Optimal high-precision shadow estimation

Sitan Chen * Jerry Li † Allen Liu ‡

July 22, 2024

Abstract

We give the first tight sample complexity bounds for shadow tomography and classical shadows in the regime where the target error is below some sufficiently small inverse polynomial in the dimension of the Hilbert space. Formally we give a protocol that, given any $m \in \mathbb{N}$ and $\varepsilon \leq O(d^{-12})$, measures $O(\log(m)/\varepsilon^2)$ copies of an unknown mixed state $\rho \in \mathbb{C}^{d \times d}$ and outputs a classical description of ρ which can then be used to estimate any collection of m observables to within additive accuracy ε. Previously, even for the simpler task of shadow tomography – where the m observables are known in advance – the best known rates either scaled benignly but suboptimally in all of m, d, ε [3, 31], or scaled optimally in ε, m but had additional polynomial factors in d for general observables [19]. Intriguingly, we also show via dimensionality reduction, that we can rescale ε and d to reduce to the regime where $\varepsilon \leq O(d^{-1/2})$. Our algorithm draws upon representation-theoretic tools recently developed in the context of full state tomography [11].

*SEAS, Harvard University. Email: sitan@seas.harvard.edu;

†Microsoft Research. Email: jerrl@microsoft.com;

‡MIT. Email: cliu568@mit.edu; This work was supported in part by an NSF Graduate Research Fellowship and a Fannie and John Hertz Foundation Fellowship

Contents

1 Introduction 3

2 Our contributions 4

3 Technical overview 6

4 Outlook 7

A Representation Theory and Keyl's POVM 11

B Basic Facts 12

C Balanced Case 12

D Splitting Reduction 15

E Reducing to Small ε via Random Projection 17

1 Introduction

In this paper, we consider the well-studied and related problems of shadow tomography [1, 2, 3, 31] and learning classical shadows [19, 18, 14]. In both problems, there is an unknown quantum state, and the goal is to simultaneously estimate as many linear properties of this state as possible, using as few copies of ρ as possible. Such problems come up naturally in a variety of real-world laboratory settings, see e.g. [28, 5, 4, 24, 21]. The power of these frameworks is that these shadow estimation tasks can be performed *efficiently*. That is, to simultaneously predict m properties of the state, one only requires $\mathrm{poly}(\log m)$ many copies of the state, and $\mathrm{poly}(\log m)$ time. Indeed, algorithms for classical shadows have already shown immense potential in real-world evaluations [33, 30, 23].

Formally, we let $\rho \in \mathbb{C}^{d\times d}$ be an arbitrary and unknown d-dimensional mixed state. In shadow tomography, we are given a set of m linear observables $\{O_i\}_{i=1}^m \subset \mathbb{C}^{d\times d}$ satisfying $0 \preceq O_i \preceq I$, and given n copies of ρ, our goal is to estimate $\mathrm{tr}(O_i\rho)$ to accuracy ε, for all $i = 1, \ldots, m$, with high probability.

In classical shadows, the setup is the same, except that the measurements must be chosen obliviously with respect to the collection of observables $\{O_i\}_{i=1}^m$. An algorithm for learning classical shadows can be broken down into two phases. First, in the measurement phase, the algorithm performs measurements on n copies of ρ, to obtain a classical representation of ρ, which is referred to the classical shadow of ρ. Then, in the estimation phase, the algorithm is given m observables $\{O_i\}_{i=1}^m$, and it must output estimates of $\mathrm{tr}(O_i\rho)$ to accuracy ε for all $i = 1, \ldots, m$, based solely on the observables and the classical shadow of ρ.

Despite significant research interest in the area, prior to our work, the exact complexity of these tasks for general ρ and $\{O_i\}_{i=1}^m$ was unknown, for any nontrivial regime of m, d, ε. For shadow tomography, the best known rate is due to [3, 31], who give an algorithm that uses

$$n = O\left(\frac{\log^2(m) \cdot \log(d)}{\varepsilon^4}\right)$$

copies of ρ, whereas the best known lower bound is the "trivial" bound of $n = \Omega(\frac{\log m}{\varepsilon^2})$ from the classical setting. For classical shadows, the best known rate is due to [19], who give an algorithm that (for general observables) requires $n = O(\frac{d \log m}{\varepsilon^2})$ copies of ρ. The best lower bound for classical shadows is due to [14], who obtain a lower bound of

$$n = \Omega\left(\left(\frac{\sqrt{d}}{\varepsilon} + \frac{1}{\varepsilon^2}\right)\log m\right),$$

which holds even when ρ is pure. The lack of tight rates for these problems is in stark contrast to other quantum learning settings such as full state tomography [26, 16, 10], state certification [25, 9], and Pauli channel estimation [6, 7], where asymptotically tight sample complexities are known.

In this work, we initiate the study of these problems in what we call the *high-accuracy* regime, that is, when $\varepsilon = O(d^{-c})$ for some c sufficiently large. This is in contrast to much prior theoretical work on these problems, which primarily focused on the "low-accuracy" regime where m and d are large compared to ε^{-1}. Our primary interest in this setting is two-fold. First, from a practical point of view, this is an important setting: in practice, it is very often the case that we wish to obtain detailed information about relatively small quantum systems [33, 30, 17]. Second, from a mathematical point of view, we find that these problems exhibit new and very interesting properties within this regime.

Indeed, informally stated, our main algorithmic result (see Theorem 2.1) is a new, statistically optimal estimator in this regime for both classical shadows and shadow tomography for general observables. To our knowledge, this is the first time that tight rates have been established for these problems, in any nontrivial regime of m, d, ε. Qualitatively speaking, our results uncover the following, previously unknown phenomena for these problems in the high-accuracy regime:

- **Shadow estimation no harder than classical counterpart.** In the high-accuracy regime, we show that the rate for shadow estimation matches the corresponding "trivial" lower bound in the classical case, where the observables are diagonal matrices. To our knowledge, this is the first time where tight rates have been demonstrated for either problem, and also the only known regime where the quantum and classical rates match for general observables.
- **Oblivious protocols can match adaptive ones.** We show that in this regime, shadow tomography and classical shadows *are statistically equivalent*. In other words, there is no statistical advantage for the measurements to be chosen adaptively based on the set of linear observables of interest. This is perhaps surprising, as all previously known statistically efficient algorithms for shadow tomography crucially required measurements to be chosen based on the set of linear observables of interest.
- **Representation theory for shadow estimation.** From a technical point of view, one interesting aspect of our work is that it deviates heavily from previous techniques on shadow estimation, and instead builds upon representation theoretic techniques previously used for full state tomography [11]. This adds to a growing literature on the power of such representation theoretic tools for shadow estimation [14, 13]; however, we emphasize that beyond this similarity, our techniques are quite distinct from these works.

Finally, we also demonstrate a formal reduction to the "medium-accuracy" regime (see Lemma E.7). That is, we show that without loss of generality, for the task of classical shadow estimation, one may assume that $\varepsilon \leq O(d^{-1/2})$. This reduction works by demonstrating that by leveraging classical ideas from dimensionality reduction [20], one can linearly trade off d and ε in the sample complexity, as long as $\varepsilon \geq O(d^{-1/2})$. While this still leaves a gap in the parameter landscape where we do not know the correct rates, as our analysis currently requires $\varepsilon \leq O(d^{-12})$, at the very least this demonstrates that all of the interesting action for this problem is in the setting where ε and d^{-1} are polynomially related. In fact, we conjecture that the correct rate for classical shadows is

$$\Theta\left(\left(\frac{d}{\varepsilon} + \frac{1}{\varepsilon^2}\right) \cdot \log m\right), \tag{1}$$

as we conjecture that (1) the reduction holds up to the threshold of $\varepsilon \geq \Omega(d^{-1})$, and (2) our rate of $O(\log m/\varepsilon^2)$ holds as long as $\varepsilon \leq O(d^{-1})$. However, it seems that improving the thresholds on both fronts requires additional ideas, and we leave establishing this rate as interesting future work.

2 Our contributions

In this work, we give new algorithms for shadow tomography and for learning classical shadows. We demonstrate that these algorithms obtain optimal sample complexities for both problems in the high-accuracy regime. In fact, our rates match the "trivial" lower bounds for these problems, demonstrating that in this regime, the quantum task is no harder than the classical one. To our knowledge, this is the first time where tight rates have been demonstrated for either problem.

