstein distance. In *Proceedings of the 31st Annual Conference on Learning Theory*, pages 3189–3221, 2018.

[HKOT23] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. In *Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science*, pages 363–390, 2023.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[HM02] Masahito Hayashi and Keiji Matsumoto. Quantum universal variable-length source coding. *Physical Review A*, 66(2):022311, 2002.

[Hol11] Alexander Holevo. *Probabilistic and statistical aspects of quantum theory*. Springer Science & Business Media, 2011.

[HW23] Jonas Helsen and Michael Walter. Thrifty shadow estimation: Reusing quantum circuits and bounding tails. *Physical Review Letters*, 131(24):240602, 2023.

[Key06] Michael Keyl. Quantum state estimation and large deviations. *Reviews in Mathematical Physics*, 18(01):19–60, 2006.

[KG09] Jonas Kahn and Mădălin Guţă. Local asymptotic normality for finite dimensional quantum systems. *Communications in Mathematical Physics*, 289(2):597–652, 2009.

[KW01] Michael Keyl and Reinhard Werner. Estimating the spectrum of a density operator. *Physical Review A*, 64(5):052311, 2001.

[LFIC24] Zhaoyi Li, Honghao Fu, Takuya Isogawa, and Isaac Chuang. Optimal quantum purity amplification. Technical report, arXiv:2409.18167, 2024.

[Mat02] Keiji Matsumoto. A new approach to the Cramér-Rao-type bound of the pure-state model. *Journal of Physics A: Mathematical and General*, 35(13):3111, 2002.

[NC10] Michael Nielsen and Isaac Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2010.

[Ngu24] Quynh Nguyen. The mixed Schur transform: efficient quantum circuit and applications. In *27th Conference on Quantum Information Processing*, 2024.

[OW15a] Ryan O'Donnell and John Wright. A note on the Haah et al. tomography algorithm, 2015. https://people.eecs.berkeley.edu/~jswright/papers/tomography-note.pdf.

[OW15b] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, 2015.

[OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, 2016.

[OW17] Ryan O'Donnell and John Wright. Efficient quantum tomography II. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, 2017.

[OW25] Ryan O'Donnell and Chirag Wadhwa. Instance-optimal quantum state certification with entangled measurements. Technical report, arXiv:2507.06010, 2025.

[Pil90] Shaiy Pilpel. Descending subsequences of random permutations. *Journal of Combinatorial Theory, Series A*, 53(1):96–116, 1990.

[PTTW25] Angelos Pelecanos, Xinyu Tan, Ewin Tang, and John Wright. Beating full state tomography for unentangled spectrum estimation. *Technical report, arXiv:2504.02785*, 2025.

[Sag01] Bruce E Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*. Springer, 2001.

[SSW25] Thilo Scharnhorst, Jack Spilecki, and John Wright. Optimal lower bounds for quantum state tomography. Manuscript, 2025.

[Sta99] Richard P Stanley. *Enumerative combinatorics Volume 2*. Cambridge University Press, Cambridge, 1999.

[VK85] Anatoly Vershik and Sergei Kerov. Asymptotic of the largest and the typical dimensions of irreducible representations of a symmetric group. *Functional Analysis and its Applications*, 19(1):21–31, 1985.

[VK92] Naum Vilenkin and Anatoli Klimy. *Representation of Lie Groups and Special Functions. Volume 3: Classical and Quantum Groups and Special Functions*. Springer, 1992.

[VN96] Vasilii Voinov and Mikhail Nikulin. *Unbiased estimators and their applications: volume 2: multivariate case*. Springer Science & Business Media, 1996.

[VN12] Vasilii Voinov and Mikhail Nikulin. *Unbiased estimators and their applications: volume 1: univariate case*. Springer Science & Business Media, 2012.

[VO05] Anatoli Vershik and Andrei Okounkov. A new approach to the representation theory of the symmetric groups. II. *Journal of Mathematical Sciences*, 131:5471–5494, 2005.

[VV11] Gregory Valiant and Paul Valiant. Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pages 685–694, 2011.

[VV17] Gregory Valiant and Paul Valiant. Estimating the unseen: improved estimators for entropy and other properties. *Journal of the ACM*, 64(6):1–41, 2017.

[Wat18] John Watrous. *The theory of quantum information*. Cambridge University Press, 2018.

[Win02] Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on information theory*, 45(7):2481–2485, 2002.

[Wri16] John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016.

[Wri24a] John Wright. Lecture 10 from CS 294: Quantum learning theory. Found at https://people.eecs.berkeley.edu/~jswright/quantumlearningtheory24/scribe%20notes/lecture10.pdf, 2024.

[Wri24b] John Wright. Lecture 12 from CS 294: Quantum learning theory. Found at https://people.eecs.berkeley.edu/~jswright/quantumlearningtheory24/scribe%20notes/lecture12.pdf, 2024.

[YCH19] Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi. Attaining the ultimate precision limit in quantum state estimation. *Communications in Mathematical Physics*, 368(1):223–293, 2019.

[YFG13] Koichi Yamagata, Akio Fujiwara, and Richard Gill. Quantum local asymptotic normality based on a new quantum likelihood ratio. *The Annals of Statistics*, 41(4):2197–2217, 2013.

[Yue23] Henry Yuen. An improved sample complexity lower bound for (fidelity) quantum state tomography. *Quantum*, 7:890, 2023.

[ZC25] Sisi Zhou and Senrui Chen. Randomized measurements for multi-parameter quantum metrology. Technical report, arXiv:2502.03536, 2025.

Part V
Appendix

A Obtaining Young's orthogonal basis

The purpose of this section is to show the following theorem.

Theorem A.1 (Theorem 11.3, restated)**.** *The unitary in Theorem 11.2 is a Schur transform. That is, it satisfies*

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \left(\mathcal{P}^{(n)}(\pi)\mathcal{Q}^{(n)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} = \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\pi) \otimes \nu_\lambda(U), \tag{131}$$

for all $\pi \in S_n$ and $U \in U(d)$.

We also restate the recursive construction of the Schur transform, for the reader's convenience.

Definition A.2 (Schur transform; Theorem 11.2, restated)**.** We define the *Schur transform*, $\mathcal{U}_{\mathrm{SW}}^{(n)} : (\mathbb{C}^d)^{\otimes n} \to \bigoplus_{\lambda \vdash (n,d)} \mathrm{Sp}_\lambda \otimes V_\lambda^d$, by the following recursive construction. We explicitly define $\mathcal{U}_{\mathrm{SW}}^{(1)}$ as the unitary such that

$$\mathcal{U}_{\mathrm{SW}}^{(1)} |i\rangle \coloneqq |\ydiagram{1}\rangle \otimes |\boxed{1}\rangle \otimes |i\rangle,$$

for all $i \in [d]$. Then, for $n > 1$,

$$\mathcal{U}_{\mathrm{SW}}^{(n)} \coloneqq \mathcal{U}_{\mathrm{R}}^{(n)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n)} \cdot (\mathcal{U}_{\mathrm{SW}}^{(n-1)} \otimes I).$$

We start by proving the base case, $n = 1$.

Lemma A.3. *Theorem A.1 holds for $n = 1$.*

Proof. Note that for $n = 1$, $S_n = \{e\}$ so that we just need to check that unitaries are correctly transformed. Moreover, there is only one Young diagram of size 1, $\lambda = \ydiagram{1}$, and one SYT of that shape, $S = \boxed{1}$. We have:

$$\mathcal{U}_{\mathrm{SW}}^{(1)} \cdot U \cdot \mathcal{U}_{\mathrm{SW}}^{(1)\dagger} = \sum_{i,j \in [d]} U_{ij} \cdot \left(\mathcal{U}_{\mathrm{SW}}^{(1)} \cdot |i\rangle\langle j| \cdot \mathcal{U}_{\mathrm{SW}}^{(1)\dagger}\right) = |\ydiagram{1}\rangle\langle\ydiagram{1}| \otimes |\boxed{1}\rangle\langle\boxed{1}| \otimes \sum_{i,j \in [d]} U_{ij} \cdot |i\rangle\langle j| = |\ydiagram{1}\rangle\langle\ydiagram{1}| \otimes |\boxed{1}\rangle\langle\boxed{1}| \otimes U.$$

Since $\nu_{(1)}(U) = U$, Theorem 11.3 holds for $n = 1$. □

We now prove the following partial result, which nevertheless will be a useful stepping stone towards the complete result.

Lemma A.4. *Suppose Theorem A.1 holds for some $n \in \mathbb{N}$. Then $\mathcal{U}_{\mathrm{SW}}^{(n+1)}$ transforms correctly, up to phase. More precisely:*

$$\mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \left(\mathcal{P}^{(n+1)}(\sigma) \cdot \mathcal{Q}^{(n+1,d)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} = \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \widetilde{\kappa}_\mu(\pi) \otimes \nu_\mu(U),$$

where $\widetilde{\kappa}_\mu$ is a representation of S_{n+1} isomorphic to κ_μ, such that $\widetilde{\kappa}_\mu{\downarrow_{S_n}} = \kappa_\mu{\downarrow_{S_n}}$, and $\widetilde{\kappa}_\mu = W_\mu \cdot \kappa_\mu \cdot W_\mu^\dagger$, for some diagonal unitary W_μ.

