Quantum tomography using state-preparation unitaries*

Joran van Apeldoorn† Arjan Cornelissen‡ András Gilyén§ Giacomo Nannicini¶

Abstract

We describe algorithms to obtain an approximate classical description of a d-dimensional quantum state when given access to a unitary (and its inverse) that prepares it. For pure states we characterize the query complexity for ℓ_q-norm error up to logarithmic factors. As a special case, we show that it takes $\widetilde{\Theta}(d/\varepsilon)$ applications of the unitaries to obtain an ε-ℓ_2-approximation of the state.

For mixed states we consider a similar model, where the unitary prepares a purification of the state. We characterize the query complexity for obtaining Schatten q-norm estimates of a rank-r mixed state, up to polylogarithmic factors. In particular, we show that a trace-norm ($q = 1$) estimate can be obtained with $\widetilde{\Theta}(dr/\varepsilon)$ queries. This improves (assuming our stronger input model) the ε-dependence over the works of O'Donnell and Wright (*STOC 2016*) and Haah et al. (*IEEE Trans. Inf. Theory, 63.9, 2017*), that use a joint measurement on $\widetilde{\mathcal{O}}(dr/\varepsilon^2)$ copies of the state.

To our knowledge, the most sample-efficient results for pure-state tomography come from setting the rank to 1 in generic mixed-state tomography algorithms, which can require a large amount of computing resources. We describe sample-optimal algorithms for pure states that are simple and fast to implement.

Along the way we show that an ℓ_∞-norm estimate of a normalized vector induces a (slightly worse) ℓ_q-norm estimate for that vector, without losing a dimension-dependent factor in the precision. We also develop an unbiased and symmetric version of phase estimation, where the probability distribution of the estimate is centered around the true value. Finally, we give an efficient method for estimating multiple expectation values, improving over the recent result by Huggins et al. (*arXiv:2111.09283*) when the measurement operators do not fully overlap. More specifically, we show that for $E_1, \ldots, E_m$ normalized measurement operators, all expectation values $\mathrm{Tr}(E_j\rho)$ can be efficiently learned up to error ε with $\widetilde{\mathcal{O}}(\sqrt{\|\sum_j E_j^2\|}/\varepsilon)$ applications of a state-preparation unitary for a purification of ρ.

1 Introduction

Quantum state tomography is the process of obtaining a classical description of a quantum state. Tomography is a fundamental tool in quantum information science, where it finds numerous applications. In the context of quantum algorithms, pure quantum state tomography can be used to retrieve a classical description of the final state of the algorithm, e.g., the solution of a linear system [26] or the evolution of a quantum system [37]. The more general mixed quantum state tomography finds applications in quantum information theory, and in the simulation of quantum many-body and thermodynamic systems. In some settings we are not interested in the full state, but only in its expectation value under a certain set of (possibly overlapping) measurements. This was first introduced by Aaronson [1] under the name shadow tomography, and has since received a lot of attention in the literature, e.g., [30, 2, 29].

Most of the existing work on this topic has focused on the sample complexity of these problems: how many copies of the state are needed to perform tomography? In this paper we consider the problem under a different input model: we assume access to a unitary (and its inverse) that prepares the state. This model is very natural when the state is the output of a quantum algorithm. In fact, it is conceivable that in most practical situations one has a quantum circuit to prepare the quantum state of interest, and in this setting our assumptions are satisfied.

*The full version of the paper can be accessed at `https://arxiv.org/abs/2207.08800`

†QuSoft and IViR, UvA, Amsterdam, the Netherlands.

‡QuSoft, UvA, Amsterdam, the Netherlands.

§Alfréd Rényi Institute of Mathematics, Budapest, Hungary.

¶IBM Quantum, IBM T.J. Watson research center, Yorktown Heights, NY, USA. Current affiliation: University of Southern California.

Suprisingly, this model has received little attention so far. The main improvements in this model come from the ability to use techniques related to amplitude estimation to reduce the dependence on the error parameter, but attaining such quadratic improvements requires the development of several new tools, and the analysis does not follow from a simple application of amplitude estimation.

Throughout the paper we consider either a d-dimensional pure state $|\psi\rangle = \sum_{j=0}^{d-1} \alpha_j |j\rangle$ or a rank-r mixed state $\rho \in \mathbb{C}^{d\times d}$. We are interested in learning the state up to error ε in some ℓ_q-norm or Schatten q-norm, often with some probability of failure $\leq \delta$. In the introduction we often use $\widetilde{\mathcal{O}}(\cdots)$ notation to hide polylogarithmic factors in the parameters d, r, $1/\varepsilon$, and $1/\delta$, even if these parameters do not appear polynomially in the $\widetilde{\mathcal{O}}(\cdots)$. For more precise complexity statements we refer to the relevant theorems in the main text.

Related work. Classical algorithms that estimate probabilities generally depend quadratically on $1/\varepsilon$, as that many samples are required to bring down the variance. In certain settings quantum algorithms can improve on this classical complexity. Brassard et al. [9] introduced the amplitude estimation algorithm, and showed that it can estimate an amplitude (or probability) with a $1/\varepsilon$ dependence, if a state-preparation unitary and its inverse are available.

Van Apeldoorn [4] generalized this for finding an ℓ_∞-norm estimate of a discrete probability distribution. In the model of van Apeldoorn, access to the distribution is given by a state-preparation oracle (and its inverse), such that the probability distribution corresponds to computational-basis measurements of the prepared state. Van Apeldoorn [4] showed that $\widetilde{\mathcal{O}}(1/\varepsilon)$ applications of the input unitary are sufficient to compute the desired ℓ_∞-norm estimate. In the same paper the question was posed whether one can also speed-up the estimation of multiple expectation values over the same distribution. A lower bound of $\Omega(\min\{\sqrt{m}/\varepsilon, 1/\varepsilon^2\})$ was given when m expectation values need to be estimated op to precision ε. It was later shown by Huggins et al. [31] that $\widetilde{\mathcal{O}}(\sqrt{m}/\varepsilon)$ queries are sufficient even when estimating expectation values of observables on a pure quantum state.

Kerenidis and Prakash [33] gave a sampling-based approach for estimating the real-valued amplitudes resulting from a quantum linear system solver, including their sign, taking $\widetilde{\mathcal{O}}\big(d/\varepsilon^2\big)$ applications of a (controlled) state-preparation unitary to compute an ℓ_2-norm estimate. We subsume their approach, and show that besides estimating real-valued amplitudes, one can even estimate complex amplitudes with the same sample complexity.

Besides these few results for pure quantum state tomography, the most frequently studied setting is that of mixed-state tomography. In this setting we want to determine how many copies are necessary to obtain a classical description with a given maximum error ε in trace-norm; it is often assumed that some upper bound r on the rank of the state is known (if the state is pure, $r = 1$). An algorithm of Gross et al. [23], that applies measurements on one copy of the state at once, achieves $\mathcal{O}\big(dr^2/\varepsilon^2\big)$ sample complexity, see also [24]. Haah et al. [25] show that the bound is optimal when the measurements are on a single copy at a time, and Chen et al. [11] complete our understanding of this setting by showing that the bound cannot be improved even with adaptive measurements schemes, as long as we require single-copy measurements. A better sample complexity can be achieved if we allow joint measurements on multiple copies of the state: with this more powerful access model, the best algorithm for tomography is due to O'Donnell and Wright [42] and Haah et al. [25], and it requires $\widetilde{\mathcal{O}}\big(dr/\varepsilon^2\big)$ copies of the quantum state (in fact, the analysis of [42] eliminates polylogarithmic factors). Haah et al. also show matching lower bounds up to polylogarithmic factors, therefore their algorithm is essentially optimal (Yuen [49] eliminates the polylogarithmic factors from the lower bound, if the distance is computed in terms of fidelity). The main drawback of their approach is that it does not only requires joint measurements on many states at once, but it also has time complexity exponential in d.

Our contribution. Our results on quantum state tomography for pure states can be divided into two settings: the sampling-based setting, in which copies of the state are available, and the state-preparation unitary setting, in which we require controlled access to a state-preparation unitary and its inverse.

To give optimal algorithms for ℓ_q-norms in general, we prove a norm-conversion lemma relating estimates in different ℓ_q-norms. The standard approach for norm conversion is to decrease the allowed error ε by a factor $d^{1/q}$, but this introduces a dependence on the dimension that can be suboptimal. We show that a dimension-independent norm conversion is possible for normalized vectors, and therefore for quantum states. We also relate estimates of the amplitudes to estimates of the corresponding probability distribution.

For pure quantum state tomography using samples, the best sample complexity is obtained by setting $r = 1$ in general mixed-state tomography algorithms, but these methods are highly impractical from a computational standpoint; i.e., even though the sample complexity is optimal, the time complexity is much worse (exponential

in d for [42, 25]; if we are willing to accept a worse sample complexity of $\widetilde{\mathcal{O}}(dr^2/\varepsilon^2)$, then Guţă et al. [24] give an algorithm with $\widetilde{\mathcal{O}}(d^3)$ time complexity, which is large but polynomial in d). We cover three different models with our sampling-based pure-state tomography results, and for each give an easy to implement tomography algorithm:

1. *Classical samples.* In this model we are given classical samples from computational-basis measurements. As we cannot recover information about the phases, we aim to produce an estimate of $|\alpha|$, the vector of absolute values of the amplitudes.

2. *Copies of the state.* In this model we are given copies of the quantum state, and aim to give an estimate up to a global phase. Our algorithm does not require joint measurements on different copies, but the algorithm is adaptive in the sense that it proceeds in two phases, where the outcomes of the first phase are used to transform the state before subsequent measurements.

3. *Conditional copies of the state.* In this model we are given copies of $(|0\rangle|\psi\rangle + |1\rangle|0\rangle)/\sqrt{2}$, and aim to give an estimate of $|\psi\rangle$ including the global phase. This is inspired by controlled usage of a state-preparation unitary (but not its inverse), as that allows producing such samples.

Our algorithms for these three models all give the same sample complexity, up to polylogarithmic factors:

THEOREM 1.1. *(Informal) In all three sampling input models $\widetilde{\mathcal{O}}(1/\varepsilon^2)$ samples are sufficient to obtain an ℓ_∞-norm estimate with error at most ε. For ℓ_q-norm error ($q \geq 2$) the sample complexity*[1] *is $\widetilde{\mathcal{O}}\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{2}{q}}}{\varepsilon^2}\right\}\right)$. See Thorems 4.1, 4.2, 4.3.*

All our sampling-based approaches require a number of samples that scales quadratically with $1/\varepsilon$ to obtain an ℓ_∞-norm estimate. For sampling approaches this error dependence is optimal even when estimating a single amplitude only. However, when estimating a single amplitude with access to a state-preparation unitary and its inverse, amplitude estimation can be used to improve this dependence to linear [9]. Van Apeldoorn [4] shows that this can be generalized to estimate all probabilities in the corresponding distribution with linear dependence. Unfortunately, for amplitudes it is impossible to get an $\mathcal{O}(1/\varepsilon)$ error dependence that is independent of the dimension. However, for the high-precision regime there is still an improvement.

THEOREM 1.2. *(Informal) Given controlled access to a state-preparation unitary for $|\psi\rangle$ and its inverse, $\widetilde{\mathcal{O}}\left(\min\left\{\frac{\sqrt{d}}{\varepsilon}, \frac{1}{\varepsilon^2}\right\}\right)$ uses of these unitaries are sufficient to estimate the vector α with ℓ_∞-norm error at most ε. For ℓ_q-norm error ($q \geq 2$) this bound becomes $\widetilde{\mathcal{O}}\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{1}{2}+\frac{1}{q}}}{\varepsilon}\right\}\right)$. See Theorem 5.1.*

In the final section of our paper we show matching lower bounds for the above sample and query complexities. We show an $\tilde{\Omega}(d/\varepsilon^2)$ bound for ℓ_1-norm estimation of the probability distribution induced by a state $|\psi\rangle$ given access to copies of $\frac{|0\rangle|\psi\rangle+|1\rangle|0\rangle}{\sqrt{2}}$, using a communication complexity argument. We also show that with access to a state-preparation unitary, this requires $\Omega\left(\frac{d}{\varepsilon}\right)$ applications of the input unitary, with a reduction from the problem of determining an unknown bit string via queries to a fractional phase oracle. Using our results on the relation between different norms (and between probability estimates and amplitude estimates), we obtain the following result.

THEOREM 1.3. *(Informal) All the upper bounds on pure-state tomography given in this paper are optimal, up to polylogarithmic factors. See Theorems 9.1, 9.2.*

We then turn to mixed quantum states of rank at most r. We show how to find an entry-wise ε-approximation using $\widetilde{\mathcal{O}}\left(\frac{\sqrt{d}}{\varepsilon}\right)$ queries with entry-wise independent error, and that this yields an ε-operator-norm estimate if we set the entry-wise error to $\varepsilon/\sqrt{d}$. This leads to the following result.

[1]Here, and in the rest of the paper, when working with norms we use $1/0 = \infty$ and $1/\infty = 0$. If one of the terms in the $\min\{\dots\}$ goes to ∞ due to this, then the complexity is simply the other term.

	Sampling models	Unitary model
ℓ_∞-norm	$\frac{1}{\varepsilon^2}$	$\min\{\frac{1}{\varepsilon^2}, \frac{\sqrt{d}}{\varepsilon}\}$
ℓ_2-norm	$\frac{d}{\varepsilon^2}$	$\frac{d}{\varepsilon}$
ℓ_q-norm	$\min\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{2}{q}}}{\varepsilon^2}\}$	$\min\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{1}{2}+\frac{1}{q}}}{\varepsilon}\}$

Table 1: Sample and query complexities of recovering a d-dimensional pure quantum state up to error ε in the specified norm, for the different models. All results are $\tilde{\Theta}$, i.e., they are tight up to polylogarithmic factors in d, $1/\varepsilon$ and $1/\delta$, where δ is the maximum failure probability.

THEOREM 1.4. *(Informal) Given access to a state-preparation unitary (and its inverse) for a purification of a rank-r quantum state $\rho \in \mathbb{C}^{d\times d}$, $\widetilde{\mathcal{O}}\left(\frac{d}{\varepsilon}\right)$ uses of these unitaries are sufficient to estimate ρ in operator norm. For trace norm error $\widetilde{\mathcal{O}}\left(\frac{dr}{\varepsilon}\right)$ uses suffice. See Theorem 8.2.*

To obtain this we develop two new intermediate results of independent interest: unbiased and symmetric phase estimation, and a computationally improved shadow tomography with state-preparation unitaries. The unbiased version of phase estimation is required for the conversion between entry-wise error and operator-norm error mentioned above (if all entry-wise errors go in the same direction then the best possible conversion would give a factor d, not $\sqrt{d}$). We show that phase estimation can be made unbiased and symmetric by adding a random phase before applying the inverse quantum Fourier transform, then removing this phase from the estimate.

THEOREM 1.5. *(Informal) Quantum phase estimation can be used to give an unbiased and symmetric estimator of the phase. See Theorem 6.4.*

Second, we implement a version of shadow tomography when given access to a state-preparation unitary for a purification of the state. Huggins et al. [31] show that we can learn the expectation values of m normalized measurement operators using $\widetilde{\mathcal{O}}(\sqrt{m}/\varepsilon)$ queries to the state-preparation unitary. (They prove a slightly more general results if the operators are not normalized, see Section 7.) We improve on this for the case where the measurement operators do not fully overlap, while recovering the same bound for the general case.

THEOREM 1.6. *(Informal) Let $E_1, \dots E_m$ be measurement operators with operator norm at most 1. Given controlled access to a state-preparation unitary (and its inverse) for a purification of a quantum state $\rho \in \mathbb{C}^{d\times d}$, $\widetilde{\mathcal{O}}\left(\sqrt{\|\sum_j E_j^2\|}/\varepsilon\right)$ uses of these unitaries are sufficient to estimate all $\mathrm{Tr}(\rho E_j)$ up to error ε. See Theorem 7.3.*

	Unitary model	
	Upper bound	Lower bound
Max-norm	$\frac{\sqrt{d}}{\varepsilon}$	$\frac{1}{\varepsilon}$
Operator norm	$\frac{d}{\varepsilon}$	$\frac{d}{\varepsilon}$
Frobenius norm	$\min\{\frac{d\sqrt{r}}{\varepsilon}, \frac{d}{\varepsilon^2}\}$	$\min\{\frac{d\sqrt{r}}{\varepsilon}, \frac{d}{\varepsilon^2}\}$
Trace norm	$\frac{dr}{\varepsilon}$	$\frac{dr}{\varepsilon}$
Schatten-q-norm	$\min\{\frac{dr^{\frac{1}{q}}}{\varepsilon}, \frac{d}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}\}$	$\min\{\frac{dr^{\frac{1}{q}}}{\varepsilon}, \frac{d}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}\}$

Table 2: Our upper and lower bounds on the query complexities of recovering a d-dimensional, rank-r mixed quantum state up to error ε in the specified norm, when given access to unitary that prepares its purification. All upper bound results are $\widetilde{\mathcal{O}}$, i.e., they are given up to polylogarithmic factors in d, $1/\varepsilon$ and $1/\delta$, where δ is the maximum failure probability. The lower bound results are Ω, and we observe that our results are tight up to polylogarithmic factors everywhere, except for the max-norm.