Perhaps surprisingly, we do this by giving a new algorithm for learning classical shadows which matches the lower bound for the ostensibly easier task of shadow tomography. That is, we show that in the high-accuracy regime, learning classical shadows is no harder than shadow tomography. More formally, we show:

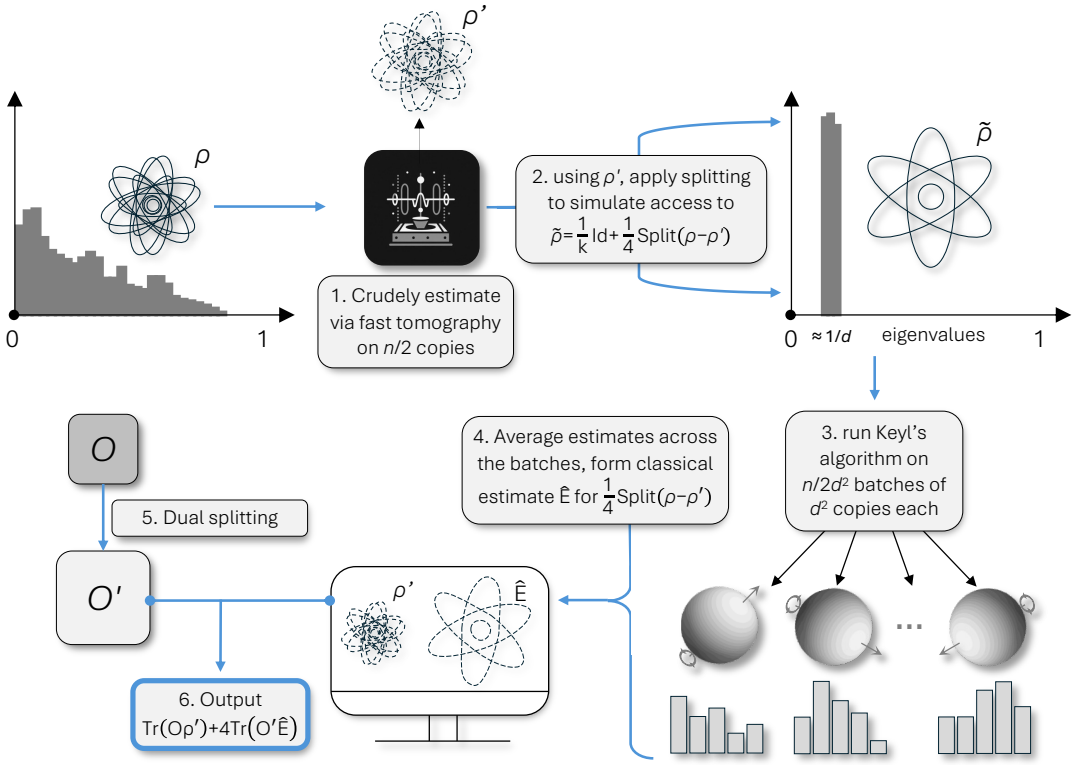

Figure 1: Overview of our shadow estimation algorithm

Theorem 2.1. *Let $m, d \in \mathbb{Z}$ be fixed, and let $\varepsilon = O(d^{-12})$. Then, there is an estimator which takes n copies of an unknown mixed state $\rho \in \mathbb{C}^{d\times d}$, where*

$$n = O\left(\frac{\log m}{\varepsilon^2}\right),$$

and outputs a classical function $F : \mathbb{C}^{d\times d} \to [0,1]$ so that for any fixed collection of m observables $\{O_i\}_{i=1}^m$, the function satisfies $|F(O_i) - \mathrm{tr}(O_i\rho)| \le \varepsilon$ for all $i = 1, \ldots, m$ with high probability.

We pause here to make a couple of remarks on this result. First, as alluded to above, this clearly also implies the same upper bound for shadow tomography, and moreover, this rate is tight, as it matches the lower bound for shadow tomography (and estimating statistical queries).

Second, our rate is dimension independent. This is in contrast to prior rates for learning classical shadows, and indeed, even the lower bound [14]. Note that our result does not violate this lower bound because the dimension-dependent term vanishes in the high-accuracy regime, and indeed our result suggests that for classical shadows the optimal dependence on d ought to be a lower order term in ε^{-1}.

Thirdly, we obtain the optimal quadratic scaling in $1/\varepsilon$, whereas the aforementioned upper bounds for shadow tomography scale with $1/\varepsilon^4$. A recent work [8] shows that with polylog(d)-copy measurements, $\Omega(1/\varepsilon^4)$ copies are necessary even in the special case when the observables are Pauli operators. We show that if the number of copies that can be measured at once scales polynomially in the dimension, then this lower bound no longer applies.

A reduction to the medium-accuracy regime As mentioned preivously, we also demonstrate the following reduction to the medium-accuracy regime:

Theorem 2.2 (informal, see Theorem E.7)**.** *Suppose that for all k, ε' satisfying $\varepsilon' \le k^{-1/2}$, there is an algorithm that solves classical shadows for k-dimensional states to error ε' (and constant failure probability) with $f(k, \varepsilon')$ copies. Let d, ε satisfy $\varepsilon \ge 100/\sqrt{d}$. Then, there is an algorithm that solves classical shadows for d-dimensional states to error ε (and constant failure probability) that uses $f(100\sqrt{d}/\varepsilon, 1/\sqrt{d})$ copies.*

Here, we briefly pause to show how this translates to a reduction to the medium-accuracy regime. The theorem states that to solve classical shadows to error ε for d dimensional states, it suffices to obtain an estimator for classical shadows in $k = 100\sqrt{d}/\varepsilon$ dimensions to error $\varepsilon' = 1/\sqrt{d}$. Notice that by the choice of parameters, we have that $k \le d$, and thus consequently $\varepsilon' \le d^{-1/2} \le k^{-1/2}$, so this new problem is indeed in the medium-accuracy regime. In fact, in Section E we show a more general reduction which allows us to linearly trade off d and ε.

Moreover, this reduction indeed recovers a linear tradeoff between ε and d that we believe is optimal. For instance, if we believe the conjectured rate (1) holds for all $\varepsilon \le d^{-1/2}$ then a straightforward computation demonstrates that this reduction yields the same rate holds for $\varepsilon \ge 100d^{-1/2}$ as well.

3 Technical overview

Our approach is a departure from the aforementioned approaches to shadow tomography, which automatically lose an extra $\log d/\varepsilon^2$ factor because they involve an outer routine based on online learning. Instead, our starting point is the recent approach of [11] for full *state tomography*.

Overview of [11]. Roughly speaking, the approach in that work consisted of two components: 1) a reduction from tomography of arbitrary mixed states to tomography of mixed states which are a small perturbation of the maximally mixed state, and 2) an analysis of Keyl's estimator for learning the *perturbation* that gives rise to such a state, rather than learning the state directly.

Let us first focus on step 2). We say call states that are small perturbations of I_d/d *balanced states*. If we express an unknown balanced state ρ as $\rho = I_d/d + E$ for some perturbation E, then they observed that the tensor product of t copies of ρ is close to the "linearized state"

$$\rho' \triangleq (I_d/d)^{\otimes t} + \sum_{\text{sym}} E \otimes (I_d/d)^{\otimes t-1}$$

where $\sum_{\text{sym}}$ denotes the sum over all tensor products of one copy of E and $t-1$ copies of I_d/d. Technically this linearized state is not necessarily a density matrix, but for any POVM $\{M_z\}_{z\in\mathcal{Z}}$, one can still consider measurement statistics of the form $\int_{\mathcal{Z}} f(M_z)\langle \rho', M_z\rangle \, \mathrm{d}z$, which correspond to the expectation of any estimator f applied to the outcome of "measuring" the linearized state.

Now consider a natural choice of f in the context of state tomography: Keyl's estimator. Whereas the expected result of applying Keyl's estimator to the actual state ρ is hard to characterize, the upshot of working with the linearization ρ' is that it is much more amenable to calculations, and we can actually explicitly compute the analogous result for ρ'. In particular, by taking f and $\{M_z\}_{z\in\mathcal{Z}}$ to be given by Keyl's estimator, the above measurement statistic $\int_{\mathcal{Z}} f(M_z)\langle \rho', M_z\rangle \, \mathrm{d}z$ turns out to be exactly given by a perturbation of the maximally mixed state by some *known multiple* of E (Corollary C.5), i.e. $I_d/d + cE$ for some known factor c.

This means that if E has sufficiently small norm, the expected result $\mathbb{E}[\widehat{\rho}]$ of applying Keyl's estimator to the actual state ρ is sufficiently close to this (Lemma C.4) that if we had access to $\mathbb{E}[\widehat{\rho}]$, we could simply estimate the perturbation E via $c^{-1}(\mathbb{E}[\widehat{\rho}] - I_d/d)$. Of course in reality we only have access to *realizations* of the state $\widehat{\rho}$ obtained by Keyl's estimator, rather than their expectation, but using existing bounds on the variance of Keyl's estimator [26], we can control the deviation between $\widehat{\rho}$ and $\mathbb{E}[\widehat{\rho}]$.