Proof. First, for all $\sigma \in S_n$ and $U \in U(d)$

$$\begin{aligned}(\mathcal{U}_{\mathrm{SW}}^{(n)} \otimes I) \cdot \left(\mathcal{P}^{(n+1)}(\sigma) \cdot \mathcal{Q}^{(n+1,d)}(U)\right) \cdot (\mathcal{U}_{\mathrm{SW}}^{(n)\dagger} \otimes I) &= \left(\mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \left(\mathcal{P}^{(n)}(\sigma) \cdot \mathcal{Q}^{(n,d)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger}\right) \otimes U \\ &= \left(\sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\sigma) \otimes \nu_\lambda(U)\right) \otimes U.\end{aligned}$$

Here, we have used that $\mathcal{P}^{(n+1)}(\sigma) = \mathcal{P}^{(n)}(\sigma) \otimes I$, and $\mathcal{Q}^{(n+1)}(U) = \mathcal{Q}^{(n)}(U) \otimes U$. We can then apply the Clebsch-Gordan transform, obtaining the following, where we have defined $\mathcal{V} \coloneqq \mathcal{U}_{\mathrm{CG}}^{(n+1)} \cdot (\mathcal{U}_{\mathrm{SW}}^{(n)} \otimes I)$:

$$\begin{aligned}\mathcal{V} \cdot \left(\mathcal{P}^{(n+1)}(\sigma) \cdot \mathcal{Q}^{(n+1,d)}(U)\right) \cdot \mathcal{V}^\dagger &= \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\sigma) \otimes \left(\mathcal{U}_{\mathrm{CG}}^{(\lambda)} \cdot (\nu_\lambda(U) \otimes U) \cdot \mathcal{U}_{\mathrm{CG}}^{(\lambda)\dagger}\right) \\ &= \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq d}} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\sigma) \otimes \left(\sum_{\substack{\lambda \nearrow \mu \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \nu_\mu(U)\right).\end{aligned} \tag{132}$$

Finally, applying the rearranging unitary $\mathcal{U}_{\mathrm{R}}^{(n+1)}$, we have:

$$\mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \left(\mathcal{P}^{(n+1)}(\sigma) \cdot \mathcal{Q}^{(n+1,d)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} = \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \Big(\sum_{\lambda \nearrow \mu} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\sigma) \Big) \otimes \nu_\mu(U). \tag{133}$$

Recall that for $\sigma \in S_n$, we have $\kappa_\mu(\sigma) = \sum_{\lambda \nearrow \mu} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\sigma)$, the symmetric group branching rule and the construction of the Young-Yamanouchi basis (see Section 2.4.3). Thus, $\mathcal{U}_{\mathrm{SW}}^{(n+1)}$ correctly implements the Schur transform for $\sigma \in S_n$. What about $\pi \in S_{n+1}$, more generally? Firstly, we claim

$$\mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \left(\mathcal{P}^{(n+1)}(\pi) \cdot \mathcal{Q}^{(n+1,d)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} = \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes X_\mu(\pi) \otimes \nu_\mu(U), \tag{134}$$

for some matrices $\{X_\mu(\pi)\}$. To see this, note that because $\mathcal{P}^{(n+1)}(\pi)$ and $\mathcal{Q}^{(n+1,d)}(U)$ commute for all $U \in U(d)$, the matrix $\mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \mathcal{P}^{(n+1)}(\pi) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger}$ commutes with all matrices of the form

$$\mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \mathcal{Q}^{(n+1,d)}(U) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} = \mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \left(\mathcal{P}^{(n+1)}(e) \cdot \mathcal{Q}^{(n+1,d)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} = \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes I_{\dim(\mu)} \otimes \nu_\mu(U). \tag{135}$$

Since the $\{\nu_\mu\}$ are non-isomorphic, irreducible representations, we must have that [Sag01, Theorem 1.7.8, Item (2)]

$$\mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \mathcal{P}^{(n+1)}(\pi) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} = \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes X_\mu(\pi) \otimes I_{\dim(V_\mu^d)}, \tag{136}$$

for some matrices $\{X_\mu(\pi)\}$. However, these matrices form a representation of S_{n+1}, and by the uniqueness of the decomposition in Schur-Weyl duality, X_μ is isomorphic as a representation to κ_μ. We relabel $X_\mu \to \widetilde{\kappa}_\mu$. Combining Equations (135) and (136) gives us Equation (134).

By Theorem 2.17, for each μ there exists a fixed unitary W_μ such that $W_\mu \cdot \widetilde{\kappa}_\mu(\pi) \cdot W_\mu^\dagger = \kappa_\mu(\pi)$ for all $\pi \in S_{n+1}$. Restricting to $\sigma \in S_n$, we obtain

$$W_\mu \cdot \Big(\sum_{\lambda \nearrow \mu} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\sigma) \Big) \cdot W_\mu^\dagger = \sum_{\lambda \nearrow \mu} |\lambda\rangle\langle\lambda| \otimes \kappa_\lambda(\sigma).$$

Since the $\{\kappa_\lambda\}$ are non-isomorphic and irreducible, W_μ, lying in the commutant of the sum, is necessarily constant on each λ-subspace (as in Equation (136), except now the multiplicity space is one-dimensional). Each constant must be a phase, since each λ-subspace is orthogonal, and W_μ is unitary. That is, W_μ is of the form

$$W_\mu = \sum_{\lambda \nearrow \mu} e^{i\theta_{\lambda\mu}} \cdot |\lambda\rangle\langle\lambda| \otimes I_{\dim(\lambda)},$$

for some angles $\{\theta_{\lambda\mu}\}_{\lambda \nearrow \mu}$. Therefore,

$$\mathcal{U}_{\mathrm{SW}}^{(n+1)} \cdot \left(\mathcal{P}^{(n+1)}(\pi) \cdot \mathcal{Q}^{(n+1,d)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+1)\dagger} = \sum_{\substack{\mu \vdash n+1 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \widetilde{\kappa}_\mu(\pi) \otimes \nu_\mu(U),$$

with $\widetilde{\kappa}_\mu$ isomorphic to κ_μ, such that $\widetilde{\kappa}_\mu{\downarrow_{S_n}} = \kappa_\mu{\downarrow_{S_n}}$, and $\widetilde{\kappa}_\mu = W_\mu \cdot \kappa_\mu \cdot W_\mu^\dagger$, for some diagonal unitary W_μ. □

This partial result implies the $n = 2$ case as well.

Corollary A.5. *Theorem A.1 holds for $n = 2$.*

Proof. By Theorem A.3 and Theorem A.4, we know that for all $\pi \in S_2$ and $U \in U(d)$:

$$\mathcal{U}_{\mathrm{SW}}^{(2)} \cdot \left(\mathcal{P}^{(2)}(\pi) \cdot \mathcal{Q}^{(2)}(U)\right) \cdot \mathcal{U}_{\mathrm{SW}}^{(2)\dagger} = \sum_{\substack{\mu \vdash 2 \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \widetilde{\kappa}(\pi) \otimes \nu_\mu(U),$$

where $\widetilde{\kappa}_\mu(\pi) = W_\mu \cdot \kappa_\mu(\pi) \cdot W_\mu^\dagger$, for some unitaries W_μ. However, for $|\mu| = 2$, we either have $\mu = (2)$ or $\mu = (1,1)$. In either case, $\dim(\mu) = 1$, as there is a unique SSYT for any Young diagram with a single row, or with a single column. Since all matrices of size 1 commute:

$$\widetilde{\kappa}_\mu(\pi) = W_\mu \cdot \kappa_\mu(\pi) \cdot W_\mu^\dagger = \kappa_\mu(\pi). \qquad \square$$

We are now ready to prove Theorem 11.3.

Proof of Theorem 11.3. The proof will follow by induction on the number of qudits n. In particular, we show that if Theorem A.1 holds for n and $n+1$ qudits, then it also holds for $n+2$. Given the base cases of $n = 1$ (Theorem A.3) and $n = 2$ (Theorem A.5), this will show the result for all $n \geq 1$.

Given the inductive hypothesis, Theorem A.4, states that for $\mu \vdash n+2$,

$$\widetilde{\kappa}_\mu{\downarrow_{S_{n+1}}} = \kappa_\mu{\downarrow_{S_{n+1}}}.$$

This implies that for any permutation $\pi \in S_{n+2}$ that satisfies $\pi(n+2) = n+2$, it holds that $\widetilde{\kappa}_\mu(\pi) = \kappa_\mu(\pi)$. Thus, since the set of permutations $\{(n+1, n+2)\} \cup \{\sigma \in S_{n+2} \mid \sigma(n+2) = n+2\}$ generates the entirety of S_{n+2}, it suffices to prove that

$$\widetilde{\kappa}_\mu(n+1, n+2) = \kappa_\mu(n+1, n+2),$$

to conclude that $\widetilde{\kappa}_\mu(\pi) = \kappa_\mu(\pi)$ for all π.

We restrict our attention to the permutation $(n+1, n+2)$. Let $\lambda \vdash n$ be a Young diagram. We know that both $\kappa_{\lambda_{ij}}(n+1, n+2)$ and $\widetilde{\kappa}_{\lambda_{ij}}(n+1, n+2)$ have the same block-diagonal structure in Young's orthogonal basis. In particular, each matrix has a 2×2 submatrix whenever (S_{ij}, S_{ji}) are both valid, and a 1×1 submatrix whenever only one is valid. Here, we use S_{ij} to denote the SYT we obtain when we start from a valid SYT S with n boxes, and add an $n+1$ at the end of row i, and an $n+2$ at the end of row j.

