Efficient quantum tomography

Ryan O'Donnell* John Wright*

September 12, 2015

Abstract

In the quantum state tomography problem, one wishes to estimate an unknown d-dimensional mixed quantum state ρ, given few copies. We show that $O(d/\epsilon)$ copies suffice to obtain an estimate $\hat{\rho}$ that satisfies $\|\hat{\rho}-\rho\|_F^2 \leq \epsilon$ (with high probability). An immediate consequence is that $O(\text{rank}(\rho)\cdot d/\epsilon^2) \leq O(d^2/\epsilon^2)$ copies suffice to obtain an ϵ-accurate estimate in the standard trace distance. This improves on the best known prior result of $O(d^3/\epsilon^2)$ copies for full tomography, and even on the best known prior result of $O(d^2 \log(d/\epsilon)/\epsilon^2)$ copies for spectrum estimation. Our result is the first to show that nontrivial tomography can be obtained using a number of copies that is just *linear* in the dimension.

Next, we generalize these results to show that one can perform efficient principal component analysis on ρ. Our main result is that $O(kd/\epsilon^2)$ copies suffice to output a rank-k approximation $\hat{\rho}$ whose trace distance error is at most ϵ more than that of the best rank-k approximator to ρ. This subsumes our above trace distance tomography result and generalizes it to the case when ρ is not guaranteed to be of low rank. A key part of the proof is the analogous generalization of our spectrum-learning results: we show that the largest k eigenvalues of ρ can be estimated to trace-distance error ϵ using $O(k^2/\epsilon^2)$ copies. In turn, this result relies on a new coupling theorem concerning the Robinson–Schensted–Knuth algorithm that should be of independent combinatorial interest.

1 Introduction

Quantum state tomography refers to the task of estimating an unknown d-dimensional quantum mixed quantum state, ρ, given the ability to prepare and measure n copies, $\rho^{\otimes n}$. It is of enormous practical importance for experimental detection of entanglement and the verification of quantum technologies. For an anthology of recent advances in the area, the reader may consult [BCG13]. As stated in its introduction,

> *The bottleneck limiting further progress in estimating the states of [quantum] systems has shifted from physical controllability to the problem of handling... the exponential scaling of the number of parameters describing quantum many-body states.*

Indeed, a system consisting of b qubits has dimension $d = 2^b$ and is described by a density matrix with $d^2 = 4^b$ complex parameters. For practical experiments with, say, $b \leq 10$, it is imperative to use tomographic methods in which n grows as slowly as possible with d. For 20 years or so, the best known method used $n = O(d^4)$ copies to estimate ρ to constant error; just recently this was improved [KRT14] to $n = O(d^3)$. Despite the practical importance and mathematical elegance

*Department of Computer Science, Carnegie Mellon University. Supported by NSF grants CCF-0747250 and CCF-1116594. The second-named author is also supported by a Simons Fellowship in Theoretical Computer Science.
{odonnell,jswright}@cs.cmu.edu

of the quantum tomography problem, the optimal dependence of n on d remained "shockingly unknown" [Har15] as of early 2015.

In this work we analyze known measurements arising from the representation theory of the symmetric and general linear groups $\mathfrak{S}(n)$ and $\mathrm{GL}_d = \mathrm{GL}_d(\mathbb{C})$ — specifically, the "Empirical Young Diagram (EYD)" measurement considered by [ARS88, KW01], followed by Keyl's [KW01, Key06] state estimation measurement based on projection to highest weight vectors. The former produces a random height-d partition $\boldsymbol{\lambda} \vdash n$ according to the *Schur–Weyl distribution* $\mathrm{SW}^n(\alpha)$, which depends only on the spectrum $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_d$ of ρ; the latter produces a random d-dimensional unitary $\boldsymbol{U}$ according to what may be termed the *Keyl distribution* $\mathrm{K}_\lambda(\rho)$. Writing $\underline{\lambda}$ for $(\lambda_1/n, \ldots, \lambda_d/n)$, we show the following results:

Theorem 1.1. $\displaystyle \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim\mathrm{SW}^n(\alpha)} \|\underline{\boldsymbol{\lambda}} - \alpha\|_2^2 \leq \frac{d}{n}.$

Theorem 1.2. $\displaystyle \mathop{\mathbf{E}}_{\substack{\boldsymbol{\lambda}\sim\mathrm{SW}^n(\alpha)\\ \boldsymbol{U}\sim\mathrm{K}_{\boldsymbol{\lambda}}(\rho)}} \|\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho\|_F^2 \leq \frac{4d-3}{n}.$

In particular, up to a small constant factor, full tomography is no more expensive than spectrum estimation. These theorems have the following straightforward consequences:

Corollary 1.3. *The spectrum of an unknown rank-r mixed state $\rho \in \mathbb{C}^{d\times d}$ can be estimated to error ϵ in ℓ_2-distance using $n = O(r/\epsilon^2)$ copies, or to error ϵ in total variation distance using $n = O(r^2/\epsilon^2)$ copies.*

Corollary 1.4. *An unknown rank-r mixed state $\rho \in \mathbb{C}^{d\times d}$ may be estimated to error ϵ in Frobenius distance using $n = O(d/\epsilon^2)$ copies, or to error ϵ in trace distance using $n = O(rd/\epsilon^2)$ copies.*

(These bounds are with high probability; confidence $1-\delta$ may be obtained by increasing the copies by a factor of $\log(1/\delta)$.)

The previous best result for spectrum estimation [HM02, CM06] used $O(r^2\log(r/\epsilon)/\epsilon)$ copies for an ϵ-accurate estimation in KL-divergence, and hence $O(r^2\log(r/\epsilon)/\epsilon^2)$ copies for an ϵ-accurate estimation in total variation distance. The previous best result for tomography is the very recent [KRT14, Theorem 2], which uses $n = O(rd/\epsilon^2)$ for an ϵ-accurate estimation in Frobenius distance, and hence $n = O(r^2 d/\epsilon^2)$ for trace distance.

As for lower bounds, it follows immediately from [FGLE12, Lemma 5] and Holevo's bound that $\widetilde{\Omega}(rd)$ copies are necessary for tomography with trace-distance error ϵ_0, where ϵ_0 is a universal constant. (Here and throughout $\widetilde{\Omega}(\cdot)$ hides a factor of $\log d$.) Also, Holevo's bound combined with the existence of $2^{\Omega(d)}$ almost-orthogonal pure states shows that $\widetilde{\Omega}(d)$ copies are necessary for tomography with Frobenius error ϵ_0, even in the rank-1 case. Thus our tomography bounds are optimal up to at most an $O(\log d)$ factor when ϵ is a constant. (Conversely, for constant d, it is easy to show that $\Omega(1/\epsilon^2)$ copies are necessary even just for spectrum estimation.) Finally, we remark that $\widetilde{\Omega}(d^2)$ is a lower bound for tomography with Frobenius error $\epsilon = \Theta(1/\sqrt{d})$; this also matches our $O(d/\epsilon^2)$ upper bound. This last lower bound follows from Holevo and the existence [Sza82] of $2^{\Omega(d^2)}$ normalized rank-$d/2$ projectors with pairwise Frobenius distance at least $\Omega(1/\sqrt{d})$.

1.1 Principal component analysis

Our next results concern principal component analysis (PCA), in which the goal is to find the best rank-k approximator to a mixed state $\rho \in \mathbb{C}^{d\times d}$, given $1 \leq k \leq d$. Our algorithm is identical to the Keyl measurement from above, except rather than outputting $\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger$, it outputs $\boldsymbol{U}\mathrm{diag}^{(k)}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger$ instead, where $\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}})$ means $\mathrm{diag}(\underline{\boldsymbol{\lambda}}_1, \dots, \underline{\boldsymbol{\lambda}}_k, 0, \dots, 0)$. Writing $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_d$ for the spectrum of ρ, our main result is:

Theorem 1.5. $\displaystyle \mathop{\mathbf{E}}_{\substack{\boldsymbol{\lambda}\sim\mathrm{SW}^n(\alpha)\\ \boldsymbol{U}\sim\mathrm{K}_{\boldsymbol{\lambda}}(\rho)}} \|\boldsymbol{U}\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho\|_1 \leq \alpha_{k+1} + \dots + \alpha_d + 6\sqrt{\frac{kd}{n}}.$

As the best rank-k approximator to ρ has trace-distance error $\alpha_{k+1} + \dots + \alpha_d$, we may immediately conclude:

Corollary 1.6. *Using $n = O(kd/\epsilon^2)$ copies of an unknown mixed state $\rho \in \mathbb{C}^{d\times d}$, one may find a rank-$k$ mixed state $\hat{\rho}$ such that the trace distance of $\hat{\rho}$ from ρ is at most ϵ more than that of the optimal rank-k approximator.*

Since $\alpha_{k+1} = \dots = \alpha_d = 0$ when ρ has rank k, Corollary 1.6 strictly generalizes the trace-distance tomography result from Corollary 1.4. We also remark that one could consider performing Frobenius-norm PCA on ρ, but it turns out that this is unlikely to give any improvement in copy complexity over full tomography; see Section 6 for details.

As a key component of our PCA result, we investigate the problem of estimating just the largest k eigenvalues, $\alpha_1, \dots, \alpha_k$, of ρ. The goal here is to use a number of copies depending only on k and not on d or $\mathrm{rank}(\rho)$. We show that the standard EYD algorithm achieves this:

Theorem 1.7. $\displaystyle \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim\mathrm{SW}^n(\alpha)} d_{\mathrm{TV}}^{(k)}(\underline{\boldsymbol{\lambda}}, \alpha) \leq \frac{1.92\,k + .5}{\sqrt{n}}$, *where $d_{\mathrm{TV}}^{(k)}(\beta, \alpha)$ denotes $\frac{1}{2}\sum_{i=1}^k |\beta_i - \alpha_i|$.*

From this we immediately get the following strict generalization of (the total variation distance result in) Corollary 1.3:

Corollary 1.8. *The largest k eigenvalues of an unknown mixed state $\rho \in \mathbb{C}^{d\times d}$ can be estimated to error ϵ in in total variation distance using $n = O(k^2/\epsilon^2)$ copies.*

The fact that this result has no dependence on the ambient dimension d or the rank of ρ may make it particularly interesting in practice.

1.2 A coupling result concerning the RSK algorithm

For our proof of Theorem 1.7, we will need to establish a new combinatorial result concerning the Robinson–Schensted–Knuth (RSK) algorithm applied to random words. We assume here the reader is familiar with the RSK correspondence; see Section 2 for a few basics and, e.g., [Ful97] for a comprehensive treatment.

Notation 1.9. Let α be a probability distribution on $[d] = \{1, 2, \dots, d\}$, and let $\boldsymbol{w} \in [d]^n$ be a random word formed by drawing each letter $\boldsymbol{w}_i$ independently according to α. Let $\boldsymbol{\lambda}$ be the shape of the Young tableaus obtained by applying the RSK correspondence to $\boldsymbol{w}$. We write $\mathrm{SW}^n(\alpha)$ for the resulting probability distribution on $\boldsymbol{\lambda}$.

Notation 1.10. For $x, y \in \mathbb{R}^d$, we say x *majorizes* y, denoted $x \succ y$, if $\sum_{i=1}^k x_{[i]} \geq \sum_{i=1}^k y_{[i]}$ for all $k \in [d] = \{1, 2, \dots, d\}$, with equality for $k = d$. Here the notation $x_{[i]}$ means the ith largest value among $x_1, \dots, x_d$. We also use the traditional notation $\lambda \trianglerighteq \mu$ instead when λ and μ are partitions of n (Young diagrams).

In Section 7 we prove the following theorem. The proof is entirely combinatorial, and can be read independently of the quantum content in the rest of the paper.

Theorem 1.11. *Let α, β be probability distributions on $[d]$ with $\beta \succ \alpha$. Then for any $n \in \mathbb{N}$ there is a coupling $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ of $\mathrm{SW}^n(\alpha)$ and $\mathrm{SW}^n(\beta)$ such that $\boldsymbol{\mu} \trianglerighteq \boldsymbol{\lambda}$ always.*

1.3 Independent and simultaneous work.

Independently and simultaneously of our work, Haah et al. [HHJ+15] have given a slightly different measurement that also achieves Corollary 1.4, up to a log factor. More precisely, their measurement achieves error ϵ in infidelity with $n = O(rd/\epsilon) \cdot \log(d/\epsilon)$ copies, or error ϵ in trace distance with $n = O(rd/\epsilon^2) \cdot \log(d/\epsilon)$ copies. They also give a lower bound of $n \geq \Omega(rd/\epsilon^2)/\log(d/r\epsilon)$ for quantum tomography with trace distance error ϵ. After seeing a draft of their work, we observed that their measurement can also be shown to achieve expected squared-Frobenius error $\frac{4d-3}{n}$, using the techniques in this paper; the brief details appear at [Wri15].

1.4 Acknowledgments.

We thank Jeongwan Haah and Aram Harrow (and by transitivity, Vlad Voroninski) for bringing [KRT14] to our attention. We also thank Aram Harrow for pointing us to [Key06]. The second-named author would also like to thank Akshay Krishnamurthy and Ashley Montanaro for helpful discussions.

2 Preliminaries

We write $\lambda \vdash n$ to denote that λ is a *partition* of n; i.e., λ is a finite sequence of integers $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots$ summing to n. We also say that the *size* of λ is $|\lambda| = n$. The *length* (or *height*) of λ, denoted $\ell(\lambda)$, is the largest d such that $\lambda_d \neq 0$. We identify partitions that only differ by trailing zeroes. A *Young diagram* of *shape* λ is a left-justified set of boxes arranged in rows, with λ_i boxes in the ith row from the top. We write $\mu \nearrow \lambda$ to denote that λ can be formed from μ by the addition of a single box to some row. A *standard Young tableau* T of *shape* λ is a filling of the boxes of λ with $[n]$ such that the rows and columns are strictly increasing. We write $\lambda = \mathrm{sh}(T)$. Note that T can also be identified with a chain $\emptyset = \lambda^{(0)} \nearrow \lambda^{(1)} \nearrow \cdots \nearrow \lambda^{(n)} = \lambda$, where $\lambda^{(t)}$ is the shape of the Young tableau formed from T by entries $1 \mathinner{..} t$. A *semistandard Young tableau* of shape λ and alphabet $\mathcal{A}$ is a filling of the boxes with letters from $\mathcal{A}$ such that rows are increasing and columns are strictly increasing. Here an *alphabet* means a totally ordered set of "letters", usually $[d]$.