Adapting to the shadow estimation setting. In this work, we observe that even though the above analysis was originally implemented in [11] to bound the accuracy of the estimate $\widehat{E}$ for perturbation E obtained in Frobenius norm, essentially the same analysis translates naturally to bounding accuracy as quantified by how close $\langle O, E\rangle$ is to $\langle O, \widehat{E}\rangle$ for an arbitrary observable O. The only place where we need to be somewhat careful is in controlling the variance of the estimator: instead of directly bounding the expected squared distance between $\widehat{\rho}$ and $\mathbb{E}[\widehat{\rho}]$, we need to bound the expected squared discrepancy $\mathbb{E}[\langle O, \widehat{\rho} - \mathbb{E}[\widehat{\rho}]\rangle^2]$. While it may be tempting to simply bound this by applying Cauchy-Schwarz and appealing to the existing bound on $\mathbb{E}[\|\widehat{\rho} - \mathbb{E}[\widehat{\rho}]\|_F^2]$, this is lossy by dimension-dependent factors. Instead, we need to exploit the rotation-invariance of Keyl's estimator to get a tighter bound on this variance (see Eq. (C) onwards).

The above discussion is already sufficient to prove our main result in the special case where ρ is balanced. It gives rise to an estimate for E, and thus for ρ, which is *oblivious* in the sense that it does not depend on the choice of observable O above. Furthermore, our algorithm only needs $O(\log(1/\delta)/\varepsilon^2)$ samples to produce an estimate $\widehat{\rho}$ for which $|\langle O, \widehat{\rho} - \rho\rangle| \leq \varepsilon$ with probability $1 - \delta$. As $\widehat{\rho}$ is oblivious to O, it can be used to estimate any set of m observables $O_1, \ldots, O_m$ with probability $1 - m\delta$. By taking $\delta = O(1/m)$, we obtain Theorem 2.1 in the special case of balanced states. We then appeal to part 1) of the analysis in [11], which allows us to effectively reduce from the case of arbitrary mixed states to the case of balanced states.

Reduction to the balanced case. Here we summarize how we adapt their approach in this step to our shadow estimation setting. At the end we comment on how it compares to the implementation in [11] for full state tomography.

Roughly speaking, this part of the proof is based on a certain "splitting" operation Split (see Definition D.1) that linearly maps any d-dimensional mixed state to an $O(d)$-dimensional one whose eigenvalues are upper bounded by $1/d$. Importantly, given measurement access to $\rho^{\otimes t}$, one can simulate measurement access to $\mathsf{Split}(\rho)^{\otimes t}$, and furthermore there is a *dual operation* DSplit (see Definition D.2) that can be applied to any observable O such that the expectation value $\langle O, \rho\rangle$ is equal to $\langle \mathsf{DSplit}(O), \mathsf{Split}(\rho)\rangle$.

We can then reduce to the case that ρ is balanced as follows: we obtain a crude estimate $\tilde{\rho}$ for ρ using *low-accuracy state tomography* (Theorem D.5), and then we "recenter" $\mathsf{Split}(\rho)$ around $\mathsf{Split}(\tilde{\rho})$ to get an $O(d)$-dimensional state which is sufficiently close to maximally mixed. Importantly, using the description of $\tilde{\rho}$ and by simulating measurement access to $\mathsf{Split}(\rho')$, we can simulate measurement access to this recentered state and thus reduce to the case where the unknown state is balanced (in $O(d)$ dimensions).

We remark that the primary difference between our implementation of this splitting technique and the one in [11] is our use of DSplit, which is specific to the shadow estimation setting. For state tomography, [11] considered a different operation. Roughly speaking, they defined a procedure Rec which *inverts* the mapping given by Split, so that once one has an estimate for the perturbation corresponding to the recentered state, their final estimator is given by applying Rec to this estimate. In contrast, because our goal is not to estimate the state in Frobenius norm, but rather to estimate it well enough to answer expectation value queries, we need an operation *dual* to Split instead of an operation *inverse* to it. This leads us to consider the operation DSplit sketched above.

4 Outlook

In this work we gave the first algorithm for classical shadows to achieve optimal sample complexity $O(\log m/\varepsilon^2)$ for general observables in some nontrivial regime, namely when the target accuracy ε is inverse polynomial in the dimension d of the Hilbert space. In contrast, prior work either suffered from extraneous logarithmic factors in m and d and polynomial factors in $1/\varepsilon$, or required polynomial factors

in d . Interestingly, our proof leverages ideas from the recent work of [11] on *full* state tomography. The central idea is to formulate an (approximately) unbiased estimate of the state by analyzing the behavior of Keyl's estimator on a certain linearization of the batch of copies of the unknown state.

The natural question left open by our work is to handle the low-accuracy regime, that is, to lift the assumption that $\varepsilon \leq 1/\mathrm{poly}(d)$. Unfortunately in the low-accuracy regime, the linearization trick mentioned above no longer applies. Resolving this would settle the main open question of [1], i.e. showing that in all parameter regimes, the sample complexity of shadow tomography is no worse than that of its classical analogue.

Another interesting direction for future work is to understand how the sample complexity of classical shadows changes under additional constraints on ρ, e.g. if it has low rank or is preparable with a shallow quantum circuit. In the special case where ρ is rank-1, this was settled in the work of [14]. The recent work of [13] obtained an improved upper bound for general low-rank states compared to the result of [19].

Acknowledgments.

The authors thank Ainesh Bakshi, Jaume de Dios Pont, Ryan O'Donnell, and Ewin Tang for illuminating discussions about shadow tomography.

References

[1] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 325–338, 2018.

[2] Scott Aaronson and Guy N Rothblum. Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333, 2019.

[3] Costin Bădescu and Ryan O'Donnell. Improved quantum data analysis. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1398–1411, 2021.

[4] Gregory Bentsen, Ionut-Dragos Potirniche, Vir B Bulchandani, Thomas Scaffidi, Xiangyu Cao, Xiao-Liang Qi, Monika Schleier-Smith, and Ehud Altman. Integrable and chaotic dynamics of spins coupled to an optical cavity. *Physical Review X*, 9(4):041011, 2019.

[5] Dolev Bluvstein, Simon J Evered, Alexandra A Geim, Sophie H Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, 2024.

[6] Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. Quantum advantages for pauli channel estimation. *Physical Review A*, 105(3):032435, 2022.

[7] Sitan Chen and Weiyuan Gong. Efficient pauli channel estimation with logarithmic quantum memory. In *arXiv:2309.14326*, 2023.

[8] Sitan Chen, Weiyuan Gong, and Qi Ye. Optimal tradeoffs for estimating pauli observables. *arXiv preprint arXiv:2404.19105*, 2024.

[9] Sitan Chen, Brice Huang, Jerry Li, and Allen Liu. Tight bounds for quantum state certification with incoherent measurements. *arXiv preprint arXiv:2204.07155*, 2022.

[10] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning?, 2023.

[11] Sitan Chen, Jerry Li, and Allen Liu. An optimal tradeoff between entanglement and copy complexity for state tomography, 2024.

[12] Roe Goodman, Nolan R Wallach, et al. *Symmetry, representations, and invariants*, volume 255. Springer, 2009.

[13] Daniel Grier, Sihan Liu, and Gaurav Mahajan. Improved classical shadows from local symmetries in the schur basis. *arXiv preprint arXiv:2405.09525*, 2024.

[14] Daniel Grier, Hakop Pashayan, and Luke Schaeffer. Sample-optimal classical shadows for pure states. *arXiv preprint arXiv:2211.11810*, 2022.

[15] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020.

[16] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 913–925, 2016.

[17] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 391(3):951–967, 2022.

[18] Hsin-Yuan Huang. Learning quantum states from their classical shadows. *Nature Reviews Physics*, 4(2):81–81, 2022.

[19] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[20] William B Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into banach spaces. *Israel Journal of Mathematics*, 54(2):129–138, 1986.

[21] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *nature*, 549(7671):242–246, 2017.

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

[23] Ryan Levy, Di Luo, and Bryan K Clark. Classical shadows for quantum process tomography on near-term quantum computers. *Physical Review Research*, 6(1):013029, 2024.

[24] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016.

[25] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 529–538, 2015.

[26] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 899–912, 2016.

[27] Ryan O'Donnell and John Wright. Efficient quantum tomography ii. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 962–974, 2017.