Theorem A.4 states that

$$\widetilde{\kappa}_{\lambda_{ij}}(n+1, n+2) = W_{\lambda_{ij}} \cdot \kappa_{\lambda_{ij}}(n+1, n+2) \cdot W_{\lambda_{ij}}^\dagger,$$

for $W_{\lambda_{ij}}^\dagger$ a diagonal unitary. Since matrices of size 1 commute, we can deduce that $\widetilde{\kappa}_{\lambda_{ij}}(n+1, n+2)$ and $\kappa_{\lambda_{ij}}(n+1, n+2)$ have the same 1×1 submatrices in that basis.

We now focus on a 2×2 submatrix indexed by the SYTs S_{ij} and S_{ji}. Let $e^{i\theta_1}, e^{i\theta_2}$ be the entries of $W_{\lambda_{ij}}^\dagger$ on $|S_{ij}\rangle\langle S_{ij}|$ and $|S_{ji}\rangle\langle S_{ji}|$ respectively. Then $\widetilde{\kappa}_{\lambda_{ij}}$ and $\kappa_{\lambda_{ij}}$ agree on the diagonal entries of the submatrix, since

$$\langle S_{ij}| \widetilde{\kappa}_{\lambda_{ij}}(n+1, n+2) |S_{ij}\rangle = e^{i\theta_1} \cdot \langle S_{ij}| \kappa_{\lambda_{ij}}(n+1, n+2) |S_{ij}\rangle \cdot e^{-i\theta_1} = \langle S_{ij}| \kappa_{\lambda_{ij}}(n+1, n+2) |S_{ij}\rangle,$$

and similarly for S_{ji}. On the other hand, the off-diagonal entries satisfy

$$\begin{aligned} \langle S_{ij}| \widetilde{\kappa}_{\lambda_{ij}}(n+1, n+2) |S_{ji}\rangle &= e^{i\theta_1} \cdot \langle S_{ij}| \kappa_{\lambda_{ij}}(n+1, n+2) |S_{ji}\rangle \cdot e^{-i\theta_2} \\ &= e^{i(\theta_1 - \theta_2)} \cdot \langle S_{ij}| \kappa_{\lambda_{ij}}(n+1, n+2) |S_{ji}\rangle. \end{aligned}$$

Since the off-diagonal entries of $\kappa_{\lambda_{ij}}(n+1, n+2)$ are given by the expression[3]

$$\langle S_{ij}| \kappa_{\lambda_{ij}}(n+1, n+2) |S_{ji}\rangle = \sqrt{1 - \frac{1}{\Delta_{ji}^2}},$$

it suffices to show that $\langle S_{ij}| \widetilde{\kappa}_{\lambda_{ij}}(n+1, n+2) |S_{ji}\rangle$ is real and positive for all pairs of SYTs of the form (S_{ij}, S_{ji}).

[3] Recall that Δ_{ji} is the difference in contents of the boxes with $n+2$ and $n+1$ in S_{ij}.

The expression $\langle S_{ij}| \widetilde{\kappa}_{\lambda_{ij}}(n+1,n+2) |S_{ji}\rangle$ can be written to involve the SWAP $= \mathcal{P}(n+1,n+2)$ operator on qudits $n+1$ and $n+2$, by noting that

$$\langle S_{ij}| \widetilde{\kappa}_{\lambda_{ij}}(n+1,n+2) |S_{ji}\rangle = \langle \lambda_{ij}, S_{ij}, T| \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot \mathcal{P}(n+1,n+2) \cdot \mathcal{U}_{\mathrm{SW}}^{(n+2)\dagger} \cdot |\lambda_{ij}, S_{ji}, T\rangle\,, \tag{137}$$

for any $T \in \mathrm{SSYT}(\lambda_{ij})$. The vectors on the left and right of $\mathcal{P}(n+1,n+2)$ are written in "our" Schur basis, which contains a Young-Yamanouchi basis in the permutation register that is a priori related by phases to Young's orthogonal basis. We will now "undo" our Schur transform to obtain the Schur basis on n qudits, which, by induction, we will assume is the true Schur basis. In particular, we follow the original steps from Section 11, in reverse. As each step is unitary, we can undo each step by applying the conjugate transpose of the relevant unitary. We have

$$\begin{aligned}
|\lambda_{ij}, S_{ji}, T\rangle &= \mathcal{U}_{\mathrm{R}}^{(n+2)} \cdot |\lambda + e_j, S_j, \lambda_{ij}, T\rangle \\
&= \mathcal{U}_{\mathrm{R}}^{(n+2)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n+2)} \cdot \Big(\sum_k \sum_{T' \in \lambda + e_j} \langle T', k | T\rangle \cdot |\lambda + e_j, S_j, T'\rangle \otimes |k\rangle \Big) \\
&= \mathcal{U}_{\mathrm{R}}^{(n+2)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n+2)} \cdot \mathcal{U}_{\mathrm{R}}^{(n+1)} \cdot \Big(\sum_k \sum_{T' \in \lambda + e_j} \langle T', k | T\rangle \cdot |\lambda, S, \lambda + e_j, T'\rangle \otimes |k\rangle \Big) \\
&= \mathcal{U}_{\mathrm{R}}^{(n+2)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n+2)} \cdot \mathcal{U}_{\mathrm{R}}^{(n+1)} \cdot \mathcal{U}_{\mathrm{CG}}^{(n+1)} \cdot \Big(\sum_{k,\ell} \sum_{T' \in \lambda + e_j} \sum_{T \in \lambda} \langle T', k | T\rangle \cdot \langle T'', \ell | T'\rangle \cdot |\lambda, S, T''\rangle \otimes |\ell\rangle \otimes |k\rangle \Big) \\
&= \mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot (\mathcal{U}_{\mathrm{SW}}^{(n)\dagger} \otimes I \otimes I) \cdot \Big(\sum_{k,\ell} \sum_{T' \in \lambda + e_j} \sum_{T \in \lambda} \langle T', k | T\rangle \cdot \langle T'', \ell | T'\rangle \cdot |\lambda, S, T''\rangle \otimes |\ell\rangle \otimes |k\rangle \Big).
\end{aligned}$$

Throughout this section, we use $T \in \lambda$ to denote in a succinct way that the SSYT T has shape λ. Similarly, we can write

$$|\lambda_{ij}, S_{ij}, T\rangle = \mathcal{U}_{\mathrm{SW}}^{(n+2)} \cdot (\mathcal{U}_{\mathrm{SW}}^{(n)\dagger} \otimes I \otimes I) \cdot \Big(\sum_{k',\ell'} \sum_{Y' \in \lambda + e_i} \sum_{Y'' \in \lambda} \langle Y', \ell' | T\rangle \cdot \langle Y'', k' | Y'\rangle \cdot |\lambda, S, Y''\rangle \otimes |k'\rangle \otimes |\ell'\rangle \Big).$$

Then the off-diagonal entry that we are trying to compute is equal to

$$\begin{aligned}
&\sum_{k,\ell,k',\ell'} \sum_{\substack{T'' \in \lambda \\ T' \in \lambda + e_j}} \sum_{\substack{Y'' \in \lambda \\ Y' \in \lambda + e_i}} \langle Y'', k' | Y'\rangle \, \langle Y', \ell' | T\rangle \, \langle T'', \ell | T'\rangle \, \langle T', k | T\rangle \cdot \langle \lambda, S, T'', \ell, k| \cdot \mathcal{U}_{\mathrm{SW}}^{(n)} \cdot \mathrm{SWAP} \cdot \mathcal{U}_{\mathrm{SW}}^{(n)\dagger} \cdot |\lambda, S, Y'', k', \ell'\rangle \\
&= \sum_{k,\ell,k',\ell'} \sum_{\substack{T'' \in \lambda \\ T' \in \lambda + e_j}} \sum_{\substack{Y'' \in \lambda \\ Y' \in \lambda + e_i}} \langle Y'', k' | Y'\rangle \, \langle Y', \ell' | T\rangle \, \langle T'', \ell | T'\rangle \, \langle T', k | T\rangle \cdot \langle \lambda, S, T'', \ell, k | \lambda, S, Y'', \ell', k'\rangle \\
&= \sum_{k,\ell} \sum_{\substack{T'' \in \lambda \\ T' \in \lambda + e_j}} \sum_{\substack{Y'' \in \lambda \\ Y' \in \lambda + e_i}} \langle Y'', k | Y'\rangle \, \langle Y', \ell | T\rangle \, \langle T'', \ell | T'\rangle \, \langle T', k | T\rangle \cdot \langle T'' | Y''\rangle\,,
\end{aligned} \tag{138}$$

where the last equality follows because each term is zero unless $(\ell, k) = (\ell', k')$. Since Equation (137) holds for any SSYT T, we will choose T to make the above summation easier to compute (its sign, at least). In particular, the one-step and two-step Clebsch-Gordan coefficients that we have developed so far gave us a solid understanding of how a highest-weight SSYT changes when we tensor onto it one or two elements of the fundamental representation $V_{(1)}^d$, like $|k\rangle$ and $|\ell\rangle$. Here, we are instead starting from an SSYT T, and want to understand which SSYTs T'' and Y'' can take us to T when we tensor on $|k\rangle$'s and $|\ell\rangle$'s. Pictorially, the SSYTs in Equation (138) satisfy the following:

$$\underbrace{T''}_{\lambda} \xrightarrow{\otimes|\ell\rangle} \underbrace{T'}_{\lambda + e_j} \xrightarrow{\otimes|k\rangle} \underbrace{T}_{\lambda_{ij}} \xleftarrow{\otimes|\ell\rangle} \underbrace{Y'}_{\lambda + e_i} \xleftarrow{\otimes|k\rangle} \underbrace{Y''}_{\lambda}$$

Since T'' and Y'' are not necessarily highest-weight SSYTs, it feels that our work from the previous section is not helpful here. Fortunately, we can relate the Clebsch-Gordan coefficients that arise in Equation (137) to the one-step and two-step Clebsch-Gordan coefficients, by setting T to be the *lowest*-weight SSYT of shape λ_{ij}, and then considering the *complement* tableaux of T, T', T'', and Y', Y''. We introduce these notions below.