The quantum measurements we analyze involve the *Schur–Weyl duality theorem.* The symmetric group $\mathfrak{S}(n)$ acts on $(\mathbb{C}^d)^{\otimes n}$ by permuting factors, and the general linear group GL_d acts on it diagonally; furthermore, these actions commute. Schur–Weyl duality states that as an $\mathfrak{S}(n) \times \mathrm{GL}_d$ representation, we have the following unitary equivalence:

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

Here we are using the following notation: The *Specht modules* Sp_λ are the irreducible representations spaces of $\mathfrak{S}(n)$, indexed by partitions $\lambda \vdash n$. We will use the abbreviation $\dim(\lambda)$ for $\dim(\mathrm{Sp}_\lambda)$; recall this equals the number of standard Young tableaus of shape λ. The *Schur (Weyl) modules* V_λ^d are the irreducible polynomial representation spaces of GL_d, indexed by partitions (highest weights) λ of length at most d. (For more background see, e.g., [Har05].) We will write $\pi_\lambda : \mathrm{GL}_d \to \mathrm{End}(\mathrm{V}_\lambda^d)$ for the (unitary) representation itself; the domain of π_λ naturally extends to all of $\mathbb{C}^{d\times d}$ by continuity. We also write $|T_\lambda\rangle$ for the highest weight vector in V_λ^d; it is characterized by the property that $\pi_\lambda(A)\,|T_\lambda\rangle = (\prod_{k=1}^d A_{kk}^{\lambda_i})\,|T_\lambda\rangle$ if $A = (A_{ij})$ is upper-triangular.

The character of V_λ^d is the *Schur polynomial* $s_\lambda(x_1, \ldots, x_d)$, a symmetric, degree-$|\lambda|$, homogeneous polynomial in $x = (x_1, \ldots, x_d)$ defined by $s_\lambda(x) = a_{\lambda+\delta}(x)/a_\delta(x)$, where $\delta = (d-1, d-$

$2,\ldots,1,0)$ and $a_\mu(x) = \det(x_i^{\mu_j})$. Alternatively, it may be defined as $\sum_T \prod_{i=1}^d x_i^{\#_i T}$, where T ranges over all semistandard tableau of shape λ and alphabet $[d]$, and $\#_i T$ denotes the number of occurrences of i in T. We have $\dim(\mathrm{V}_\lambda^d) = s_\lambda(1,\ldots,1)$, the number of semistandard Young tableaus in the sum. We'll write $\Phi_\lambda(x)$ for the *normalized Schur polynomial* $s_\lambda(x_1,\ldots,x_d)/s_\lambda(1,\ldots,1)$. Finally, we recall the following two formulas, the first following from Stanley's hook-content formula and the Frame–Robinson–Thrall hook-length formula, the second being the Weyl dimension formula:

$$s_\lambda(1,\ldots,1) = \frac{\dim(\lambda)}{|\lambda|!} \prod_{(i,j)\in\lambda} (d+j-i) = \prod_{1\le i<j\le d} \frac{(\lambda_i - \lambda_j) + (j-i)}{j-i}. \tag{1}$$

Given a positive semidefinite matrix $\rho \in \mathbb{C}^d$, we typically write $\alpha \in \mathbb{R}^d$ for its *sorted spectrum*; i.e., its eigenvalues $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_d \ge 0$. When ρ has trace 1 it is called a *density matrix* (or *mixed state*), and in this case α defines a (sorted) probability distribution on $[d]$.

We will several times use the following elementary majorization inequality:

$$\text{If } c,x,y \in \mathbb{R}^d \text{ are sorted (decreasing) and } x \succ y \text{ then } c\cdot x \ge c \cdot y. \tag{2}$$

Recall [Ful97] that the Robinson–Schensted–Knuth correspondence is a certain bijection between strings $w \in \mathcal{A}^n$ and pairs (P,Q), where P is a semistandard *insertion tableau* filled by the multiset of letters in w, and Q is a standard *recording tableau*, satisfying $\mathrm{sh}(Q) = \mathrm{sh}(P)$. We write $\mathrm{RSK}(w) = (P,Q)$ and write $\mathrm{shRSK}(w)$ for the common shape of P and Q, a partition of n of length at most $|\mathcal{A}|$. One way to characterize $\lambda = \mathrm{shRSK}(w)$ is by *Greene's Theorem* [Gre74]: $\lambda_1 + \cdots + \lambda_k$ is the length of the longest disjoint union of k increasing subsequences in w. In particular, $\lambda_1 = \mathrm{LIS}(w)$, the length of the longest increasing (i.e., nondecreasing) subsequence in w. We remind the reader here of the distinction between a *subsequence* of a string, in which the letters need not be consecutive, and a *substring*, in which they are. We use the notation $w[i \mathinner{..} j]$ for the substring $(w_i, w_{i+1}, \ldots, w_j) \in \mathcal{A}^{j-i+1}$.

Let $\alpha = (\alpha_1,\ldots,\alpha_d)$ denote a probability distribution on alphabet $[d]$, let $\alpha^{\otimes n}$ denote the associated product probability distribution on $[d]^n$, and write $\alpha^{\otimes\infty}$ for the product probability distribution on infinite sequences. We define the associated *Schur–Weyl growth process* to be the (random) sequence

$$\emptyset = \boldsymbol{\lambda}^{(0)} \nearrow \boldsymbol{\lambda}^{(1)} \nearrow \boldsymbol{\lambda}^{(2)} \nearrow \boldsymbol{\lambda}^{(3)} \nearrow \cdots \tag{3}$$

where $\boldsymbol{w} \sim \alpha^{\otimes\infty}$ and $\boldsymbol{\lambda}^{(t)} = \mathrm{shRSK}(\boldsymbol{w}[1 \mathinner{..} t])$. Note that the marginal distribution on $\boldsymbol{\lambda}^{(n)}$ is what we call $\mathrm{SW}^n(\alpha)$. The Schur–Weyl growth process was studied in, e.g., [O'C03], wherein it was noted that the RSK correspondence implies

$$\mathbf{Pr}[\boldsymbol{\lambda}^{(t)} = \lambda^{(t)} \quad \forall t \le n] = s_{\lambda^{(n)}}(\alpha) \tag{4}$$

for any chain $\emptyset = \lambda^{(0)} \nearrow \cdots \nearrow \lambda^{(n)}$. (Together with the fact that $s_\lambda(\alpha)$ is homogeneous of degree $|\lambda|$, this gives yet another alternate definition of the Schur polynomials.) One consequence of this is that for any $i \in [d]$ we have

$$\mathbf{Pr}[\boldsymbol{\lambda}^{(n+1)} = \lambda + e_i \mid \boldsymbol{\lambda}^{(n)} = \lambda] = \frac{s_{\lambda+e_i}(\alpha)}{s_\lambda(\alpha)}. \tag{5}$$

(This formula is correct even when $\lambda + e_i$ is not a valid partition of $n+1$; in this case $s_{\lambda+e_i} \equiv 0$ formally under the determinantal definition.) The above equation is also a probabilistic interpretation of the following special case of *Pieri's rule*:

$$(x_1 + \cdots + x_d) s_\lambda(x_1,\ldots,x_d) = \sum_{i=1}^d s_{\lambda+e_i}(x_1,\ldots,x_d). \tag{6}$$

We will need the following consequence of (5):

Proposition 2.1. *Let $\lambda \vdash n$ and let $\alpha \in \mathbb{R}^d$ be a sorted probability distribution. Then*

$$\left(\frac{s_{\lambda+e_1}(\alpha)}{s_\lambda(\alpha)}, \dots, \frac{s_{\lambda+e_d}(\alpha)}{s_\lambda(\alpha)}\right) \succ (\alpha_1, \dots, \alpha_d). \tag{7}$$

Proof. Let β be the reversal of α (i.e. $\beta_i = \alpha_{d-i+1}$) and let $(\boldsymbol{\lambda}^{(t)})_{t\geq 0}$ be a Schur–Weyl growth process corresponding to β. By (5) and the fact that the Schur polynomials are symmetric, we conclude that the vector on the left of (7) is $(p_1, \dots, p_d)$, where $p_i = \mathbf{Pr}[\boldsymbol{\lambda}^{(n+1)} = \lambda + e_i \mid \boldsymbol{\lambda}^{(n)} = \lambda]$. Now $p_1 + \cdots + p_k$ is the probability, conditioned on $\boldsymbol{\lambda}^{(n)} = \lambda$, that the $(n+1)$th box in the process enters into one of the first k rows. But this is indeed at least $\alpha_1 + \cdots + \alpha_k = \beta_d + \cdots + \beta_{d-k+1}$, because the latter represents the probability that the $(n+1)$th letter is $d-k+1$ or higher, and such a letter will always be inserted within the first k rows under RSK. □

A further consequence of (4) (perhaps first noted in [ITW01]) is that for $\lambda \vdash n$,

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda}\sim \mathrm{SW}^n(\alpha)}[\boldsymbol{\lambda} = \lambda] = \dim(\lambda)s_\lambda(\alpha). \tag{8}$$

At the same time, as noted in [ARS88] (see also [Aud06, Equation (36)]) it follows from Schur–Weyl duality that if $\rho \in \mathbb{C}^{d\times d}$ is a density matrix with spectrum α then

$$\mathrm{tr}(\Pi_\lambda \rho^{\otimes n}) = \dim(\lambda)s_\lambda(\alpha),$$

where Π_λ denotes the isotypic projection onto $\mathrm{Sp}_\lambda \otimes \mathrm{V}^d_\lambda$. Thus we have the identity

$$\mathrm{tr}(\Pi_\lambda \rho^{\otimes n}) = \mathop{\mathbf{Pr}}_{\boldsymbol{\lambda}\sim \mathrm{SW}^n(\alpha)}[\boldsymbol{\lambda} = \lambda]. \tag{9}$$

3 Spectrum estimation

Several groups of researchers suggested the following method for estimating the sorted spectrum α of a quantum mixed state $\rho \in \mathbb{C}^{d\times d}$: measure $\rho^{\otimes n}$ according to the isotypic projectors $\{\Pi_\lambda\}_{\lambda\vdash n}$; and, on obtaining $\boldsymbol{\lambda}$, output the estimate $\hat{\alpha} = \underline{\boldsymbol{\lambda}} = (\boldsymbol{\lambda}_1/n, \dots, \boldsymbol{\lambda}_d/n)$. The measurement is sometimes called "weak Schur sampling" [CHW07] and we refer to the overall procedure as the "Empirical Young Diagram (EYD)" algorithm. We remark that the algorithm's behavior depends only on the rank r of ρ; it is indifferent to the ambient dimension d. So while we will analyze the EYD algorithm in terms of d, we will present the results in terms of r.

In [HM02, CM06] it is shown that $n = O(r^2 \log(r/\epsilon)/\epsilon^2)$ suffices for EYD to obtain $d_{\mathrm{KL}}(\underline{\boldsymbol{\lambda}}, \alpha) \leq 2\epsilon^2$ and hence $d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \alpha) \leq \epsilon$ with high probability. However we give a different analysis. By equation (9), the expected ℓ_2^2-error of the EYD algorithm is precisely $\mathbf{E}_{\boldsymbol{\lambda}\sim\mathrm{SW}^n(\alpha)} \|\underline{\boldsymbol{\lambda}} - \alpha\|_2^2$. Theorem 1.1, which we prove in this section, bounds this quantity by $\frac{r}{n}$. Thus

$$\mathbf{E}\, d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \alpha) = \tfrac{1}{2}\,\mathbf{E}\,\|\underline{\boldsymbol{\lambda}} - \alpha\|_1 \leq \tfrac{1}{2}\sqrt{r}\,\mathbf{E}\,\|\underline{\boldsymbol{\lambda}} - \alpha\|_2 \leq \tfrac{1}{2}\sqrt{r}\sqrt{\mathbf{E}\,\|\underline{\boldsymbol{\lambda}} - \alpha\|_2^2} \leq \frac{r}{2\sqrt{n}},$$

which is bounded by $\epsilon/4$, say, if $n = 4r^2/\epsilon^2$. Thus in this case $\mathbf{Pr}[d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \alpha) > \epsilon] < 1/4$. By a standard amplification (repeating the EYD algorithm $O(\log 1/\delta)$ times and outputting the estimate which is within 2ϵ total variation distance of the most other estimates), we obtain Corollary 1.3.

We give two lemmas, and then the proof of Theorem 1.1.