[28] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.

[29] Bruce E Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*, volume 203. Springer Science & Business Media, 2013.

[30] GI Struchalin, Ya A Zagorovskii, EV Kovlakov, SS Straupe, and SP Kulik. Experimental estimation of quantum state properties from classical shadows. *PRX Quantum*, 2(1):010307, 2021.

[31] Adam Bene Watts and John Bostanci. Quantum event learning and gentle random measurements. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2024.

[32] John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016.

[33] Ting Zhang, Jinzhao Sun, Xiao-Xu Fang, Xiao-Ming Zhang, Xiao Yuan, and He Lu. Experimental quantum state measurement with classical shadows. *Physical Review Letters*, 127(20):200501, 2021.

A Representation Theory and Keyl's POVM

Our algorithm will be based on Keyl's POVM [22] which is at the heart of most quantum state tomography algorithms that use entangled measurements [32, 26, 27, 11]. Defining Keyl's POVM requires some basic concepts from representation theory. In the following, we list the ones relevant to our discussion here; see [32] for a more detailed exposition.

Definition A.1. *[Young Tableaux] We have the following standard definitions:*

- *Given a partition $\lambda \vdash n$, 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.*
- *A standard Young tableaux (SYT) T of shape λ is a Young diagram of shape λ where each box is filled with some integer in $[n]$ such that the rows are strictly increasing from left to right and the columns are strictly increasing from top to bottom.*
- *A semistandard Young tableaux (SSYT) T of shape λ is a Young diagram of shape λ where each box is filled with some integer in $[d]$ for some d and the rows are weakly increasing from left to right and the columns are strictly increasing from top to bottom.*

We recall the correspondence between Young tableaux and representations of the symmetric and general linear groups:

Definition A.2. *We say a representation μ of GL_d over a complex vector space $\mathbb{C}^m$ is a polynomial representation if for any $U \in \mathbb{C}^{d\times d}$, $\mu(U) \in \mathbb{C}^{m\times m}$ is a polynomial in the entries of U.*

Fact A.3 ([29])**.** *The irreducible representations of the symmetric group S_n are exactly indexed by the partitions $\lambda \vdash n$ and have dimensions $\dim(\lambda)$ equal to the number of standard Young tableaux of shape λ. We denote the corresponding vector space Sp_λ.*

Fact A.4 ([12])**.** *For each $\lambda \vdash n$, there is a (unique) irreducible polynomial representation of GL_d corresponding to λ. We denote the corresponding map and vector space $(\pi_\lambda, V_\lambda^d)$. The dimension $\dim(V_\lambda^d)$ is equal to the number of semistandard Young tableaux of shape λ with entries in $[d]$. This representation, restricted to U_d is also an irreducible representation.*

Theorem A.5 (Schur-Weyl Duality [12])**.** *Consider the representation of $S_n \times GL_d$ on $(\mathbb{C}^d)^{\otimes n}$ where the action of the permutation $\pi \in S_n$ permutes the different copies of $\mathbb{C}^d$ and the action of $U \in GL_d$ is applied independently to each copy. This representation can be decomposed as a direct sum*

$$(\mathbb{C}^d)^{\otimes n} = \bigoplus_{\substack{\lambda \vdash n \\ \ell(\lambda) \le d}} \mathsf{Sp}_\lambda \otimes V_\lambda^d \,.$$

Definition A.6 (Schur Subspace)**.** *We call $\mathsf{Sp}_\lambda \otimes V_\lambda^d$ the λ-Schur subspace. Given integers n, d and $\lambda \vdash n$, we define $\Pi_\lambda^d : (\mathbb{C}^d)^{\otimes n} \to \mathsf{Sp}_\lambda \otimes V_\lambda^d$ to project onto the λ-Schur subspace.*

Theorem A.7 (Gelfand-Tsetlin Basis [12])**.** *Let n, d be positive integers. For each partition $\lambda \vdash n$ where λ has at most d parts, there is a basis $v_1, \ldots, v_m$ of V_λ^d with $m = \dim(V_\lambda^d)$ such that for any matrix $D_\alpha = \mathrm{diag}(\alpha_1, \ldots, \alpha_d)$, we have $v_i^\dagger \pi_\lambda(D_\alpha) v_i = \alpha^{f^{(i)}}$ for all i where $f^{(i)}$ are each d-tuples that give the frequencies of $1, 2, \ldots, d$ in each of the different semi-standard tableaux of shape λ.*

Definition A.8 (Maximal-weight Vector)**.** *For a partition $\lambda \vdash n$, we define the maximal weight vector $v_\lambda \in V_\lambda^d$ to be the vector given by Theorem A.7 with $f^{(i)} = \lambda_i$ for all i.*

Remark A.9. *Note that for a given partition $\lambda = (\lambda_1, \ldots, \lambda_d)$, there is a semi-standard tableaux of shape λ with frequencies $\lambda_1, \ldots, \lambda_d$ (where we just fill the ith row with all entries equal to i) so the vector defined above indeed exists.*

Definition A.10 (Weak Schur Sampling)**.** *We use the term weak Schur sampling to refer to the POVM on $\mathbb{C}^{d^n \times d^n}$ with elements given by Π_λ^d for λ ranging over all partitions of n into at most d parts.*

Definition A.11 (Keyl's POVM [22])**.** *We define the following POVM on $\mathbb{C}^{d^n \times d^n}$: first perform weak Schur sampling to obtain $\lambda \vdash n$. Then discard the permutation register (corresponding to the subspace Sp_λ). Within the remaining subspace V_λ^d, measure according to*

$$\dim(V_\lambda^d)\{\pi_\lambda(U) v_\lambda v_\lambda^\dagger \pi_\lambda(U)^\dagger\}_U$$

where U ranges over Haar random unitaries. Note that the outcome of the measurement consists of a partition $\lambda \vdash n$ and a unitary $U \in \mathbb{C}^{d \times d}$.

Definition A.12 (Schur Weyl Distribution)**.** *Given integers n, d and a tuple $(\alpha_1, \ldots, \alpha_d)$ with $\alpha_i \geq 0$ and $\alpha_1 + \cdots + \alpha_d = 1$, the Schur-Weyl distribution $\mathrm{SW}^n(\alpha)$ is a distribution over partitions $\lambda \vdash n$ into at most d parts obtained by measuring the state $\mathrm{diag}(\alpha_1, \ldots, \alpha_d)^{\otimes n}$ via weak Schur sampling. When α is uniform, we may write SW_d^n instead.*

B Basic Facts

Claim B.1. *Let $X, Y \in \mathbb{C}^{d \times d}$ be Hermitian matrices. Then*

$$\mathbb{E}_U[(U^\dagger X U)\langle U^\dagger X U, Y\rangle] = \frac{1}{d^2-1}\left(\|X\|_F^2 - \frac{\mathrm{tr}(X)^2}{d}\right)\left(Y - \frac{\mathrm{tr}(Y)I}{d}\right) + \frac{\mathrm{tr}(X)^2\mathrm{tr}(Y)I}{d^2}$$

where the expectation is over a Haar random unitary U.

Claim B.2. *Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be a vector of nonnegative weights summing to 1. Then for $\lambda \sim \mathrm{SW}^n(\alpha)$, with probability at least $1/2$,*

$$\sum_{i=1}^d \lambda_i^2 \geq \frac{n^{1.5}}{4}.$$

Claim B.3. *Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be a vector of nonnegative weights summing to 1. Then for $\lambda \sim \mathrm{SW}^n(\alpha)$,*

$$\mathbb{E}\left[\sum_{i=1}^d \lambda_i^2\right] \leq 2((\alpha_1^2 + \cdots + \alpha_d^2)n^2 + n^{1.5})$$

Lemma B.4. *Let $0 \leq \lambda \leq 1$. Given t copies of an unknown state ρ, and given a description of a density matrix σ, it is possible to simulate any measurement of $(\lambda\rho + (1-\lambda)\sigma)^{\otimes t}$ using a measurement of $\rho^{\otimes t}$.*

C Balanced Case

We begin by presenting our algorithm for the case when ρ is close to maximally mixed. In this case, given n total copies of ρ, our algorithm sets $t = 0.01d^2$ and measures n/t copies of $\rho^{\otimes t}$, each using Keyl's POVM. Recall that Keyl's POVM involves first obtaining a partition λ and then obtaining a unitary U. The estimator that we construct after measuring according to Keyl's POVM will be $U\mathrm{diag}(\lambda_1/t, \ldots, \lambda_d/t)U^\dagger$,

which we call Keyl's estimator. We then average this estimator (with some appropriate linear rescaling) over all n/t batches to construct our final estimate for ρ. In Theorem C.6, we prove that this estimator successfully solves classical shadows.