Finally, we prove lower bounds on the estimation of a density matrix given access to a unitary (and its inverse) that prepares a purification of it. The lower bound proof on high level consists of three steps. First,

we embed a bit string of length dr into a family of density matrices. Then, we quantify how much information about the embedded bit string can be obtain by an algorithm that recovers any of these density matrices up to the specified precision. We conclude by arguing that obtaining this amount of information about the bit string requires a particular number of queries to the state-preparing unitary. Our results show a $\Omega(dr/\varepsilon)$ lower bound for all $\varepsilon = O(1)$, and Table 2 gives an overview of the other lower bound results that follow from this.

2 Preliminaries

In this section we introduce some notation, our computational model, and give a brief overview of some important technical tools that we use. In particular we introduce some important results in the block-encoding framework [20], and a version of Jordan's gradient algorithm [32].

2.1 Notation and computational model For any integer j, we define $[j] := \{0, \dots, j-1\}$. Let $\oplus$ denote the direct sum, i.e., $A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$. We write $\vec{1}$ for the all-ones vector and J for the all-ones matrix, with dimensions that will be clear from context. We write Δ^d for the set of d-dimensional probability distributions. We write $[a, \infty]$ for the set $[a, \infty) \cup \{\infty\}$. All logarithms are base 2 unless otherwise indicated.

Given a vector v, we write $\|v\|_q$ for the standard ℓ_q-norm. We use the convention that $1/0 = \infty$ and $1/\infty = 0$ in calculations involving the value of q for an ℓ_q-norm. Although the letter p is commonly used to denote norms (i.e., ℓ_p-norms), in this paper we use p to denote vectors containing the entries of a discrete probability distribution; hence, we use different letters for norms. For a matrix M we use write $\|M\|_q$ for the Schatten q-norm, i.e., the ℓ_q-norm of the vector of singular values. For operator norm (Schatten ∞-norm) we just write $\|M\|$. We write that $\tilde{\alpha}$ is an ε-ℓ_q-norm estimate of α if $\|\alpha - \tilde{\alpha}\|_q \leq \varepsilon$. For a vector α, we denote by $|\alpha|$ the vector with entries given by the modulus of the entries of α.

We assume that the quantum computer is controlled by a classical computer (with a RAM) that can choose the gates to run depending on intermediate measurement results. For simplicity, we neglect the cost of any classical computation as long as it is only a polylogarithmic factor (in all input parameters) slower than the quantum gate complexity. Our gate set consists of all single-qubit gates and CNOT. To simplify the statements of our results we assume access to a QRAM-like gate, the indexed-SWAP gate. This gate acts on a state with many qubits as follows:

$$\text{indexed-SWAP}|i\rangle|j\rangle|x_1\rangle \dots |x_d\rangle = |i\rangle|j\rangle \text{SWAP}_{i,j}(|x_1\rangle \dots |x_d\rangle)$$

where $\text{SWAP}_{i,j}$ swaps the i-th and j-th qubit. Such a gate can be built using $\mathcal{O}(d)$ gates, and $\log(d)$ depth, see Appendix B for details on this implementation. We always state the number of calls to such a gate and the size of the memory it acts on.

2.2 Block-encodings We list the technical results that we need to efficiently manipulate matrices given via block-encoding circuits. For more background see [17]. First we define a block-encoding as follows.

Definition 2.1. (Block-encoding) *A unitary U is an a-qubit block-encoding of A if the top-left block of the unitary U is A:*

$$A = \big(\langle 0|^{\otimes a} \otimes I\big) U \big(|0\rangle^{\otimes a} \otimes I\big) \iff U = \begin{bmatrix} A & \cdot \\ \cdot & \cdot \end{bmatrix}.$$

Note that we are simplifying the block-encoding framework: traditionally, block-encodings are defined with three parameters (normalization factor, number of additional qubits, error of the implementation), but in this paper the normalization factor and error of the implementation are easily tracked without additional notation. Thus, we use a simpler presentation. Readers familiar with block-encodings can easily restate our results using the more familiar notation.

Although we do not use POVMs (positive-operator-valued measures) directly, we mention the following lemma to showcase that the block-encoding framework is applicable in large generality. In particular, thanks to the following lemma, some of our results in the block-encoding framework are directly applicable to POVMs.

Lemma 2.1. *[5] If a two-outcome POVM denoted by E can be coherently implemented on a quantum computer using a ancillary qubits via the unitary U, then an $(a+1)$-block-encoding of E can be implemented using a single call to U, $U^\dagger$, and a CNOT gate.*

The following two lemmas show how to add and amplify block-encodings, which we use repeatedly for mixed-state tomography.

LEMMA 2.2. (LINEAR COMBINATION OF BLOCK ENCODINGS) *[20, 5] Let $E = \sum_{j=0}^{m} y_j E_j$ be a w-qubit operator for $y \in \mathbb{R}^m$, and let $\beta \geq \|y\|_1$. If U_y is a state-preparation oracle for $\frac{1}{\sqrt{\beta}} \sum_j \sqrt{y_j}|j\rangle|0\rangle + |\psi\rangle|1\rangle$ for some unnormalized state $|\psi\rangle$, and U_E implements an a-block-encoding of E_j conditioned on j, then an $(a+\lceil\log(m)\rceil+1)$-block-encoding of E/β can be implemented with a single use of U_y, $U_y^\dagger$, and U_E, and a single two-qubit gate.*

LEMMA 2.3. (UNIFORM AMPLIFICATION OF BLOCK-ENCODINGS, [39],[19, THEOREM 33]) *Let U be an a-block-encoding of A, and let $\|A\| \leq \beta \leq 1$. Then an $(a+1)$-block-encoding of $A/(2\beta)$ can be implemented, up to operator norm error ε, using $\mathcal{O}(\beta \log(\beta/\varepsilon))$ applications of U and $U^\dagger$, and $\mathcal{O}(a\beta \log(\beta/\varepsilon))$ additional gates with circuit depth $\mathcal{O}(\log(a)\beta \log(\beta/\varepsilon))$.*

One of the main motivations for defining block-encodings is the following Hamiltonian simulation result, that we use for implementing "phase oracles" required for gradient computation.

LEMMA 2.4. (HAMILTONIAN SIMULATION USING BLOCK-ENCODINGS, [40],[19, COROLLARY 63]) *Let U be an a-block-encoding of A. Then an $(a+2)$-block-encoding of $e^{\mathbf{i}tA}$ can be implemented, up to operator norm error ε, using $\mathcal{O}(t + \log(1/\varepsilon))$ applications of (controlled) U and $U^\dagger$, and $\mathcal{O}(a(t + \log(1/\varepsilon)))$ additional gates with circuit depth $\mathcal{O}(\log(a)(t + \log(1/\varepsilon)))$.*

Finally, we will use the following lemma to construct block-encodings for gradient computation:

LEMMA 2.5. (BLOCK-ENCODING INNER PRODUCTS WITH CONTROLLED STATE-PREPARATION UNITARIES) *Let $U := \sum_x U_x \otimes |x\rangle\langle x|$ and $V := \sum_x V_x \otimes |x\rangle\langle x|$ be controlled (by the second register) state-preparation unitaries, where $U_x \colon |0\rangle|0\rangle^{\otimes a} \mapsto |0\rangle|\psi_x\rangle + |1\rangle|\tilde{\psi}_x\rangle$ and $V_x \colon |0\rangle|0\rangle^{\otimes a} \mapsto |0\rangle|\phi_x\rangle + |1\rangle|\tilde{\phi}_x\rangle$ are $(a+1)$-qubit state-preparation unitaries for some (subnormalized) a-qubit quantum states $|\psi_x\rangle, |\phi_x\rangle$. Then $(I_1 \otimes V^\dagger) \cdot (SWAP \otimes I) \cdot (I_1 \otimes U)$ is an $(a+2)$-block-encoding of the diagonal matrix $\mathrm{diag}(\{\langle\phi_x, \psi_x\rangle\})$, where I_1 acts on a single qubit and the SWAP gate acts on the first two qubits.*

Proof.

$$\begin{aligned}
&\langle 0|^{\otimes a+2}\langle x|(I \otimes V^\dagger) \cdot (\mathrm{SWAP} \otimes I) \cdot (I \otimes U)|0\rangle^{\otimes a+2}|y\rangle \\
&\qquad = \langle 0|\Big(\langle 0|\langle\phi_x| + \langle 1|\langle\tilde{\phi}_x|\Big)\langle x|(\mathrm{SWAP} \otimes I)|0\rangle\Big(|0\rangle|\psi_x\rangle + |1\rangle|\tilde{\psi}_x\rangle\Big)|y\rangle \\
&\qquad = (\langle 00|\langle\phi_x| + \langle 01|\langle\tilde{\phi}_x|)\langle x|(|00\rangle|\psi_x\rangle + |10\rangle|\tilde{\psi}_x\rangle)|y\rangle \\
&\qquad = \langle\phi_x, \psi_y\rangle\delta_{xy}
\end{aligned}$$

□

2.3 Quantum gradient computation We now briefly review Jordan's algorithm for estimating the gradient and provide a generic analysis of its behavior. Before describing the algorithm, we introduce appropriate representation of our qubit strings suitable for fixed-point arithmetics.

DEFINITION 2.2. ([18, DEFINITION 5.1]) *For every $b \in \{0,1\}^n$, let $j^{(b)} \in \{0, \ldots, 2^n - 1\}$ be the integer corresponding to the binary string $b = (b_1, \ldots, b_n)$. We label the n-qubit basis state $|b_1\rangle|b_2\rangle \cdots |b_n\rangle$ by $|x^{(b)}\rangle$, where*

$$x^{(b)} = \frac{j^{(b)}}{2^n} - \frac{1}{2} + 2^{-n-1}.$$

We denote the set of corresponding labels as $G_n := \left\{\frac{j^{(b)}}{2^n} - \frac{1}{2} + 2^{-n-1} : j^{(b)} \in \{0, \ldots, 2^n - 1\}\right\}$. Note that there is a bijection between $\{j^{(b)}\}_{b\in\{0,1\}^n}$ and $\{x^{(b)}\}_{b\in\{0,1\}^n}$, so we will use $|x^{(b)}\rangle$ and $|j^{(b)}\rangle$ interchangeably.

Following [18, Definition 5.2] for $x \in G_n$ we define the Fourier transform of a state $|x\rangle$ as

$$QFT_{G_n} : |x\rangle \mapsto \frac{1}{\sqrt{2^n}} \sum_{k\in G_n} e^{2\pi i 2^n xk}|k\rangle.$$

In [18, Claim 5.1] it is shown that this unitary is the same as the usual quantum Fourier transform up to conjugation with a tensor product of n single-qubit unitaries.

We now prove a simplified version of [18, Lemma 5.1] in order to give some intuition about Jordan's gradient computation algorithm, which can be viewed as a continuous extension of the Bernstein-Vazirani algorithm [8].

LEMMA 2.6. (THE CORE OF JORDAN'S GRADIENT COMPUTATION ALGORITHM) *Let $N = 2^n$, and $g \in \mathbb{R}^d$ such that $\|g\|_\infty \leq 1/3$. If $\left\| |\psi\rangle - \frac{1}{\sqrt{N^d}} \sum_{x \in G_n^d} e^{2\pi \mathbf{i} N \langle g, x\rangle} |x\rangle \right\| + \left\| \tilde{Q} - QFT_{G_n}^{-1} \right\| \leq \frac{1}{12}$ then measuring the state $\tilde{Q}^{\otimes d}|\psi\rangle$ in the computational basis yields an estimate $k \in G_n^d$ such that*

$$\Pr[|k_j - g_j| > 3/N] \leq 1/3 \quad \textit{for every } j \in [d].$$

Proof. The proof is analogous to that of [18, Lemma 5.1]. Observe that the "ideal" state is a product state

$$\bigotimes_{j=1}^{d} \frac{1}{\sqrt{N}} \sum_{x_j \in G_n} e^{2\pi \mathbf{i} N g_j x_j} |x_j\rangle.$$

If we apply $\mathrm{QFT}_{G_n}^{-1}$ to the j-th component of the state we get

$$\frac{1}{N} \sum_{x_j, k_j \in G_n} e^{2\pi \mathbf{i} N x_j (g_j - k_j)} |k_j\rangle.$$

In the analysis of phase estimation [41], it can be shown[2] that for every $i \in [d]$, the we have

$$\Pr\left[|k_j - g_j| > \frac{3}{N}\right] \leq \frac{1}{4}. \tag{2.1}$$

Since we work with the state $|\psi\rangle$ instead of the "ideal" state, and apply $\tilde{Q}$ as opposed to $\mathrm{QFT}_{G_n}^{-1}$ the measurement statistics might differ. On the other hand the closeness condition implies that

$$\left\| \left(I_{G_b}^{\otimes [j-1]} \otimes \widetilde{Q} \otimes I_{G_b}^{\otimes [d] \setminus [j]} \right) |\psi\rangle - \left(I_{G_b}^{\otimes [j-1]} \otimes \mathrm{QFT}_{G_n}^{-1} \otimes I_{G_b}^{\otimes [d] \setminus [j]} \right) \frac{1}{\sqrt{N^d}} \sum_{x \in G_n^d} e^{2\pi i N \langle g, x\rangle} |x\rangle \right\| \leq \frac{1}{12}.$$

This means that the probability of the above event Eq. (2.1) changes by at most $\frac{1}{12}$ (see for example [48, Exercise 4.3]). Even though we analysed the procedure only for a single coordinate, the measurement statistics of the j-th coordinate is unaffected by operations that act on the other qubits. Therefore, the error analysis still holds if apply $\tilde{Q}$ to all coordinates and then measure. ∎

Now we derive the corollary that we use for estimating various quantities. Before proving it we invoke two technical results that we use in the proof.

THEOREM 2.1. (CHERNOFF-HOEFFDING [28, THEOREM 1]) *Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $0 \leq X_i \leq 1$ for every $i \in [n]$, and let $\overline{X} := \frac{1}{n} \sum_{i \in [n]} X_i$. If $p := \mathbb{E}[\overline{X}] \in (0, 1)$, then for $0 \leq \varepsilon \leq 1 - p$ we have that*

$$\Pr(\overline{X} \geq p + \varepsilon) \leq \exp(-D(p + \varepsilon \| p) n) \leq \exp(-2n\varepsilon^2),$$

where $D(x\|y) = x \ln(\frac{x}{y}) + (1 - x) \ln(\frac{1-x}{1-y})$.

Let QFT_n denote the standard quantum Fourier transform over n qubits. We rely on the following efficient implementations devised by Cleve and Watrous.

[2]Note that this is where we use the assumption $|g_j| \leq 1/3$ in order to convert the phases to the intervals $[-\frac{1}{3}, \frac{1}{3}]$. Also note that the Fourier transform we use is slightly altered, but the same argument still holds as in [41, (5.34)]. One can also directly translate the result by considering the conjugation of the ordinary quantum Fourier transform with a tensor product of n single-qubit unitaries derived in [18].

THEOREM 2.2. (PARALLEL QFT CIRCUITS [12, THEOREMS 1-2]) *For any $n \in \mathbb{N}$ there is a quantum circuit of depth $\mathcal{O}(n)$ implementing QFT_n with $\mathcal{O}\big(n(\log n)^2 \log\log n\big)$ gates. Moreover, for any $\varepsilon \in (0,1)$ there is a quantum circuit $\widetilde{Q}$ of depth $\mathcal{O}(\log n + \log\log(1/\varepsilon))$ such that $\left\|\widetilde{Q} - QFT_n\right\| \leq \varepsilon$ and $\widetilde{Q}$ consists of $\mathcal{O}(n\log(n/\varepsilon))$ gates.*

Since QFT_{G_n} can be obtained from QFT_n via conjugation by a layer of single-qubit gates [18, Claim 5.1], the above complexity bounds also apply for implementing QFT_{G_n}. Now we are ready to prove the main corollary of this preliminary section.