Lemma 3.1. *Let $\alpha \in \mathbb{R}^d$ be a probability distribution. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}^n(\alpha)} \sum_{i=1}^d \boldsymbol{\lambda}_i^2 \le \sum_{i=1}^d (n\alpha_i)^2 + dn.$$

Proof. Define the polynomial function

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

By Proposition 2.34 and equation (12) of [OW15], $\mathbf{E}_{\boldsymbol{\lambda}\sim \mathrm{SW}^n(\alpha)}[p_2^*(\boldsymbol{\lambda})] = n(n-1)\cdot \sum_{i=1}^d \alpha_i^2$. Hence,

$$\mathbf{E}\sum_{i=1}^d \boldsymbol{\lambda}_i^2 = \mathbf{E}\left[p_2^*(\boldsymbol{\lambda}) + \sum_{i=1}^d (2i-1)\boldsymbol{\lambda}_i\right] \le \mathbf{E}\, p_2^*(\boldsymbol{\lambda}) + \sum_{i=1}^d (2i-1)(n/d) \le n^2\cdot \sum_{i=1}^d \alpha_i^2 + dn.$$

Here the first inequality used inequality (2) and $\boldsymbol{\lambda} \succ (n/d,\dots,n/d)$. □

Lemma 3.2. *Let $\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha)$, where $\alpha \in \mathbb{R}^d$ is a sorted probability distribution. Then $(\mathbf{E}\,\boldsymbol{\lambda}_1,\dots,\mathbf{E}\,\boldsymbol{\lambda}_d) \succ (\alpha_1 n,\dots,\alpha_d n)$.*

Proof. Let $\boldsymbol{w} \sim \alpha^{\otimes n}$, so $\boldsymbol{\lambda}$ is distributed as $\mathrm{shRSK}(\boldsymbol{w})$. The proof is completed by linearity of expectation applied to the fact that $(\boldsymbol{\lambda}_1,\dots,\boldsymbol{\lambda}_d) \succ (\#_1\boldsymbol{w},\dots,\#_d\boldsymbol{w})$ always, where $\#_k\boldsymbol{w}$ denotes the number of times letter k appears in $\boldsymbol{w}$. In turn this fact holds by Greene's Theorem: we can form k disjoint increasing subsequences in $\boldsymbol{w}$ by taking all its 1's, all its 2's, ..., all its k's. □

Proof of Theorem 1.1. We have

$$\begin{aligned} n^2\cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}^n(\alpha)} \|\underline{\boldsymbol{\lambda}} - \alpha\|_2^2 &= \mathbf{E}\sum_{i=1}^d (\boldsymbol{\lambda}_i - \alpha_i n)^2 = \mathbf{E}\sum_{i=1}^d (\boldsymbol{\lambda}_i^2 + (\alpha_i n)^2) - 2\sum_{i=1}^d (\alpha_i n)\cdot \mathbf{E}\,\boldsymbol{\lambda}_i \\ &\le dn + 2\sum_{i=1}^d (\alpha_i n)^2 - 2\sum_{i=1}^d (\alpha_i n)\cdot \mathbf{E}\,\boldsymbol{\lambda}_i \le dn + 2\sum_{i=1}^d (\alpha_i n)^2 - 2\sum_{i=1}^d (\alpha_i n)\cdot(\alpha_i n) = dn, \end{aligned}$$

where the first inequality used Lemma 3.1 and the second used Lemma 3.2 and inequality (2) (recall that the coefficients $\alpha_i n$ are decreasing). Dividing by n^2 completes the proof. □

4 Quantum state tomography

In this section we analyze the tomography algorithm proposed by Keyl [Key06] based on projection to the highest weight vector. Keyl's method, when applied to density matrix $\rho \in \mathbb{C}^{d\times d}$ with sorted spectrum α, begins by performing weak Schur sampling on $\rho^{\otimes n}$. Supposing the partition thereby obtained from $\mathrm{SW}^n(\alpha)$ is $\lambda \vdash n$, the state collapses to $\frac{1}{s_\lambda(\alpha)}\pi_\lambda(\rho) \in \mathrm{V}_\lambda^d$. The main step of Keyl's algorithm is now to perform a normalized POVM within V_λ^d whose outcomes are unitary matrices in $\mathrm{U}(d)$. Specifically, his measurement maps a (Borel) subset $F \subseteq \mathrm{U}(d)$ to

$$M(F) := \int_F \pi_\lambda(U)\,|T_\lambda\rangle\,\langle T_\lambda|\,\pi_\lambda(U)^\dagger \cdot \dim(\mathrm{V}_\lambda^d)\,dU,$$

where dU denotes Haar measure on $\mathrm{U}(d)$. (To see that this is indeed a POVM — i.e., that $M := M(\mathrm{U}(d)) = I$ — first note that the translation invariance of Haar measure implies $\pi_\lambda(V)M\pi_\lambda(V)^\dagger =$

M for any $V \in \mathrm{U}(d)$. Thinking of π_λ as an irreducible representation of the unitary group, Schur's lemma implies M must be a scalar matrix. Taking traces shows M is the identity.)

We write $\mathrm{K}_\lambda(\rho)$ for the probability distribution on $\mathrm{U}(d)$ associated to this POVM; its density with respect to the Haar measure is therefore

$$\mathrm{tr}\left(\pi_\lambda(\tfrac{1}{s_\lambda(\alpha)}\rho)\pi_\lambda(U)\left|T_\lambda\right\rangle\left\langle T_\lambda\right|\pi_\lambda(U)^\dagger \cdot \dim(\mathrm{V}_\lambda^d)\right) = \Phi_\lambda(\alpha)^{-1}\cdot\left\langle T_\lambda\right|\pi_\lambda(U^\dagger\rho U)\left|T_\lambda\right\rangle. \tag{10}$$

Supposing the outcome of the measurement is U, Keyl's final estimate for ρ is $\hat{\rho} = U\mathrm{diag}(\underline{\lambda})U^\dagger$. Thus the expected Frobenius-squared error of Keyl's tomography algorithm is precisely

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{\lambda}\sim\mathrm{SW}^n(\alpha)\\ \boldsymbol{U}\sim\mathrm{K}_{\boldsymbol{\lambda}}(\rho)}} \|\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho\|_F^2.$$

Theorem 1.2, which we prove in this section, bounds the above quantity by $\frac{4d-3}{n}$. Let us assume now that $\mathrm{rank}(\rho) \le r$. Then $\ell(\boldsymbol{\lambda}) \le r$ always and hence the estimate $\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger$ will also have rank at most r. Thus by Cauchy–Schwarz applied to the singular values of $\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho$,

$$\mathbf{E}\, d_{\mathrm{tr}}(\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger, \rho) = \tfrac{1}{2}\,\mathbf{E}\,\|\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho\|_1 \le \tfrac{1}{2}\sqrt{2r}\,\mathbf{E}\,\|\underline{\boldsymbol{\lambda}} - \alpha\|_F \le \sqrt{r/2}\sqrt{\mathbf{E}\,\|\underline{\boldsymbol{\lambda}} - \alpha\|_F^2} \le \sqrt{\tfrac{O(rd)}{n}},$$

and Corollary 1.4 follows just as Corollary 1.3 did.

The remainder of this section is devoted to the proof of Theorem 1.2.

4.1 Integration formulas

Notation 4.1. Let $Z \in \mathbb{C}^{d\times d}$ and let λ be a partition of length at most d. The *generalized power function* Δ_λ is defined by

$$\Delta_\lambda(Z) = \prod_{k=1}^{d} \mathrm{pm}_k(Z)^{\lambda_k - \lambda_{k+1}},$$

where $\mathrm{pm}_k(Z)$ denotes the kth principal minor of Z (and $\lambda_{d+1} = 0$).

As noted by Keyl [Key06, equation (141)], when Z is positive semidefinite we have $\left\langle T_\lambda\right|\pi_\lambda(Z)\left|T_\lambda\right\rangle = \Delta_\lambda(Z)$; this follows by writing $Z = LL^\dagger$ for $L = (L_{ij})$ lower triangular with nonnegative diagonal and using the fact that $\Delta_\lambda(Z) = \Delta_\lambda(L^\dagger)^2 = \prod_{k=1}^d L_{kk}^{2\lambda_k}$. Putting this into (10) we have an alternate definition for the distribution $\mathrm{K}_\lambda(\rho)$:

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{K}_\lambda(\rho)} f(\boldsymbol{U}) = \Phi_\lambda(\alpha)^{-1}\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{U}(d)}\left[f(\boldsymbol{U})\cdot\Delta_\lambda(\boldsymbol{U}^\dagger\rho\boldsymbol{U})\right], \tag{11}$$

where $\boldsymbol{U} \sim \mathrm{U}(d)$ denotes that $\boldsymbol{U}$ has the Haar measure. For example, taking $f \equiv 1$ yields the identity

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{U}(d)} \Delta_\lambda(\boldsymbol{U}^\dagger\rho\boldsymbol{U}) = \Phi_\lambda(\alpha); \tag{12}$$

this expresses the fact that the spherical polynomial of weight λ for $\mathrm{GL}_d/\mathrm{U}(d)$ is precisely the normalized Schur polynomial (see, e.g., [Far15]). For a further example, taking $f(U) = \Delta_\mu(U^\dagger\rho U)$ and using the fact that $\Delta_\lambda\cdot\Delta_\mu = \Delta_{\lambda+\mu}$, we obtain

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{K}_\lambda(\rho)} \Delta_\mu(\boldsymbol{U}^\dagger\rho\boldsymbol{U}) = \frac{\Phi_{\lambda+\mu}(\alpha)}{\Phi_\lambda(\alpha)}; \quad \text{in particular,} \quad \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{K}_\lambda(\rho)} (\boldsymbol{U}^\dagger\rho\boldsymbol{U})_{1,1} = \frac{\Phi_{\lambda+e_1}(\alpha)}{\Phi_\lambda(\alpha)}. \tag{13}$$

For our proof of Theorem 1.2, we will need to develop and analyze a more general formula for the expected diagonal entry $\mathbf{E}(\boldsymbol{U}^\dagger\rho\boldsymbol{U})_{k,k}$. We begin with some lemmas.

Definition 4.2. For λ a partition and m a positive integer we define the following partition of height (at most) m:

$$\lambda^{[m]} = (\lambda_1 - \lambda_{m+1}, \dots, \lambda_m - \lambda_{m+1}).$$

We also define the following "complementary" partition $\lambda_{[m]}$ satisfying $\lambda = \lambda^{[m]} + \lambda_{[m]}$:

$$(\lambda_{[m]})_i = \begin{cases} \lambda_{m+1} & i \leq m, \\ \lambda_i & i \geq m+1. \end{cases}$$

Lemma 4.3. *Let $\rho \in \mathbb{C}^{d\times d}$ be a density matrix with spectrum α and let $\lambda \vdash n$ have height at most d. Let $m \in [d]$ and let f_m be an m-variate symmetric polynomial. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathrm{K}_\lambda(\rho)} f_m(\boldsymbol{\beta}) = \Phi_\lambda(\alpha)^{-1} \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathrm{U}(d)} \left[f_m(\boldsymbol{\beta}) \cdot \Phi_{\lambda^{[m]}}(\boldsymbol{\beta}) \cdot \Delta_{\lambda_{[m]}}(\boldsymbol{U}^\dagger \rho \boldsymbol{U}) \right],$$

where we write $\boldsymbol{\beta} = \mathrm{spec}_m(\boldsymbol{U}^\dagger\rho\boldsymbol{U})$ for the spectrum of the top-left $m\times m$ submatrix of $\boldsymbol{U}^\dagger\rho\boldsymbol{U}$.

Proof. Let $\boldsymbol{V} \sim \mathrm{U}(m)$ and write $\overline{\boldsymbol{V}} = \boldsymbol{V} \oplus I$, where I is the $(d-m)$-dimensional identity matrix. By translation-invariance of Haar measure we have $\boldsymbol{U}\overline{\boldsymbol{V}} \sim \mathrm{U}(d)$, and hence from (11),

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathrm{K}_\lambda(\rho)} f_m(\boldsymbol{\beta}) = \Phi_\lambda(\alpha)^{-1} \mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathrm{U}(d), \boldsymbol{V}\sim\mathrm{U}(m)} \left[f_m(\mathrm{spec}_m(\overline{\boldsymbol{V}}^\dagger \boldsymbol{U}^\dagger \rho \boldsymbol{U}\overline{\boldsymbol{V}})) \cdot \Delta_\lambda(\overline{\boldsymbol{V}}^\dagger \boldsymbol{U}^\dagger \rho \boldsymbol{U}\overline{\boldsymbol{V}}) \right]. \tag{14}$$

Note that conjugating a matrix by $\overline{\boldsymbol{V}}$ does not change the spectrum of its upper-left $k\times k$ block for any $k \geq m$. Thus $\mathrm{spec}_m(\overline{\boldsymbol{V}}^\dagger \boldsymbol{U}^\dagger \rho \boldsymbol{U}\overline{\boldsymbol{V}})$ is identical to β, and $\mathrm{pm}_k(\overline{\boldsymbol{V}}^\dagger \boldsymbol{U}^\dagger \rho \boldsymbol{U}\overline{\boldsymbol{V}}) = \mathrm{pm}_k(\boldsymbol{U}^\dagger\rho\boldsymbol{U})$ for all $k \geq m$. Thus using $\Delta_\lambda = \Delta_{\lambda_{[m]}} \cdot \Delta_{\lambda^{[m]}}$ we have

$$(14) = \Phi_\lambda(\alpha)^{-1} \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{U}(d)} \left[f_m(\boldsymbol{\beta}) \cdot \Delta_{\lambda_{[m]}}(\boldsymbol{U}^\dagger\rho\boldsymbol{U}) \cdot \mathop{\mathbf{E}}_{\boldsymbol{V}\sim\mathrm{U}(m)} \left[\Delta_{\lambda^{[m]}}(\overline{\boldsymbol{V}}^\dagger \boldsymbol{U}^\dagger \rho \boldsymbol{U}\overline{\boldsymbol{V}}) \right]\right].$$

But the inner expectation equals $\Phi_{\lambda^{[m]}}(\boldsymbol{\beta})$ by (12), completing the proof. ☐

Lemma 4.4. *In the setting of Lemma 4.3,*

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathrm{K}_\lambda(\rho)} \mathop{\mathrm{avg}}_{i=1}^{m} \left\{ (\boldsymbol{U}^\dagger\rho\boldsymbol{U})_{i,i} \right\} = \sum_{i=1}^m \frac{s_{\lambda^{[m]}+e_i}(1/m)}{s_{\lambda^{[m]}}(1/m)} \cdot \frac{\Phi_{\lambda+e_i}(\alpha)}{\Phi_\lambda(\alpha)}, \tag{15}$$

where $1/m$ abbreviates $1/m, \dots, 1/m$ (repeated m times).

Remark 4.5. The right-hand side of (15) is also a weighted average — of the quantities $\Phi_{\lambda+e_i}(\alpha)/\Phi_\lambda(\alpha)$ — by virtue of (5). The lemma also generalizes (13), as $s_{\lambda^{[1]}+e_1}(1)/s_{\lambda^{[1]}}(1)$ is simply 1.