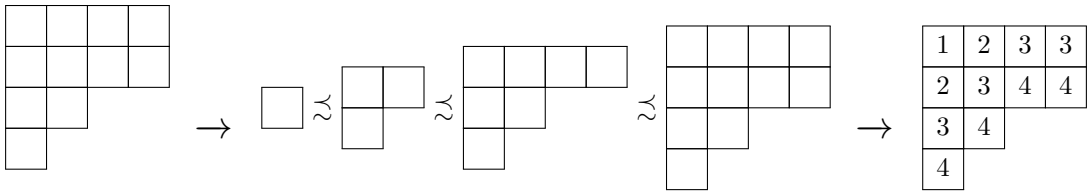


Figure 24: An example illustrating lowest-weight SSYTs. Take $d = 4$. On the left, we have the Young diagram $\lambda = (4, 4, 2, 1)$. In the middle, we have the sequence $T_\lambda^{\leq p}$, for $p \in [d]$. Note that $T_\lambda^{\leq p}$ consists of the lowest p rows of λ, shifted upwards by $(d - p)$. Thus, $T_\lambda^{\leq p}$ contains the cells of λ directly above at least $(d - p)$ boxes. On the right is T_λ, the lowest-weight SSYT, corresponding to the chain $T_\lambda^{\leq 1} \preceq \cdots \preceq T_\lambda^{\leq d}$. The boxes filled with symbol p are the boxes in λ that are above exactly $(d - p)$ boxes. This SSYT can also be obtained by filling the lowest box in each column with d, then the lowest remaining box in each column with $(d - 1)$, and so on.

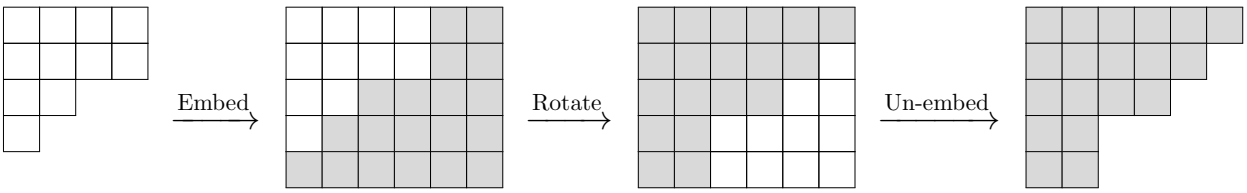


Figure 25: An example illustrating complementary Young diagrams. Take $d = 5$, and $m = 6$. On the left, we have the Young diagram $\lambda = (4, 4, 2, 1, 0)$. On the right, we have the Young diagram $\mu = \text{compl}_{(d,m)}(\lambda) = (6, 5, 4, 2, 2)$. In between, we illustrate how we can obtain μ from λ by first "embedding" λ into a $d \times m$ grid of boxes, then rotating the diagram 180°, and finally "un-embedding" the shaded boxes, which are the boxes that were not originally in λ.

Definition A.6 (Lowest-weight SSYTs)**.** Let $\lambda = (\lambda_1, \ldots, \lambda_d) \vdash n$ be a Young diagram. The *lowest weight SSYT* of shape λ and alphabet $[d]$ is denoted by T_λ and satisfies:

$$T_\lambda^{\leq p} = (\lambda_{d-p+1}, \ldots, \lambda_d).$$

The *lowest-weight vector* in V_λ^d is the vector $|T_\lambda\rangle$.

This formal definition can be used to show straightforwardly that T_λ is, in fact, an SSYT, using the interlacing property described in Section 2.4.4. Indeed, note that $(T_\lambda)_i^{\leq p} = \lambda_{d-p+i}$. Then $T_\lambda^{\leq p} \preceq T_\lambda^{\leq (p+1)}$ since $(T_\lambda)_{i+1}^{\leq (p+1)} \leq (T_\lambda)_i^{\leq p} \leq (T_\lambda)_i^{\leq (p+1)}$, which holds because $\lambda_{d-(p+1)+(i+1)} = \lambda_{d-p+i} \leq \lambda_{d-(p+1)+i}$, as the rows of λ are weakly decreasing. However, we can also give a more intuitive picture.

Note that the rows of $T_\lambda^{\leq p}$ are the same as the bottom p rows of λ. As illustrated in Figure 24, this means that a box in T_λ is filled with the symbol p when it is above exactly $(d - p)$ boxes. Thus, the lowest-weight SSYT has a simple, alternate characterization: T_λ is the SSYT of shape λ obtained by filling the lowest box in each column with the value d, and then repeating this procedure with the unfilled boxes using the values $d - 1, \ldots, 1$. Intuitively, T_λ is the SSYT of shape λ with the maximum number of d's, then the maximum number of $(d - 1)$'s, and so on.

Definition A.7 (Complementary Young diagrams)**.** Let $\lambda \in \mathbb{Z}^d$ be a Young diagram, and let m be an integer at least as large as the number of columns in λ. The Young diagram which *complements* λ with respect to m columns, $\mu \in \mathbb{Z}^d$, is the Young diagram such that

$$\mu_i = m - \lambda_{d+1-i}.$$

We write $\mu =: \text{compl}_{(d,m)}(\lambda)$, though we will drop (d, m) when these parameters are clear. We also say μ is the *complement* of λ, or μ is *complementary* to λ.

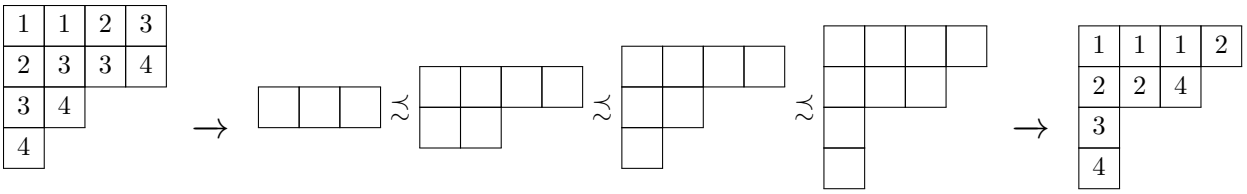


Figure 26: An example illustrating complementing SSYTs. Take $d = 4$ and $m = 5$. On the left, we have an SSYT S of shape $\lambda = (4, 4, 2, 1)$ and alphabet $[d]$. In the middle, we have the sequence $\mathrm{compl}_{(p,m)}(S^{\leq p})$, for $p \in [d]$. On the right, we have the SSYT $T = \mathrm{compl}_{(d,m)}(S)$ corresponding to this chain.

We note that $\mathrm{compl}(\lambda)$ is, indeed, a Young diagram, since $\mu_{i+1} = m - \lambda_{p-i} \leq m - \lambda_{p+1-i} = \mu_i$. The complementary Young diagram can be obtained by a three-step geometric process illustrated in Figure 25. It is also clear from the figure that $\mathrm{compl}_{(d,m)}(\mathrm{compl}_{(d,m)}(\lambda)) = \lambda$, so that complementation with fixed parameters is bijective.

For S an SSYT, we can consider the complements of the Young diagrams $S^{\leq p} \in \mathbb{Z}^p$, with respect to a common value of m, for $p = 1, \ldots, d$.

Lemma A.8. *Let S be an SSYT with alphabet $[d]$. Let m be an integer at least as large as the number of columns of S. For each $p \in [d]$, construct the complementary tableau $\mathrm{compl}_{(p,m)}(S^{\leq p}) \in \mathbb{Z}^p$. Then, for each $p \in [d-1]$, $\mathrm{compl}_{(p,m)}(S^{\leq p}) \precsim \mathrm{compl}_{(p+1,m)}(S^{\leq p+1})$.*

Proof. Let $A, B \in \mathbb{Z}^d$ be Young diagrams. Recall that $A \preceq B$ iff the row-lengths of B *interlace* the row-lengths of A: $A_{i+1} \leq B_{i+1} \leq A_i$ for all $i \in [d-1]$.