We will rely on a few of the intermediate lemmas from [11]. We begin with a few definitions. Note that Keyl's POVM is symmetric over the unitary in the following sense.

Definition C.1. *We say a POVM $\{M_z\}_{z\in\mathcal{Z}}$ in $\mathbb{C}^{d^t\times d^t}$ is copy-wise rotationally invariant if it is equivalent to*

$$\{U^{\otimes t} M_z (U^\dagger)^{\otimes t} dU\}_{z\in\mathcal{Z}}$$

where $U \in \mathbb{C}^{d\times d}$ is a random unitary drawn from the Haar measure.

Definition C.2. *Let $\{M_z\}_{z\in\mathcal{Z}}$ be a POVM in $\mathbb{C}^{d^t\times d^t}$ that is copywise rotationally invariant. We say a function $f : \{M_z\}_{z\in\mathcal{Z}} \to \mathbb{C}^{d\times d}$ is rotationally compatible with the POVM if*

$$f(U^{\otimes t} M_z (U^\dagger)^{\otimes t}) = U f(M_z) U^\dagger$$

for all $z \in \mathcal{Z}$ and unitary U.

Fact C.3. *Keyl's POVM is copy-wise rotationally invariant and the estimator $(\lambda, U) \to U\mathrm{diag}(\lambda_1/t, \dots, \lambda_d/t)U^\dagger$ is rotationally compatible with Keyl's POVM.*

Proof. This follows immediately from Theorem A.5. □

When the state ρ is close to maximally mixed, we can bound the mean of Keyl's estimator $U\mathrm{diag}(\lambda_1/t, \dots, \lambda_d/t)U^\dagger$ as follows.

Lemma C.4. *[11] Let $\{M_z\}_{z\in\mathcal{Z}}$ be a POVM in in $\mathbb{C}^{d^t\times d^t}$ that is copywise rotationally invariant. Let $f : \{M_z\}_{z\in\mathcal{Z}} \to \mathbb{C}^{d\times d}$ be a rotationally compatible estimator such that $\mathrm{tr}(f(M_z)) = 0$ for all $z \in \mathcal{Z}$. Let $X = (I_d/d + E)^{\otimes t}$ and let $X' = (I_d/d)^{\otimes t} + \sum_{\mathrm{sym}} E \otimes (I_d/d)^{\otimes t-1}$. Assume that $\|E\|_F \le \left(\frac{0.01}{t}\right)^4$. Then*

$$\left\| \int_{\mathcal{Z}} f(M_z) \langle X - X', M_z\rangle \, \mathrm{d}z \right\|_F \le \frac{10^5 t^2 \|E\|_F^2}{d} \sqrt{\int_{\mathcal{Z}} \frac{\|f(M_z)\|_F^2 \,\mathrm{tr}(M_z)}{d^t} \, \mathrm{d}z} \,.$$

Corollary C.5. *[11] Let $\{M_{\lambda,U}\}_{\lambda,U}$ be Keyl's POVM where λ ranges over partitions of t and U ranges over unitaries in $\mathbb{C}^{d\times d}$. Let $X' = (I_d/d)^{\otimes t} + \sum_{\mathrm{sym}} E \otimes (I_d/d)^{\otimes t-1}$. Then*

$$\sum_{\lambda\vdash t} \int U\mathrm{diag}(\lambda_1/t, \dots, \lambda_d/t)U^\dagger \cdot \langle M_{\lambda,U}, X'\rangle \, \mathrm{d}U = \frac{I_d}{d} + \frac{dE}{t(d^2-1)} \mathbb{E}_{\lambda\sim \mathrm{SW}_d^t} \left[\sum_{j=1}^d \lambda_j^2 - (t^2/d) \right].$$

We can now prove the main theorem in the balanced case.

Theorem C.6. *Let $\rho = \frac{I_d}{d} + E$ be an unknown quantum state in $\mathbb{C}^{d\times d}$. Assume that $\|E\|_F \le \sqrt{\varepsilon}/d^2$. Then for any target accuracy $\varepsilon \le 1/d^{12}$, there is an algorithm that measures $O(1/\varepsilon^2)$ copies of ρ and returns $\widehat{E} \in \mathbb{C}^{d\times d}$ such that for any Hermitian matrix $O \in \mathbb{C}^{d\times d}$,*

$$\Pr\left[\left| \langle O, E\rangle - \langle O, \widehat{E}\rangle \right| \ge \varepsilon \cdot \frac{\|O\|_F}{\sqrt{d}} \right] \le 0.1 \,.$$

Algorithm 1 Shadows for balanced states

Input: m copies of $\rho^{\otimes t}$ for some unknown quantum state $\rho \in \mathbb{C}^{d\times d}$
for $j \in [m]$ **do**
  Measure $\rho^{\otimes t}$ according to Keyl's POVM
  Let $\lambda \vdash t$ be the partition and U be the unitary obtained from the measurement
  Set $D_j = U\mathrm{diag}(\lambda_1/t, \ldots, \lambda_d/t)U^\dagger$
end for
Compute $\theta = \mathbb{E}_{\lambda\sim \mathrm{SW}_d^t}[\sum_j \lambda_j^2] - (t^2/d)$
Compute $\widehat{E} = \frac{t(d^2-1)}{d\theta}\left(\frac{D_1+\cdots+D_m}{m} - \frac{I_d}{d}\right)$
Output: $\widehat{E}$

Proof. We run Algorithm 1 with $t = 0.01d^2$ and $m = 10^6/(\varepsilon^2 d^2)$ (note that the total number of copies used is then indeed $O(1/\varepsilon^2)$). The POVM in Algorithm 1 is clearly copywise rotationally invariant and the estimator is rotationally compatible with it. Let us use the shorthand $\{M_z\}_{z\in\mathcal{Z}}$ to denote this POVM and for M_z corresponding to unitary U and partition λ, we let $f(M_z) = U\mathrm{diag}(\lambda_1/t, \ldots, \lambda_d/t)U^\dagger$. We have

$$\mathbb{E}\left[\frac{D_1+\cdots+D_m}{m}\right] = \int_{\mathcal{Z}} f(M_z)\langle M_z, (I_d/d + E)^{\otimes t}\rangle \,\mathrm{d}z$$

where the expectation is over the randomness of the quantum measurement in Algorithm 1. We can make the estimator D_j have trace 0 by simply subtracting out I_d/d and adding it back at the end. Thus, by Lemma C.4 and Corollary C.5, recalling the definition of θ in Line 7 of Algorithm 1, we have

$$\begin{aligned}\left\|\mathbb{E}\left[\frac{D_1+\cdots+D_m}{m}\right] - \frac{I_d}{d} - \frac{d\theta E}{t(d^2-1)}\right\|_F &\le \frac{10^5 t^2 \|E\|_F^2}{d}\sqrt{\int_{\mathcal{Z}} \frac{\|f(M_z)\|_F^2 \,\mathrm{tr}(M_z)}{d^t}\,\mathrm{d}z} \\ &\le \frac{10^5 t^2 \|E\|_F^2}{d}.\end{aligned}$$

Thus, if $\widehat{E}$ is the output of Algorithm 1, then

$$\left\|\mathbb{E}[\widehat{E}] - E\right\|_F \le \frac{10^5 t^3 \|E\|_F^2}{\theta}.$$

Next, we compute the variance of the estimator. Note that WLOG, we can assume the observable O has