Proof. On the left-hand side of (15) we have $\frac{1}{m}$ times the expected trace of the upper-left $m\times m$ submatrix of $\boldsymbol{U}^\dagger\rho\boldsymbol{U}$. So by applying Lemma 4.3 with $f_m(\beta) = \frac{1}{m}(\beta_1 + \cdots + \beta_m)$, it is equal to

$$\begin{aligned}
&\Phi_\lambda(\alpha)^{-1} \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{U}(d)} \left[\frac{1}{m}(\boldsymbol{\beta}_1 + \cdots + \boldsymbol{\beta}_m) \cdot \frac{s_{\lambda^{[m]}}(\boldsymbol{\beta})}{s_{\lambda^{[m]}}(1,\dots,1)} \cdot \Delta_{\lambda_{[m]}}(\boldsymbol{U}^\dagger\rho\boldsymbol{U}) \right] \\
&= \Phi_\lambda(\alpha)^{-1} \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{U}(d)} \left[\frac{1}{m} \sum_{i=1}^m \frac{s_{\lambda^{[m]}+e_i}(\boldsymbol{\beta})}{s_{\lambda^{[m]}}(1,\dots,1)} \cdot \Delta_{\lambda_{[m]}}(\boldsymbol{U}^\dagger\rho\boldsymbol{U}) \right] \qquad \text{(by Pieri (6))} \\
&= \Phi_\lambda(\alpha)^{-1} \cdot \sum_{i=1}^m \frac{s_{\lambda^{[m]}+e_i}(1,\dots,1)}{m \cdot s_{\lambda^{[m]}}(1,\dots,1)} \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathrm{U}(d)} \left[\Phi_{\lambda^{[m]}+e_i}(\boldsymbol{\beta}) \cdot \Delta_{\lambda_{[m]}}(\boldsymbol{U}^\dagger\rho\boldsymbol{U}) \right] \\
&= \Phi_\lambda(\alpha)^{-1} \cdot \sum_{i=1}^m \frac{s_{\lambda^{[m]}+e_i}(1,\dots,1)}{m \cdot s_{\lambda^{[m]}}(1,\dots,1)} \cdot \Phi_{\lambda+e_i}(\alpha),
\end{aligned}$$

where in the last step we used Lemma 4.3 again, with $f_m \equiv 1$ and $\lambda + e_i$ in place of λ. But this is equal to the right-hand side of (15), using the homogeneity of Schur polynomials. $\square$

Lemma 4.6. *Assume the setting of Lemma 4.3. Then* $\eta_i := \mathbf{E}_{\boldsymbol{U}\sim \mathrm{K}_\lambda(\rho)}(\boldsymbol{U}^\dagger \rho \boldsymbol{U})_{m,m}$ *is a convex combination of the quantities* $R_i := \Phi_{\lambda+e_i}(\alpha)/\Phi_\lambda(\alpha)$, $1 \le i \le m$.[1]

Proof. This is clear for $m = 1$. For $m > 1$, Remark 4.5 implies

$$\operatorname*{avg}_{i=1}^{m}\{\eta_i\} = p_1 R_1 + \cdots + p_m R_m, \quad \operatorname*{avg}_{i=1}^{m-1}\{\eta_i\} = q_1 R_1 + \cdots + q_m R_m,$$

where $p_1+\cdots+p_m = q_1+\cdots+q_m = 1$ and $q_m = 0$. Thus $\eta_i = \sum_{i=1}^m r_i R_i$, where $r_i = (mp_i - (m-1)q_i)$, and evidently $\sum_{i=1}^m r_i = m - (m-1) = 1$. It remains to verify that each $r_i \ge 0$. This is obvious for $i = m$; for $i < m$, we must check that

$$\frac{s_{\lambda^{[m]}+e_i}(1,\ldots,1)}{s_{\lambda^{[m]}}(1,\ldots,1)} \ge \frac{s_{\lambda^{[m-1]}+e_i}(1,\ldots,1)}{s_{\lambda^{[m-1]}}(1,\ldots,1)}. \tag{16}$$

Using the Weyl dimension formula from (1), one may explicitly compute that the ratio of the left side of (16) to the right side is precisely $1 + \frac{1}{(\lambda_i - \lambda_m)+(m-i)} \ge 1$. This completes the proof. $\square$

We will in fact only need the following corollary:

Corollary 4.7. *Let* $\rho \in \mathbb{C}^{d\times d}$ *be a density matrix with spectrum* α *and let* $\lambda \vdash n$ *have height at most* d. *Then* $\mathbf{E}_{\boldsymbol{U}\sim \mathrm{K}_\lambda(\rho)}(\boldsymbol{U}^\dagger \rho \boldsymbol{U})_{m,m} \ge \Phi_{\lambda+e_m}(\alpha)/\Phi_\lambda(\alpha)$ *for every* $m \in [d]$,

Proof. This is immediate from Lemma 4.6 and the fact that $\Phi_{\lambda+e_i}(\alpha) \ge \Phi_{\lambda+e_m}(\alpha)$ whenever $i < m$ (assuming $\lambda + e_i$ is a valid partition). This latter fact was recently proved by Sra [Sra15], verifying a conjecture of Cuttler et al. [CGS11]. $\square$

4.2 Proof of Theorem 1.2

Throughout the proof we assume $\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha)$ and $\boldsymbol{U} \sim \mathrm{K}_{\boldsymbol{\lambda}}(\rho)$. We have

$$\begin{aligned} &n^2 \cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho\|_F^2 = n^2 \cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\mathrm{diag}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{U}^\dagger\rho\boldsymbol{U}\|_F^2 \\ &= \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^d \boldsymbol{\lambda}_i^2 + \sum_{i=1}^d (\alpha_i n)^2 - 2n \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \sum_{i=1}^d \boldsymbol{\lambda}_i (\boldsymbol{U}^\dagger\rho\boldsymbol{U})_{i,i} \le dn + 2\sum_{i=1}^d(\alpha_i n)^2 - 2n \mathop{\mathbf{E}}_{\lambda} \sum_{i=1}^d \boldsymbol{\lambda}_i \mathop{\mathbf{E}}_{\boldsymbol{U}} (\boldsymbol{U}^\dagger\rho\boldsymbol{U})_{i,i}, \end{aligned} \tag{17}$$

using Lemma 3.1. Then by Corollary 4.7,

$$\begin{aligned} &\mathop{\mathbf{E}}_{\lambda} \sum_{i=1}^d \boldsymbol{\lambda}_i \mathop{\mathbf{E}}_{\boldsymbol{U}} (\boldsymbol{U}^\dagger\rho\boldsymbol{U})_{i,i} \ge \mathop{\mathbf{E}}_{\lambda} \sum_{i=1}^d \boldsymbol{\lambda}_i \frac{\Phi_{\boldsymbol{\lambda}+e_i}(\alpha)}{\Phi_{\boldsymbol{\lambda}}(\alpha)} = \mathop{\mathbf{E}}_{\lambda} \sum_{i=1}^d \boldsymbol{\lambda}_i \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \frac{s_{\boldsymbol{\lambda}}(1,\ldots,1)}{s_{\boldsymbol{\lambda}+e_i}(1,\ldots,1)} \\ &\ge \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^d \boldsymbol{\lambda}_i \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \left(2 - \frac{s_{\boldsymbol{\lambda}+e_i}(1,\ldots,1)}{s_{\boldsymbol{\lambda}}(1,\ldots,1)}\right) = 2\mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^d \boldsymbol{\lambda}_i \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} - \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^d \boldsymbol{\lambda}_i \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \frac{s_{\boldsymbol{\lambda}+e_i}(1,\ldots,1)}{s_{\boldsymbol{\lambda}}(1,\ldots,1)}, \end{aligned} \tag{18}$$

[1] To be careful, we may exclude all those i for which $\lambda + e_i$ is an invalid partition and thus $R_i = 0$.

where we used $r \geq 2 - \frac{1}{r}$ for $r > 0$. We lower-bound the first term in (18) by first using the inequality (2) and Proposition 2.1, and then using inequality (2) and Lemma 3.2 (as in the proof of Theorem 1.1):

$$2 \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^{d} \boldsymbol{\lambda}_i \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \geq 2 \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^{d} \boldsymbol{\lambda}_i \alpha_i \geq 2n \sum_{i=1}^{d} \alpha_i^2. \tag{19}$$

As for the second term in (18), we use (8) and the first formula in (1) to compute

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^{d} \boldsymbol{\lambda}_i \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \frac{s_{\boldsymbol{\lambda}+e_i}(1,\dots,1)}{s_{\boldsymbol{\lambda}}(1,\dots,1)} &= \sum_{i=1}^{d} \sum_{\lambda \vdash n} \dim(\lambda) s_{\lambda}(\alpha) \cdot \lambda_i \cdot \frac{s_{\lambda+e_i}(\alpha)}{s_{\lambda}(\alpha)} \frac{\dim(\lambda+e_i)(d+\lambda_i-i+1)}{\dim(\lambda)(n+1)} \\
&= \sum_{i=1}^{d} \sum_{\lambda \vdash n} \dim(\lambda+e_i) s_{\lambda+e_i}(\alpha) \cdot \frac{\lambda_i(d-i+\lambda_i+1)}{n+1} \\
&\leq \sum_{i=1}^{d} \mathop{\mathbf{E}}_{\boldsymbol{\lambda}' \sim \mathrm{SW}^{n+1}(\alpha)} \frac{(\boldsymbol{\lambda}'_i - 1)(d-i+\boldsymbol{\lambda}'_i)}{n+1} \qquad \text{(by (8) again)} \\
&\leq \frac{1}{n+1} \left(\mathop{\mathbf{E}}_{\boldsymbol{\lambda}' \sim \mathrm{SW}^{n+1}(\alpha)} \sum_{i=1}^{d} (\boldsymbol{\lambda}'_i)^2 + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}' \sim \mathrm{SW}^{n+1}(\alpha)} \sum_{i=1}^{d} (d-i-1)\boldsymbol{\lambda}'_i \right) \\
&\leq \frac{1}{n+1} \left((n+1)n \sum_{i=1}^{d} \alpha_i^2 + \sum_{i=1}^{d} (d+i-2)((n+1)/d) \right) \\
&= n \sum_{i=1}^{d} \alpha_i^2 + \frac{3}{2} d - \frac{3}{2}
\end{aligned} \tag{20}$$

where the last inequality is deduced exactly as in the proof of Lemma 3.1. Finally, combining (17)–(20) we get

$$n^2 \cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda}, \boldsymbol{U}} \|\boldsymbol{U} \mathrm{diag}(\underline{\boldsymbol{\lambda}}) \boldsymbol{U}^\dagger - \rho\|_F^2 \leq 4dn - 3n.$$

Dividing both sides by n^2 completes the proof. □

5 Truncated spectrum estimation

In this section we prove Theorem 1.7, from which Corollary 1.8 follows in the same way as Corollary 1.3. The key lemma involved is the following:

Lemma 5.1. *Let $\alpha \in \mathbb{R}^d$ be a sorted probability distribution. Then for any $k \in [d]$,*

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha)} \sum_{i=1}^{k} \boldsymbol{\lambda}_i \leq \sum_{i=1}^{k} \alpha_i n + 2\sqrt{2} k \sqrt{n}.$$

We remark that it is easy to *lower*-bound this expectation by $\sum_{i=1}^{k} \alpha_i n$ via Lemma 3.2. We now show how to deduce Theorem 1.7 from Lemma 5.1. Then in Section 5.1 we prove the lemma.

Proof of Theorem 1.7. Let $\boldsymbol{w} \sim \alpha^{\otimes n}$, let $\mathrm{RSK}(\boldsymbol{w}) = (\boldsymbol{P}, \boldsymbol{Q})$, and let $\boldsymbol{\lambda} = \mathrm{sh}(\boldsymbol{P})$, so $\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha)$. Write $\boldsymbol{w}'$ for the string formed from $\boldsymbol{w}$ by deleting all letters bigger than k. Then it is a basic property of the RSK algorithm that $\mathrm{RSK}(\boldsymbol{w}')$ produces the insertion tableau $\boldsymbol{P}'$ formed from $\boldsymbol{P}$ by deleting all boxes with labels bigger than k. Thus $\boldsymbol{\lambda}' = \mathrm{sh}(\boldsymbol{P}') = \mathrm{shRSK}(\boldsymbol{w}')$. Denoting

$\alpha_{[k]} = \alpha_1 + \cdots + \alpha_k$, we have $\boldsymbol{\lambda}' \sim \mathrm{SW}^{\boldsymbol{m}}(\alpha')$, where $\boldsymbol{m} \sim \mathrm{Binomial}(n, \alpha_{[k]})$ and α' denotes α conditioned on the first k letters; i.e., $\alpha' = (\alpha_i/\alpha_{[k]})_{i=1}^k$. Now by the triangle inequality,

$$2n \cdot \mathbf{E}\, d_{\mathrm{TV}}^{(k)}(\underline{\boldsymbol{\lambda}}, \alpha) = \mathbf{E} \sum_{i=1}^k |\boldsymbol{\lambda}_i - \alpha_i n| \le \mathbf{E} \sum_{i=1}^k (\boldsymbol{\lambda}_i - \boldsymbol{\lambda}'_i) + \mathbf{E} \sum_{i=1}^k \left|\boldsymbol{\lambda}'_i - \alpha'_i \boldsymbol{m}\right| + \sum_{i=1}^k \left|\alpha'_i \boldsymbol{m} - \alpha_i n\right|. \quad (21)$$

The first quantity in (21) is at most $2\sqrt{2}k\sqrt{n}$, using Lemma 5.1 and the fact that $\mathbf{E}[\sum_{i=1}^k \boldsymbol{\lambda}'_i] = \mathbf{E}[\boldsymbol{m}] = \sum_{i=1}^k \alpha_i n$. The second quantity in (21) is at most $k\sqrt{n}$ using Theorem 1.1:

$$\mathbf{E} \sum_{i=1}^k \left|\boldsymbol{\lambda}'_i - \alpha'_i \boldsymbol{m}\right| = \mathop{\mathbf{E}}_{\boldsymbol{m}} \boldsymbol{m} \cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda}'} \|\underline{\boldsymbol{\lambda}}' - \alpha'\|_1 \le \mathop{\mathbf{E}}_{\boldsymbol{m}} \boldsymbol{m}\sqrt{k}\sqrt{\mathop{\mathbf{E}}_{\boldsymbol{\lambda}'} \|\underline{\boldsymbol{\lambda}}' - \alpha'\|_2^2} \le k \mathop{\mathbf{E}}_{\boldsymbol{m}} \sqrt{\boldsymbol{m}} \le k\sqrt{n}.$$

And the third quantity in (21) is at most $\sqrt{n}$:

$$\mathop{\mathbf{E}}_{\boldsymbol{m}} \sum_{i=1}^k \left|\alpha'_i \boldsymbol{m} - \alpha_i n\right| = \mathop{\mathbf{E}}_{\boldsymbol{m}} \sum_{i=1}^k \tfrac{\alpha_i}{\alpha_{[k]}} \left|\boldsymbol{m} - \alpha_{[k]} n\right| = \mathop{\mathbf{E}}_{\boldsymbol{m}} |\boldsymbol{m} - \alpha_{[k]} n| \le \mathbf{stddev}(\boldsymbol{m}) \le \sqrt{n}.$$

Thus $2n \cdot \mathbf{E}\, d_{\mathrm{TV}}^{(k)}(\underline{\boldsymbol{\lambda}}, \alpha) \le ((2\sqrt{2}+1)k + 1)\sqrt{n}$, and dividing by $2n$ completes the proof. $\square$

5.1 Proof of Lemma 5.1

Our proof of Lemma 5.1 is essentially by reduction to the case when α is the uniform distribution and $k = 1$. We thus begin by analyzing the uniform distribution.