Since S is a SSYT, we have $S^{\leq p} \preceq S^{\leq (p+1)}$ for each $p \in [d-1]$. Thus, for each $p \in [d-1]$ and $j \in [p-1]$, we have $(S^{\leq p})_{j+1} \leq (S^{\leq (p+1)})_{j+1} \leq (S^{\leq p})_j$. Now, from $(S^{\leq p})_{j+1} \leq (S^{\leq (p+1)})_{j+1}$ with $j = p - i$, we have

$$\mathrm{compl}(S^{\leq (p+1)})_{i+1} = m - (S^{\leq (p+1)})_{p+1-i} \leq m - (S^{\leq (p)})_{p+1-i} = \mathrm{compl}(S^{\leq p})_i. \tag{139}$$

Moreover, from $(S^{\leq (p+1)})_{j+1} \leq (S^{\leq p})_j$ with $j = p + 1 - i$, we have

$$\mathrm{compl}(S^{\leq p})_i = m - (S^{\leq p})_{p+1-i} \leq m - (S^{\leq (p+1)})_{p+2-i} = \mathrm{compl}(S^{\leq (p+1)})_i. \tag{140}$$

Combining Equations (139) and (140) shows the interlacing condition:

$$\mathrm{compl}(S^{\leq p})_{i+1} \leq \mathrm{compl}(S^{\leq p})_i \leq \mathrm{compl}(S^{\leq (p+1)})_i.$$

We conclude that $\mathrm{compl}(S^{\leq p}) \precsim \mathrm{compl}(S^{\leq (p+1)})$. □

Thus, $\mathrm{compl}_{(1,m)}(S^{\leq 1}) \precsim \cdots \precsim \mathrm{compl}_{(d,m)}(S^{\leq d})$. This chain is associated with an SSYT with alphabet $[d]$, by the correspondence described in Section 2.4.4.

Corollary A.9. *Let S be an SSYT with alphabet $[d]$, and let m be an integer at least as large as the number of columns of S. Then, there exists a unique SSYT with alphabet $[d]$, T, such that, for each $p \in [d]$, $T^{\leq p} = \mathrm{compl}_{(p,m)}(S^{\leq p})$.*

Definition A.10 (Complementary SSYTs)**.** Let S be an SSYT, and let m be an integer larger than the number of columns in S. The SSYT which *complements* S to be the SSYT T that satisfies

$$T^{\leq p} = \mathrm{compl}_{(p,m)}(S^{\leq p}).$$

Thus, $T_i^{\leq p} = m - S_{p+1-i}^{\leq p}$. We write $T = \mathrm{compl}_{(d,m)}(S)$, dropping (d, m) when these parameters are clear from context. We also T is the *complement* of S, or T is *complementary* to S.

Note that if $T = \mathrm{compl}_{(d,m)}(S)$, then $\mathrm{compl}_{(p,m)}(T^{\leq p}) = S^{\leq p}$ for all $p \in [d]$, so that $S = \mathrm{compl}_{(d,m)}(T)$ as well. That is, $S = \mathrm{compl}(\mathrm{compl}(S))$. See Figure 26 for an example of complementary SSYTs.

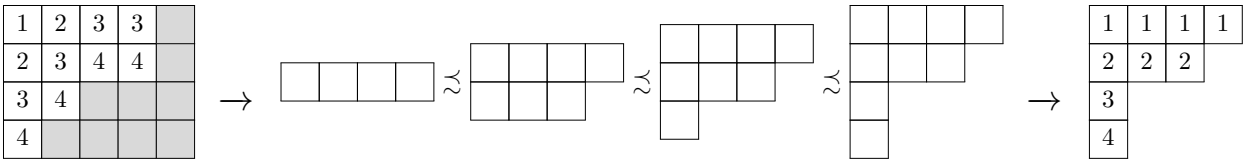


Figure 27: Another example of complementing SSYTs, here illustrating how the complements of lowest-weight SSYTs are highest-weight SSYTs. As in the previous figure, take $d = 4$, $m = 5$, and $\lambda = (4, 4, 2, 1)$. On the left, we have the lowest-weight SSYT S_λ of shape λ and alphabet $[d]$, which we have depicted as embedded into a $d \times m$ grid. The gray cells are those that do not belong to λ. In the middle, we have the sequence $\mathrm{compl}_{(p,m)}(S_\lambda^{\leq p})$, for $p \in [d]$. Note that $\mathrm{compl}_{(p,m)}(S_\lambda^{\leq p})$ is the complement of the bottom p rows, regarded as a Young diagram. Thus, $\mathrm{compl}_{(p,m)}(S_\lambda^{\leq p})$ consists of the gray cells in the bottom p rows of λ, rotated 180°. Therefore, the p-th element in the chain adds boxes only in the p-th row. This means that the SSYT corresponding to this chain, $T^\lambda = \mathrm{compl}_{(d,m)}(S)$ is highest-weight. On the right is T^λ, and indeed it is highest weight.

Lemma A.11. *Let $\lambda \in \mathbb{Z}^d$ be a Young diagram, and let m be an integer at least as large as the number of columns of λ. Let $\mu \coloneqq \mathrm{compl}_{(d,m)}(\lambda)$. Let S_λ be the lowest-weight SSYT of shape λ, and let T^μ be the highest-weight SSYT of shape μ, both with alphabet $[d]$. Then S_λ and T^μ are complementary SSYTs.*

It is perhaps easiest to understand this lemma pictorially; see Figure 27. We now give a formal argument.

Proof. Recall $(S_\lambda^{\leq p})_i = \lambda_{d-p+i}$. Then

$$\mathrm{compl}_{(p,m)}(S_\lambda^{\leq p})_i = m - (S_\lambda^{\leq p})_{p+1-i} = m - \lambda_{d-p+(p+1-i)} = m - \lambda_{d+1-i},$$

so that $\mathrm{compl}_{(p,m)}(S_\lambda^{\leq p}) = (m - \lambda_d, \ldots, m - \lambda_{d+1-p})$. Since $\mathrm{compl}_{(p+1,m)}(S_\lambda^{\leq (p+1)})_i = \mathrm{compl}_{(p,m)}(S_\lambda^{\leq p})_i$ for $i \in [p]$, the only boxes in $\mathrm{compl}_{(p+1,m)}(S_\lambda^{\leq (p+1)}) \setminus \mathrm{compl}_{(p,m)}(S_\lambda^{\leq p})$ are those in the $(p+1)$-th row. Therefore, all boxes labelled p in $\mathrm{compl}_{(d,m)}(S_\lambda)$ occur in the p-th row, and therefore this SSYT is highest-weight. Since $\mathrm{compl}_{(d,m)}(S_\lambda)$ is an SSYT of shape $\mathrm{compl}_{(d,m)}(S_\lambda^{\leq d}) = \mathrm{compl}_{(d,m)}(\lambda) = \mu$, we have $\mathrm{compl}_{(d,m)}(S_\lambda) = T^\mu$. □

The reason why we set T to be the lowest-weight SSYT, is because its complement is a highest-weight SSYT, and the following lemma allows us to relate the Clebsch-Gordan coefficients in Equation (138) to the one- and two-step Clebsch-Gordan coefficients.

Lemma A.12. *Let S, S' be SSYTs with shapes λ and $\lambda + e_i$ respectively, and m be an integer larger than the number of columns in both S and S'. Moreover, let $T = \mathrm{compl}(S')$ and $T' = \mathrm{compl}(S)$. Then, it holds that*

$$\langle S, k | S' \rangle = (-1)^{d-k} \cdot \left(\frac{\dim(V^d_{\lambda+e_i})}{\dim(V^d_\lambda)} \right)^{1/2} \cdot \langle T, k | T' \rangle$$

The proof of this lemma is deferred to the end of this section. We now use this lemma to rewrite the coefficients from Equation (138) in the following way:

$$\langle Y'', k | Y' \rangle = (-1)^{d-k} \left(\frac{\dim(V^d_{\lambda+e_i})}{\dim(V^d_\lambda)} \right)^{1/2} \langle \widehat{Y'}, k | \widehat{Y''} \rangle, \qquad \langle Y', \ell | T \rangle = (-1)^{d-\ell} \left(\frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_{\lambda+e_i})} \right)^{1/2} \langle \widehat{T}, \ell | \widehat{Y'} \rangle,$$

$$\langle T'', \ell | T' \rangle = (-1)^{d-\ell} \left(\frac{\dim(V^d_{\lambda+e_j})}{\dim(V^d_\lambda)} \right)^{1/2} \langle \widehat{T'}, \ell | \widehat{T''} \rangle, \qquad \langle T', k | T \rangle = (-1)^{d-k} \left(\frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_{\lambda+e_j})} \right)^{1/2} \langle \widehat{T}, k | \widehat{T'} \rangle.$$