COROLLARY 2.1. (ALMOST-LINEAR BLOCK-HAMILTONIAN TO GRADIENT) *Let $\varepsilon, \delta \in (0, \frac{1}{6}]$, $b := \lceil \log_2(\frac{24}{\varepsilon})\rceil$, $B = 2^b$ and $\beta := \frac{1}{48}$. Suppose that we have an a-block-encoding W of a diagonal matrix with diagonal entries $f(x) \in \mathbb{R}$ for $x \in G_b^d$ satisfying $|f(x) - \langle x, g\rangle| \leq \frac{\varepsilon\beta}{6\pi}$ for at least a $(1-\beta^2)$ fraction of the points in G_b^d. Then with $\mathcal{O}\big(\frac{1}{\varepsilon}\log(\frac{d}{\delta})\big)$ (controlled) uses of W (and its inverse) and $\mathcal{O}\big(\big(d\log(\frac{1}{\varepsilon})\log\log(\frac{1}{\varepsilon}) + \frac{a}{\varepsilon}\big)\log(\frac{d}{\delta})\big)$ other gates with circuit depth $\mathcal{O}\Big(\frac{\log(a)}{\varepsilon}\Big)$ we can compute a vector $k \in [-\frac{8}{3}, \frac{8}{3}]^d$ such that $\Pr[\|k-g\|_\infty > \varepsilon] \leq \delta$.*

Proof. The main idea is to apply Lemma 2.6 with preparing the (approximate) initial state via block-Hamiltonian simulation Lemma 2.4. The first step is to prepare a uniform superposition over the grid G_b^d by applying a Hadamard gate to all $d \cdot b$ qubits, that are initially in the $|0\rangle$ state.

Note that due to the assumptions in the statement we have that $|f(x)| \leq 1$ and so $|\langle x, g\rangle| \leq 1 + \varepsilon \leq \frac{7}{6}$ for at least $1 - 2^{-10}$ fraction of points in G_b^d since $\beta \leq 2^{-5}$, in turn implying that $\|g\|_\infty \leq \frac{8}{3}$. Indeed, let us assume that $g_j > \frac{8}{3}$, we show that this would imply that for at least half of the points with $x_j \geq \frac{7}{16}$ we have that $\langle x, g\rangle > \frac{7}{6}$. First, clearly $x_j \cdot g_j > \frac{7}{6}$. Let $\bar{g} \in \mathbb{R}^{d-1}$ be the vector we get from g by removing its j-th coordinate. Then for any $\bar{x} \in G_b^{d-1}$ we have that $\langle \bar{x}, \bar{g}\rangle = -\langle -\bar{x}, \bar{g}\rangle$ so at least one of $\langle \bar{x}, \bar{g}\rangle, \langle -\bar{x}, \bar{g}\rangle$ is greater than or equal zero. Since $b \geq 4$ at least a $\frac{1}{16}$ fraction of points $x \in G_b$ satisfy $x \geq \frac{7}{16}$, and so for at least a $\frac{1}{32}$ fraction of points $x \in G_b^d$ we would get $\langle x, g\rangle > \frac{7}{6}$. Therefore, we will apply Lemma 2.6 to the gradient $\frac{g}{8}$ with precision $\frac{\varepsilon}{8}$.

First let us assume that we have access to a perfect phase oracle $P := \sum_{x \in G_b^d} |x\rangle\langle x| e^{2\pi i \frac{B}{8} f(x)}$ so that we can prepare the sate $|\psi'\rangle = \frac{1}{\sqrt{B^d}} \sum_{x \in G_b^d} |x\rangle e^{2\pi i \frac{B}{8} f(x)}$. We can bound the difference from the ideal state $|\phi\rangle = \frac{1}{\sqrt{N^d}} \sum_{x \in G_b^d} e^{2\pi i N \langle g, x\rangle} |x\rangle$ analogously to the proof of [18, Lemma 5.1]. Let $S \subseteq G_b^d$ be the set of points for which $|f(x) - \langle x, g\rangle| \leq \frac{\varepsilon\beta}{4\pi}$ holds, then

$$
\begin{aligned}
\||\psi'\rangle - |\phi\rangle\|^2 &= \frac{1}{B^d} \sum_{x \in G_b^d} \left| e^{2\pi i \frac{B}{8} f(x)} - e^{2\pi i \frac{B}{8} \langle x, g\rangle} \right|^2 \\
&= \frac{1}{B^d} \sum_{x \in S} \left| e^{2\pi i \frac{B}{8} f(x)} - e^{2\pi i \frac{B}{8} \langle x, g\rangle} \right|^2 + \frac{1}{B^d} \sum_{x \in G_b^d \setminus S} \left| e^{2\pi i \frac{B}{8} f(x)} - e^{2\pi i \frac{B}{8} \langle x, g\rangle} \right|^2 \\
(|e^{iz} - e^{iy}| \leq |z - y|) \qquad &\leq \frac{1}{B^d} \sum_{x \in S} \left| 2\pi \frac{B}{8} f(x) - 2\pi \frac{B}{8} \langle x, g\rangle \right|^2 + \frac{1}{B^d} \sum_{x \in G_b^d \setminus S} 4 \\
&= \frac{1}{B^d} \sum_{x \in S} (2\pi \frac{B}{8})^2 |f(x) - \langle x, g\rangle|^2 + 4 \frac{|G_b^d \setminus S|}{B^d} \\
(\text{by the assumptions of the corollary}) \qquad &\leq \frac{1}{B^d} \sum_{x \in S} 4\beta^2 + 4\beta^2 \\
&\leq 8\beta^2.
\end{aligned}
$$

We can implement a $(3 - 2\sqrt{2})\beta$-approximation $\widetilde{P}$ of the perfect phase oracle P by applying block-Hamiltonian simulation Lemma 2.4 to W.[3] This enables us to prepare an approximate state $|\tilde{\psi}\rangle$ such that $\left\||\tilde{\psi}\rangle - |\psi'\rangle\right\| \leq (3 - 2\sqrt{2})\beta$ and so in turn $\left\||\tilde{\psi}\rangle - |\phi\rangle\right\| \leq 3\beta$. If we apply an approximate (inverse) quantum Fourier transform $\widetilde{Q}$

[3] An ε-precise $(a+2)$-block-encoding of e^{itH} is $\mathcal{O}(\sqrt{\varepsilon})$-close in operator norm to a perfect Hamiltonian simulation unitary U of the form $|0\rangle\langle 0|^{\otimes a+2} \otimes e^{\mathbf{i}tA} + V$, where $V(|0\rangle^{a+2} \otimes I) = 0$.

[6] such that $\left\|\tilde{Q} - \mathrm{QFT}_{G_n}^{-1}\right\| \leq \beta$ then the conditions of Lemma 2.6 hold, therefore measuring the state yields and estimate satisfying for each $j \in [d]$

$$\Pr\left[|k_j - g_j| > \frac{3}{B}\right] \leq \frac{1}{3}.$$

Finally, we repeat the entire procedure $2m$-times for $m := \lceil 10 \ln(\frac{d}{\delta}) \rceil$ and take the median of the estimates for each coordinate $j \in [d]$. If $|k_j - g_j| \leq \frac{3}{B}$ holds for at least $m+1$ estimates, then the median will give a $\frac{3}{B} \leq \frac{\varepsilon}{8}$-precise estimate for g_j. We use the Chernoff-Hoeffding theorem to bound the probability of failure. Indeed, Theorem 2.1 shows that the probability that $|k_j - g_j| > \frac{3}{B}$ holds for at least m out of $2m$ estimates is at most $\exp(-D(\frac{1}{2}\|\frac{1}{3})2m)) \leq \exp(-\frac{m}{10}) \leq \frac{\delta}{d}$. By the union bound this implies that $\Pr[\|k-g\|_\infty > \varepsilon] \leq \delta$. Since $\|g\|_\infty \leq \frac{8}{3}$ we can always truncate each coordinate into $[-\frac{8}{3}, \frac{8}{3}]$ without affecting this failure probability.

The query complexity follows from the fact that we prepare the state $|\tilde{\psi}\rangle$ a total of $\mathcal{O}\big(\log(\frac{d}{\delta})\big)$ times, each time making $\mathcal{O}(B) = \mathcal{O}\big(\frac{1}{\varepsilon}\big)$ (controlled) queries to W. The additional gate complexity of preparing $|\tilde{\psi}\rangle$ is $\mathcal{O}(a)$ times the query complexity (coming from Lemma 2.4) plus the number of initial Hadamard gates. The biggest contribution to the gate complexity comes from the implementation of the approximate (inverse) quantum Fourier transform $\widetilde{Q}$. By Theorem 2.2 the gate complexity of $\widetilde{Q}$ can be bounded by $\mathcal{O}(b\log(b/\beta)) = \mathcal{O}\big(\log(\frac{1}{\varepsilon})\log\log(\frac{1}{\varepsilon})\big)$ while its depth by $\mathcal{O}(\log(b) + \log\log(1/\beta)) = \mathcal{O}\big(\log\log(\frac{1}{\varepsilon})\big)$. The additional classical computation can also be performed in parallel with depth $\mathcal{O}(\mathrm{poly}(b,m))$ which is $\mathcal{O}\big(\mathrm{polylog}(\frac{d}{\delta\varepsilon})\big)$, since $m = \mathcal{O}\big(\log(\frac{d}{\delta})\big)$, and $b = \mathcal{O}\big(\log(\frac{1}{\varepsilon})\big)$. ∎

In Section 6 we improve upon the above Lemma 2.6 and Corollary 2.1 by making them (essentially) unbiased, by using our new unbiased phase estimation subroutine instead of just applying $\big(\mathrm{QFT}_{G_n}^{-1}\big)^{\otimes n}$ to each coordinate in Jordan's algorithm. Those improvements play a vital role for our mixed-state tomography results, but they are not necessary for pure-state tomography. Since the unbiased version has some additional $\log\log$ factors, we use the simpler routine for now.

3 Relations between vector estimates

In this section we prove two lemmas that relate different types of estimates for vectors. The first lemma shows a relation between estimates of the vector of amplitudes, and of the vector of corresponding probabilities. The second lemma relates ℓ_q-estimates for different values of q when the vector is normalized (as is the case with amplitudes and probabilities). Together, these two lemmas allow us to upper bound the complexity of giving ℓ_q-norm estimates for both amplitudes and probabilities, starting from an ℓ_∞-norm estimate for amplitudes. Similarly, with these lemmas a lower bound on the complexity of finding an ℓ_1-norm estimate for probabilities translates to a lower bound on all other cases.

3.1 Relation between amplitude and probability estimation

For a classical probability distribution, we learn all aspects of the distribution by estimating it in ℓ_1-norm error, i.e., total variation distance. For pure quantum states the ℓ_2-norm error plays a similar role. It is natural to ask how an ℓ_2-norm estimate of a quantum state relates to an ℓ_1-norm estimate of the probability distribution given by computational-basis measurements on that state. In a similar fashion, we want to understand this question also when using ℓ_∞-norm error on the quantum state. We provide answers to these questions by showing a relation between ℓ_q-norm error on a state and ℓ_r-norm error on the corresponding probability distribution; the special case $q = 2, r = 1$ is also discussed in [8, Lemma 3.6].

LEMMA 3.1. *Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state. Let $p \in \mathbb{R}^d$ given by $p_j = |\alpha_j|^2$ be the probability distribution arising from a computational-basis measurement. Let $q \in [2,\infty]$ and let $t := \frac{1}{\frac{1}{q}+\frac{1}{2}} \in [1,2]$. An ε-ℓ_q-norm estimate $|\tilde{\alpha}|$ of $|\alpha|$ can be used to compute a 4ε-t-norm estimate $\tilde{p}$ of p, using $\mathcal{O}(d)$ gates.*

Proof. We first note that for any ℓ_q-norm estimate $|\tilde{\alpha}|$ of a unit vector $|\alpha|$ we may assume $\||\tilde{\alpha}|\|_2 = 1$, that is, the vector represents a pure state. Indeed, if this is not the case, we can instead use $|\tilde{\alpha}|/\||\tilde{\alpha}|\|_q$, which we can

compute using $\mathcal{O}(d)$ gates, and which satisfies

$$\begin{aligned}\left\| |\alpha| - \frac{|\tilde{\alpha}|}{\||\tilde{\alpha}|\|_q} \right\|_q &\leq \||\alpha| - |\tilde{\alpha}|\|_q + \left\| |\tilde{\alpha}| - \frac{|\tilde{\alpha}|}{\||\tilde{\alpha}|\|_q} \right\|_q \\ &\leq \varepsilon + \||\tilde{\alpha}|\|_q \cdot \left(1 - \frac{1}{\||\tilde{\alpha}|\|_q}\right) \\ &\leq \varepsilon + (1+\varepsilon) \cdot \left(1 - \frac{1}{1+\varepsilon}\right) \\ &= 2\varepsilon.\end{aligned}$$

By assumption, $q = \frac{t}{1-t/2}$. Using Hölder's inequality for $\lambda_1, \lambda_2 \geq 1$ with $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} = 1$, we obtain the following upper bound on the ℓ_t-norm error in an estimate of p:

$$\begin{aligned}\|p - \tilde{p}\|_t &= \left(\sum_{j\in[d]} |p_j - \tilde{p}_j|^t\right)^{1/t} \\ &= \left(\sum_{j\in[d]} ||\alpha_j|^2 - |\tilde{\alpha}|_j^2|^t\right)^{1/t} \\ &\leq \left(\sum_{j\in[d]} ||\alpha_j| - |\tilde{\alpha}|_j|^t ||\alpha_j| + |\tilde{\alpha}|_j|^t\right)^{1/t} \\ &\leq \left(\sum_{j\in[d]} ||\alpha_j| - |\tilde{\alpha}||^{t\lambda_1}\right)^{\frac{1}{t\lambda_1}} \left(\sum_{j\in[d]} ||\alpha_j| + |\tilde{\alpha}|_j|^{t\lambda_2}\right)^{\frac{1}{t\lambda_2}} \\ &= \||\alpha| - |\tilde{\alpha}|\|_{t\lambda_1} \||\alpha| + |\tilde{\alpha}|\|_{t\lambda_2}.\end{aligned}$$

(Recall that $\lambda_1, \lambda_2, t \geq 1$, so $t\lambda_1 \geq 1$ and $t\lambda_2 \geq 1$.) Pick $\lambda_2 = 2/t$. We then have

$$\||\alpha| + |\tilde{\alpha}|\|_{t\lambda_2} = \||\alpha| + |\tilde{\alpha}|\|_2 \leq \||\alpha|\|_2 + \||\tilde{\alpha}|\|_2 \leq 2.$$

Combining this with $\lambda_1 = \frac{1}{1-\frac{1}{\lambda_2}} = \frac{1}{1-\frac{t}{2}}$, we get

$$\begin{aligned}\|p - \tilde{p}\|_t &\leq \||\alpha| - |\tilde{\alpha}|\|_{t\lambda_1} \||\alpha| + |\tilde{\alpha}|\|_{t\lambda_2} \\ &\leq \||\alpha| - |\tilde{\alpha}|\|_{\frac{t}{1-t/2}} \cdot 2 \\ &= 2\||\alpha| - |\tilde{\alpha}|\|_q \\ &\leq 4\varepsilon.\end{aligned}$$

□

Note that the reverse does not hold, and in particular ε-ℓ_1-norm estimates of p are not equivalent to $\Theta(\varepsilon)$-ℓ_2-norm estimates of α . This is not only due to the information about the phases being lost: even for the case $d = 2$ where the α_j are positive reals, estimating the probabilities is not enough to learn the amplitudes to a similar error. In particular, let $p_0 = \alpha_0^2 = \varepsilon \leq 1$ and let $\tilde{p}_0 = p_0 + \varepsilon$ be an ε-estimate for the probability. The amplitude satisfies

$$\tilde{\alpha}_0 = \sqrt{\tilde{p}_0} = \sqrt{p_0 + \varepsilon} = \sqrt{2\varepsilon} \geq \left(1 + \frac{1}{4}\right)\sqrt{\varepsilon} = \alpha_0 + \frac{1}{4}\sqrt{\varepsilon}.$$

Thus, the precision gets quadratically worse for small amplitudes.