5.1.1 The uniform distribution case

In this subsection we will use the abbreviation $(1/d)$ for the uniform distribution $(1/d, \dots, 1/d)$ on $[d]$. Our goal is the following fact, which is of independent interest:

Theorem 5.2. $\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n(1/d)} \boldsymbol{\lambda}_1 \le n/d + 2\sqrt{n}$.

We remark that Theorem 5.2 implies Lemma 5.1 (with a slightly better constant) in the case of $\alpha = (1/d, \dots, 1/d)$, since of course $\boldsymbol{\lambda}_i \le \boldsymbol{\lambda}_1$ for all $i \in [k]$. Also, by taking $d \to \infty$ we recover the well known fact that $\mathbf{E}\,\boldsymbol{\lambda}_1 \le 2\sqrt{n}$ when $\boldsymbol{\lambda}$ has the Plancherel distribution. Indeed, our proof of Theorem 5.2 extends the original proof of this fact by Vershik and Kerov [VK85] (cf. the exposition in [Rom14]).

Proof. Consider the Schur–Weyl growth process under the uniform distribution $(1/d, \dots, 1/d)$ on $[d]$. For $m \ge 1$ we define

$$\delta_m = \mathbf{E}[\boldsymbol{\lambda}_1^{(m)} - \boldsymbol{\lambda}_1^{(m-1)}] = \mathbf{Pr}[\text{the } m\text{th box enters into the 1st row}] = \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^{m-1}(1/d)} \frac{s_{\boldsymbol{\lambda}+e_1}(1/d)}{s_{\boldsymbol{\lambda}}(1/d)},$$

where we used (5). By Cauchy–Schwarz and identity (8),

$$\begin{aligned}
\delta_m^2 &\le \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^{m-1}(1/d)} \left(\frac{s_{\boldsymbol{\lambda}+e_1}(1/d)}{s_{\boldsymbol{\lambda}}(1/d)}\right)^2 = \sum_{\lambda \vdash m-1} \dim(\lambda) s_\lambda(1/d) \cdot \left(\frac{s_{\lambda+e_1}(1/d)}{s_\lambda(1/d)}\right)^2 \\
&= \sum_{\lambda \vdash m-1} \dim(\lambda) s_{\lambda+e_1}(1/d) \cdot \left(\frac{s_{\lambda+e_1}(1/d)}{s_\lambda(1/d)}\right) = \sum_{\lambda \vdash m-1} \dim(\lambda+e_1) s_{\lambda+e_1}(1/d) \cdot \left(\frac{d+\lambda_1}{dm}\right) \qquad (22) \\
&\le \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^{m}(1/d)} \left(\frac{d+\boldsymbol{\lambda}_1}{dm}\right) = \left(\frac{d+\delta_1+\ldots+\delta_m}{dm}\right),
\end{aligned}$$

where the ratio in (22) was computed using the first formula of (1) (and the homogeneity of Schur polynomials). Thus we have established the following recurrence:

$$\delta_m \leq \frac{1}{\sqrt{dm}}\sqrt{d + \delta_1 + \cdots + \delta_m}. \tag{23}$$

We will now show by induction that $\delta_m \leq \frac{1}{d} + \frac{1}{\sqrt{m}}$ for all $m \geq 1$. Note that this will complete the proof, by summing over $m \in [n]$. The base case, $m = 1$, is immediate since $\delta_1 = 1$. For general $m > 1$, think of $\delta_1, \ldots, \delta_{m-1}$ as fixed and δ_m as variable. Now if δ_m satisfies (23), it is bounded above by the (positive) solution δ^* of

$$\delta = \frac{1}{\sqrt{dm}}\sqrt{c + \delta}, \qquad \text{where } c = d + \delta_1 + \cdots + \delta_{m-1}.$$

Note that if $\delta > 0$ satisfies

$$\delta \geq \frac{1}{\sqrt{dm}}\sqrt{c + \delta} \tag{24}$$

then it must be that $\delta \geq \delta^* \geq \delta_m$. Thus it suffices to show that (24) holds for $\delta = \frac{1}{d} + \frac{1}{\sqrt{m}}$. But indeed,

$$\begin{aligned}
\frac{1}{\sqrt{dm}}\sqrt{c + \frac{1}{d} + \frac{1}{\sqrt{m}}} &= \frac{1}{\sqrt{dm}}\sqrt{d + \delta_1 + \cdots + \delta_{m-1} + \frac{1}{d} + \frac{1}{\sqrt{m}}} \\
&\leq \frac{1}{\sqrt{dm}}\sqrt{d + \sum_{i=1}^{m}\left(\frac{1}{d} + \frac{1}{\sqrt{i}}\right)} \leq \frac{1}{\sqrt{dm}}\sqrt{d + \frac{m}{d} + 2\sqrt{m}} = \frac{1}{\sqrt{dm}}\left(\sqrt{d} + \sqrt{\frac{m}{d}}\right) = \frac{1}{d} + \frac{1}{\sqrt{m}},
\end{aligned}$$

where the first inequality used induction. The proof is complete. □

5.1.2 Reduction to the uniform case

Proof of Lemma 5.1. Given the sorted distribution α on $[d]$, let β be the sorted probability distribution on $[d]$ defined, for an appropriate value of m, as

$$\beta_1 = \alpha_1, \ldots, \beta_k = \alpha_k, \quad \beta_{k+1} = \ldots = \beta_m = \alpha_{k+1} > \beta_{m+1} \geq 0, \quad \beta_{m+2} = \ldots = \beta_d = 0.$$

In other words, β agrees with α on the first k letters and is otherwise uniform, except for possibly a small "bump" at β_{m+1}. By construction we have $\beta \succ \alpha$. Thus it follows from our coupling result, Theorem 1.11, that

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha)} \sum_{i=1}^{k} \boldsymbol{\lambda}_i \leq \mathop{\mathbf{E}}_{\boldsymbol{\mu} \sim \mathrm{SW}^n(\beta)} \sum_{i=1}^{k} \boldsymbol{\mu}_i,$$

and hence it suffices to prove the lemma for β in place of α. Observe that β can be expressed as a mixture

$$\beta = p_1 \cdot \mathcal{D}_1 + p_2 \cdot \mathcal{D}_2 + p_3 \cdot \mathcal{D}_3, \tag{25}$$

of a certain distribution $\mathcal{D}_1$ supported on $[k]$, the uniform distribution $\mathcal{D}_2$ on $[m]$, and the uniform distribution $\mathcal{D}_3$ on $[m+1]$. We may therefore think of a draw $\boldsymbol{\mu} \sim \mathrm{SW}^n(\beta)$ occurring as follows. First, $[n]$ is partitioned into three subsets $\boldsymbol{I}_1, \boldsymbol{I}_2, \boldsymbol{I}_3$ by including each $i \in [n]$ into $\boldsymbol{I}_j$ independently with probability p_j. Next we draw strings $\boldsymbol{w}^{(j)} \sim \mathcal{D}_j^{\otimes \boldsymbol{I}_j}$ independently for $j \in [3]$. Finally, we let

$\boldsymbol{w} = (\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}, \boldsymbol{w}^{(3)}) \in [d]^n$ be the natural composite string and define $\boldsymbol{\mu} = \mathrm{shRSK}(\boldsymbol{w})$. Let us also write $\boldsymbol{\mu}^{(j)} = \mathrm{shRSK}(\boldsymbol{w}^{(j)})$ for $j \in [3]$. We now claim that

$$\sum_{i=1}^{k} \boldsymbol{\mu}_i \leq \sum_{i=1}^{k} \boldsymbol{\mu}_i^{(1)} + \sum_{i=1}^{k} \boldsymbol{\mu}_i^{(2)} + \sum_{i=1}^{k} \boldsymbol{\mu}_i^{(3)}$$

always holds. Indeed, this follows from Greene's Theorem: the left-hand side is $|\boldsymbol{s}|$, where $\boldsymbol{s} \in [d]^n$ is a maximum-length disjoint union of k increasing subsequences in $\boldsymbol{w}$; the projection of $\boldsymbol{s}^{(j)}$ onto coordinates $\boldsymbol{I}_j$ is a disjoint union of k increasing subsequences in $\boldsymbol{w}^{(j)}$ and hence the right-hand side is at least $|\boldsymbol{s}^{(1)}| + |\boldsymbol{s}^{(2)}| + |\boldsymbol{s}^{(3)}| = |\boldsymbol{s}|$. Thus to complete the proof of the lemma, it suffices to show

$$\mathbf{E} \sum_{i=1}^{k} \boldsymbol{\mu}_i^{(1)} + \mathbf{E} \sum_{i=1}^{k} \boldsymbol{\mu}_i^{(2)} + \mathbf{E} \sum_{i=1}^{k} \boldsymbol{\mu}_i^{(3)} \leq \sum_{i=1}^{k} \alpha_i n + 2\sqrt{2}\, k\sqrt{n}. \tag{26}$$

Since $\mathcal{D}_1$ is supported on $[k]$, the first expectation above is equal to $\mathbf{E}[|\boldsymbol{w}^{(1)}|] = p_1 n$. By (the remark just after) Theorem 5.2, we can bound the second expectation as

$$\mathbf{E} \sum_{i=1}^{k} \boldsymbol{\mu}_i^{(2)} \leq k\, \mathbf{E}\, \boldsymbol{\mu}_1^{(2)} \leq k\, \mathbf{E}\, |\boldsymbol{w}^{(2)}|/m + 2k\, \mathbf{E} \sqrt{|\boldsymbol{w}^{(2)}|} \leq k(p_2 n)/m + 2k\sqrt{p_2 n}.$$

Similarly the third expectation in (26) is bounded by $k(p_3 n)/(m+1) + 2k\sqrt{p_3 n}$. Using $\sqrt{p_2} + \sqrt{p_3} \leq \sqrt{2}$, we have upper-bounded the left-hand side of (26) by

$$(p_1 + p_2 \tfrac{k}{m} + p_3 \tfrac{k}{m+1})n + 2\sqrt{2}\, k\sqrt{n} = \left(\sum_{i=1}^{k} \beta_i\right) n + 2\sqrt{2}\, k\sqrt{n},$$

as required. $\square$

6 Principal component analysis

In this section we analyze a straightforward modification to Keyl's tomography algorithm that allows us to perform principal component analysis on an unknown density matrix $\rho \in \mathbb{C}^{d\times d}$. The PCA algorithm is the same as Keyl's algorithm, except that having measured $\boldsymbol{\lambda}$ and $\boldsymbol{U}$, it outputs the rank-k matrix $\boldsymbol{U}\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger$ rather than the potentially full-rank matrix $\boldsymbol{U}\mathrm{diag}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger$. Here we recall the notation $\mathrm{diag}_{(k)}(\lambda)$ for the $d \times d$ diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_k$ followed by $d - k$ diagonal entries equal to 0.

Before giving the proof of Theorem 1.5, let us show why the case of Frobenius-norm PCA appears to be less interesting than the case of trace-distance PCA. The goal for Frobenius PCA would be to output a rank-k matrix $\widetilde{\rho}$ satisfying

$$\|\widetilde{\rho} - \rho\|_F \leq \sqrt{\alpha_{k+1}^2 + \ldots + \alpha_d^2} + \epsilon,$$

with high probability, while trying to minimize the number of copies n as a function of k, d, and ϵ. However, even when ρ is guaranteed to be of rank 1, it is likely that any algorithm will require $n = \Omega(d/\epsilon^2)$ copies to output an ϵ-accurate rank-1 approximator $\widetilde{\rho}$. This is because such an approximator will satisfy $\|\widetilde{\rho} - \rho\|_1 \leq \sqrt{2} \cdot \|\widetilde{\rho} - \rho\|_F = O(\epsilon)$, and it is likely that $n = \Omega(d/\epsilon^2)$ copies of ρ are required for such a guarantee (see, for example, the lower bounds of [HHJ+15], which show that $n = \Omega(\frac{d}{\epsilon^2 \log(d/\epsilon)})$ copies are necessary for tomography of rank-1 states.). Thus, even in the simplest case of rank-1 PCA of rank-1 states, we probably cannot improve on the $n = O(d/\epsilon^2)$ copy complexity for full tomography given by Corollary 1.4.

Now we prove Theorem 1.5. We note that the proof shares many of its steps with the proof of Theorem 1.2.