$\mathrm{tr}(O) = 0$ since our estimator $\widehat{E}$ is always traceless. We have

$$\begin{aligned}
\mathbb{E}\left[\langle O, \widehat{E} - \mathbb{E}[\widehat{E}]\rangle^2\right] &\le \frac{d^2t^2}{m\theta^2}\,\mathbb{E}\left[\langle O, D_1 - \mathbb{E}[D_1]\rangle^2\right] \\
&\le \frac{4d^2t^2}{m\theta^2}\left(\mathbb{E}\left[\left\langle O, D_1 - \frac{I_d}{d}\right\rangle^2\right]\right) \\
&\le \frac{8d^2t^2}{m\theta^2}\left(\mathbb{E}\left[\mathbb{E}_U\left[\left\langle U^\dagger O U, D_1 - \frac{I_d}{d}\right\rangle^2\right]\right]\right) \\
&\le \frac{8d^2t^2}{m\theta^2}\,\frac{\|O\|_F^2\,\mathbb{E}[\|D_1\|_F^2]}{d^2} \\
&= \frac{8\,\|O\|_F^2}{m\theta^2}\,\mathbb{E}_{\lambda\sim \mathrm{SW}^t(\rho)}[\sum_j \lambda_j^2]
\end{aligned}$$

where in the above, we used that $\|E\|_F \le 1/(10d)^8$ so

$$0.9(I_d/d)^{\otimes t} \preceq \rho^{\otimes t} \preceq 1.1(I_d/d)^{\otimes t}$$

and thus when measuring $\rho^{\otimes t}$ with Keyl's POVM (which is rotationally invariant), the rotation of the outcome is approximately uniform up to a factor of 2. Now by Claim B.3, we can upper bound $\mathbb{E}_{\lambda\sim \mathrm{SW}^t(\rho)}[\sum_j \lambda_j^2] \le 2(\|\rho\|_F^2\, t^2 + t^{1.5}) \le 4t^{1.5}$ where recall that we set $t \le 0.01d^2$. Also by Claim B.2, we have $\theta \ge t^{1.5}/4$. Thus, putting everything together, we conclude

$$\mathbb{E}\left[\langle O, \widehat{E} - E\rangle^2\right] \le 2\cdot 10^{10} t^3\,\|E\|_F^4\,\|O\|_F^2 + \frac{(10\,\|O\|_F)^2}{mt^{1.5}} \le \frac{\varepsilon^2\,\|O\|_F^2}{10^2 d}\,.$$

The desired statement then follows from Chebyshev's inequality. □

D Splitting Reduction

As in [11], when ρ is not balanced, we reduce to the balanced case via a splitting reduction.

Definition D.1. *Let $b_1, \dots, b_d \in \mathbb{Z}_{\ge 0}$. We define $\mathsf{Split}_{b_1,\dots,b_d}$ to be a linear map that sends any $M \in \mathbb{C}^{d\times d}$ to a square matrix with dimension $2^{b_1} + \cdots + 2^{b_d}$ defined as follows. The rows and columns of $\mathsf{Split}_{b_1,\dots,b_d}(M)$ are indexed by pairs (j, s) where $j \in [d]$ and $s \in \{0,1\}^{b_j}$ and these are sorted first by j and then lexicographically according to s. Now the entry indexed by row (j_1, s_1) and column (j_2, s_2) is defined as*

- *If $b_{j_1} \le b_{j_2}$ then the entry is $M_{j_1 j_2}/2^{b_{j_2}}$ if s_1 is a prefix of s_2 and is 0 otherwise*
- *If $b_{j_1} > b_{j_2}$ then the entry is $M_{j_1 j_2}/2^{b_{j_1}}$ if s_2 is a prefix of s_1 and is 0 otherwise*

As an example, we have the following splitting of a 2×2 matrix with $b_1 = 2, b_2 = 1$.

$$\mathsf{Split}_{2,1}\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{bmatrix}\right) = \begin{bmatrix}
0.25a_{11} & 0 & 0 & 0 & 0.25a_{12} & 0 \\
0 & 0.25a_{11} & 0 & 0 & 0.25a_{12} & 0 \\
0 & 0 & 0.25a_{11} & 0 & 0 & 0.25a_{12} \\
0 & 0 & 0 & 0.25a_{11} & 0 & 0.25a_{12} \\
0.25a_{21} & 0.25a_{21} & 0 & 0 & 0.5a_{22} & 0 \\
0 & 0 & 0.25a_{21} & 0.25a_{21} & 0 & 0.5a_{22}
\end{bmatrix}$$

We can also define a dual map to Split.

Definition D.2. *Let $b_1, \dots, b_d \in \mathbb{Z}_{\geq 0}$. We define $\mathsf{DSplit}_{b_1,\dots,b_d}$ to be a linear map that sends any $M \in \mathbb{C}^{d\times d}$ to a square matrix with dimension $2^{d_1}+\cdots+2^{b_d}$ defined as follows. The rows and columns of $\mathsf{DSplit}_{b_1,\dots,b_d}(M)$ are indexed by pairs (j,s) where $j \in [d]$ and $s \in \{0,1\}^{b_j}$ and these are sorted first by j and then lexicographically according to s. Now the entry indexed by row (j_1,s_1) and column (j_2,s_2) is defined as*

- *$M_{j_1 j_2}$ if s_1 is a prefix of s_2 and or s_2 is a prefix of s_1 and 0 otherwise*

We have the following basic properties.

Claim D.3. *Let $b_1, \dots, b_d \in \mathbb{Z}_{\geq 0}$. We have the following statements for any $M, N \in \mathbb{C}^{d\times d}$:*

- $\left\|\mathsf{Split}_{b_1,\dots,b_d}(M)\right\|_F \leq \|M\|_F$
- $\langle \mathsf{Split}_{b_1,\dots,b_d}(M), \mathsf{DSplit}_{b_1,\dots,b_d}(N)\rangle = \langle M, N\rangle$
- $\left\|\mathsf{DSplit}_{b_1,\dots,b_d}(N)\right\|_F \leq 2\sqrt{2^{b_1}+\cdots+2^{b_d}}\,\|N\|$

Proof. The first two statements follow immediately from the definitions. To verify the third, for each integer k, let $S_k \subseteq [d]$ denote the set of indices j such that $b_j = k$. Note that the entries of $\mathsf{DSplit}_{b_1,\dots,b_d}(N)$ are obtained by taking the entries of N and duplicating each a certain number of times – an entry $N_{j_1 j_2}$ appears $2^{\max(b_{j_1}, b_{j_2})}$ times. Thus,

$$\begin{aligned}
\left\|\mathsf{DSplit}_{b_1,\dots,b_d}(N)\right\|_F^2 &= \sum_{j_1,j_2\in[d]} 2^{\max(b_{j_1},b_{j_2})} N_{j_1 j_2}^2 \\
&\leq \sum_{k=0}^{\infty} 2^k \sum_{j_1\in S_k \text{ or } j_2 \in S_k} N_{j_1 j_2}^2 \\
&\leq \sum_{k=0}^{\infty} 2^k (2|S_k|\,\|N\|^2) \\
&= 2(2^{b_1}+\cdots+2^{b_d})\,\|N\|^2
\end{aligned}$$

and this gives the desired inequality. □

Claim D.4 ([11])**.** *Given measurement access to $\rho^{\otimes t}$ where $\rho \in \mathbb{C}^{d\times d}$ is a state, $\mathsf{Split}_{b_1,\dots,b_d}(\rho)$ is a valid state and we can simulate measurement access to access to $\mathsf{Split}_{b_1,\dots,b_d}(\rho)^{\otimes t}$.*

To complete the reduction, our full algorithm first obtains a rough estimate ρ' of ρ via tomography and then applies the splitting reduction in the eigenbasis of ρ'

Theorem D.5 ([15])**.** *For any $\delta, \varepsilon < 1$ and unknown state $\rho \in \mathbb{C}^{d\times d}$, there is an algorithm that makes unentangled measurements on $O(d^2\log(1/\delta)/\varepsilon^2)$ copies of ρ and with $1-\delta$ probability outputs a state $\widehat{\rho}$ such that $\|\rho - \widehat{\rho}\|_F \leq \varepsilon$.*

Theorem D.6. *Let ρ be an unknown quantum state in $\mathbb{C}^{d\times d}$. Then for any target accuracy $\varepsilon \leq 1/d^{12}$ and failure probability δ, there is an algorithm that measures $O(\log(1/\delta)/\varepsilon^2)$ copies of ρ and stores classical information (of $O(d^2\log(1/\delta))$ real numbers) such that given any Hermitian matrix $O \in \mathbb{C}^{d\times d}$, it can produce an estimate τ (from only this classical information) with*

$$\Pr\left[|\langle O,\rho\rangle - \tau| \geq \varepsilon\,\|O\|\right] \leq \delta\,.$$

Proof. First, we apply Theorem D.5 with half of the total copies to learn a state ρ' such that

$$\|\rho' - \rho\|_F \le d\varepsilon \le \frac{\sqrt{\varepsilon}}{d^5}\,.$$

Now let U be the matrix that diagonalizes ρ'. We will work in the U-basis where say $\rho' = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. For each $j \in [d]$ let b_j be the smallest nonnegative integer such that $2^{b_j} \ge d\lambda_j$. Note that we must have $2^{b_1} + \cdots + 2^{b_d} \le 4d$. Also, $\mathsf{Split}_{b_1,\ldots,b_d}(\rho')$ is a diagonal matrix with all entries at most $1/d$ so $\left\|\mathsf{Split}_{b_1,\ldots,b_d}(\rho')\right\| \le 1/d$. Let $k = 2^{b_1} + \cdots + 2^{b_d}$. Now let

$$\sigma = \frac{1}{3}\left(\frac{4I_k}{k} - \mathsf{Split}_{b_1,\ldots,b_d}(\rho')\right)\,.$$