3.2 Dimension-independent norm conversion for normalized vectors If we have an estimate of a vector with error at most ε in, for example, the ℓ_∞-norm, we can use norm conversion to show that this is also an estimate with error at most $\varepsilon d^{1/q}$ in the ℓ_q-norm. However, this bound is poor for large d. Here, we show that we can do better if we know that the vector we are estimating is normalized in some ℓ_s-norm, using the fact that such a vector cannot have too many large entries. In fact, we obtain a norm conversion lemma that does not depend on the dimension at all. We first prove a very general version of the following lemma; for results in subsequent sections of the paper we always use $s = 2$ (for quantum states) or $s = 1$ (for probability distributions), and set $\gamma = 1$.

LEMMA 3.2. *Let $\alpha \in \mathbb{C}^d$ be such that $\|\alpha\|_s \leq \gamma$. Let $\tilde{\alpha} \in \mathbb{C}^d$ be such that $\|\alpha - \tilde{\alpha}\|_\infty \leq \eta$. Let $\bar{\alpha} \in \mathbb{C}^d$ be the vector defined as $\bar{\alpha}_j = \tilde{\alpha}_j$ if $|\tilde{\alpha}_j| \geq 2\eta$, $\bar{\alpha}_j = 0$ otherwise. Then for all $q \in (s, \infty)$ we have $\|\alpha - \bar{\alpha}\|_q \leq \min\{4\eta^{\frac{q-s}{q}}\gamma^{\frac{s}{q}}, 3d^{1/q}\eta\}$.*

Proof. The second term in the min follows from the standard norm conversion and the fact that $\bar{\alpha}$ is an 3η-ℓ_∞-approximation; thus, we only need to prove the first term.

We know that $\|\alpha - \tilde{\alpha}\|_\infty \leq \eta$. Let $J = \{j \in [d] : |\tilde{\alpha}_j| \geq 2\eta\}$. Then

$$\{j \in [d] : |\alpha_j| \geq 3\eta\} \subseteq J \subseteq \{j \in [d] : |\alpha_j| \geq \eta\}.$$

And, as $\|\alpha\|_s \leq \gamma$, we have $|J| \leq \frac{\gamma^s}{\eta^s}$.

Now, let $\bar{\alpha}$ be $\tilde{\alpha}$ on all $j \in J$ and 0 everywhere else. On the indices in J we know that $\bar{\alpha}$ is an η estimate of α. Then

$$\begin{aligned}
\|\alpha - \bar{\alpha}\|_q &= \left(\sum_{j\in[d]} |\alpha_j - \bar{\alpha}_j|^q\right)^{1/q} \\
&\leq \left(\sum_{j\notin J} |\alpha_j - \bar{\alpha}_j|^q\right)^{1/q} + \left(\sum_{j\in J} |\alpha_j - \bar{\alpha}_j|^q\right)^{1/q} \\
&\leq \left(\sum_{j\notin J} |\alpha_j|^{q-s}|\alpha_j|^s\right)^{1/q} + \left(|J| \max_j |\alpha_j - \bar{\alpha}_j|^q\right)^{1/q} \\
&\leq \left(\sum_{j\notin J} (3\eta)^{q-s}|\alpha_j|^s\right)^{1/q} + \left(\frac{\gamma^s}{\eta^s}\eta^q\right)^{1/q} \\
&= (3\eta)^{\frac{q-s}{q}} \left(\sum_{j\notin J} |\alpha_j|^s\right)^{1/q} + \eta^{\frac{q-s}{q}}\gamma^{\frac{s}{q}} \\
&\leq 4\eta^{\frac{q-s}{q}}\gamma^{\frac{s}{q}}.
\end{aligned}$$

□

The lemma stated above is very general, but we only use it with several very specific parameter settings. Thus, we present the following simplified statement.

COROLLARY 3.1. *Let $\varepsilon \in (0, 1]$, $s \leq q$, and let y be an ℓ_s-normalized complex vector. In order to obtain an ε-ℓ_q-norm estimate of y, an η-ℓ_∞-norm estimate suffices for*

$$\eta = \max\left\{\frac{1}{3}\left(\frac{\varepsilon}{3}\right)^{\frac{1}{1-\frac{s}{q}}}, \frac{\varepsilon}{d^{\frac{1}{q}}}\right\}.$$

Proof. The first term follows from Lemma 3.2 by letting $\gamma = 1$. The second term in the max comes from a standard norm-conversion on the vector of errors, as a $v \in [-\varepsilon, \varepsilon]^d$ has q-norm at most $\varepsilon d^{1/q}$. □

This corollary has an immediate consequence. If one is only interested in finding an ℓ_q-estimate of the vector of probabilities p_j prepared by some unitary operation $U : |0\rangle \mapsto \sum_{j=1}^{d} \sqrt{p_j}|j\rangle$, then one can directly apply Corollary 3.1 in conjunction with the $\widetilde{\mathcal{O}}(1/\varepsilon)$-query algorithm for ℓ_∞-algorithm from [4]. The number of controlled and inverse calls to U then becomes

$$\widetilde{\mathcal{O}}(1/\eta) = \widetilde{\mathcal{O}}\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{1-\frac{1}{q}}}, \frac{d^{\frac{1}{q}}}{\varepsilon}\right\}\right),$$

which we show to be optimal up to polylogarithmic factors in Section 9.2.

4 Pure-state tomography using copies

In this section we present and analyze pure-state tomography algorithms that use very little quantum power. The first algorithm that we describe, in Section 4.1, is part of the folklore: we just take measurements in the computational basis, and obtain the absolute values of the amplitudes from the measurement outcomes. We are not aware of a specific reference for the sample complexity of this method, hence we provide a proof for completeness. Then, we add the ability to perform some operations on the quantum state, in Section 4.2 and Section 4.3: the first section simplifies the analysis of the tomography algorithm given in [33], the second one relaxes some of the assumptions with a slight increase in the sample complexity. The strongest model in this section relates to the case where we have access to a state-preparation unitary and its controlled version, but not its inverse. We discuss the setting where the inverse is available in Section 5.

4.1 Absolute values using computational-basis measurements Given classical samples via computational-basis measurements, how many samples do we need for an ℓ_q-norm estimate of α? Clearly we cannot learn the phases, so we have to limit ourselves to the absolute values of the amplitudes. Even then, the remark at the end of Section 3.1 seems discouraging: on the surface, estimation of the related distribution p seems the best that we can do with computational-basis measurements, and converting the error bound from probabilities to amplitudes makes the precision quadratically worse. However, as we discuss next, $\widetilde{\mathcal{O}}\big(\frac{1}{\varepsilon^2}\big)$ samples suffice and are optimal to give an ℓ_∞-norm estimate of $|\alpha|$.

PROPOSITION 4.1. *Let $0 < \varepsilon, \delta < 1$. Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j|j\rangle$ be a quantum state with $\alpha_j \in \mathbb{C}$, and let $p \in \mathbb{R}^d$, defined by $p_j = |\alpha_j|^2$, be the probability distribution of the outcomes of a computational-basis measurement. Then, $\mathcal{O}\big(\log(d/\delta)/\varepsilon^2\big)$ measurements of $|\psi\rangle$ in the computational basis suffice to learn an ε-ℓ_∞-norm estimate $|\tilde{\alpha}|$ of $|\alpha|$, with success probability at least $1-\delta$.*

Proof. Let us consider a single coordinate α_j with associated probability $p_j = |\alpha_j|^2$. Our goal is to estimate $|\alpha_j|$. We take k samples to find an estimate $\tilde{p}_j$ of p_j. The Chernoff bound tells us that for the error ε_j in this coordinate we have

$$\mathbb{P}[\tilde{p}_j > p_j + \varepsilon_j] \le e^{-D(p_j+\varepsilon_j||p_j)k}, \qquad \text{if } p_j + \varepsilon_j < 1,$$
$$\mathbb{P}[\tilde{p}_j > p_j - \varepsilon_j] \le e^{-D(p_j-\varepsilon_j||p_j)k}, \qquad \text{if } p_j - \varepsilon_j > 0,$$

where $D(x||y)$ is Kullback–Leibler divergence. We need the conditions shown on the right-hand side because the Kullback-Leibler divergence is only defined for $x, y \in (0,1)$, but it is easily observed that if the conditions on the right are not satisfied, the probabilities on the left-hand side trivially become 0. Since $D(x||y) \ge \frac{(x-y)^2}{2\max\{x,y\}}$, for all $x, y \in (0,1)$, we get

$$\mathbb{P}[|\tilde{p}_j - p_j| > \varepsilon_j] \le 2e^{-\frac{\varepsilon_j^2}{2(p_j+\varepsilon_j)}k},$$

and it is easily checked that this bound also holds whenever $p_j + \varepsilon_j \ge 1$, or $p_j - \varepsilon_j \le 0$. Hence, picking $k \ge \frac{2(p_j+\varepsilon_j)\ln(2/\delta')}{\varepsilon_j^2} = \frac{2(|\alpha_j|^2+\varepsilon_j)\ln(2/\delta')}{\varepsilon_j^2}$ ensures that $P[|\tilde{p}_j - p_j| \ge \varepsilon_j] \le \delta'$.

We now pick $\varepsilon_j = \varepsilon|\alpha_j|/2 + (\varepsilon/2)^2$. Note that we do not actually know this value, as it depends on the yet-to-be-estimated α_j, but with this choice we find that

$$\frac{2(|\alpha_j|^2+\varepsilon_j)\ln(\frac{2}{\delta'})}{\varepsilon_j^2} = \frac{2(|\alpha_j|^2+\frac{\varepsilon|\alpha_j|}{2}+(\frac{\varepsilon}{2})^2)\ln(\frac{2}{\delta'})}{(\frac{\varepsilon|\alpha_j|}{2}+(\frac{\varepsilon}{2})^2)^2} \le \frac{2(|\alpha_j|^2+\varepsilon|\alpha_j|+(\frac{\varepsilon}{2})^2)\ln(\frac{2}{\delta'})}{(\frac{\varepsilon}{2})^2(|\alpha_j|+\frac{\varepsilon}{2})^2} = \frac{8\ln(\frac{2}{\delta'})}{\varepsilon^2}.$$

Thus, it suffices to choose $k = 8\ln(2/\delta')/\varepsilon^2$. Letting $\delta' = \delta/d$ and applying the union bound, we have that, with probability $1-\delta$, for all $j \in [d]$, the resulting estimates $\tilde{p}_j$ satisfy $|\tilde{p}_j - p_j| \le \varepsilon_j$. First, this implies

$$\begin{aligned}
|\tilde{\alpha}_j| - |\alpha_j| &\le \sqrt{p_j + \varepsilon_j} - |\alpha_j| \\
&= \sqrt{|\alpha_j|^2 + \frac{\varepsilon|\alpha_j|}{2} + \left(\frac{\varepsilon}{2}\right)^2} - |\alpha_j| \\
&\le \sqrt{|\alpha_j|^2 + \varepsilon|\alpha_j| + \left(\frac{\varepsilon}{2}\right)^2} - |\alpha_j| \\
&= |\alpha_j| + \frac{\varepsilon}{2} - |\alpha_j| \\
&= \frac{\varepsilon}{2} \\
&< \varepsilon.
\end{aligned}$$

Next we show that $|\alpha_j| - |\tilde{\alpha}_j| < \varepsilon$. First consider the case where $p_j \le \varepsilon_j$. In that case, we have

$$|\alpha_j|^2 = p_j \le \frac{\varepsilon|\alpha_j|}{2} + \left(\frac{\varepsilon}{2}\right)^2 \qquad \Leftrightarrow \qquad \left(\frac{2|\alpha_j|}{\varepsilon}\right)^2 \le \frac{2|\alpha_j|}{\varepsilon} + 1 \qquad \Leftrightarrow \qquad |\alpha_j| \le \frac{1+\sqrt{5}}{4}\varepsilon.$$

Hence, we find that

$$|\alpha_j| - |\tilde{\alpha}_j| \le |\alpha_j| \le \frac{1+\sqrt{5}}{4}\varepsilon < \varepsilon.$$

On the other hand, if $p_j > \varepsilon_j$, we have

$$\begin{aligned}
|\alpha_j| - |\tilde{\alpha}_j| \le |\alpha_j| - \sqrt{p_j - \varepsilon_j} &= |\alpha_j| - \sqrt{|\alpha_j|^2 - \frac{\varepsilon|\alpha_j|}{2} - \left(\frac{\varepsilon}{2}\right)^2} \\
&= |\alpha_j| - \left(|\alpha_j| - \frac{\varepsilon}{4}\right)\sqrt{1 - \frac{5(\frac{\varepsilon}{2})^2}{4(|\alpha_j| - \frac{\varepsilon}{4})^2}} \\
&\le |\alpha_j| - \left(|\alpha_j| - \frac{\varepsilon}{4}\right)\left(1 - \frac{5(\frac{\varepsilon}{2})^2}{4(|\alpha_j| - \frac{\varepsilon}{4})^2}\right) \\
&= \frac{\varepsilon}{4} + \frac{5(\frac{\varepsilon}{2})^2}{4(|\alpha_j| - \frac{\varepsilon}{4})} \le \frac{\varepsilon}{4} + \frac{5\varepsilon^2}{16\frac{\sqrt{5}}{4}\varepsilon} \\
&= \frac{\varepsilon}{4} + \frac{\sqrt{5}\varepsilon}{4} \\
&< \varepsilon,
\end{aligned}$$

where in the last line, we used that $|\tilde{\alpha}_j| > (1+\sqrt{5})\varepsilon/4$. Thus, we can compute a vector $|\tilde{\alpha}|$ that is an ε-ℓ_∞-norm estimate of $|\alpha|$ with $O\big(\log(d/\delta)/\varepsilon^2\big)$ samples. ∎

We can use the above theorem to approximate the vector of absolute values of the amplitudes in other norms as well.

THEOREM 4.1. *Let $0 < \delta < 1$, $\varepsilon > 0$, $d \in \mathbb{N}$ and $q \in [2, \infty]$. Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state with $\alpha_j \in \mathbb{C}$. Then,*

$$O\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{2}{q}}}{\varepsilon^2}\right\} \cdot \log\left(\frac{d}{\delta}\right)\right)$$

computational-basis measurements of $|\psi\rangle$ suffice to learn an ε-ℓ_q-norm estimate $|\tilde{\alpha}|$ of $|\alpha|$, with success probability $1-\delta$.

Proof. Since the vector $|\alpha|$ is normalized in ℓ_2-norm, we know from Corollary 3.1, that in order to obtain an ε-ℓ_q-norm estimate of $|\alpha|$, it suffices to find an η-ℓ_∞-estimate of $|\alpha|$, where

$$\eta = \max\left\{\frac{1}{3}\left(\frac{\varepsilon}{3}\right)^{\frac{1}{1-\frac{2}{q}}}, \frac{\varepsilon}{d^{\frac{1}{q}}}\right\}.$$

From Lemma 4.1, we now find that this can be done using

$$O\left(\frac{\log\left(\frac{d}{\delta}\right)}{\eta^2}\right) = O\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{2}{q}}}{\varepsilon^2}\right\} \cdot \log\left(\frac{d}{\delta}\right)\right).$$

computational-basis measurements. ∎

4.2 Recovering the phase information using conditional samples Our discussion above shows how to estimate the vector of the absolute values of the amplitudes α_j with $\mathcal{O}\big(\log d/\varepsilon^2\big)$ copies of the quantum state. In this section, we consider having conditional samples of the state $|\psi\rangle$, by which we mean states of the form

$$\frac{|0\rangle|\psi\rangle + |1\rangle|0\rangle}{\sqrt{2}},$$

and we consider the problem of recovering all complex amplitudes of $|\psi\rangle$, including the phases.

We note here that if we have access to a controlled state-preparation unitary, we can prepare such a conditional sample of the state with one call to this operation. Crucially, we do not need the inverse of the state-preparation unitary – if we have access to that as well, then the results from Section 5 improve over those presented here.

The algorithm is based on the Hadamard test, as described in the next result.

LEMMA 4.1. *Let $|\psi_0\rangle = \sum_{j\in[d]} \alpha_j|j\rangle$, $|\psi_1\rangle = \sum_{j\in[d]} \beta_j|j\rangle$, and let*

$$|\phi\rangle = \frac{|0\rangle|\psi_0\rangle + |1\rangle|\psi_1\rangle}{\sqrt{2}}.$$

Using $\mathcal{O}\big(\log(d/\delta)/\varepsilon^2\big)$ copies of $|\phi\rangle$, we can, with success probability at least $1-\delta$, compute an ε-ℓ_∞-norm estimate of the $2d$-dimensional vector containing entries $|\alpha_j \pm \beta_j|$, for all $j = 0, \ldots, d-1$.