Proof of Theorem 1.5. Throughout the proof we assume $\boldsymbol{\lambda} \sim \mathrm{SW}^n(\alpha)$ and $\boldsymbol{U} \sim \mathrm{K}_{\boldsymbol{\lambda}}(\rho)$. We write $\boldsymbol{R}$ for the lower-right $(d-k) \times (d-k)$ submatrix of $\boldsymbol{U}^\dagger \rho \boldsymbol{U}$ and we write $\boldsymbol{\Gamma} = \boldsymbol{U}^\dagger \rho \boldsymbol{U} - \boldsymbol{R}$. Then

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\boldsymbol{U} \mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}}) \boldsymbol{U}^\dagger - \rho\|_1 = \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{U}^\dagger \rho \boldsymbol{U}\|_1 \leq \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{\Gamma}\|_1 + \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\boldsymbol{R}\|_1. \tag{27}$$

We can upper-bound the first term in (27) using

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{\Gamma}\|_1 \leq \sqrt{2k} \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{\Gamma}\|_F \leq \sqrt{2k} \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\mathrm{diag}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{U}^\dagger \rho \boldsymbol{U}\|_F \leq \sqrt{\frac{8kd}{n}}. \tag{28}$$

The first inequality is Cauchy–Schwarz together with the fact that $\mathrm{rank}(\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{\Gamma}) \leq 2k$ (since the matrix is nonzero only in its first k rows and columns). The second inequality uses that $\mathrm{diag}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{U}^\dagger \rho \boldsymbol{U}$ is formed from $\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}}) - \boldsymbol{\Gamma}$ by subtracting a matrix, $\boldsymbol{R}$, of disjoint support; this can only increase the squared Frobenius norm (sum of squares of entries). Finally, the third inequality uses Theorem 1.2. To analyze the second term in (27), we note that $\boldsymbol{R}$ is a principal submatrix of $\boldsymbol{U}^\dagger \rho \boldsymbol{U}$, and so it is positive semidefinite. As a result,

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\boldsymbol{R}\|_1 = \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \mathrm{tr}(\boldsymbol{R}) = 1 - \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \mathrm{tr}(\boldsymbol{\Gamma}). \tag{29}$$

By Corollary 4.7,

$$\begin{aligned} \mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \mathrm{tr}(\boldsymbol{\Gamma}) &= \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^k \mathop{\mathbf{E}}_{\boldsymbol{U}} (\boldsymbol{U}^\dagger \rho \boldsymbol{U})_{i,i} \geq \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^k \frac{\Phi_{\boldsymbol{\lambda}+e_i}(\alpha)}{\Phi_{\boldsymbol{\lambda}}(\alpha)} = \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^k \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \frac{s_{\boldsymbol{\lambda}}(1,\dots,1)}{s_{\boldsymbol{\lambda}+e_i}(1,\dots,1)} \\ &\geq \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^k \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \left(2 - \frac{s_{\boldsymbol{\lambda}+e_i}(1,\dots,1)}{s_{\boldsymbol{\lambda}}(1,\dots,1)}\right) = 2 \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^k \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} - \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^k \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \frac{s_{\boldsymbol{\lambda}+e_i}(1,\dots,1)}{s_{\boldsymbol{\lambda}}(1,\dots,1)}, \end{aligned} \tag{30}$$

where we used $r \geq 2 - \frac{1}{r}$ for $r > 0$. The first term here is lower-bounded using Proposition 2.1:

$$2 \mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^k \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \geq 2 \sum_{i=1}^k \alpha_i. \tag{31}$$

As for the second term in (30), we use (8) and the first formula in (1) to compute

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{\lambda}} \sum_{i=1}^{k} \frac{s_{\boldsymbol{\lambda}+e_i}(\alpha)}{s_{\boldsymbol{\lambda}}(\alpha)} \frac{s_{\boldsymbol{\lambda}+e_i}(1,\dots,1)}{s_{\boldsymbol{\lambda}}(1,\dots,1)} &= \sum_{i=1}^{k} \sum_{\lambda \vdash n} \dim(\lambda) s_\lambda(\alpha) \cdot \frac{s_{\lambda+e_i}(\alpha)}{s_\lambda(\alpha)} \frac{\dim(\lambda+e_i)(d+\lambda_i-i+1)}{\dim(\lambda)(n+1)} \\
&= \sum_{i=1}^{k} \sum_{\lambda \vdash n} \dim(\lambda+e_i) s_{\lambda+e_i}(\alpha) \cdot \frac{(d-i+\lambda_i+1)}{n+1} \\
&\le \sum_{i=1}^{k} \mathop{\mathbf{E}}_{\boldsymbol{\lambda}' \sim \mathrm{SW}^{n+1}(\alpha)} \frac{(d-i+\boldsymbol{\lambda}'_i)}{n+1} && \text{(by (8) again)} \\
&\le \frac{1}{n+1} \cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda}' \sim \mathrm{SW}^{n+1}(\alpha)} \sum_{i=1}^{k} \boldsymbol{\lambda}'_i + \frac{kd}{n} \\
&\le \sum_{i=1}^{k} \alpha_i + \frac{2\sqrt{2}k}{\sqrt{n}} + \frac{kd}{n}, && (32)
\end{aligned}$$

where the last step is by Lemma 5.1. Combining (27)–(32) we get

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\boldsymbol{U}\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho\|_1 \le \left(1 - \sum_{i=1}^{k} \alpha_i\right) + \sqrt{\frac{8kd}{n}} + \frac{2\sqrt{2}k}{\sqrt{n}} + \frac{kd}{n} \le \sum_{i=k+1}^{d} \alpha_i + \sqrt{\frac{32kd}{n}} + \frac{kd}{n},$$

where the second inequality used $k \le \sqrt{kd}$. Finally, as the expectation is also trivially upper-bounded by 2, we may use $6\sqrt{r} \ge \min(2, \sqrt{32r} + r)$ (which holds for all $r \ge 0$) to conclude

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda},\boldsymbol{U}} \|\boldsymbol{U}\mathrm{diag}_{(k)}(\underline{\boldsymbol{\lambda}})\boldsymbol{U}^\dagger - \rho\|_1 \le \sum_{i=k+1}^{d} \alpha_i + 6\sqrt{\frac{kd}{n}}. \qquad \square$$

7 Majorization for the RSK algorithm

In this section we prove Theorem 1.11. The key to the proof will be the following strengthened version of the $d = 2$ case, which we believe is of independent interest.

Theorem 7.1. *Let $0 \le p, q \le 1$ satisfy $|q - \frac{1}{2}| \ge |p - \frac{1}{2}|$; in other words, the q-biased probability distribution $(q, 1-q)$ on $\{1,2\}$ is "more extreme" than the p-biased distribution $(p, 1-p)$. Then for any $n \in \mathbb{N}$ there is a coupling $(\boldsymbol{w}, \boldsymbol{x})$ of the p-biased distribution on $\{1,2\}^n$ and the q-biased distribution on $\{1,2\}^n$ such that for all $1 \le i \le j \le n$ we have $\mathrm{LIS}(\boldsymbol{x}[i \mathinner{..} j]) \ge \mathrm{LIS}(\boldsymbol{w}[i \mathinner{..} j])$ always.*

We now show how to prove Theorem 1.11 given Theorem 7.1. Then in the following subsections we will prove Theorem 7.1.

Proof of Theorem 1.11 given Theorem 7.1. A classic result of Muirhead [Mui02] (see also [MOA11, B.1 Lemma]) says that $\beta \succ \alpha$ implies there is a sequence $\beta = \gamma_0 \succ \gamma_1 \succ \cdots \succ \gamma_t = \alpha$ such γ_i and γ_{i+1} differ in at most 2 coordinates. Since the $\trianglerighteq$ relation is transitive, by composing couplings it suffices to assume that α and β themselves differ in at most two coordinates. Since the Schur–Weyl distribution is symmetric with respect to permutations of $[d]$, we may assume that these two coordinates are 1 and 2. Thus we may assume $\alpha = (\alpha_1, \alpha_2, \beta_3, \beta_4, \dots, \beta_d)$, where $\alpha_1 + \alpha_2 = \beta_1 + \beta_2$ and α_1, α_2 are between β_1, β_2.

We now define the coupling $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ as follows: We first choose a string $\boldsymbol{z} \in (\{*\} \cup \{3, 4, \dots, d\})^n$ according to the product distribution in which symbol j has probability β_j for $j \geq 3$ and symbol $*$ has the remaining probability $\beta_1 + \beta_2$. Let $\boldsymbol{n}_*$ denote the number of $*$'s in $\boldsymbol{z}$. Next, we use Theorem 7.1 to choose coupled strings $(\boldsymbol{w}, \boldsymbol{x})$ with the p-biased distribution on $\{1, 2\}^{\boldsymbol{n}_*}$ and the q-biased distribution on $\{1, 2\}^{\boldsymbol{n}_*}$ (respectively), where $p = \frac{\alpha_1}{\beta_1 + \beta_2}$ and $q = \frac{\beta_1}{\beta_1 + \beta_2}$. Note indeed that $|q - \frac{1}{2}| \geq |p - \frac{1}{2}|$, and hence $\mathrm{LIS}(\boldsymbol{x}[i \mathinner{..} j]) \geq \mathrm{LIS}(\boldsymbol{w}[i \mathinner{..} j])$ for all $1 \leq i \leq \boldsymbol{n}_*$. Now let "$\boldsymbol{z} \cup \boldsymbol{w}$" denote the string in $[d]^n$ obtained by filling in the $*$'s in $\boldsymbol{z}$ with the symbols from $\boldsymbol{w}$, in the natural left-to-right order; similarly define "$\boldsymbol{z} \cup \boldsymbol{x}$". Note that $\boldsymbol{z} \cup \boldsymbol{w}$ is distributed according to the product distribution $\alpha^{\otimes n}$ and likewise for $\boldsymbol{z} \cup \boldsymbol{x}$ and $\beta^{\otimes n}$. Our final coupling is now obtained by taking $\boldsymbol{\lambda} = \mathrm{shRSK}(\boldsymbol{z} \cup \boldsymbol{w})$ and $\boldsymbol{\mu} = \mathrm{shRSK}(\boldsymbol{z} \cup \boldsymbol{x})$. We need to show that $\boldsymbol{\mu} \trianglerighteq \boldsymbol{\lambda}$ always.

By Greene's Theorem, it suffices to show that if $s_1, \dots, s_k$ are disjoint increasing subsequences in $\boldsymbol{z} \cup \boldsymbol{w}$ of total length S, we can find k disjoint increasing subsequences $s'_1, \dots, s'_k$ in $\boldsymbol{z} \cup \boldsymbol{x}$ of total length at least S. We first dispose of some simple cases. If none of $s_1, \dots, s_k$ contains any 1's or 2's, then we may take $s'_i = s_i$ for $i \in [k]$, since these subsequences all still appear in $\boldsymbol{z} \cup \boldsymbol{x}$. The case when exactly one of $s_1, \dots, s_k$ contains any 1's or 2's is also easy. Without loss of generality, say that s_k is the only subsequence containing 1's and 2's. We may partition it as (t, u), where t is a subsequence of $\boldsymbol{w}$ and u is a subsequence of the non-$*$'s in $\boldsymbol{z}$ that follow $\boldsymbol{w}$. Now let t' be the longest increasing subsequence in $\boldsymbol{x}$. As t is an increasing subsequence of $\boldsymbol{w}$, we know that t' is at least as long as t. Further, (t', u) is an increasing subsequence in $\boldsymbol{z} \cup \boldsymbol{x}$. Thus we may take $s'_i = s_i$ for $i < k$, and $s'_k = (t', u)$.

We now come to the main case, when at least two of $s_1, \dots, s_k$ contain 1's and/or 2's. Let's first look at the position $j \in [n]$ of the rightmost 1 or 2 among $s_1, \dots, s_k$. Without loss of generality, assume it occurs in s_k. Next, look at the position $i \in [n]$ of the rightmost 1 or 2 among $s_1, \dots, s_{k-1}$. Without loss of generality, assume it occurs in s_{k-1}. We will now modify the subsequences $s_1, \dots, s_k$ as follows:

- all 1's and 2's are deleted from $s_1, \dots, s_{k-2}$ (note that these all occur prior to position i);
- s_{k-1} is changed to consist of all the 2's within $(\boldsymbol{z} \cup \boldsymbol{w})[1 \mathinner{..} i]$;
- the portion of s_k to the right of position i is unchanged, but the preceding portion is changed to consist of all the 1's within $(\boldsymbol{z} \cup \boldsymbol{w})[1 \mathinner{..} i]$.

It is easy to see that the new $s_1, \dots, s_k$ remain disjoint subsequences of $\boldsymbol{z} \cup \boldsymbol{w}$, with total length at least S. We may also assume that the portion of s_k between positions $i+1$ and j consists of a longest increasing subsequence of $\boldsymbol{w}$.

Since the subsequences $s_1, \dots, s_{k-2}$ don't contain any 1's or 2's, they still appear in $\boldsymbol{z} \cup \boldsymbol{x}$, and we may take these as our $s'_1, \dots, s'_{k-2}$. We will also define s'_{k-1} to consist of all 2's within $(\boldsymbol{z} \cup \boldsymbol{x})[1 \mathinner{..} i]$. Finally, we will define s'_k to consist of all 1's within $(\boldsymbol{z} \cup \boldsymbol{z})[1 \mathinner{..} i]$, followed by the longest increasing subsequence of $\boldsymbol{x}$ occurring within positions $(i+1) \mathinner{..} j$ in $\boldsymbol{z} \cup \boldsymbol{x}$, followed by the portion of s_k to the right of position j (which does not contain any 1's or 2's and hence is still in $\boldsymbol{z} \cup \boldsymbol{x}$). It is clear that $s'_1, \dots, s'_k$ are indeed disjoint increasing subsequences of $\boldsymbol{z} \cup \boldsymbol{x}$. Their total length is the sum of four quantities:

- the total length of $s_1, \dots, s_{k-2}$;
- the total number of 1's and 2's within $(\boldsymbol{z} \cup \boldsymbol{x})[1 \mathinner{..} i]$;
- the length of the longest increasing subsequence of $\boldsymbol{x}$ occurring within positions $(i+1) \mathinner{..} j$ in $\boldsymbol{z} \cup \boldsymbol{x}$;

- the length of the portion of s_k to the right of position j.

By the coupling property of $(\boldsymbol{w}, \boldsymbol{x})$, the third quantity above is at least the length of the longest increasing subsequence of $\boldsymbol{w}$ occurring within positions $(i+1)\,..\,j$ in $\boldsymbol{z} \cup \boldsymbol{w}$. But this precisely shows that the total length of $s'_1, \dots, s'_k$ is at least that of $s_1, \dots, s_k$, as desired. $\square$

7.1 Substring-LIS-dominance: RSK and Dyck paths

In this subsection we make some preparatory definitions and observations toward proving Theorem 7.1. We begin by codifying the key property therein.

Definition 7.2. Let $w, w' \in \mathcal{A}^n$ be strings of equal length. We say w' *substring-LIS-dominates* w, notated $w' \ggg w$, if $\mathrm{LIS}(w'[i\,..\,j]) \geq \mathrm{LIS}(w[i\,..\,j])$ for all $1 \leq i \leq j \leq n$. (Thus the coupling in Theorem 7.1 satisfies $\boldsymbol{w} \ggg \boldsymbol{v}$ always.) The relation $\ggg$ is reflexive and transitive. If we have the substring-LIS-dominance condition just for $i = 1$ we say that w' *prefix-LIS-dominates* w. If we have it just for $j = n$ we say that w' *suffix-LIS-dominates* w.

Definition 7.3. For a string $w \in \mathcal{A}^n$ we write $\mathrm{behead}(w)$ for $w[2\,..\,n]$ and $\mathrm{curtail}(w)$ for $w[1\,..\,n-1]$.

Remark 7.4. We may equivalently define substring-LIS-dominance recursively, as follows. If w' and w have length 0 then $w' \ggg w$. If w' and w have length $n > 0$, then $w' \ggg w$ if and only if $\mathrm{LIS}(w') \geq \mathrm{LIS}(w)$ and $\mathrm{behead}(w') \ggg \mathrm{behead}(w)$ and $\mathrm{curtail}(w') \ggg \mathrm{curtail}(w)$. By omitting the second/third condition we get a recursive definition of prefix/suffix-LIS-dominance.