For brevity, we use $\widehat{T}$ to mean $\mathrm{compl}(T)$, and similarly for the SSYTs T', T'', Y', Y''. Moreover, since

complementation is injective (as shown above), $\langle T''|Y''\rangle = \langle \widehat{T''}|\widehat{Y''}\rangle$. Hence Equation (138) becomes

$$\sum_{k,\ell} \sum_{\substack{T''\in\lambda \\ T'\in\lambda+e_j}} \sum_{\substack{Y''\in\lambda \\ Y'\in\lambda+e_i}} \langle Y'', k|Y'\rangle \langle Y', \ell|T\rangle \langle T'', \ell|T'\rangle \langle T', k|T\rangle \cdot \langle T''|Y''\rangle$$

$$= \sum_{k,\ell} (-1)^{d-k}(-1)^{d-k}(-1)^{d-\ell}(-1)^{d-\ell} \cdot \frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_\lambda)}$$

$$\cdot \sum_{\substack{\widehat{T''}\in\mathrm{compl}(\lambda) \\ \widehat{T'}\in\mathrm{compl}(\lambda+e_j)}} \sum_{\substack{\widehat{Y''}\in\mathrm{compl}(\lambda) \\ \widehat{Y'}\in\mathrm{compl}(\lambda+e_i)}} \langle \widehat{Y'}, k|\widehat{Y''}\rangle \langle \widehat{T}, \ell|\widehat{Y'}\rangle \langle \widehat{T'}, \ell|\widehat{T''}\rangle \langle \widehat{T}, k|\widehat{T'}\rangle \cdot \langle \widehat{T''}|\widehat{Y''}\rangle$$

$$= \frac{\dim(V^d_{\lambda_{ij}})}{\dim(V^d_\lambda)} \sum_{k,\ell} \sum_{\substack{\widehat{T''}\in\mathrm{compl}(\lambda) \\ \widehat{T'}\in\mathrm{compl}(\lambda+e_j)}} \sum_{\substack{\widehat{Y''}\in\mathrm{compl}(\lambda) \\ \widehat{Y'}\in\mathrm{compl}(\lambda+e_i)}} \langle \widehat{Y'}, k|\widehat{Y''}\rangle \langle \widehat{T}, \ell|\widehat{Y'}\rangle \langle \widehat{T'}, \ell|\widehat{T''}\rangle \langle \widehat{T}, k|\widehat{T'}\rangle \cdot \langle \widehat{T''}|\widehat{Y''}\rangle\,. \tag{141}$$

Now the nonzero terms in the summation above correspond to SSYTs that look as follows:

$$\underbrace{\widehat{T''}}_{\mu+e_{i'}+e_{j'}} \xleftarrow{\otimes|\ell\rangle} \underbrace{\widehat{T'}}_{\mu+e_{i'}} \xleftarrow{\otimes|k\rangle} \underbrace{\widehat{T}}_{\mu} \xrightarrow{\otimes|\ell\rangle} \underbrace{\widehat{Y'}}_{\mu+e_{j'}} \xrightarrow{\otimes|k\rangle} \underbrace{\widehat{Y''}}_{\mu+e_{i'}+e_{j'}}\,,$$

where $\widehat{T}$ is a highest-weight SSYT of shape $\mu = \mathrm{compl}(\lambda_{ij})$. We set $i' = d+1-i$, $j' = d+1-j$, and we used the fact that $\mathrm{compl}(\mu + e_{i'}) = \mathrm{compl}(\mu) - e_i = \lambda + e_j$. Thus, now the terms that appear correspond to the coefficients we get when we add boxes to a highest-weight SSYT, something that we understand much better.

It suffices to show that Equation (141) is positive. The factor before the summation is always positive, and can be thus ignored. Then, we observe that the terms that appear in this expression are exactly the two-step Clebsch Gordan coefficients:

$$\langle \widehat{T'}, \ell|\widehat{T''}\rangle \langle \widehat{T}, k|\widehat{T'}\rangle = \begin{cases} a^\mu_{k\ell\to i'j'}, \\ b^\mu_{k\ell\to i'j'} \end{cases} \quad \text{or} \qquad \text{and} \qquad \langle \widehat{Y'}, k|\widehat{Y''}\rangle \langle \widehat{T}, \ell|\widehat{Y'}\rangle = \begin{cases} a^\mu_{\ell k\to j'i'}, \\ b^\mu_{\ell k\to j'i'} \end{cases} \quad \text{or}$$

From Theorem 10.18, we know that the a coefficients are always nonnegative. Thus, any negative terms in Equation (141) must come when at least one of the terms is a b coefficient.

We consider the two cases. First, let $\langle \widehat{T'}, \ell|\widehat{T''}\rangle \langle \widehat{T}, k|\widehat{T'}\rangle = b^\mu_{k\ell\to i'j'}$ have value strictly less than zero. From Theorem 10.16, this must mean that $k > \ell$, and that $\widehat{T''} = \widehat{T}_{k\ell\to j'i'}$, that is, $\widehat{T''}$ has a k at the end of the j'-th row, and an ℓ at the end of the i'-th row. Since $k > \ell$, $b^\mu_{\ell k\to j'i'} = 0$, and thus the other term $\langle \widehat{Y'}, k|\widehat{Y''}\rangle \langle \widehat{T}, \ell|\widehat{Y'}\rangle$ is either 0, or $a^\mu_{\ell k\to j'i'}$. When $\langle \widehat{Y'}, k|\widehat{Y''}\rangle \langle \widehat{T}, \ell|\widehat{Y'}\rangle = a^\mu_{\ell k\to j'i'}$, this must mean that $\widehat{Y''} = \widehat{T}_{\ell k\to j'i'}$, and thus $\widehat{Y''}$ has an ℓ at the end of the j'-th row, and a k at the end of the i'-th row. Since $i' \neq j'$, and $k > \ell$, the SSYTs $\widehat{Y''}$ and $\widehat{T''}$ are different, and thus $\langle \widehat{T''}|\widehat{Y''}\rangle = 0$. Hence, we can only have positive or zero contributions to Equation (141) in that case.

We proceed in the same fashion in the second case, when we let $\langle \widehat{Y'}, k|\widehat{Y''}\rangle \langle \widehat{T}, \ell|\widehat{Y'}\rangle = b^\mu_{\ell k\to j'i'}$ have value strictly less than zero. From Theorem 10.16, this must mean that $\ell > k$, and that $\widehat{Y''} = \widehat{Y}_{\ell k\to i'j'}$, that is, $\widehat{Y''}$ has an ℓ at the end of the i'-th row, and a k at the end of the j'-th row. Since $\ell > k$, $b^\mu_{k\ell\to i'j'} = 0$, and thus the other term $\langle \widehat{T'}, k|\widehat{T''}\rangle \langle \widehat{T}, \ell|\widehat{T'}\rangle$ is either 0, or $a^\mu_{k\ell\to i'j'}$. When $\langle \widehat{T'}, k|\widehat{T''}\rangle \langle \widehat{T}, \ell|\widehat{T'}\rangle = a^\mu_{k\ell\to i'j'}$, this must mean that $\widehat{T''} = \widehat{T}_{k\ell\to i'j'}$, and thus $\widehat{T''}$ has a k at the end of the i'-th row, and an ℓ at the end of the j'-th row. Since $i' \neq j'$, and $\ell > k$, the SSYTs $\widehat{T''}$ and $\widehat{Y''}$ are different, and thus $\langle \widehat{T''}|\widehat{Y''}\rangle = 0$. Hence, we can only have positive or zero contributions to Equation (141) in that case as well.

We conclude that this off-diagonal term must be real and positive, and thus equal to $\sqrt{1 - \frac{1}{\Delta^2_{ji}}}$. □

Proof of Theorem A.12. As per Equation (58), we decompose the left-hand side as a product of scalar factors

$$\langle S, k|S'\rangle = \prod_{p=1}^{d} (S^{=p}, \boxed{k}^{=p} \mid (S')^{=p}) = \left(S^{=k}, e_{10}^{=k} \mid (S')^{=k}\right) \cdot \prod_{p=k+1}^{d} (S^{=p}, e_{11}^{=p} \mid (S')^{=p})\,. \tag{142}$$

As in Equation (63), the only way for the Clebsch-Gordan coefficient $\langle S, k | S' \rangle$ to be nonzero, is if there exist indices $i_k, \ldots, i_d$, such that $i_d = i$, and

$$(S')^{=p} = \begin{cases} S^{=p} & p < k \\ S^{=p}_{+i_p} & p = k \\ S^{=p}_{+i_p,-i_{p-1}} & p > k \end{cases} \implies (S')^{\leq p} = \begin{cases} S^{\leq p} & p < k \\ S^{\leq p} + e_{i_p} & p \geq k \end{cases}.$$

Taking the complements of the tableaux in the right-hand side expression, we deduce that

$$(T')^{\leq p} = \begin{cases} T^{\leq p} & p < k \\ T^{\leq p} + e_{j_p} & p \geq k \end{cases} \implies (T')^{=p} = \begin{cases} T^{=p} & p < k \\ T^{=p}_{+j_p} & p = k \\ T^{=p}_{+j_p,-j_{p-1}} & p > k \end{cases},$$

where we define $j_p \coloneqq p + 1 - i_p$. This, in particular, means that the tableaux $S^{\leq p}$ and $T^{\leq p}$ are related using the following two expressions:

$$S^{\leq p}_{\overline{x}} = m - (T')^{\leq p}_{p+1-x} = \begin{cases} m - T^{\leq p}_{p+1-x} & p < k \\ m - T^{\leq p}_{p+1-x} - \delta_{x(i_p)} & p \geq k \end{cases},$$

$$T^{\leq p}_{\overline{x}} = m - (S')^{\leq p}_{p+1-x} = \begin{cases} m - S^{\leq p}_{p+1-x} & p < k \\ m - S^{\leq p}_{p+1-x} - \delta_{x(j_p)} & p \geq k \end{cases}.$$