Proof. Note that

$$|\phi\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \sum_{j\in[d]} \alpha_j|j\rangle + |1\rangle \sum_{j\in[d]} \beta_j|j\rangle\right).$$

Applying a Hadamard gate to the first qubit yields

$$\frac{1}{2}\left(|0\rangle \sum_{j\in[d]} (\alpha_j + \beta_j)|j\rangle + |1\rangle \sum_{j\in[d]} (\alpha_j - \beta_j)|j\rangle\right).$$

We can now perform computational-basis measurements on this new state, and build up a histogram of the observed outcomes. The proposition then follows from Lemma 4.1, by setting the precision to $\varepsilon/2$. ∎

Inspired by the method used by Kerenidis and Prakash [33], we apply the above proposition to two states with amplitudes α and $|\alpha|$. The algorithm of Kerenidis and Prakash is only concerned with real amplitudes, hence it only needs to estimate the sign of each large α_j. Clearly this can be learned from a sufficiently precise estimate of $|\alpha_j - |\alpha_j||$. Since we consider general phases we need to be more careful, as we need to distinguish between a very small positive complex component $\varepsilon\mathbf{i}$ and a very small negative complex $-\varepsilon\mathbf{i}$. To this end we also apply the proposition to α and $\mathbf{i}|\alpha|$, and we give a more careful geometric analysis.

PROPOSITION 4.2. *Let $0 < \varepsilon, \delta < 1$, and let $|\psi\rangle = \sum_{j\in[d]} \alpha_j|j\rangle$ be a quantum state with $\alpha_j \in \mathbb{C}$. Then $\mathcal{O}\big(\log(d/\delta)/\varepsilon^2\big)$ copies of $(|0\rangle|\psi\rangle + |1\rangle|0\rangle)/\sqrt{2}$ suffice to compute an ε-ℓ_∞-norm estimate $\tilde{\alpha}$ of α, with success probability at least $1-\delta$.*

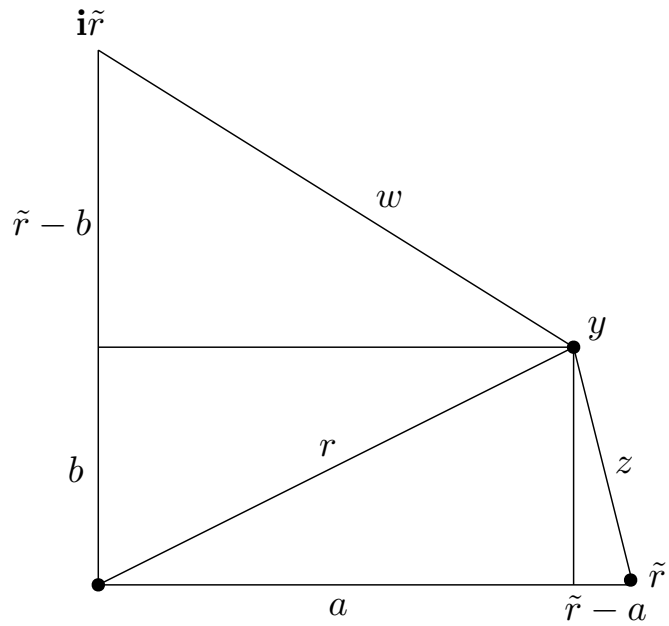

Figure 1: Geometry of the different points in $\mathbb{C}$ involved in the proof of Proposition 4.2. Subscripts "j" have been dropped for clarity.

Proof. Let us consider a single amplitude $\alpha_j = (a_j, b_j)$. Let $r_j = |\alpha_j|$. Using Lemma 4.1 we get an $\varepsilon/32$ estimate $\tilde{r}_j$ of r_j with $\mathcal{O}\big(\log(d/\delta)/\varepsilon^2\big)$ samples. We only consider the elements where $\tilde{r}_j \geq \varepsilon/2$, as we can set the rest to 0 without introducing too much error.

Let $z_j = |\alpha_j - \tilde{r}_j|$, and $w_j = |\alpha_j - \mathbf{i}\tilde{r}_j|$; see Figure 1. Using Lemma 4.1 we can find an $\varepsilon/32$-approximation of z_j with the same number of samples as before, and similarly for w_j, because we can construct a controlled unitary that transforms $(|0\rangle|\psi\rangle + |1\rangle|0\rangle)/\sqrt{2}$ into $(|0\rangle|\psi\rangle + |1\rangle \sum_j \tilde{r}_j|j\rangle)/\sqrt{2}$, and similarly with an extra phase (we assume w.l.o.g. that $\|\tilde{r}\|_2 = 1$, as in the proof of Lemma 3.1).

We now show how to find a_j from r_j and z_j, should we know them exactly; note that z_j is still defined using measured value $\tilde{r}_j$, i.e., we are in the situation of Figure 1. The squared length of the vertical line down from α_j is $b_j^2 = r_j^2 - a_j^2$ by Pythagoras. Applying Pythagoras again, we find that

$$z_j^2 = b_j^2 + (\tilde{r}_j - a_j)^2 = r_j^2 - a_j^2 + (\tilde{r}_j - a_j)^2 = r_j^2 + \tilde{r}_j^2 - 2\tilde{r}_j a_j,$$

hence $a_j = \frac{1}{2}\tilde{r}_j + \frac{1}{2}\frac{r_j^2}{\tilde{r}_j} - \frac{z_j^2}{2\tilde{r}_j}$. Thus, we have a formula for a_j, and we just need to bound the error that may affect a_j when we use the estimates $\tilde{r}_j$ and $\tilde{z}_j$ instead of r_j and z_j. Here it is important that we used our estimate $\tilde{r}_j$ for the β_j in Lemma 4.1, and hence $\tilde{r}_j$ is the exact length of the horizontal line, not an estimate of it.

To give an upper bound on the error induced by the error in our estimates $\tilde{r}_j$ and $\tilde{z}_j$, we consider the gradient of the function $f_{a_j}(r_j, z_j) = \frac{1}{2}\tilde{r}_j + \frac{1}{2}\frac{r_j^2}{\tilde{r}_j} - \frac{z_j^2}{2\tilde{r}_j}$ in terms of r_j and z_j:

$$\nabla f_{a_j} = \left(\frac{r_j}{\tilde{r}_j}, -\frac{z_j}{\tilde{r}_j}\right).$$

We bound the ℓ_1-norm of the gradient on the box defined by the constraints $|r_j - \tilde{r}_j| \leq \varepsilon/32$, $|z_j - \tilde{z}_j| \leq \varepsilon/32$. We know that $\varepsilon/2 \leq \tilde{r}_j$, and hence that $r_j + \varepsilon/32 \leq \tilde{r}_j + \varepsilon/16 \leq 2\tilde{r}_j$, so the first coordinate is upper bounded in absolute value by 2. Furthermore, $z_j \leq \tilde{r}_j + r_j \leq 3\tilde{r}_j$ by the triangle inequality, so $|-\frac{z_j+\varepsilon/32}{\tilde{r}_j}| \leq 4$. Hence, the sum of absolute values of the entries of the gradient is upper bounded by 8 over the whole box, therefore the additive error on a_j is at most 8 times the additive error on r_j and z_j. Since $|r_j - \tilde{r}_j| \leq \varepsilon/16$ and $|z_j - \tilde{z}_j| \leq \varepsilon/16$, we have $|a_j - (\tilde{r}_j - \frac{\tilde{z}_j^2}{2\tilde{r}_j})| \leq 8\varepsilon/16 = \varepsilon/2$.

With a similar argument, but using w_j instead of z_j, we also find b_j up to error $\varepsilon/2$, and hence α_j up to error ε. ∎

As a controlled state-preparation unitary can be used to prepare the conditional samples we immediately get the following corollary.

COROLLARY 4.1. *Let $0 < \varepsilon, \delta < 1$, and let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state with $\alpha_j \in \mathbb{C}$. Then $\mathcal{O}\big(\log(d/\delta)/\varepsilon^2\big)$ applications (in parallel) of a controlled state-preparation unitary for $|\psi\rangle$ suffice to compute an ε-ℓ_∞-norm estimate $\tilde{\alpha}$ of α, with success probability at least $1-\delta$.*

The sample complexity of Corollary 4.1 is asymptotically the same as in the algorithm of Kerenidis and Prakash [33], but our analysis is simpler thanks to Lemma 4.1, we estimate both the real and the imaginary part, and we directly get $\mathcal{O}\big(\log \frac{1}{\delta}\big)$ dependence on the probability of failure (as opposed to probability of success $1 - 1/d^c$ for some constant c in [33]).

Using the above result, we can also construct algorithms that approximate α in other ℓ_q-norms.

THEOREM 4.2. *Let $0 < \varepsilon, \delta < 1$, and let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state with $\alpha_j \in \mathbb{C}$. Then,*

$$O\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{2}{q}}}{\varepsilon^2}\right\} \cdot \log\frac{d}{\delta}\right)$$

copies of the state $(|0\rangle|\psi\rangle + |1\rangle|0\rangle)/\sqrt{2}$ suffice to compute an ε-ℓ_q-norm estimate $\tilde{\alpha}$ of α, with success probability at least $1-\delta$.

Proof. The proof follows from Corollary 3.1 combined with Proposition 4.2, in exactly the same way as in the proof of Theorem 4.1. ∎

4.3 Amplitudes up to a global phase with only copies of the state Finally we consider the model in which we simply have access to copies of the pure state. Although this model is conceptually simple, the estimation algorithm is more complicated than before. The number of samples required is still $\widetilde{\mathcal{O}}\big(1/\varepsilon^2\big)$, but slightly worse in polylogarithmic factors. In fact the method is very similar to the proof of Proposition 4.2, but instead of comparing α_j to $|\alpha_j|$, we have to compare the amplitudes to each other.

PROPOSITION 4.3. *Let $0 < \varepsilon, \delta < 1$, and let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state with $\alpha_j \in \mathbb{C}$. Then $\mathcal{O}\big(\log(d)\log(d/\delta)/\varepsilon^2\big)$ copies of $|\psi\rangle$, with the ability to perform unitary operations on each copy before measurement, suffice to compute an ε-ℓ_∞-norm estimate $\tilde{\alpha}$ of α, up to global phase, with success probability at least $1-\delta$.*

Proof. As before, let $r_j = |\alpha_j|$ and let $\alpha_j = a_j + b_j\mathbf{i}$. We first use Lemma 4.1 to compute an $\varepsilon/(16m)$-estimate $\tilde{r}_j$ of r_j. We only consider the coordinates where $\tilde{r}_j \geq \varepsilon/2$, and permute the basis states in all remaining copies of $|\psi\rangle$ in such a way that these form the first k coordinates of the state. Let m be the smallest value such that $k < 2^m$, i.e., we are only interested in the amplitudes for basis states where all but the last m qubits are 0. For ease of notation we consider the sub-normalized state $|\phi\rangle$ corresponding to this part and relabel the indices so the $\tilde{r}_j$ are in decreasing order. Note that the remaining part of $|\psi\rangle$ has ℓ_∞-norm less than $\varepsilon/2$, hence it can be ignored for our estimation.

For $h \in [m]$ we consider the state resulting from applying a Hadamard gate to the h-th qubit of $|\phi\rangle$ (below, for an m-digit binary string $j \in [2^m]$, we write j_h to denote the h-th binary digit):

$$I^{\otimes h-1} \otimes H \otimes I^{\otimes m-h}|\phi\rangle = \frac{1}{\sqrt{2}} \sum_{\substack{j\in[2^m]\\ j_h=0}} (\alpha_j + \alpha_{j+2^h})|j\rangle + (\alpha_j - \alpha_{j+2^h})|j+2^h\rangle.$$

Hence we can learn $\varepsilon/(16m)$-estimates of $s_{j,h} = |\alpha_j - \alpha_{j+2^h}|$ using computational-basis measurements with success probability at least $1-\delta/(2m)$. Note that this can be interpreted as an application of Lemma 4.1, where the h-th qubit is considered the flag. Repeating this with an additional phase gate also gives estimates $t_{j,h} = |\alpha_j - \mathbf{i}\alpha_{j+2^h}|$.

We now consider a single value of h and aim to learn α_{j+2^h} relative to α_j. For now, we assume that $\alpha_j \in \mathbb{R}_{\geq 0}$ and show how to give an estimate of $\alpha_j + 2^h$; we discuss how to relax the assumption subsequently. Note that:

$$s_{j,h}^2 = |\alpha_j|^2 + |\alpha_{j+2^h}|^2 - 2\Re(\alpha_j^\dagger \alpha_{j+2^h}),$$

and therefore, due to our assumption on α_j, we have:

$$a_{j+2^h} = \frac{r_j^2 + r_{j+2^h}^2 - s_{j,h}^2}{2r_j},$$

with a similar argument as in Proposition 4.2. As in Proposition 4.2, we consider this estimate of a_{j+2^h} as a function of r_j, r_{j+2^h}, and $s_{j,h}$, and compute its gradient, which now consists of three partial derivatives; we then bound the ℓ_1-norm of the gradient over the possible values for $s_{j,h}, r_j, r_{j+2^h}$:

$$\begin{aligned}
\left|\frac{\partial a_{j+2^h}}{\partial r_{j+2^h}}\right| &= \left|\frac{r_{j+2^h}}{r_j}\right| \leq 1 + \frac{3}{16m} \leq \frac{3}{2}, \\
\left|\frac{\partial a_{j+2^h}}{\partial s_{j,h}}\right| &= \left|-\frac{s_{j,h}}{r_j}\right| \leq 3, \\
\left|\frac{\partial a_{j+2^h}}{\partial r_j}\right| &= \left|\frac{1}{2} + \frac{s_{j,h}^2 - r_{j+2^h}^2}{4r_j^2}\right| \leq \frac{1}{2} + \frac{1}{4}\left|\frac{s_{j,h}}{r_j}\right|^2 + \frac{1}{4}\left|\frac{r_{j+2^h}}{r_j}\right|^2 \leq \frac{7}{2}.
\end{aligned}$$

For the first inequality above, we used the fact that $r_j, r_{j+2^h} \geq \varepsilon/3$ and $r_j + \varepsilon/(16m) \geq r_{j+2^h}$. The middle inequality follows from $s_{j,h} \leq r_j + r_{j+2^h} \leq 2r_j + \varepsilon/(16m) \leq 3r_j$. Hence the gradient's ℓ_1-norm is upper-bounded by 8, implying that our estimate for a_{j+2^h} is $\varepsilon/(2m)$-close. Repeating this argument using $t_{j,h}$ shows how to obtain an $\varepsilon/(2m)$-estimate of b_{j+2^h}, resulting in an ε/m-estimate of α_{j+2^h}. Recall that so far we assumed that $\alpha_j \in \mathbb{R}_{\geq 0}$; this implies that we learned α_{j+2^h} only up to the phase of α_j[4], and we have to reconcile the different relative phases for pairs of amplitudes computed with this procedure.

Thus, we now combine our estimates for the different values of h, to learn the entire state up to a global phase. We arbitrarily assume that one amplitude is real, say $\alpha_0 \in \mathbb{R}_{\geq 0}$. Consider some index j with Hamming weight w, i.e., there are w positions in the binary representation of j that are 1. We start at the all-zero string, and, proceeding from the most significant bit, flip bits to obtain j. This yields a path of length w over indices j', with $w \leq m$. For all j' we have $\tilde{r}_{j'} \geq \varepsilon/2$, hence we know all amplitudes $\alpha_{j'}$ up to the phase of the previous amplitude in the path, with precision ε/m each. By the triangle inequality and the union bound we can therefore estimate α_j up to the phase of α_0, with precision ε for all $j < k$. Since $k \leq d$ and hence $m \leq \log(d)$, we get the stated complexity. ∎

We can also use the above result to derive algorithms that estimate the vector α up to different norms, just like in the previous subsections. This results in the following theorem.

THEOREM 4.3. *Let $0 < \delta < 1$, $\varepsilon > 0$, $d \in \mathbb{N}$, $q \in [2, \infty]$, and let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state with $\alpha_j \in \mathbb{C}$. Then,*

$$O\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{2}{q}}}{\varepsilon^2}\right\} \cdot \log d \log \frac{d}{\delta}\right)$$

copies of $|\psi\rangle$, with the ability to perform unitary operations on each copy, suffice to find an ε-ℓ_q-norm estimate $\tilde{\alpha}$ of α, up to global phase, with success probability at least $1 - \delta$.