Definition 7.5. Let Q be a (nonempty) standard Young tableau. We define $\mathrm{curtail}(Q)$ to be the standard Young tableau obtained by deleting the box with maximum label from Q.

The following fact is immediate from the definition of the RSK correspondence:

Proposition 7.6. *Let $w \in \mathcal{A}^n$ be a nonempty string. Suppose* $\mathrm{RSK}(w) = (P, Q)$ *and* $\mathrm{RSK}(\mathrm{curtail}(w)) = (P', Q')$. *Then* $Q' = \mathrm{curtail}(Q)$.

The analogous fact for beheading is more complicated.

Definition 7.7. Let Q be a (nonempty) standard Young tableau. We define $\mathrm{behead}(Q)$ to be the standard Young tableau obtained by deleting the top-left box of Q, sliding the hole outside of the tableau according to jeu de taquin (see, e.g., [Ful97, Sag01]), and then decreasing all entries by 1. (The more traditional notation for $\mathrm{behead}(Q)$ is $\Delta(Q)$.)

The following fact is due to [Sch63]; see [Sag01, Proposition 3.9.3] for an explicit proof.[2]

Proposition 7.8. *Let $w \in \mathcal{A}^n$ be a nonempty string. Suppose* $\mathrm{RSK}(w) = (P, Q)$ *and* $\mathrm{RSK}(\mathrm{behead}(w)) = (P', Q')$. *Then* $Q' = \mathrm{behead}(Q)$.

Proposition 7.9. *Let $w, w' \in \mathcal{A}^n$ be strings of equal length and write* $\mathrm{RSK}(w) = (P, Q)$, $\mathrm{RSK}(w') = (P', Q')$. *Then whether or not $w' \ggg w$ can be determined just from the recording tableaus Q' and Q.*

Proof. This follows from the recursive definition of $\ggg$ given in Remark 7.4: whether $\mathrm{LIS}(w') \geq \mathrm{LIS}(w)$ can be determined by checking whether the first row of Q' is at least as long as the first row of Q; the recursive checks can then be performed with the aid of Propositions 7.6, 7.8. $\square$

[2] Technically, therein it is proved only for strings with distinct letters. One can recover the result for general strings in the standard manner; if the letters w_i and w_j are equal we break the tie by using the order relation on i, j. See also [vL13, Lemma].

Definition 7.10. In light of Proposition 7.9 we may define the relation $\ggg$ on standard Young tableaus.

Remark 7.11. The simplicity of Proposition 7.6 implies that it is very easy to tell, given $w, w' \in \mathcal{A}^n$ with recording tableaus Q and Q', whether w' suffix-LIS-dominates w. One only needs to check whether $Q'_{1j} \leq Q_{1j}$ for all $j \geq 1$ (treating empty entries as ∞). On the other hand, it is not particularly easy to tell from Q' and Q whether w' prefix-LIS-dominates w; one seems to need to execute all of the jeu de taquin slides.

We henceforth focus attention on alphabets of size 2. Under RSK, these yield standard Young tableaus with at most 2-rows. (For brevity, we henceforth call these *2-row Young tableaus*, even when they have fewer than 2 rows.) In turn, 2-row Young tableaus can be identified with Dyck paths (also known as ballot sequences).

Definition 7.12. We define a *Dyck path of length n* to be a path in the xy-plane that starts from $(0, 0)$, takes n steps of the form $(+1, +1)$ (an *upstep*) or $(+1, -1)$ (a *downstep*), and never passes below the x-axis. We say that the *height* of a step s, written $\mathrm{ht}(s)$, is the y-coordinate of its endpoint; the *(final) height* of a Dyck path W, written $\mathrm{ht}(W)$, is the height of its last step. We do *not* require the final height of a path to be 0; if it is we call the path *complete*, and otherwise we call it *incomplete*. A *return* refers to a point where the path returns to the x-axis; i.e., to the end of a step of height 0. An *arch* refers to a minimal complete subpath of a Dyck path; i.e., a subpath between two consecutive returns (or between the origin and the first return).

Definition 7.13. We identify each 2-row standard Young tableau Q of size n with a Dyck path W of length n. The identification is the standard one: reading off the entries of Q from 1 to n, we add an upstep to W when the entry is in the first row and a downstep when it is in the second row. The fact that this produces a Dyck path (i.e., the path does not pass below the x-axis) follows from the standard Young tableau property. Note that the final height of W is the difference in length between Q's two rows. We also naturally extend the terminology "return" to 2-row standard Young tableaus Q: a *return* is a second-row box labeled $2j$ such that boxes in Q labeled $1, \ldots, 2j$ form a rectangular $2 \times j$ standard Young tableau.

Definition 7.14. In light of Definition 7.10 and the above identification, we may define the relation $\ggg$ on Dyck paths.

Of course, we want to see how beheading and curtailment apply to Dyck paths. The following fact is immediate:

Proposition 7.15. *If W is the Dyck path corresponding to a nonempty 2-row standard Young tableau Q, then the Dyck path W' corresponding to* $\mathrm{curtail}(Q)$ *is formed from W by deleting its last segment. We write $W' =$* $\mathrm{curtail}(W)$ *for this new path.*

Again, the case of beheading is more complicated. We first make some definitions.

Definition 7.16. *Raising* refers to converting a downstep in a Dyck path to an upstep; note that this increases the Dyck path's height by 2. Conversely, *lowering* refers to converting an upstep to a downstep. Generally, we only allow lowering when the result is still a Dyck path; i.e., never passes below the x-axis.

Proposition 7.17. *Let Q be a nonempty 2-row standard Young tableau, with corresponding Dyck path W. Let W' be the Dyck path corresponding to* $\mathrm{behead}(Q)$*. Then W' is formed from W as follows: First, the initial step of W is deleted (and the origin is shifted to the new initial point).*

If W had no returns then the operation is complete and W' is the resulting Dyck path. Otherwise, if W had at least one return, then in the new path W' that step (which currently goes below the x-axis) is raised. In either case, we write $W' = \mathrm{behead}(W)$ for the resulting path.

Proof. We use Definitions 7.7 and 7.13. Deleting the top-left box of Q corresponds to deleting the first step of W, and decreasing all entries in Q by 1 corresponds to shifting the origin in W. Consider now the jeu de taquin slide in Q. The empty box stays in the first row until it first reaches a position j such that $Q_{1,j+1} > Q_{2,j}$ — if such a position exists. Such a position does exist if and only if Q contains a return (with box $(2, j)$ being the first such return). If Q (equivalently, W) has no return then the empty box slides out of the first row of Q, and indeed this corresponds to making no further changes to W. If Q has its first return at box $(2, j)$, this means the jeu de taquin will slide up the box labeled $2j$ (corresponding to raising the first return step in W); then all remaining slides will be in the bottom row of Q, corresponding to no further changes to W. □

Remark 7.18. Similar to Remark 7.11, it is easily to "visually" check the suffix-LIS-domination relation for Dyck paths: W' suffix-LIS-dominates W if and only if W' is at least as high as W throughout the length of both paths. On the other hand, checking the full substring-LIS-domination relation is more involved; we have $W' \rhd\!\!\gg W$ if and only if for any number of simultaneous beheadings to W' and W, the former path always stays at least as high as the latter.

Finally, we will require the following definition:

Definition 7.19. A *hinged range* is a sequence $(R_0, s_1, R_1, s_2, R_2, \ldots, s_k, R_k)$ (with $k \geq 0$), where each s_i is a step (upstep or downstep) called a *hinge* and each R_i is a Dyck path (possibly of length 0) called a *range*. The "internal ranges" $R_1, \ldots, R_{k-1}$ are required to be complete Dyck paths; the "external ranges" R_0 and R_k may be incomplete.

We may identify the hinged range with the path formed by concatenating its components; note that this need not be a Dyck path, as it may pass below the origin.

If H is a hinged range and H' is formed by raising zero or more of its hinges (i.e., converting downstep hinges to upsteps), we say that H' is a *raising* of H or, equivalently, that H is a *lowering* of H'. We call a hinged range *fully lowered* (respectively, *fully raised*) if all its hinges are downsteps (respectively, upsteps).

7.2 A bijection on Dyck paths

Theorem 7.20. *Fix integers $n \geq 2$ and $1 \leq \lambda_2 \leq \lfloor \frac{n}{2} \rfloor$. Define*

$$\begin{aligned} \mathcal{W} = \big\{(W, s_1) :\ & W \text{ is a length-}n \text{ Dyck path with exactly } \lambda_2 \text{ downsteps;} \\ & s_1 \text{ is a downstep in } W\big\} \end{aligned}$$

and

$$\begin{aligned} \mathcal{W}' = \bigcup_{k=1}^{\lambda_2} \big\{(W', s_1') :\ & W' \text{ is a length-}n \text{ Dyck path with exactly } \lambda_2 - k \text{ downsteps;} \\ & s_1' \text{ is an upstep in } W' \text{ with } k+1 \leq \mathrm{ht}(s_1') \leq \mathrm{ht}(W') - k + 1; \\ & s_1' \text{ is the rightmost upstep in } W' \text{ of its height}\big\}. \end{aligned}$$

Then there is an explicit bijection $f : \mathcal{W} \to \mathcal{W}'$ such that whenever $f(W, s_1) = (W', s_1')$ it holds that $W' \rhd\!\!\gg W$.

Remark 7.21. Each length-n Dyck path with exactly λ_2 downsteps occurs exactly λ_2 times in $\mathcal{W}$. Each length-n Dyck path with strictly fewer than λ_2 downsteps occurs exactly $n - 2\lambda_2 + 1$ times in $\mathcal{W}'$.

Proof of Theorem 7.20. Given any $(W, s_1) \in \mathcal{W}$, we define f's value on it as follows. Let s_2 be the first downstep following s_1 in W having height $\mathrm{ht}(s_1) - 1$; let s_3 be the first downstep following s_2 in W following s_2 having height $\mathrm{ht}(s_2) - 1$; etc., until reaching downstep s_k having no subsequent downstep of smaller height. Now decompose W as a (fully lowered) hinged range $H = (R_0, s_1, R_1, \ldots, s_k, R_k)$. Let $H' = (R'_0, s'_1, R'_1, \ldots, s'_k, R'_k)$ be the fully raised version of H (where each R'_j is just R_j and each s'_j is an upstep). Then $f(W, s_k)$ is defined to be (W', s'_1), where W' is the Dyck path corresponding to H'.

First we check that indeed $(W', s'_1) \in \mathcal{W}'$. As W' is formed from W by k raisings, it has exactly $\lambda_2 - k$ downsteps. Since $\mathrm{ht}(s_k) \geq 0$ it follows that $\mathrm{ht}(s_1) \geq k - 1$ and hence $\mathrm{ht}(s'_1) \geq k + 1$. On the other hand, $\mathrm{ht}(s'_1) + (k - 1) = \mathrm{ht}(s'_k) \leq \mathrm{ht}(W')$ and so $\mathrm{ht}(s'_1) \leq \mathrm{ht}(W') - k + 1$. Finally, s'_1 is the rightmost upstep in W' of its height because H' is fully raised.

To show that f is a bijection, we will define the function $g : \mathcal{W}' \to \mathcal{W}$ that will evidently be f's inverse. Given any $(W', s'_1) \in \mathcal{W}$, with W' having exactly $\lambda_2 - k$ downsteps, we define g's value on it as follows. Let s'_2 be the *last* (rightmost) upstep following s'_1 in W' having height $\mathrm{ht}(s'_1) + 1$; let s'_3 be the last upstep following s'_2 in W' having height $\mathrm{ht}(s'_2) + 1$; etc., until s'_k is defined. That this s'_k indeed exists follows from the fact that $\mathrm{ht}(s'_1) \leq \mathrm{ht}(W') - k + 1$. Now decompose W' as a (fully raised) hinged range $H' = (R'_0, s'_1, R'_1, \ldots, s'_k, R'_k)$. The fact that R'_k is a Dyck path (i.e., does not pass below its starting height) again follows from the fact that $\mathrm{ht}(s'_k) = \mathrm{ht}(s'_1) + k - 1 \leq \mathrm{ht}(W')$. Finally, let $H = (R_0, s_1, R_1, \ldots, s_k, R_k)$ be the fully lowered version of H', and W the corresponding path. As W has exactly λ_2 downsteps, we may define $g(W', s'_1) = (W, s_1)$ provided W is indeed a Dyck path. But this is the case, because the lowest point of W occurs at the endpoint of s_k, and $\mathrm{ht}(s_k) = \mathrm{ht}(s_1) - k + 1 = \mathrm{ht}(s'_1) - 2 - k + 1 = \mathrm{ht}(s'_1) - k - 1 \geq 0$ since $\mathrm{ht}(s'_1) \geq k + 1$.

It is fairly evident that f and g are inverses. The essential thing to check is that the sequence $s_1, \ldots, s_k$ determined from s_1 when computing $f(W, s_1)$ is "the same" (up to raising/lowering) as the sequence $s'_1, \ldots, s'_{k'}$ determined from s'_1 in computing $g(W', s'_1)$, and vice versa. The fact that the sequences have the same *length* follows, in the $g \circ f = id$ case, from the fact that $\mathrm{ht}(W') = \mathrm{ht}(W) + 2k$; it follows, in the $f \circ g = id$ case, from the fact that R'_k is a Dyck path. The fact that the hinges have the same identity is evident from the nature of fully raising/lowering hinged ranges.

It remains to show that if $f(W, s_1) = (W', s'_1)$ then $W' \ggg W$. Referring to Remark 7.18, we need to show that if W' and W are both simultaneously beheaded some number of times b, then in the resulting paths, W' is at least as high as W throughout their lengths. In turn, this is implied by the following more general statement:

Claim 7.22. *After b beheadings, W' and W may be expressed as hinged ranges $H' = (R_0, s'_1, R_1, \ldots, s'_k, R_k)$ and $H = (R_0, s_1, R_1, \ldots, s_k, R_k)$ (respectively) such that H' is the fully raised version of H (i.e., each s'_j is an upstep).*

(Note that we do not necessarily claim that H is the fully lowered version of H'.)