We use the formula in Equation (60) to write the first factor in Equation (142):

$$\begin{aligned} \left(S^{=k}, e^{=k}_{10} \mid S^{=k}_{+i_k}\right) &= \left| \frac{\prod_{x=1}^{k-1} \left(c_x(S^{\leq k-1}) - c_{i_k}(S^{\leq k}) - 1\right)}{\prod_{x=1, x\neq i_k}^{k} \left(c_x(S^{\leq k}) - c_{i_k}(S^{\leq k})\right)} \right|^{1/2} \\ &= \left| \frac{\prod_{x=1}^{k-1} \left(S^{\leq k-1}_{\overline{x}} - S^{\leq k}_{\overline{i_k}} - x + i_k - 1\right)}{\prod_{x=1, x\neq i_k}^{k} \left(S^{\leq k}_{\overline{x}} - S^{\leq k}_{\overline{i_k}} - x + i_k\right)} \right|^{1/2} \\ &= \left| \frac{\prod_{x=1}^{k-1} \left(m - T^{\leq k-1}_{\overline{k-x}} - m + T^{\leq k}_{\overline{k+1-i_k}} + 1 - x + i_k - 1\right)}{\prod_{x=1, x\neq i_k}^{k} \left(m - T^{\leq k}_{\overline{k+1-x}} - m + T^{\leq k}_{\overline{k+1-i_k}} + 1 - x + i_k\right)} \right|^{1/2} \\ &= \left| \frac{\prod_{x=1}^{k-1} - \left(T^{\leq k-1}_{\overline{k-x}} - T^{\leq k}_{\overline{k+1-i_k}} - (k-x) + (k+1-i_k) - 1\right)}{\prod_{x=1, x\neq i_k}^{k} - \left(T^{\leq k}_{\overline{k+1-x}} - T^{\leq k}_{\overline{k+1-i_k}} - (k+1-x) + (k+1-i_k) - 1\right)} \right|^{1/2}. \end{aligned}$$

We will now change the iterator of the numerator to $x' \leftarrow k - x$, and the iterator of the denominator to $x'' \leftarrow k + 1 - x$. Moreover, recall that $j_k = k + 1 - i_k$. The expression thus becomes

$$\begin{aligned} &\left(S^{=k}, e^{=k}_{10} \mid S^{=k}_{+i_k}\right) \\ &= \left| \frac{\prod_{x'=1}^{k-1} \left(T^{\leq k-1}_{\overline{x'}} - T^{\leq k}_{\overline{j_k}} - x' + j_k - 1\right)}{\prod_{x''=1, x''\neq j_k}^{k} \left(T^{\leq k}_{\overline{x''}} - T^{\leq k}_{\overline{j_k}} - x'' + j_k - 1\right)} \right|^{1/2} \\ &= \left| \frac{\prod_{x'=1}^{k-1} \left(T^{\leq k-1}_{\overline{x'}} - T^{\leq k}_{\overline{j_k}} - x' + j_k - 1\right)}{\prod_{x''=1, x''\neq j_k}^{k} \left(T^{\leq k}_{\overline{x''}} - T^{\leq k}_{\overline{j_k}} - x'' + j_k\right)} \right|^{1/2} \cdot \left| \frac{\prod_{x''=1, x''\neq j_k}^{k} \left(T^{\leq k}_{\overline{x''}} - T^{\leq k}_{\overline{j_k}} - x'' + j_k\right)}{\prod_{x''=1, x''\neq j_k}^{k} \left(T^{\leq k}_{\overline{x''}} - T^{\leq k}_{\overline{j_k}} - x'' + j_k - 1\right)} \right|^{1/2} \\ &= \left(T^{=k}, e^{=k}_{10} \mid T^{=k}_{+j_k}\right) \cdot \left| \frac{\prod_{x''=1, x''\neq j_k}^{k} \left(T^{\leq k}_{\overline{x''}} - T^{\leq k}_{\overline{j_k}} - x'' + j_k\right)}{\prod_{x''=1, x''\neq j_k}^{k} \left(T^{\leq k}_{\overline{x''}} - T^{\leq k}_{\overline{j_k}} - x'' + j_k - 1\right)} \right|^{1/2}. \end{aligned}$$

We can now undo the operations above to rewrite the multiplicative correction term in terms of the original tableau S:

$$\begin{aligned}\left(S^{=k}, e_{10}^{=k} \mid S_{+i_k}^{=k}\right) &= \left(T^{=k}, e_{10}^{=k} \mid T_{+j_k}^{=k}\right) \cdot \left| \frac{\prod_{x''=1, x'' \neq j_k}^{k} \left(T_{x''}^{\leq k} - T_{j_k}^{\leq k} - x'' + j_k\right)}{\prod_{x''=1, x'' \neq j_k}^{k} \left(T_{x''}^{\leq k} - T_{j_k}^{\leq k} - x'' + j_k - 1\right)} \right|^{1/2} \\ &= \left(T^{=k}, e_{10}^{=k} \mid T_{+j_k}^{=k}\right) \cdot \left| \frac{\prod_{x=1, x \neq i_k}^{k} \left(S_{\bar{x}}^{\leq k} - S_{\bar{i_k}}^{\leq k} - x + i_k - 1\right)}{\prod_{x=1, x \neq i_k}^{k} \left(S_{\bar{x}}^{\leq k} - S_{\bar{i_k}}^{\leq k} - x + i_k\right)} \right|^{1/2} \end{aligned} \tag{143}$$

We proceed to deal with the remaining scalar factors, which correspond to $p = k+1$ up to $p = d$. We use the formula in Equation (61) to write:

$$\begin{aligned}&\left(S^{=p}, e_{11}^{=p} \mid S_{+i_p, -i_{p-1}}^{=p}\right) \\ &= S(i_p, i_{p-1}) \left| \frac{\prod_{x \neq i_{p-1}}^{p-1} \left(c_x(S^{\leq p-1}) - c_{i_p}(S^{\leq p}) - 1\right) \prod_{x \neq i_p}^{p} \left(c_x(S^{\leq p}) - c_{i_{p-1}}(S^{\leq p-1})\right)}{\prod_{x \neq i_p}^{p} \left(c_x(S^{\leq p}) - c_{i_p}(S^{\leq p})\right) \prod_{x \neq i_{p-1}}^{p-1} \left(c_x(S^{\leq p-1}) - c_{i_{p-1}}(S^{\leq p-1}) - 1\right)} \right|^{1/2} \\ &= S(i_p, i_{p-1}) \left| \frac{\prod_{x \neq i_{p-1}}^{p-1} \left(S_{\bar{x}}^{\leq p-1} - S_{\bar{i_p}}^{\leq p} - x + i_p - 1\right) \prod_{x \neq i_p}^{p} \left(S_{\bar{x}}^{\leq p} - S_{\bar{i}_{p-1}}^{\leq p-1} - x + i_{p-1}\right)}{\prod_{x \neq i_p}^{p} \left(S_{\bar{x}}^{\leq p} - S_{\bar{i_p}}^{\leq p} - x + i_p\right) \prod_{x \neq i_{p-1}}^{p-1} \left(S_{\bar{x}}^{\leq p-1} - S_{\bar{i}_{p-1}}^{\leq p-1} - x + i_{p-1} - 1\right)} \right|^{1/2} \\ &= S(i_p, i_{p-1}) \left| \frac{\prod_{x \neq i_{p-1}}^{p-1} \left(m - T_{p-x}^{\leq p-1} - m + T_{p+1-i_p}^{\leq p} + 1 - x + i_p - 1\right)}{\prod_{x \neq i_p}^{p} \left(m - T_{p+1-x}^{\leq p} - m + T_{p+1-i_p}^{\leq p} + 1 - x + i_p\right)} \right|^{1/2} \\ &\qquad \cdot \left| \frac{\prod_{x \neq i_p}^{p} \left(m - T_{p+1-x}^{\leq p} - m + T_{p-i_{p-1}}^{\leq p-1} + 1 - x + i_{p-1}\right)}{\prod_{x \neq i_{p-1}}^{p-1} \left(m - T_{p-x}^{\leq p-1} - m + T_{p-i_{p-1}}^{\leq p-1} + 1 - x + i_{p-1} - 1\right)} \right|^{1/2} \\ &= S(i_p, i_{p-1}) \left| \frac{\prod_{x \neq i_{p-1}}^{p-1} - \left(T_{p-x}^{\leq p-1} - T_{p+1-i_p}^{\leq p} - (p-x) + (p+1-i_p) - 1\right)}{\prod_{x \neq i_p}^{p} - \left(T_{p+1-x}^{\leq p} - T_{p+1-i_p}^{\leq p} - (p+1-x) + (p+1-i_p) - 1\right)} \right|^{1/2} \\ &\qquad \cdot \left| \frac{\prod_{x \neq i_p}^{p} - \left(T_{p+1-x}^{\leq p} - T_{p-i_{p-1}}^{\leq p-1} - (p+1-x) + (p-i_{p-1})\right)}{\prod_{x \neq i_{p-1}}^{p-1} - \left(T_{p-x}^{\leq p-1} - T_{p-i_{p-1}}^{\leq p-1} - (p-x) + (p-i_{p-1})\right)} \right|^{1/2}. \end{aligned}$$