Proof. The result follows immediately from combining Corollary 3.1 and Proposition 4.3, in exactly the same way as in the proof of Theorem 4.1. ∎

[4]It is possible to fully restate our argument without this assumption in the first place, but the resulting derivation is considerably longer and less elegant.

5 Pure-state tomography using phase estimation

In this section we turn to the strongest input model of this paper, where we have access to a state-preparation unitary and its inverse. This allows us to reduce the dependence on the error parameter ε from $1/\varepsilon^2$ to $1/\varepsilon$. We rely on the framework introduced in Section 2.3.

5.1 State preparation for amplitude encoding

Let $x \in [-1,1]^d$, so $\|x\|_2 \leq \sqrt{d}$. Define

$$|\mathrm{amp}(x)\rangle := \frac{1}{\sqrt{d}} \sum_{j\in[d]} x_j |j\rangle|0\rangle + \frac{1}{\sqrt{d}} \sum_{j\in[d]} \sqrt{1-x_j^2}|j\rangle|1\rangle.$$

In this section we give a simple subroutine that as input takes a binary description of x and constructs the state $|\mathrm{amp}(x)\rangle$.

LEMMA 5.1. *Let $|x\rangle$ be a binary encoding of an $x \in [-1,1]^d$ where each x_j can be written exactly with b bits of precision. There is a quantum algorithm U_{amp} that acts as*

$$U_{amp}|x\rangle|0\rangle = |x\rangle|amp(\tilde{x})\rangle$$

where $\|\tilde{x} - x\|_\infty \leq \varepsilon$. U_{amp} uses $\mathcal{O}\big(\log(d) + \log(1/\varepsilon)\log^2\log(1/\varepsilon)\big)$ gates, and $2\min\{b, \log(2/\varepsilon)\}$ indexed-SWAP gates acting on d bits.

Proof. The algorithm is as follows, starting from $|x\rangle|0\rangle|0\rangle|0\rangle|0\rangle = |x_1\rangle\ldots|x_d\rangle|0\rangle|0\rangle|0\rangle|0\rangle$

- Use $\mathcal{O}(\log(d))$ gates to setup a uniform superposition over $[d]$:

$$|x\rangle \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle|0\rangle|0\rangle|0\rangle$$

- Swap in the first $\min\{b, \log(1/\varepsilon)\}$ bits of x_i, conditioned on the 3th to last register, using $\min\{b, \log(2/\varepsilon)\}$ indexed-SWAP gates on d bits each:

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} |\underline{x}^{(i)}\rangle|i\rangle|\bar{x}_i\rangle|0\rangle|0\rangle$$

  where $\bar{x}_i$ is the cut-off version of x_i, so $|x_i - \bar{x}_i| \leq \varepsilon/2$ and $\underline{x}^{(i)}$ is the remaining part of x.

- Approximate $\tilde{a}_i = \arcsin(\bar{x}_i)$ up to $\varepsilon/2$ precision using $\mathcal{O}\big(\log(1/\varepsilon)\log^2\log(1/\varepsilon)\big)$ gates (see [47] for the complexity):

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} |\underline{x}^{(i)}\rangle|i\rangle|\bar{x}_i\rangle|\tilde{a}_i\rangle|0\rangle$$

- Use $\mathcal{O}(\log(1/\varepsilon))$ rotations with exponentially decreasing angle, controlled on the bits of $\tilde{a}_i$ to rotate the last qubit:

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} |\underline{x}^{(i)}\rangle|i\rangle|\bar{x}_i\rangle|\tilde{a}_i\rangle\left(\tilde{a}_i|0\rangle + \sqrt{1-\tilde{a}_i^2}|1\rangle\right)$$

- Uncompute $\tilde{a}_i$ and swap back $\bar{x}_j$:

$$|x\rangle \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle\left(\sin(\tilde{a}_i)|0\rangle + \sqrt{1-\sin(\tilde{a}_i)^2}|1\rangle\right)$$

As $|\tilde{a}_i - \arcsin(\bar{x}_j)| \leq \varepsilon/2$, and the sin function is 1-Lipschitz, we get that $|\bar{x}_j - \sin(\tilde{a}_j)| \leq \varepsilon/2$. Combining this with $|x_j - \bar{x}_j| \leq \varepsilon/2$ gives the desired precision. ∎

5.2 Pure-state tomography With the quantum circuit of Lemma 5.1 for preparing $|\mathrm{amp}(\tilde{x})\rangle$ we have all the ingredients for our pure-state tomography algorithm. We only state it for the estimation of the real part of $|\psi\rangle$, but one can also extract the imaginary part with the same running time simply by applying the algorithm to the quantum state $\mathbf{i}|\psi\rangle$.

PROPOSITION 5.1. *Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state, and $U|0\rangle = |\psi\rangle$. There is a quantum algorithm that, with probability at least $1-\delta$, outputs $\tilde{\alpha} \in \mathbb{R}^d$ such that $\|\Re(\alpha) - \tilde{\alpha}\|_\infty \leq \varepsilon$, using*

$$\mathcal{O}\left(\frac{\sqrt{d}}{\varepsilon}\log(\frac{d}{\delta})\right)$$

applications of (controlled) U and $U^\dagger$, $\mathcal{O}\left(\frac{\sqrt{d}}{\varepsilon}\log(\frac{d}{\delta})\log(\frac{d}{\varepsilon})\right)$ indexed-SWAP gates acting on d bits, and $\widetilde{\mathcal{O}}\left(d + \frac{\sqrt{d}}{\varepsilon}\right)$ additional gates. If $\varepsilon \geq \frac{1}{\sqrt{d}}$, the number of applications of U can be reduced to $\widetilde{\mathcal{O}}\left(\frac{1}{\varepsilon^2}\right)$ (while potentially increasing the gate complexity to $\widetilde{\mathcal{O}}\left(\frac{1}{\varepsilon^4}\right)$).

Proof. We first describe the algorithm that calls U $\widetilde{\mathcal{O}}\left(\frac{\sqrt{d}}{\varepsilon}\right)$ times. Let $f(x) = \langle \Re(\alpha), x\rangle$. Taking $U' := (I \otimes U)$ and $V := U_{amp}$, Lemma 2.5 gives us an approximate block-encoding W of the diagonal matrix $\langle x, \alpha\rangle/\sqrt{d}$, and from this we get an approximate block-encoding of $f(x)/\sqrt{d}$ averaging W and $W^\dagger$ via Lemma 2.2 so we can apply Corollary 2.1. Then we get an ε-ℓ_∞-norm estimate of $\Re(\alpha)/\sqrt{d}$ (where the denominator $\sqrt{d}$ comes from the normalization in $|\mathrm{amp}(x)\rangle$) with $\mathcal{O}\left(\frac{1}{\varepsilon}\log(\frac{d}{\delta})\right)$ uses of the block-encoding of $(W + W^\dagger)/2$, the construction of which requires a constant number of calls to U and U_{amp}. To obtain the desired estimate of $\Re(\alpha)$ we elevate the precision to $\varepsilon/\sqrt{d}$, which brings the total number of uses of U and U_{amp} to $\mathcal{O}\left(\frac{\sqrt{d}}{\varepsilon}\log(\frac{d}{\delta})\right)$. The gate complexity is $\mathcal{O}\left(\frac{\sqrt{d}}{\varepsilon}\log(\frac{d}{\delta})\log(\frac{d}{\varepsilon})\right)$ indexed-SWAP gates acting on d bits, and

$$\mathcal{O}\left(\left(d\log(\frac{1}{\varepsilon})\log\log(\frac{1}{\varepsilon}) + \frac{\sqrt{d}}{\varepsilon}\log\frac{d}{\varepsilon}\log d\right)\log\frac{d}{\delta}\right)$$

additional gates, where the first term in the summation comes from the additional gates of Corollary 2.1, whereas the second term comes from the cost of U_{amp}.

Next, we show how to improve the algorithm when $\varepsilon \geq \frac{1}{\sqrt{d}}$. Any j such that $|\alpha_j| \leq \varepsilon$ can be ignored because of the ℓ_∞-norm objective, so we can simply set $\tilde{\alpha}_j = 0$. Since we are only interested in j such that $|\alpha_j| > \varepsilon$, we note that there are at most $1/\varepsilon^2$ such j because $\sum_j |\alpha_j|^2 = 1$. After taking $\mathcal{O}\left(\frac{\log(n/\delta)}{\varepsilon^2}\right)$ measurements of $|\psi\rangle$ in the computational basis, the probability that all such j are observed is at least $1-\delta$. We can then apply the algorithm described above to obtain α_j only for those j. As this set has cardinality $\mathcal{O}(1/\varepsilon^2)$, the quantum algorithm requires $\widetilde{\mathcal{O}}\left(\frac{1}{\varepsilon^2}\right)$ applications of U. This concludes the proof. ∎

Note that the algorithm in Proposition 5.1 with complexity $\widetilde{\mathcal{O}}\left(\frac{\sqrt{d}}{\varepsilon}\right)$ can be made essentially unbiased by using Corollary 6.1 instead of Corollary 2.1.

THEOREM 5.1. *Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state, $\alpha \in \mathbb{C}^d$ the vector with elements α_j, and $U|0\rangle = |\psi\rangle$. Then, for $q \geq 2$*

$$O\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{1}{2}+\frac{1}{q}}}{\varepsilon}\right\}\log\frac{d}{\delta}\right)$$

(controlled) applications of U and its inverse suffice to compute an ε-ℓ_q-norm estimate $\tilde{\alpha}$ of α, with success probability at least $1-\delta$.

Proof. The first term follows from Proposition 5.1 and Corollary 3.1. The second term follows from Theorem 4.2. ∎

5.3 Tomography for sparse vectors To conclude this section, we show that the tomography algorithm based on phase estimation can be improved if we know that the quantum state contains at most $s < d$ large amplitudes. We proceed by finding the large elements first, then applying the tomography algorithm only to extract a description of only the corresponding part of the quantum state.

PROPOSITION 5.2. *Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state, and $U|0\rangle = |\psi\rangle$. Let $0 < \delta < 1$, and let R be such that $\{j \in [d] : |\alpha_j| \geq \varepsilon\} \subseteq R \subseteq [d]$ be a subset of the indices that contains all large elements. Let $P := \sum_{i\notin R} |\alpha_i|^2$ and $s = |R|$. There is a quantum algorithm that, with probability at least $1-\delta$, outputs an $\mathcal{O}\big((s + P/\varepsilon^2)\log(s)\log(1/\delta)\big)$-sparse $\tilde{\alpha} \in \mathbb{R}^d$ such that $\|\alpha - \tilde{\alpha}\|_\infty \leq \varepsilon$ using*

$$\mathcal{O}\left(\left(\left(\frac{\sqrt{s}}{\varepsilon} + \frac{\sqrt{P}}{\varepsilon^2}\right) + \log\frac{\log(s+P/\varepsilon)}{\delta}\right)\log(s)\log\left(\frac{s+P/\varepsilon}{\delta}\right)\right)$$

applications of (controlled) U and its inverse, $\mathcal{O}\Big(\Big(\Big(\frac{\sqrt{s}}{\varepsilon} + \frac{\sqrt{P}}{\varepsilon^2}\Big) + \log\frac{\log(\frac{s+P}{\varepsilon})}{\delta}\Big)\log(s)\log\Big(\frac{s+P/\varepsilon}{\delta}\Big)\log\big(\frac{s+P}{\varepsilon}\big)\Big)$ indexed-SWAP gates acting on $\mathcal{O}\big((s + \frac{P}{\varepsilon^2})\log(s)\ln(1/\delta)\big)$ bits, and $\widetilde{\mathcal{O}}\Big(s + \frac{\sqrt{s}}{\varepsilon} + \frac{P}{\varepsilon^2}\Big)$ additional gates.

Proof. We start by finding all (at most s) elements in R that are at least ε in size using amplitude amplification. We can then ignore all other elements and apply our state tomography algorithm on the relevant elements.

Let $k = |\{j \in [d] : |\alpha_j| \geq \varepsilon\}|$ be the number of large elements. So there are $s-k$ elements in R that are smaller than ε. We start by simply measuring the state and observing an index j. Note that with probability at least $k\varepsilon^2$ this is one of the relevant entries, although this is unknown to us. After observing a single entry, we mark all other entries as "good" and amplify the "good" part of the state before measuring again, to avoid seeing an element twice. We repeat this until we have seen T different elements, for some T to be determined later.

If at some point we have seen j large elements (which we do not know), then the probability on the "good" elements is at most $(s-j)\varepsilon^2 + P$. Hence, this part can be amplified to find a new element with probability $\geq 2/3$ using $\mathcal{O}\left(\frac{1}{\sqrt{(s-j)\varepsilon^2+P}}\right)$ queries. The probability that this new element is one of the large ones is at least $\frac{2(k-j)\varepsilon^2}{3(s-j)\varepsilon^2+P}$. Thus, the expected number of samples before seeing a new large element is at most $\frac{3(s-j)\varepsilon^2+P}{2(k-j)\varepsilon^2}$. For the expected number of queries needed before seeing all large elements we then get

$$\begin{aligned}\mathcal{O}\left(\sum_{j=0}^{k-1}\frac{1}{\sqrt{(s-j)\varepsilon^2+P}}\frac{(s-j)\varepsilon^2+P}{(k-j)\varepsilon^2}\right) &= \mathcal{O}\left(\sum_{j=1}^{k}\left(\frac{\sqrt{(s-k+j)\varepsilon^2}}{j\varepsilon^2} + \frac{\sqrt{P}}{j\varepsilon^2}\right)\right)\\ &= \mathcal{O}\left(\left(\frac{\sqrt{s}}{\varepsilon} + \frac{\sqrt{P}}{\varepsilon^2}\right)\log(s)\right).\end{aligned}$$

By Markov's equality we can stop the algorithm after 6 times the expected number of queries and still be successful with probability $\geq 5/6$.

We also need to ensure that we do not return too many elements. The expected number of elements found is equal to the expected number of samples, hence it is

$$\sum_{j=0}^{k-1}\frac{(s-j)\varepsilon^2+P}{(k-j)\varepsilon^2} \leq \left(s + \frac{P}{\varepsilon^2}\right)\log(s).$$

Again, by Markov's inequality we can stop the algorithm if we see more then 6 times this number of samples, and still succeed with probability at least 5/6. By the union bound both conditions are met with probability at least 2/3. Repeating $\mathcal{O}(\ln(1/\delta))$ times and taking all elements seen in runs with not too many samples gives us as success probability at least $1-\delta/2$. The query complexity then becomes

$$\mathcal{O}\left(\left(\frac{\sqrt{s}}{\varepsilon} + \frac{\sqrt{P}}{\varepsilon^2}\right)\log(s)\log(1/\delta)\right)$$

for this entire procedure.

We now assume the last step was successful, so we have a set I of indices such that I contains all large elements and $|I| = \mathcal{O}\big(\big(s + \frac{P}{\varepsilon^2}\big)\log(s)\ln(1/\delta)\big)$. Applying Theorem 5.1 on just these indices gives the query and gate complexity from the lemma.

∎

The above proposition gives an improvement if the state is close to a sparse state in ℓ_2-norm. We can directly get a bound on this closeness if almost all elements are small.

COROLLARY 5.1. *Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state, and $U|0\rangle = |\psi\rangle$. Let $0 < \delta < 1$, and let s be such that $|\{j \in [d] : |\alpha_j| \geq \varepsilon\sqrt{\frac{s}{d}}\}| \leq s$. There is a quantum algorithm that, with probability at least $1-\delta$, outputs an $\mathcal{O}(s\log(s)\log(1/\delta))$-sparse $\tilde{\alpha} \in \mathbb{R}^d$ such that $\|\alpha - \tilde{\alpha}\|_\infty \leq \varepsilon$ using*

$$\mathcal{O}\left(\left(\frac{\sqrt{s}}{\varepsilon} + \log\frac{\log s}{\delta}\right)\log(s)\log\left(\frac{s}{\delta}\right)\right)$$

applications of U and its inverse, and ... additional gates.

Proof. Let $R = \{j \in [d] : |\alpha_j| \geq \varepsilon\sqrt{\frac{s}{d}}\}$. Then, $P = \sum_{i\notin R} |\alpha_i|^2 \leq d\varepsilon^2 \frac{s}{d} \leq \varepsilon^2 s$. Applying Proposition 5.2 with this choice of R and the above bound on P gives the desired result. ∎

As usual, we now convert the above bound for the ℓ_∞-norm to other norms.