Note that σ is a quantum state in $\mathbb{C}^{k\times k}$. Now by Claim D.4 and Lemma B.4, we simulate access to copies of the state

$$\tilde{\rho} = \frac{3}{4}\sigma + \frac{1}{4}\mathsf{Split}_{b_1,\ldots,b_d}(\rho)\,.$$

Note that

$$\left\|\tilde{\rho} - \frac{I_k}{k}\right\|_F = \left\|\frac{1}{4}\mathsf{Split}_{b_1,\ldots,b_d}(\rho - \rho')\right\|_F \le \frac{\sqrt{\varepsilon}}{d^5}\,.$$

Thus, we can apply Theorem C.6 using $O(1/\varepsilon^2)$ copies to obtain an estimate $\widehat{E}$ for $\tilde{\rho} - \frac{I_k}{k}$. We get that

$$\Pr\left[\left|\left\langle \mathsf{DSplit}_{b_1,\ldots,b_d}(O), \tilde{\rho} - \frac{I_k}{k}\right\rangle - \left\langle \mathsf{DSplit}_{b_1,\ldots,b_d}(O), \widehat{E}\right\rangle\right| \ge 4\varepsilon\,\|O\|\right] \le 0.1$$

where we used Claim D.3. When the above holds, then we set

$$\tau = \langle O, \rho'\rangle + 4\langle \mathsf{DSplit}_{b_1,\ldots,b_d}(O), \widehat{E}\rangle$$

and then get

$$|\langle O, \rho\rangle - \tau| \le 16\varepsilon\,\|O\|\,.$$

To complete the proof and get $1-\delta$ success probability, note that we can repeat the above $c = 10\log(1/\delta)$ times independently to obtain estimates $\widehat{E}_1, \ldots \widehat{E}_c$. Then for a given query O we compute estimates $\tau_1, \ldots, \tau_c$ as above and output the median of these estimates. □

E Reducing to Small ε via Random Projection

In this section, we show how to reduce to the case where $\varepsilon \le 1/\sqrt{d}$ by first applying a random "projection" (we will actually use matrices with Gaussian entries so it is not technically a projection). We show in Claim E.2 that we have an unbiased estimator after the projection and we bound the variance in Claim E.3.

Definition E.1. *For k matrices $V_1 \in \mathbb{C}^{d\times m}, \ldots, V_k \in \mathbb{C}^{d\times m}$ and a matrix $M \in \mathbb{C}^{d\times d}$, we write $M_{V_1,\ldots,V_k}$ to be the $k\times k$ matrix whose ij entry is $\mathrm{tr}(V_i^\dagger M V_j)$ when $i \ne j$ and is 0 otherwise.*

Claim E.2. *Let d, m be integers. Let $M, N \in \mathbb{C}^{d\times d}$ be traceless Hermitian matrices. Let $V_1, \ldots, V_k \in \mathbb{C}^{d\times m}$ be matrices with entries drawn i.i.d. from $N(0, 1/d)$. Then*

$$\mathbb{E}[\langle M_{V_1,\ldots,V_k}, N_{V_1,\ldots,V_k}\rangle] = \frac{k(k-1)m}{d^2}\langle M, N\rangle\,.$$

Proof. Let $i, j \in [k]$ with $i \neq j$. We have

$$\begin{aligned}\mathbb{E}[M_{V_1,\dots,V_k}[i,j]N_{V_1,\dots,V_k}[i,j]] &= \mathbb{E}[\mathrm{tr}(V_i^\dagger M V_j)\mathrm{tr}(V_i^\dagger N V_j)] \\ &= \frac{1}{d}\,\mathbb{E}[\langle V_i^\dagger M, V_i^\dagger N\rangle] \\ &= \frac{m}{d^2}\langle M, N\rangle\,.\end{aligned}$$

Now summing the above over all $i, j \in [k]$ with $i \neq j$ gives us

$$\mathbb{E}[\langle M_{V_1,\dots,V_k}, N_{V_1,\dots,V_k}\rangle] = \frac{k(k-1)m}{d}\langle M, N\rangle\,.$$

□

Now we bound the variance of the above quantity.

Claim E.3. *Let $M, N \in \mathbb{C}^{d\times d}$ be traceless Hermitian matrices. Let $V_1, \dots, V_k \in \mathbb{C}^{d\times m}$ be matrices whose entries are drawn i.i.d. from $N(0, 1/d)$. Then*

$$\mathsf{Var}(\langle M_{V_1,\dots,V_k}, N_{V_1,\dots,V_k}\rangle) \leq 6\left(\frac{m^2k^2\,\|M\|_F^2\,\|N\|_F^2}{d^4} + \frac{mk^2\langle M^2, N^2\rangle}{d^4} + \frac{mk^3\,\|MN\|_F^2}{d^4}\right).$$

Proof. We can write

$$\langle M_{V_1,\dots,V_k}, N_{V_1,\dots,V_k}\rangle = \sum_{i,j\in[k],i\neq j} \mathrm{tr}(V_i^\dagger M V_j)\mathrm{tr}(V_i^\dagger N V_j)\,.$$

Define

$$P_{ij} = \mathrm{tr}(V_i^\dagger M V_j)\mathrm{tr}(V_i^\dagger N V_j)\,,$$

for any pairs (i_1, j_1) and (i_2, j_2) that are disjoint, $\mathsf{Cov}(P_{i_1j_1}, P_{i_2j_2}) = 0$. Now we compute $\mathsf{Cov}(P_{ij_1}, P_{ij_2}) = 0$ where i, j_1, j_2 are all distinct. We have

$$\begin{aligned}\mathbb{E}[P_{ij_1}P_{ij_2}] &= \mathbb{E}_{V_i}[\mathbb{E}_{V_{j_1}}[\mathrm{tr}(V_i^\dagger M V_{j_1})\mathrm{tr}(V_i^\dagger N V_{j_1})]\,\mathbb{E}_{V_{j_2}}[\mathrm{tr}(V_i^\dagger M V_{j_2})\mathrm{tr}(V_i^\dagger N V_{j_2})]] \\ &= \mathbb{E}_{V_i}\left[\frac{1}{d^2}\langle V_i^\dagger M, V_i^\dagger N\rangle^2\right] \\ &= \frac{1}{d^2}\left(m\,\mathbb{E}_v[(vMNv^\dagger)^2] + m(m-1)(\mathbb{E}_v[vMNv^\dagger])^2\right) \\ &= \frac{m^2\mathrm{tr}(MN)^2}{d^4} + \frac{m\,\|MN\|_F^2}{d^4} + \frac{m\mathrm{tr}((MN)^2)}{d^4}\end{aligned}$$

where the vector v is a d-dimensional vector with entries drawn from $N(0, 1/d)$. Also

$$\begin{aligned}\mathbb{E}[P_{ij_1}]\,\mathbb{E}[P_{ij_2}] &= \left(\mathbb{E}_{V_i,V_j}[\mathrm{tr}(V_i^\dagger M V_j)\mathrm{tr}(V_i^\dagger N V_j)]\right)^2 \\ &= \left(\mathbb{E}_{V_i}\left[\frac{1}{d}\langle V_i^\dagger M, V_i^\dagger N\rangle\right]\right)^2 \\ &= \frac{m^2}{d^2}(\mathbb{E}_v[vMNv^\dagger])^2 \\ &= \frac{m^2\mathrm{tr}(MN)^2}{d^2}\end{aligned}$$

and thus,

$$\mathsf{Cov}(P_{ij_1}, P_{ij_2}) = \frac{2m \|MN\|_F^2}{d^4}$$

Next, we can compute $\mathsf{Var}(P_{ij})$ for $i \neq j$. We have

$$\begin{aligned}
E[P_{ij}^2] &= \mathbb{E}_{V_i,V_j}[\mathrm{tr}(V_i^\dagger M V_j)^2 \mathrm{tr}(V_i^\dagger N V_j)^2] \\
&= \frac{1}{d^2}\,\mathbb{E}_{V_i}\left[\left\|V_i^\dagger M\right\|_F^2 \left\|V_i^\dagger N\right\|_F^2 + 2\langle V_i^\dagger M, V_i^\dagger N\rangle^2\right] \\
&= \frac{m^2 \|M\|_F^2 \|N\|_F^2}{d^4} + \frac{2m\langle M^2, N^2\rangle}{d^4} + \frac{2m^2 \mathrm{tr}(MN)^2}{d^4} + \frac{2m\|MN\|_F^2}{d^4} + \frac{2m\mathrm{tr}((MN)^2)}{d^4}
\end{aligned}$$

and recall

$$\mathbb{E}[P_{ij}]^2 = \frac{m^2 \mathrm{tr}(MN)^2}{d^4}$$

so

$$\mathsf{Var}(P_{ij}) = \frac{m^2 \|M\|_F^2 \|N\|_F^2}{d^4} + \frac{2m\langle M^2, N^2\rangle}{d^4} + \frac{m^2 \mathrm{tr}(MN)^2}{d^4} + \frac{2m\|MN\|_F^2}{d^4} + + \frac{2m\mathrm{tr}((MN)^2)}{d^4}\,.$$