We will now change the iterators of the numerators to $x' \leftarrow p-x$, $y' \leftarrow p+1-x$, and the iterators of the denominators to $x'' \leftarrow p+1-x$, $y'' \leftarrow p-x$. Moreover, recall that $j_p = p+1-i_p$ and $j_{p-1} = p-i_{p-1}$. The expression thus becomes

$$\begin{aligned}&\left(S^{=p}, e_{11}^{=p} \mid S_{+i_p, -i_{p-1}}^{=p}\right) \\ &= S(i_p, i_{p-1}) \left| \frac{\prod_{x' \neq j_{p-1}}^{p-1} \left(T_{x'}^{\leq p-1} - T_{j_p}^{\leq p} - x' + j_p - 1\right)}{\prod_{y' \neq j_p}^{p} \left(T_{y'}^{\leq p} - T_{j_p}^{\leq p} - y' + j_p - 1\right)} \cdot \frac{\prod_{x'' \neq j_p}^{p} \left(T_{x''}^{\leq p} - T_{j_{p-1}}^{\leq p-1} - x'' + j_{p-1}\right)}{\prod_{y'' \neq j_{p-1}}^{p-1} \left(T_{y''}^{\leq p-1} - T_{j_{p-1}}^{\leq p-1} - y'' + j_{p-1}\right)} \right|^{1/2} \\ &= \frac{S(i_p, i_{p-1})}{S(j_p, j_{p-1})} \left(T^{=p}, e_{11}^{=p} \mid T_{+j_p, -j_{p-1}}^{=p}\right) \\ &\qquad \cdot \left| \prod_{y' \neq j_p}^{p} \frac{T_{y'}^{\leq p} - T_{j_p}^{\leq p} - y' + j_p}{T_{y'}^{\leq p} - T_{j_p}^{\leq p} - y' + j_p - 1} \cdot \prod_{y'' \neq j_{p-1}}^{p-1} \frac{T_{y''}^{\leq p-1} - T_{j_{p-1}}^{\leq p-1} - y'' + j_{p-1} - 1}{T_{y''}^{\leq p-1} - T_{j_{p-1}}^{\leq p-1} - y'' + j_{p-1}} \right|^{1/2}. \end{aligned}$$

We note here that

$$\frac{S(i_p, i_{p-1})}{S(j_p, j_{p-1})} = \frac{S(i_p, i_{p-1})}{S(p+1-i_p, p-i_{p-1})} = -1,$$

since whenever $i_p > i_{p-1}$, then $p+1-i_p \le p-i_{p-1}$, and whenever $i_p \le i_{p-1}$, then $p+1-i_p > p-i_{p-1}$. We will now "undo" the operations above to rewrite the multiplicative correction term in terms of the original tableau S:

$$\begin{aligned}
&\left(S^{=p}, e_{11}^{=p} \mid S^{=p}_{+i_p,-i_{p-1}}\right) \\
&= -\left(T^{=p}, e_{11}^{=p} \mid T^{=p}_{+j_p,-j_{p-1}}\right) \cdot \left| \prod_{y' \neq j_p}^{p} \frac{m - S^{\le p}_{p+1-y'} - m + S^{\le p}_{p+1-j_p} + 1 - y' + j_p}{m - S^{\le p}_{p+1-y'} - m + S^{\le p}_{p+1-j_p} + 1 - y' + j_p - 1} \right|^{1/2} \\
&\quad \cdot \left| \prod_{y'' \neq j_{p-1}}^{p-1} \frac{m - S^{\le p-1}_{p-y''} - m + S^{\le p-1}_{p-j_{p-1}} + 1 - y'' + j_{p-1} - 1}{m - S^{\le p-1}_{p-y''} - m + S^{\le p-1}_{p-j_{p-1}} + 1 - y'' + j_{p-1}} \right|^{1/2} \\
&= -\left(T^{=p}, e_{11}^{=p} \mid T^{=p}_{+j_p,-j_{p-1}}\right) \cdot \left| \prod_{x \neq i_p}^{p} \frac{S^{\le p}_{x} - S^{\le p}_{i_p} - x + i_p - 1}{S^{\le p}_{x} - S^{\le p}_{i_p} - x + i_p} \right|^{1/2} \\
&\quad \cdot \left| \prod_{x \neq i_{p-1}}^{p-1} \frac{S^{\le p-1}_{x} - S^{\le p-1}_{i_{p-1}} - x + i_{p-1}}{S^{\le p-1}_{x} - S^{\le p-1}_{i_{p-1}} - x + i_{p-1} - 1} \right|^{1/2}. \qquad (144)
\end{aligned}$$

We can thus write the original Clebsch-Gordan coefficient as

$$\begin{aligned}
&\langle S, k | S' \rangle \\
&= \left(S^{=k}, e_{10}^{=k} \mid S^{=k}_{+i_k}\right) \cdot \prod_{p=k+1}^{d} \left(S^{=p}, e_{11}^{=p} \mid S^{=p}_{+i_p,-i_{p-1}}\right) \\
&= \left(T^{=k}, e_{10}^{=k} \mid T^{=k}_{+j_k}\right) \cdot \left| \frac{\prod_{x=1, x\neq i_k}^{k} \left(S^{\le k}_{x} - S^{\le k}_{i_k} - x + i_k - 1\right)}{\prod_{x=1, x\neq i_k}^{k} \left(S^{\le k}_{x} - S^{\le k}_{i_k} - x + i_k\right)} \right|^{1/2} \cdot \prod_{p=k+1}^{d} - \left(T^{=p}, e_{11}^{=p} \mid T^{=p}_{+j_p,-j_{p-1}}\right) \\
&\quad \cdot \left| \prod_{x \neq i_p}^{p} \frac{S^{\le p}_{x} - S^{\le p}_{i_p} - x + i_p - 1}{S^{\le p}_{x} - S^{\le p}_{i_p} - x + i_p} \cdot \prod_{x \neq i_{p-1}}^{p-1} \frac{S^{\le p-1}_{x} - S^{\le p-1}_{i_{p-1}} - x + i_{p-1}}{S^{\le p-1}_{x} - S^{\le p-1}_{i_{p-1}} - x + i_{p-1} - 1} \right|^{1/2}. \quad \text{(Eqs. (143) and (144))}
\end{aligned}$$

If we telescope the above expression, all the products cancel except the one with the iterator $x \neq i_d$. Moreover, we have defined $i_d = i$, and $S^{\le d} = \mathsf{sh}(S) = \lambda$. Thus,

$$\begin{aligned}
&\langle S, k | S' \rangle \\
&= (-1)^{d-k} \left(T^{=k}, e_{10}^{=k} \mid T^{=k}_{+j_k}\right) \cdot \prod_{p=k+1}^{d} \left(T^{=p}, e_{11}^{=p} \mid T^{=p}_{+j_p,-j_{p-1}}\right) \cdot \left| \prod_{x \neq i_d}^{d} \frac{S^{\le d}_{x} - S^{\le d}_{i_d} - x + i_d - 1}{S^{\le d}_{x} - S^{\le d}_{i_d} - x + i_d} \right|^{1/2} \\
&= (-1)^{d-k} \langle T, T^k | T' \rangle \cdot \left| \prod_{x \neq i}^{d} \frac{\lambda_x - \lambda_i - x + i - 1}{\lambda_x - \lambda_i - x + i} \right|^{1/2} \\
&= (-1)^{d-k} \langle T, T^k | T' \rangle \cdot \left(\frac{\dim(V^d_{\lambda + e_i})}{\dim(V^d_\lambda)} \right)^{1/2}.
\end{aligned}$$

The last equality follows from the Weyl dimension formula in Equation (15). First, we write

$$\dim(V^d_\lambda) = \prod_{1 \le x < y \le d} \frac{(\lambda_x - \lambda_y) + (y - x)}{y - x},$$

$$\dim(V^d_{\lambda + e_i}) = \prod_{\substack{1 \le x < y \le d \\ x, y \neq i}} \frac{(\lambda_x - \lambda_y) + (y - x)}{y - x} \cdot \prod_{x < i} \frac{(\lambda_x - \lambda_i) + (i - x) - 1}{i - x} \cdot \prod_{i < x} \frac{(\lambda_i - \lambda_x) + (x - i) + 1}{x - i}.$$

Thus

$$\begin{aligned}\frac{\dim(V^d_{\lambda+e_i})}{\dim(V^d_\lambda)} &= \prod_{x<i} \frac{(\lambda_x - \lambda_i) + (i-x) - 1}{(\lambda_x - \lambda_i) + (i-x)} \cdot \prod_{i<x} \frac{(\lambda_i - \lambda_x) + (x-i) + 1}{(\lambda_i - \lambda_x) + (x-i)} \\ &= \prod_{x<i} \frac{\lambda_x - \lambda_i - x + i - 1}{\lambda_x - \lambda_i - x + i} \cdot \prod_{i<x} \left| \frac{\lambda_x - \lambda_i - x + i - 1}{\lambda_x - \lambda_i - x + i} \right| \\ &= \left| \prod_{x \neq i} \frac{\lambda_x - \lambda_i - x + i - 1}{\lambda_x - \lambda_i - x + i} \right|,\end{aligned}$$

Where the second equality holds because the left-hand side is always nonnegative. □