The claim can be proved by induction on b. The base case $b = 0$ follows by definition of f. Throughout the induction we may assume that the common initial Dyck path R_0 is nonempty, as otherwise s_1 must be an upstep, in which case we can redefine the common initial Dyck path of W and W' to be $(s_1, R_1) = (s'_1, R_1)$.

We now show the inductive step. Assume W' and W are nonempty paths as in the claim's statement, with R_0 nonempty. Suppose now that W' and W are simultaneously beheaded. The first step of W' and W (an upstep belonging to R_0) is thus deleted, and the origin shifted. If R_0 contained a downstep to height 0 then the first such downstep is raised in both behead(W') and behead(W) and the inductive claim is maintained. Otherwise, suppose R_0 contained no downsteps to height 0. It follows immediately that W' originally had no returns to height 0 at all; hence the beheading of W' is completed by the deletion of its first step. It may also be that W had no returns to height 0 at all; then the beheading of W is also completed by the deletion of its first step and the induction hypothesis is clearly maintained. On the other hand, W *may* have had some downsteps to 0 within $(s_1, R_1, \ldots, s_k, R_k)$. In this case, the first (leftmost) such downstep must occur at one of the hinges s_j, and the beheading of W is completed by raising this hinge. The inductive hypothesis is therefore again maintained. This completes the induction. $\square$

We derive an immediate corollary, after introducing a bit of notation:

Definition 7.23. We write $\mathrm{SYT}_n(=\lambda_2)$ (respectively, $\mathrm{SYT}_n(\leq\lambda_2)$) for the set of 2-row standard Young tableaus of size n with exactly (respectively, at most) λ_2 boxes in the second row.

Corollary 7.24. *For any integers $n \geq 2$ and $0 \leq \lambda_2 \leq \lfloor \frac{n}{2} \rfloor$, there is a coupling $(\boldsymbol{Q}, \boldsymbol{Q}')$ of the uniform distribution on $\mathrm{SYT}_n(=\lambda_2)$ and the uniform distribution on $\mathrm{SYT}_n(\leq\lambda_2 - 1)$ such that $\boldsymbol{Q}' \rhd\!\!\!\gg \boldsymbol{Q}$ always.*

Proof. Let $(\boldsymbol{W}, \boldsymbol{s}_1)$ be drawn uniformly at random from the set $\mathcal{W}$ defined in Theorem 7.20, and let $(\boldsymbol{W}', \boldsymbol{s}_1') = f(\boldsymbol{W}, \boldsymbol{s}_1)$. Let $\boldsymbol{Q} \in \mathrm{SYT}_n(=\lambda_2)$, $\boldsymbol{Q}' \in \mathrm{SYT}_n(\leq\lambda_2 - 1)$ be the 2-row standard Young tableaus identified with $\boldsymbol{W}$, $\boldsymbol{W}'$ (respectively). Then Theorem 7.20 tells us that $\boldsymbol{Q}' \rhd\!\!\!\gg \boldsymbol{Q}$ always, and Remark 7.21 tells us that $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ are each uniformly distributed. $\square$

Corollary 7.25. *For any integers $n \geq 0$ and $0 \leq \lambda_2' \leq \lambda_2 \leq \lfloor \frac{n}{2} \rfloor$, there is a coupling $(\boldsymbol{Q}, \boldsymbol{Q}')$ of the uniform distribution on $\mathrm{SYT}_n(\leq\lambda_2)$ and the uniform distribution on $\mathrm{SYT}_n(\leq\lambda_2')$ such that $\boldsymbol{Q}' \rhd\!\!\!\gg \boldsymbol{Q}$ always.*

Proof. The cases $n < 2$ and $\lambda_2' = \lambda_2$ are trivial, so we may assume $n \geq 2$ and $0 \leq \lambda_2' < \lambda_2 \leq \lfloor \frac{n}{2} \rfloor$. By composing couplings and using transitivity of $\rhd\!\!\!\gg$, it suffices to treat the case $\lambda_2' = \lambda_2 - 1$. But the uniform distribution on $\mathrm{SYT}_n(\leq\lambda_2)$ is a mixture of (a) the uniform distribution on $\mathrm{SYT}_n(=\lambda_2)$, (b) the uniform distribution on $\mathrm{SYT}_n(\leq\lambda_2 - 1)$; and these can be coupled to $\mathrm{SYT}_n(\leq\lambda_2 - 1)$ under the $\rhd\!\!\!\gg$ relation using (a) Corollary 7.24, (b) the identity coupling. $\square$

Before giving the next corollary, we have a definition.

Definition 7.26. Let $\mathcal{A}$ be any 2-letter alphabet. We write $\mathcal{A}^n_k$ for the set of length-n strings over $\mathcal{A}$ with exactly k copies of the larger letter, and we write $\mathcal{A}^n_{k,n-k} = \mathcal{A}^n_k \cup \mathcal{A}^n_{n-k}$.

Corollary 7.27. *For $\mathcal{A}$ a 2-letter alphabet and integers $0 \leq k' \leq k \leq \lfloor \frac{n}{2} \rfloor$, there is a coupling $(\boldsymbol{w}, \boldsymbol{w}')$ of the uniform distribution on $\mathcal{A}^n_{k,n-k}$ and the uniform distribution on $\mathcal{A}^n_{k',n-k'}$ such that $\boldsymbol{w}' \rhd\!\!\!\gg \boldsymbol{w}$ always.*

Proof. We first recall that if $\boldsymbol{x} \sim \mathcal{A}^n_k$ is uniformly random and $(\boldsymbol{P}, \boldsymbol{Q}) = \mathrm{RSK}(\boldsymbol{x})$, then the recording tableau $\boldsymbol{Q}$ is uniformly random on $\mathrm{SYT}_n(\leq k)$. This is because for each possible recording tableau $Q \in \mathrm{SYT}_n(\leq k)$ there is a unique insertion tableau P of the same shape as Q having exactly k boxes labeled with the larger letter of $\mathcal{A}$. (Specifically, if $P \vdash (\lambda_1, \lambda_2)$, then the last $k - \lambda_2$ boxes of P's first row, and all of the boxes of P's second row, are labeled with A's larger letter.) It follows that the same is true if $\boldsymbol{x} \sim \mathcal{A}^n_{k,n-k}$ is uniformly random. But now the desired coupling follows from Corollary 7.25 (recalling Definition 7.10). $\square$

In fact, Corollary 7.27 is fundamentally stronger than our desired Theorem 7.1, as we now show:

Proof of Theorem 7.1. For $r \in [0,1]$, suppose we draw an r-biased string $\boldsymbol{y} \in \{1,2\}^n$ and define the random variable $\boldsymbol{j}$ such that $\boldsymbol{y} \in \{1,2\}^n_{\boldsymbol{j},n-\boldsymbol{j}}$. (Note that given $\boldsymbol{j}$, the string $\boldsymbol{y}$ is uniformly distributed on $\{1,2\}^n_{\boldsymbol{j},n-\boldsymbol{j}}$.) Write $L_r(\ell)$ for the cumulative distribution function of $\boldsymbol{j}$; i.e., $L_r(\ell) = \mathbf{Pr}[\boldsymbol{y} \in \cup_{j\le\ell}\{1,2\}^n_{j,n-j}]$, where $\boldsymbol{y}$ is r-biased.

Claim: $L_q(\ell) \ge L_p(\ell)$ for all $0 \le \ell \le \lfloor \frac{n}{2} \rfloor$.

Before proving the claim, let us show how it is used to complete the proof of Theorem 7.1. We define the required coupling $(\boldsymbol{w}, \boldsymbol{x})$ of p-biased and q-biased distributions as follows: First we choose $\boldsymbol{\theta} \in [0,1]$ uniformly at random. Next we define $\boldsymbol{k}$ (respectively, $\boldsymbol{k}'$) to be the least integer such that $L_p(\boldsymbol{k}) \ge \boldsymbol{\theta}$ (respectively, $L_q(\boldsymbol{k}') \ge \boldsymbol{\theta}$); from the claim it follows that $\boldsymbol{k}' \le \boldsymbol{k}$ always. Finally, we let $(\boldsymbol{w}, \boldsymbol{x})$ be drawn from the coupling on $\{1,2\}^n_{\boldsymbol{k},n-\boldsymbol{k}}$ and $\{1,2\}^n_{\boldsymbol{k}',n-\boldsymbol{k}'}$ specified in Corollary 7.27. Then as required, we have that $\boldsymbol{x}' \ggg \boldsymbol{w}$ always, and that $\boldsymbol{w}$ has the p-biased distribution and $\boldsymbol{x}$ has the q-biased distribution.

It therefore remains to prove the claim. We may exclude the trivial cases $\ell = \frac{n}{2}$ or $q \in \{0,1\}$, where $L_q(\ell) = 1$. Also, since $L_r(\ell) = L_{1-r}(\ell)$ by symmetry, we may assume $0 < q \le p \le \frac{1}{2}$. Thus it suffices to show that $\frac{d}{dr}L_r(\ell) \le 0$ for $0 < r \le \frac{1}{2}$. Letting $\boldsymbol{h}$ denote the "Hamming weight" (number of 2's) in an r-biased random string on $\{1,2\}^n$, we have

$$\begin{aligned} L_r(\ell) &= \mathbf{Pr}[\boldsymbol{h} \le \ell] + \mathbf{Pr}[\boldsymbol{h} \ge n-\ell] = 1 - \mathbf{Pr}[\boldsymbol{h} > \ell] + \mathbf{Pr}[\boldsymbol{h} > n-\ell-1] \\ \Rightarrow \frac{d}{dr}L_r(\ell) &= -\frac{d}{dr}\mathbf{Pr}[\boldsymbol{h} > \ell] + \frac{d}{dr}\mathbf{Pr}[\boldsymbol{h} > n-1-\ell]. \end{aligned}$$

(The first equality used $\ell < \frac{n}{2}$.) But it is a basic fact that $\frac{d}{dr}\mathbf{Pr}[\boldsymbol{h} > t] = n\binom{n-1}{t}r^t(1-r)^{n-1-t}$. Thus

$$\frac{d}{dr}L_r(\ell) = n\binom{n-1}{\ell}\left(-r^\ell(1-r)^{n-1-\ell} + r^{n-1-\ell}(1-r)^\ell\right),$$

and we may verify this is indeed nonpositive:

$$-r^\ell(1-r)^{n-1-\ell} + r^{n-1-\ell}(1-r)^\ell \le 0 \iff 1 \le \left(\tfrac{1-r}{r}\right)^{n-1-2\ell},$$

which is true since $0 < r \le \frac{1}{2}$ and $n-1-2\ell \ge 0$ (using $\ell < \frac{n}{2}$ again). $\square$

References

[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. 1, 2

[Aud06] Koenraad Audenaert. A digest on representation theory of the symmetric group. Found at http://personal.rhul.ac.uk/usah/080/qitnotes_files/irreps_v06.pdf, 2006. 2

[BCG13] Konrad Banaszek, Marcus Cramer, and David Gross. Focus on quantum tomography. *New Journal of Physics*, 15(12):125020, 2013. 1

[CGS11] Allison Cuttler, Curtis Greene, and Mark Skandera. Inequalities for symmetric means. *European Journal of Combinatorics*, 32(6):745–761, 2011. 4.1

[CHW07] Andrew Childs, Aram Harrow, and Paweł Wocjan. Weak Fourier-Schur sampling, the hidden subgroup problem, and the quantum collision problem. In *24th Annual Symposium on Theoretical Aspects of Computer Science*, pages 598–609, 2007. 3

[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. 1, 3

[Far15] Jacques Faraut. Rayleigh theorem, projection of orbital measures and spline functions. *Advances in Pure and Applied Mathematics*, 2015. 4.1

[FGLE12] Steven Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012. 1

[Ful97] William Fulton. *Young tableaux: with applications to representation theory and geometry*. Cambridge University Press, 1997. 1.2, 2, 7.7

[Gre74] Curtis Greene. An extension of Schensted's theorem. *Advances in Mathematics*, 14:254–265, 1974. 2

[Har05] Aram Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, Massachusetts Institute of Technology, 2005. 2

[Har15] Aram Harrow, 2015. http://dabacon.org/pontiff/?p=10785. 1

[HHJ+15] Jeongwan Haah, Aram Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. Preprint, August 2015. 1.3, 6

[HM02] Masahito Hayashi and Keiji Matsumoto. Quantum universal variable-length source coding. *Physical Review A*, 66(2):022311, 2002. 1, 3

[ITW01] Alexander Its, Craig Tracy, and Harold Widom. Random words, Toeplitz determinants and integrable systems I. In *Random Matrices and their Applications*, pages 245–258. Cambridge University Press, 2001. 2

[Key06] Michael Keyl. Quantum state estimation and large deviations. *Reviews in Mathematical Physics*, 18(01):19–60, 2006. 1, 1.4, 4, 4.1

[KRT14] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. Technical report, arXiv:1410.6913, 2014. 1, 1, 1.4

[KW01] Michael Keyl and Reinhard Werner. Estimating the spectrum of a density operator. *Physical Review A*, 64(5):052311, 2001. 1

[MOA11] Albert W Marshall, Ingram Olkin, and Barry Arnold. *Inequalities: theory of majorization and its applications*. Springer Series in Statistics, 2011. 7

[Mui02] Robert Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proceedings of the Edinburgh Mathematical Society*, 21:144–162, 1902. 7

[O'C03] Neil O'Connell. Conditioned random walks and the RSK correspondence. *Journal of Physics A: Mathematical and General*, 36(12):3049, 2003. 2

[OW15] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, 2015. 3

[Rom14] Dan Romik. *The surprising mathematics of longest increasing subsequences*. Cambridge University Press, 2014. 5.1.1

[Sag01] Bruce Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*. Springer, 2001. 7.7, 7.1

[Sch63] Marcel-Paul Schützenberger. Quelques remarques sur une construction de Schensted. *Mathematica Scandinavica*, 12:117–128, 1963. 7.1

[Sra15] Suvrit Sra. On inequalities for normalized Schur functions. *European Journal of Combinatorics*, 2015. 4.1

[Sza82] Stanisław Szarek. Nets of Grassmann manifold and orthogonal group. In *Proceedings of research workshop on Banach space theory*, pages 169–185. University of Iowa, Iowa City, IA, 1982. 1

[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. 5.1.1

[vL13] Mark van Leeuwen, 2013. `http://mathoverflow.net/a/140739/658`. 2

[Wri15] `http://www.cs.cmu.edu/~jswright`, 2015. 1.3