Thus, we can bound

$$\mathsf{Var}(\langle M_{V_1,\ldots,V_k}, N_{V_1,\ldots,V_k}\rangle) \leq 6\left(\frac{m^2k^2 \|M\|_F^2 \|N\|_F^2}{d^4} + \frac{mk^2\langle M^2, N^2\rangle}{d^4} + \frac{mk^3\|MN\|_F^2}{d^4}\right)$$

as desired. □

Corollary E.4. *Let $M, N \in \mathbb{C}^{d\times d}$ be Hermitian matrices such that $\|M\|_1 \leq 1$ and $\|N\|_{op} \leq 1$. Let k be a parameter and let $V_1, \ldots, V_k \in \mathbb{C}^{d\times 2d/k}$ be matrices whose entries are drawn i.i.d. from $N(0, 1/d)$. Then for any parameter $\gamma > 0$,*

$$\Pr\left[\left|\langle M, N\rangle - \frac{d}{2(k-1)}\langle M_{V_1,\ldots,V_k}, N_{V_1,\ldots,V_k}\rangle\right| \geq \gamma\right] \leq \frac{20d}{k^2\gamma^2}\,.$$

Proof. This follows from combining Claim E.2 and Claim E.3 and applying Chebyshev's inequality. □

Before we can complete the reduction, we need a few basic matrix concentration bounds. The following statements follow from standard matrix concentration inequalities.

Claim E.5. *Let $V_1, \ldots, V_k \in \mathbb{C}^{d\times 2d/k}$ be matrices whose entries are drawn i.i.d. from $N(0, 1/d)$. Then with probability 0.99,*

$$0.1 I_d \leq \sum_{i=1}^{k} V_i V_i^\dagger \leq 10 I_d\,.$$

Also let $N \in \mathbb{C}^{d\times d}$ be a fixed matrix with $\|N\|_{op} \leq 1$. Then with probability 0.99

$$\left\|N_{V_1,\ldots,V_k}\right\|_{op} \leq 10\,.$$

Before we formalize the reduction, we first need the following definition:

Definition E.6. *We say that an algorithm solves classical shadows with parameters d, ε with probability p using $f(d, \varepsilon, p)$ copies if for an arbitrary unknown state ρ, the algorithm makes measurements on $f(d, \varepsilon, p)$ copies of ρ and stores classical information such that for any observable $O \in \mathbb{C}^{d\times d}$, the algorithm can access only the classical information and produce an estimate τ such that with probability p,*

$$|\langle O, \rho\rangle - \tau| \leq \varepsilon \|O\| .$$

With this, we can now show:

Theorem E.7. *Assume we are given parameters d, ε, k such that $d \geq k \geq 100\sqrt{d}/\varepsilon$. Then if there is an algorithm for solving classical shadows with probability 0.9 for parameters $k, 0.1\varepsilon k/d$ using $f(k, 0.01\varepsilon k/d, 0.9)$ samples, then for any $\delta > 0$, there is also an algorithm for solving classical shadows with parameters d, ε that succeeds with probability $1 - \delta$ and uses $f(d, \varepsilon, 1 - \delta) \leq O(f(k, 0.01\varepsilon k/d, 0.9) + 1/\varepsilon^2) \cdot \log(1/\delta)$ samples.*

Proof. It suffices to prove the statement with $\delta = 1/3$ and then we can simply take $\log(1/\delta)$ independent runs of the algorithm and output the median of the estimates.

Now draw $V_1, \ldots, V_k \in \mathbb{C}^{d\times 2d/k}$ with i.i.d. entries drawn from $N(0, 1/d)$. Recall by Claim E.5, with probability 0.99,

$$0.1 I_d \preceq \sum_{i=1}^{k} V_i V_i^\dagger \preceq 10 I_d . \tag{2}$$

Assuming the above holds, we define

$$Y = I_d - 0.1 \sum_{i=1}^{k} V_i V_i^\dagger$$

and construct the following quantum channel. We map a matrix $M \in \mathbb{C}^{d\times d}$ to a block diagonal matrix with a $d \times d$ block consisting of $Y^{1/2} M Y^{1/2}$ and a $k \times k$ block with entries given by $0.1\mathrm{tr}(V_i^\dagger M V_j)$ for all $i, j \in [k]$ (this is similar to the matrix $M_{V_1,\ldots,V_k}$ except the diagonal is included as well). It is clear that this map is completely positive and trace preserving. We run all copies of ρ through this channel and then measure them according to the POVM $(\Pi, I_{d+k} - \Pi)$ where Π is the projector onto the $k \times k$ block. We keep all of the copies for which the measurement outcome is Π and discard the rest. Let

$$\alpha = \mathrm{tr}(V_1^\dagger \rho V_1) + \cdots + \mathrm{tr}(V_k^\dagger \rho V_k) .$$

Note that by (2), $\alpha \geq 0.1$. Let β be the fraction of samples that we actually keep. Since we have more than $O(1/\varepsilon^2)$ samples, with 0.99 probability, $0.1\alpha - 0.001\varepsilon \leq \beta \leq 0.1\alpha + 0.001\varepsilon$. Note that all of these copies are now in the state

$$\rho' = \frac{\rho_{V_1,\ldots,V_k} + \mathrm{diag}(\mathrm{tr}(V_1^\dagger \rho V_1), \ldots, \mathrm{tr}(V_k^\dagger \rho V_k))}{\mathrm{tr}(V_1^\dagger \rho V_1) + \cdots + \mathrm{tr}(V_k^\dagger \rho V_k)} .$$

We now run the classical shadows algorithm with parameters $k, 0.01\varepsilon k/d$ on this $k \times k$ matrix. As long as enough samples are kept in the previous step i.e. $\beta \geq 0.1\alpha - 0.001\varepsilon$, we have enough copies remaining to run this algorithm. Given a query, O, we construct the matrix $O_{V_1,\ldots,V_k}$ and then compute an estimate τ' for $\langle \rho', O_{V_1,\ldots,V_k}\rangle$.

WLOG assume $\|O\|_{\mathsf{op}} \leq 1$. Then by the assumption about the classical shadows algorithm, and as long as $\left\|O_{V_1,\ldots,V_k}\right\|_{\mathsf{op}} \leq 10$ (which happens with 0.99 probability by Claim E.5), our estimate τ' satisfies

$$|\tau' - \langle \rho', O_{V_1,\ldots,V_k}\rangle| \leq \frac{0.1\varepsilon k}{d}$$

with probability at least 0.9. Now we output the estimate

$$\tau = \frac{5d\beta}{(k-1)}\tau' .$$

Assuming that the previous inequality holds,

$$\begin{aligned}\left|\tau - \frac{10\beta}{\alpha}\langle\rho, O\rangle\right| &\leq \left|\frac{10\beta}{\alpha}\langle\rho, O\rangle - \frac{5d\beta}{(k-1)}\langle\rho', O_{V_1,\ldots,V_k}\rangle\right| + 0.6\varepsilon \\ &= \frac{10\beta}{\alpha}\left|\langle\rho, O\rangle - \frac{d}{2(k-1)}\langle\rho_{V_1,\ldots,V_k}, O_{V_1,\ldots,V_k}\rangle\right| + 0.6\varepsilon\end{aligned}$$

where the last step uses that $O_{V_1,\ldots,V_k}$ is 0 on the diagonal by definition. Now by Corollary E.4, with probability 0.9, the above quantity is at most 0.8ε. Also assuming $0.1\alpha - 0.001\varepsilon \leq \beta \leq 0.1\alpha + 0.001\varepsilon$, we have

$$\left|\frac{10\beta}{\alpha} - 1\right| \leq 0.1\varepsilon .$$

This immediately implies that our estimate τ has $|\tau - \langle\rho, O\rangle| \leq \varepsilon$. Overall, combining the failure probabilities over all of the steps, the total failure probability is less than $1/3$, so with $2/3$ probability, the estimate τ is ε-accurate and we are done. □