THEOREM 5.2. *Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state, $\alpha \in \mathbb{C}^d$ the vector with elements α_j, and $U|0\rangle = |\psi\rangle$. Let $0 < \delta < 1$, let $q \geq 2$ and let s be such that $|\{j \in [d] : |\alpha_j| \geq \varepsilon\sqrt{\frac{s}{d}}\}| \leq s$. Then we can compute an ε-ℓ_q-norm estimate $\tilde{\alpha}$ of α using*

$$O\left(\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{s^{\frac{1}{2}+\frac{1}{q}}}{\varepsilon}\right\} + \log\frac{\log s}{\delta}\right)\log s \log\frac{s}{\delta}\right)$$

conditional applications of U and its inverse, with success probability at least $1-\delta$.

Proof. Follows from Proposition 5.2 and Corollary 5.1. ∎

6 First intermezzo: unbiased phase estimation

We now describe a method for phase estimation that is unbiased, more precisely symmetric in the sense that for a phase ϕ it provides an estimate φ such that the probability of getting estimate $\phi + \epsilon$ is the same as getting estimate $\phi - \epsilon$ (modulo 2π) for all ε. Note that this is not satisfied by ordinary phase estimation, but this property is highly desirable, as we showcase in our applications. In particular, we need unbiased phase estimation to recover unbiased estimates of the entries of a density matrix, allowing us to give tighter error bounds with high probability.

Our method is based on adding and later subtracting a random phase shift; this idea is not new, see, e.g., [36, Section 3.2]. The first step in our analysis is to show that the resulting estimator is symmetric. We subsequently show how to boost the precision of this symmetric estimator in a symmetric way. Since the problem is invariant under shifting by 2π we can always interpret phases ϕ, φ modulo 2π; in particular for phases we define the distance modulo 2π introducing the notation $|\phi - \varphi|_{2\pi} := \min\{|\phi - \varphi - 2\pi\ell| : \ell \in \mathbb{Z}\}$. In this section, "digit" always refers to "binary digit", i.e., all numbers are expressed in fixed-point binary encoding; for example, $0.b_1b_2\ldots b_n$ where $b_i \in \{0,1\}$ for $i \in [n]$ is the n-digit encoding of b. Recall that the function $\mathrm{sinc}(x)$ is a complex entire function defined as $\sin(x)/x$ for $x \neq 0$ and $\mathrm{sinc}(0) = 1$.

THEOREM 6.1. (UNBIASED PHASE ESTIMATION) *If we run Algorithm 1 with $n = \infty$ in Line 1, then it returns a random phase $\varphi \in [0, 2\pi)$ with probability density function*

$$f(\varphi) := \frac{M}{2\pi}\frac{\mathrm{sinc}^2(\frac{M}{2}|\phi-\varphi|_{2\pi})}{\mathrm{sinc}^2(\frac{1}{2}|\phi-\varphi|_{2\pi})}. \tag{6.2}$$

Algorithm 1 Suppressed-Bias Phase Estimation

Input: $|\psi(\phi)\rangle = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} e^{i\phi k}|k\rangle$ (for unknown ϕ), and a parameter $n \in \mathbb{N}$

1: Sample a uniformly random n-digit binary number $u \in [0, 1)$ and define $\xi := \frac{2\pi u}{M}$

2: Apply multi-phase gate $\sum_{k=0}^{M-1} e^{-i\xi k}|k\rangle\langle k|$ to $|\psi(\phi)\rangle$

3: Perform inverse Fourier transform over $\mathbb{Z}_M$ and measure the state, yielding outcome j

4: **Return** $\varphi := \frac{2\pi j}{M} + \xi = \frac{2\pi}{M}(j + u)$

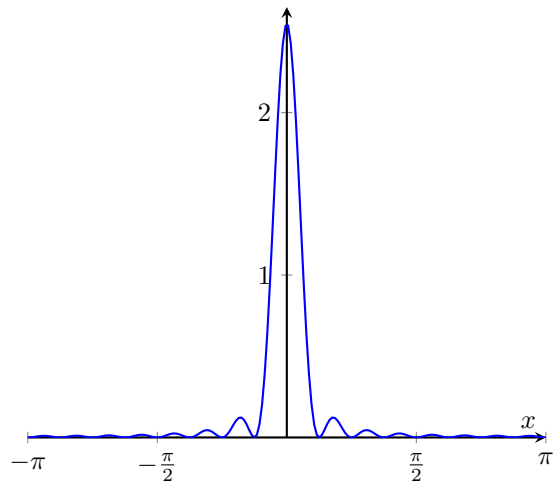

Figure 2: Plot of Eq. (6.2) for $x = \phi - \varphi$ and $M = 16$.

This probability density function is normalized so that $\int_0^{2\pi} f(\varphi)d\varphi = 1$, moreover it only depends on $|\phi - \varphi|_{2\pi}$ showing that this procedure satisfies our criterion for unbiasedness, see Figure 2.

Proof. Suppose we have a quantum state $\frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} e^{i\phi k}|k\rangle$ and we wish to estimate the phase ϕ. Then by applying the inverse quantum Fourier transform over $\mathbb{Z}_M$ and measuring we get outcome j, giving rise to estimate $\varphi = \omega_j = \frac{2\pi j}{M}$ with probability (using [9, Lemma 10]):

$$\left|\frac{1}{M}\sum_{k=0}^{M-1} e^{-i\omega_j k} e^{i\phi k}\right|^2 = \left|\frac{1}{M}\sum_{k=0}^{M-1} e^{i(\phi-\omega_j)k}\right|^2 = \left|\frac{1}{M}\sum_{k=0}^{M-1} e^{i|\phi-\omega_j|k}\right|^2 = \frac{\operatorname{sinc}^2(\frac{M}{2}|\phi-\omega_j|_{2\pi})}{\operatorname{sinc}^2(\frac{1}{2}|\phi-\omega_j|_{2\pi})}. \tag{6.3}$$

Now let us modify this procedure by first choosing a uniformly random phase $\xi \in [0, \frac{2\pi}{M})$ and applying phase estimation to the state $\frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} e^{i(\phi k - \xi k)}|k\rangle$ then outputting $\varphi = \omega_j + \xi$ for the resulting j. Then, for a fixed ξ the probability of outputting $\varphi = \omega_j + \xi$ is

$$\frac{\operatorname{sinc}^2(\frac{M}{2}|\phi-\xi-\omega_j|_{2\pi})}{\operatorname{sinc}^2(\frac{1}{2}|\phi-\xi-\omega_j|_{2\pi})} = \frac{\operatorname{sinc}^2(\frac{M}{2}|\phi-\varphi|_{2\pi})}{\operatorname{sinc}^2(\frac{1}{2}|\phi-\varphi|_{2\pi})}. \tag{6.4}$$

Since the choice of ξ is uniformly random over $[0, \frac{2\pi}{M})$ this implies that the probability density function of getting estimate $\varphi \in [0, 2\pi)$ is given by Eq. (6.2). ∎

According to Eq. (6.2) the probability of getting an outcome φ with error at most $\frac{c}{M}$ for some $c \leq \pi M$ is

substitute $y = \frac{M}{2}x \Rightarrow$
$$\Pr[|\phi - \varphi|_{2\pi} \leq \frac{c}{M}] = \int_{-\frac{c}{M}}^{\frac{c}{M}} \frac{M}{2\pi} \frac{\operatorname{sinc}^2(Mx/2)}{\operatorname{sinc}^2(x/2)} dx$$

use $|\operatorname{sinc}(z)| \leq 1 \Rightarrow$
$$= \frac{1}{\pi} \int_{-\frac{c}{2}}^{\frac{c}{2}} \frac{\operatorname{sinc}^2(y)}{\operatorname{sinc}^2(y/M)} dy$$

$$\geq \frac{1}{\pi} \int_{-\frac{c}{2}}^{\frac{c}{2}} \operatorname{sinc}^2(y) dy. \tag{6.5}$$

In particular one can compute the value of this bound for $c = 1, 2, 3$ resulting in:

$$\Pr[|\phi - \varphi|_{2\pi} \leq \frac{1}{M}] \geq 0.30\ldots \qquad \Pr[|\phi - \varphi|_{2\pi} \leq \frac{2}{M}] \geq 0.57\ldots \qquad \Pr[|\phi - \varphi|_{2\pi} \leq \frac{3}{M}] \geq 0.75\ldots \tag{6.6}$$

Note the increased accuracy compared to ordinary phase estimation: the difference between two distinct phase estimates is at least $\frac{2\pi}{M}$, so in case the true phase we try to estimate is, say, $\frac{\pi}{M}$, then ordinary phase estimation always has an error at least $\frac{\pi}{M} > \frac{3}{M}$. On the other hand, here we get $\frac{3}{M}$-accuracy with probability greater than $\frac{3}{4}$.

6.1 Unbiased boosting Now we show how to boost this procedure so that it gives an unbiased estimate that is also $\frac{6}{M}$-accurate with exponentially high probability. We achieve this essentially by the usual median trick, except some care is needed because the median is ill-defined modulo 2π.

Algorithm 2 Boosted Unbiased Phase Estimation

Input: $2m$ copies of $|\psi(\phi)\rangle = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} e^{i\phi k}|k\rangle$ (for unknown ϕ)

1: **For** $j = 1$ to $2m$

2: Run Algorithm 1 setting $n = \infty$ on the j-th copy of $|\psi(\phi)\rangle$ and record the estimate φ_j

3: Find the shortest interval $I = [a, b] \subseteq [-2\pi, 2\pi]$ such that $I \cup (I + 2\pi)$ contains at least $m+1$ of the estimates φ_j

4: **If** $a + b \geq 0$ **then return** $\overline{\varphi} := \frac{a+b}{2}$ **else return** $\overline{\varphi} := \frac{a+b}{2} + 2\pi$

THEOREM 6.2. *Algorithm 2 returns an unbiased* $\overline{\varphi} \in [0, 2\pi]$ *s.t.* $\Pr[|\phi - \overline{\varphi}|_{2\pi} \leq \frac{6}{M}] \geq 1 - \exp(-\frac{m}{4})$.

Proof. In Line 3 almost surely there is a unique shortest interval (modulo 2π), since the endpoints a, b of the shortest interval must come from the $2m$ estimates (modulo 2π), which themselves come from the continuous distribution of Eq. (6.2). Alternatively, if there are multiple shortest intervals (modulo 2π) we can just choose one uniformly at random. This algorithm is naturally unbiased as the distribution of the shortest intervals (modulo 2π) is symmetric with respect to ϕ.

Moreover, the probability that there are at least $m + 1$ estimates φ_j such that $|\phi - \varphi_j|_{2\pi} \leq \frac{3}{M}$ is at least $1 - \exp(-\frac{m}{4})$ due to the Chernoff bound. Indeed, the probability that $|\phi - \varphi|_{2\pi} > \frac{3}{M}$ is at most $\frac{1}{4}$ by Eq. (6.6). So by the Chernoff-Hoeffding Theorem 2.1 the probability that $|\phi - \varphi|_{2\pi} > \frac{3}{M}$ holds for at least m out of $2m$ estimates is at most $\exp(-D(\frac{1}{2}\|\frac{1}{4})2m) \leq \exp(-\frac{m}{4})$. This implies that the shortest interval has length at most $\frac{6}{M}$ and it also must overlap with the interval $[\phi - \frac{3}{M}, \phi + \frac{3}{M}]$ (modulo 2π), so in particular $|\phi - \overline{\varphi}|_{2\pi} \leq \frac{6}{M}$. ∎

6.2 Unbiased estimators of $e^{i\phi}$ Our unbiased phase estimators can be used for constructing unbiased estimators of the complex number $e^{i\phi}$. The unbiased nature of our phase estimates φ means that $\mathbb{E}[e^{i\varphi}] = \lambda e^{i\phi}$, for some $\lambda \in [-1, 1]$. Moreover, due to the shift invariance of $f(\varphi)$, i.e., the fact that $f(\varphi)$ depends only on $|\phi - \varphi|_{2\pi}$, we have that λ only depends on M (and m in the boosted case). Therefore $e^{i\varphi}/\lambda$ is an unbiased

estimator of $e^{i\phi}$. One can also compute the value

$$\begin{aligned}\lambda(M) = \int_{-\pi}^{\pi} \cos(x) \cdot \frac{M}{2\pi} \frac{\operatorname{sinc}^2(Mx/2)}{\operatorname{sinc}^2(x/2)} dx &= \int_{-\pi}^{\pi} (1 - 2\sin^2(x/2)) \cdot \frac{1}{2M\pi} \frac{\sin^2(Mx/2)}{\sin^2(x/2)} dx \\ &= 1 - \frac{1}{M\pi} \int_{-\pi}^{\pi} \sin^2(Mx/2) dx \\ &= 1 - \frac{1}{M}.\end{aligned}$$

To compute the variance it is useful to note that for $M \geq 2$

$$\begin{aligned}\int_{-\pi}^{\pi} \sin^2(x) \cdot \frac{M}{2\pi} \frac{\operatorname{sinc}^2(Mx/2)}{\operatorname{sinc}^2(x/2)} dx &= \int_{-\pi}^{\pi} 4\sin^2(x/2)\cos^2(x/2) \cdot \frac{1}{2M\pi} \frac{\sin^2(Mx/2)}{\sin^2(x/2)} dx \\ &= \frac{2}{M\pi} \int_{-\pi}^{\pi} \cos^2(x/2)\sin^2(Mx/2) dx \\ &= \frac{1}{M}.\end{aligned}$$

Thus we have that $(1 + \frac{1}{M-1})e^{i\varphi}$ is an unbiased estimator of $e^{i\phi}$ with variance $\Theta(\frac{1}{M})$.

For the boosted version it is harder to compute the value of $\lambda(M, m)$, but due to the concentration proven in Theorem 6.2 we know that its value must be $\lambda(M, m) = 1 - \mathcal{O}\big(\frac{1}{M^2} + \exp(-\frac{m}{4})\big)$. So $e^{i\overline{\varphi}}/\lambda(M, m)$ is an unbiased estimator of $e^{i\phi}$ with variance $\mathcal{O}\big(\frac{1}{M^2} + \exp(-\frac{m}{4})\big)$.

6.3 Unbiased probability estimation Unbiased estimators for $\lambda e^{i\phi}$ give us the possibility of modifying the standard amplitude estimation algorithm [9], so that we estimate the squared amplitude without bias. To that end, suppose that we have access to a state-preparation unitary that prepares the state $\sqrt{1-p}|\psi_0\rangle|0\rangle + \sqrt{p}|\psi_1\rangle|1\rangle$, and our goal is to estimate p. Recall that the amplitude estimation algorithm runs phase estimation on the Grover iterate, which is a 2-dimensional rotation with eigenvalues $e^{\pm 2i\theta}$, where $\theta = \arcsin\sqrt{p}$. Consequently, it obtains an estimate for 2θ or -2θ, both with probability 1/2.

If we now substitute our unbiased phase estimation algorithm into this procedure, we obtain an unbiased estimate of either $\lambda e^{2i\theta}$ or $\lambda e^{-2i\theta}$, both with probability 1/2. In either case, taking the real part of our estimate now estimates $\lambda\cos(2\theta) = \lambda\cos(2\arcsin\sqrt{p}) = \lambda(1-2p)$ without bias. Thus, if we denote the outcome of the unbiased phase estimation algorithm by $Z = e^{2\pi i\varphi}$, then

$$\mathbb{E}\left[\frac{1}{2} - \frac{\operatorname{Re}[Z]}{2\lambda}\right] = \frac{1}{2} - \frac{\lambda(1-2p)}{2\lambda} = p.$$

We can crudely bound the variance of this estimator to be $\operatorname{Var}[\operatorname{Re}[Z]]/(2\lambda)^2 \leq \operatorname{Var}[Z]/(2\lambda)^2 = \mathcal{O}(1/M^2 + \exp(-m/4))$. However, if p is very close to 0 or 1, then the probability distribution of Z will be very tightly concentrated around 1 or -1 on the unit circle in the complex plane, where the unit circle runs perpendicular to the real axis. Thus, in this regime taking the real part of Z intuitively squashes samples much closer together, and as a result the variance of $\operatorname{Re}[Z]$ can be much smaller than that of Z.

Quantitatively, if $\theta \leq 3/(2M)$, then the endpoints of the interval of concentration for $\operatorname{Re}[Z]$, as derived in Theorem 6.2, are $\cos(2\theta + 3/M)$ and 1, which means that the length of the interval is $1 - \cos(2\theta + 3/M) = \mathcal{O}((\theta + 1/M)^2) = \mathcal{O}(1/M^2)$. Thus, the variance in this case is $\mathcal{O}(1/M^4 + \exp(-m/4))$. A similar analysis holds true in the case where $\theta \geq \pi/2 - 3/(2M)$.

On the other hand, if $3/(2M) < \theta < \pi/2 - 3/(2M)$, then the endpoints of the interval of concentration for $\operatorname{Re}[Z]$ are $\cos(2\theta + 3/M)$ and $\cos(2\theta - 3/M)$. This implies that the length of the concentration interval is $\cos(2\theta - 3/M) - \cos(2\theta + 3/M) = 2\sin(3/M)\sin(2\theta) \leq 12\sin(\theta)\cos(\theta)/M = \mathcal{O}(\sqrt{p(1-p)}/M)$. Thus, the variance becomes $\mathcal{O}(p(1-p)/M^2 + \exp(-m/4))$.

Putting both cases together, we obtain that the variance for unbiased probability estimation using boosted unbiased phase estimation is $\mathcal{O}(p(1-p)/M^2 + 1/M^4 + \exp(-m/4))$. Note that from this variance bound, one can essentially recover the precision that is obtained by Brassard et al. [9], up to constant factors. Thus, this

way of estimating the probability gives one an unbiased estimator, while maintaining the precision attained by traditional techniques.

Notice that this estimation procedure for the probability might output estimates that are outside the interval $[0,1]$. Indeed, for example if $p=1$ any non-trivial unbiased estimator must eventually produce estimates that are larger than 1.

Finally, note that computing the value of $\lambda(M,m)$ to high precision might be computationally difficult for larger m values. However, we can compute the value approximately by Monte Carlo simulation. One can generate $2m$ samples corresponding to $\phi=0$ using the density function Eq. (6.2), and run Algorithm 2 finally outputting $\Re(e^{i\overline{\varphi}})$. Clearly $\mathbb{E}[\Re(e^{i\overline{\varphi}})]=\mathbb{E}[e^{i\overline{\varphi}}]=\lambda(M,m)$. On the other hand as we have shown above the variance of $\Re(e^{i\overline{\varphi}})$ is $\mathcal{O}(1/M^4+\exp(-m/4))$. Intuitively speaking this means that the computation of $\lambda(M,m)$ should not prohibit applications of this result, especially considering that $\lambda(M,m)$ can be pre-computed ahead of time.

6.4 Implementation with finite precision

THEOREM 6.3. (SUPPRESSED-BIASED PHASE ESTIMATION) *If we run Algorithm 1 with some finite n in Line 1, then it returns a random phase $\varphi\in[0,2\pi)$ of the form $\frac{2\pi}{M}\left(j+\frac{\ell}{2^n}\right)$ for some $j\in\{0,1,\dots,M-1\}$ and $\ell\in\{0,1,2,3,\dots,2^n-1\}$ such that the distribution of the outcome is $\frac{2\pi}{2^n}$-close in total variation distance to the distribution*

$$\Pr\left[\varphi=\frac{2\pi}{M}\left(j+\frac{\ell}{2^n}\right)\right]=\int_{\frac{2\pi}{M}\left(j+\frac{\ell}{2^n}\right)}^{\frac{2\pi}{M}\left(j+\frac{\ell+1}{2^n}\right)} f(x)dx, \tag{6.7}$$

where $f(x)$ is defined in Eq. (6.2).

Proof. First let us consider running Algorithm 1 with $n=\infty$, except in Line 4 truncating u to have only n binary digits. Then it follows from Theorem 6.1 that the output distribution is given by Eq. (6.7).

Next consider further modifying this algorithm by truncating u to n binary digits in Line 2 as well, bringing us to the finite-n version of Algorithm 1. This introduces a change in the applied unitary with magnitude (in terms of operator norm) no greater than $\frac{2\pi}{2^n}$. Thus a perturbation is induced on the state in the algorithm of magnitude (in the ℓ_2-norm) no greater than $\frac{2\pi}{2^n}$, ultimately changing the measurement statistics by no more than $\frac{2\pi}{2^n}$ in total variation distance, cf. [48, Exercise 4.3]. ∎

Overall we can conclude that the output distribution of the discretized Algorithm 1 gets a perturbation that is at most $\mathcal{O}(2^{-n})$ in the Wasserstein-1 distance compared to the infinite-precision version of Algorithm 1.

Similarly, we believe that using the discretized version of Algorithm 1 within Algorithm 2 would exponentially suppresses the bias, and the conclusion about boosting should not be affected. However, discretizing the proof appears to be difficult, because it heavily relies on a symmetry argument – ultimately breaking due to the non-symmetric discretization errors. The main difficulty is that even small perturbations to the φ_j values can induce some large jumps in the shortest interval in some edge cases. For this reason below we introduce a slightly more complicated version of Algorithm 2 that avoids such large jumps, and so we can formally analyze its discretized version.

Algorithm 3 Boosted Suppressed-Bias Phase Estimation

Input: $2m$ copies of $|\psi(\phi)\rangle=\frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}e^{i\phi k}|k\rangle$ (for unknown ϕ)

1: **For each** $j\in[2m]$
2: Run Algorithm 1 with a fixed n on the j-th copy of $|\psi(\phi)\rangle$ and record the estimate φ_j
3: **For each** $j\in[2m]$
4: Compute d_j the m-th smallest distance in the (multi)set $\left\{|\varphi_j-\varphi_k|_{2\pi}\colon k\in[2m]\setminus\{j\}\right\}$
5: Define $w_j:=\exp(-\frac{mM}{\pi}d_j)$
6: **Return** $\overline{\varphi}:=\varphi_j$ with probability $\frac{w_j}{W}$ where $W=\sum_{j\in[2m]}w_j$

THEOREM 6.4. (BOOSTED SUPPRESSED-BIASED PHASE ESTIMATION) *Let $\beta\in(0,\frac{1}{32}]$ and $b,m\in\mathbb{N}$. We can implement an approximate version of Algorithm 3 using a (classically controlled) quantum circuit of depth*

$\mathcal{O}(\log(b)+\log\log(1/\beta))$ containing $\mathcal{O}(mb\log(b/\beta))$ two-qubit gates such that if the circuit is run on $2m$ copies of an input state ρ such that $\frac{1}{2}\|\rho-|\psi(\phi)\rangle\langle\psi(\phi)|\|_1 \le \beta$, then we get an outcome $\overline{\varphi}\in[0,2\pi]$ satisfying

$$\Pr[|\phi-\overline{\varphi}|_{2\pi} \le \frac{4\pi}{2^b}] \ge 1-2e^{-\frac{m}{4}}, \tag{6.8}$$

and

$$|\mathbb{E}[\phi-\overline{\varphi}]| \le 2\pi\min\{6m\beta, e^{-\frac{m}{4}}+2^{-b+1}\}, \tag{6.9}$$

where we interpret $\phi-\overline{\varphi}$ as a number in $[-\pi,\pi)$. Moreover, the required classical computations can be performed in time $\mathcal{O}\Big(m\mathrm{poly}\Big(b\log\Big(\frac{m}{\beta}\Big)\Big)\Big)$.

Proof. Our circuit runs Algorithm 3 setting $n=\max\{\lceil\log(10/\beta)\rceil,\lceil\log(4m)\rceil\}$, $M=2^b$ and implementing the quantum Fourier transform in Algorithm 1 by a β-approximate version given by Theorem 2.2.

To analyze our finite-precision version of Algorithm 3, we proceed similarly to the proof of Theorem 6.3: we gradually degrade the "ideal" algorithm that uses infinite precision $n=\infty$, has the "ideal" input state $|\psi(\phi)\rangle$, and uses the exact quantum Fourier transform. We do the degradation in several steps until we get to our actual algorithm.

First we prove Eq. (6.8). We start by analysing Algorithm 1. Let $\eta := \frac{5\pi}{4}$, then its "ideal" version satisfies $\Pr[|\phi-\varphi|_{2\pi}\le\frac{\eta-2^{-8}}{2^b}]\ge 0.85\ldots$ according to Eq. (6.5). So even if we truncate u to n binary digits in Line 4 of Algorithm 1, then we have that $\Pr[|\phi-\varphi|_{2\pi}\le\frac{\eta}{2^b}]\ge 0.85$. If we already truncate u to n binary digits in Line 2 of Algorithm 1, then this introduces a change in the applied unitary with magnitude (in terms of operator norm) no greater than $\frac{2\pi}{2^n}$. Similarly, using a β-approximate implementation of the quantum Fourier transform changes the applied unitary by at most β. Thus a perturbation is induced on the state in the algorithm of magnitude (in the ℓ_2-norm) no greater than $\frac{2\pi}{2^n}+\beta$, ultimately changing the measurement statistics by no more than $\frac{2\pi}{2^n}+\beta$ in total variation distance, cf. [48, Exercise 4.3]. Similarly, if we replace the ideal input state $|\psi(\phi)\rangle$ by ρ, that induces an additional change in total variation distance no more than β. So the overall change can be bounded by $\frac{2\pi}{2^n}+2\beta\le\frac{2\pi}{2^8}+\frac{1}{16}<0.088$. We can conclude that if we run Algorithm 1 with $n\ge 8$ on the input state ρ with a β-approximate quantum Fourier transform, then its output φ satisfies $\Pr[|\phi-\varphi|_{2\pi}\le\frac{\eta}{2^b}]\ge\frac{3}{4}$.

Therefore, as in the proof of Theorem 6.2, the probability that there are at least $m+1$ estimates φ_j such that $|\phi-\varphi_j|_{2\pi}\le\frac{\eta}{2^b}$ is at least $1-\exp(-\frac{m}{4})$ due to the Chernoff bound. In that case for each $\varphi_j\in[\phi-\frac{\eta}{2^b},\phi+\frac{\eta}{2^b}]$ (modulo 2π) we have that $d_j\le\frac{2\eta}{2^b}$ so $w_j\ge\exp(-\frac{2\eta m}{\pi})$ and thus $W\ge(m+1)\exp(-\frac{2\eta m}{\pi})$. Further, there are at most $m-1$ remaining $\varphi_i\notin[\phi-\frac{\eta}{2^b},\phi+\frac{\eta}{2^b}]$, and for any such i the m shortest distances $|\varphi_i-\varphi_k|_{2\pi}$ must include some $\varphi_k\in[\phi-\frac{\eta}{2^b},\phi+\frac{\eta}{2^b}]$. Thus, if $|\varphi_i-\phi|_{2\pi}>\frac{4\pi}{2^b}$ then $d_i>\frac{4\pi-\eta}{2^b}$ and $w_i\le\exp(-(4\pi-\eta)\frac{m}{\pi})$. We can conclude the proof of Eq. (6.8) by using the union bound, observing that

$$\Pr\Big[|\phi-\overline{\varphi}|_{2\pi}>\frac{4\pi}{2^b}\Big]\le\exp\Big(-\frac{m}{4}\Big)+(m-1)\frac{\exp(-(4\pi-\eta)\frac{m}{\pi})}{W}\le 2\exp\Big(-\frac{m}{4}\Big). \tag{6.10}$$

The above equation implies the trivial bound on the bias. The total bias coming from estimates $\overline{\varphi}$ such that $|\phi-\overline{\varphi}|_{2\pi}\le\frac{4\pi}{2^b}$ is at most $\frac{4\pi}{2^b}$. The total contribution from the other estimates is at most $\pi\cdot 2\exp\big(-\frac{m}{4}\big)$ due to Eq. (6.10). So, indeed $|\mathbb{E}[\phi-\overline{\varphi}]|\le 4\pi 2^{-b}+2\pi e^{-\frac{m}{4}}$.

Now we finish proving Eq. (6.9). Due to symmetry, the estimate $\overline{\varphi}$ outputted by the "ideal" algorithm is unbiased; interpreting $\phi-\overline{\varphi}$ as a number in $[-\pi,\pi)$ (rather than, say $(-\pi,\pi]$) does not introduce bias either, since the probability density of the estimate $\overline{\varphi}$ is continuous in the case $n=\infty$, and so $\Pr[\overline{\varphi}=\phi\pm\pi]=0$. Thus, Eq. (6.9) trivially holds for the "ideal" algorithm.

We gradually transform the "ideal" algorithm to our approximate version. For this let us introduce the notation $\overline{\varphi}^{(\ell)}$ for denoting the output state when Algorithm 3 is only performed "ideally" until just before executing Line ℓ, from which point our approximate version is executed. First, let us consider truncating u to n binary digits in Line 6, thus returning $\overline{\varphi}^{(6)}$, i.e., the sampled φ_j represented with only finite precision. This introduces a change in the output $|\overline{\varphi}^{(6)}-\overline{\varphi}|_{2\pi}$ that is at most $\frac{2\pi}{M}2^{-n}$. Also, the value of $\phi-\overline{\varphi}^{(6)}\in[-\pi,\pi)$ changes by at most $\frac{2\pi}{M}2^{-n}$ compared to $\phi-\overline{\varphi}\in[-\pi,\pi)$, unless $\overline{\varphi}\in[\phi+\pi,\phi+\pi+\frac{2\pi}{M}2^{-n}]$ modulo 2π (when the change

might be as large as 2π). The probability of the latter happening can be bounded by

$$\begin{aligned}\Pr[\overline{\varphi} \in [\phi+\pi, \phi+\pi+\frac{2\pi}{M}2^{-n}]] &\leq \Pr[\exists j\colon \varphi_j \in [\phi+\pi, \phi+\pi+\frac{2\pi}{M}2^{-n}]] \\ &\leq 2m \Pr[\varphi \in [\phi+\pi, \phi+\pi+\frac{2\pi}{M}2^{-n}]] \\ &\leq 2m\frac{2\pi}{M}2^{-n}\sup_{\varphi} f(\varphi) \\ \text{(by Eq. (6.2) \& (6.4))} \qquad &\leq 2m2^{-n},\end{aligned}$$

thus we get

$$|\mathbb{E}[\phi - \overline{\varphi}^{(6)}]| \leq \frac{2\pi}{M}2^{-n} + 2\pi \Pr[\overline{\varphi} \in [\phi+\pi, \phi+\pi+\frac{2\pi}{M}2^{-n}]] \leq 2\pi(2m+1)2^{-n}. \tag{6.11}$$

Second, let us consider truncating u in Line 3. This will change the distances d_j by at most $\frac{2\pi}{M}2^{-n}$ since every pair of distances $|\varphi_j - \varphi_k|_{2\pi}$ is changed by no more than $\frac{2\pi}{M}2^{-n}$. Thus every weight w_j is perturbed by a multiplicative factor $\exp(\pm\frac{mM}{\pi}\frac{2\pi}{M}2^{-n}) = \exp(\pm 2m2^{-n})$ and consequently W gets a multiplicative perturbation up to $\exp(\pm 2m2^{-n})$. This induces a perturbation of the probabilities $p_j := w_j/W$ by a multiplicative factor up to $\exp(\pm 4m2^{-n})$, resulting in an up to $(\exp(4m2^{-n}) - 1)/2$-perturbation in total variation distance to the sampling distribution in Line 6. As $n \geq \log_2(4m)$ this can be upper bounded by $4m2^{-n}$. Since the only change between the outputs $\overline{\varphi}^{(6)}$ and $\overline{\varphi}^{(3)}$ is due to the change in the sampling distribution in Line 6 we get by Eq. (6.11)

$$|\mathbb{E}[\phi - \overline{\varphi}^{(3)}]| \leq |\mathbb{E}[\phi - \overline{\varphi}^{(6)}]| + 2\pi \cdot 4m2^{-n} \leq 2\pi(6m+1)2^{-n} = 4\pi m(3+\frac{1}{2m})2^{-n}. \tag{6.12}$$

Finally, let us switch completely to our approximate algorithm. This affects the unitary applied in Line 2 of Algorithm 1, switches to the approximate Fourier transform, and replaces the "ideal" input state by ρ. As we already showed this introduces an up to $\frac{2\pi}{2^n} + 2\beta$ change in the measurement statistics of Algorithm 1 in total variation distance. Since Algorithm 1 is repeated $2m$ times, the overall perturbation in total variation distance can be bounded by $2m(\frac{2\pi}{2^n} + 2\beta)$. Thus we get by Eq. (6.12)

$$|\mathbb{E}[\phi - \overline{\varphi}]| \leq |\mathbb{E}[\phi - \overline{\varphi}^{(3)}]| + 2\pi \cdot 2m\left(\frac{2\pi}{2^n} + 2\beta\right) \leq 4\pi m\left(\frac{2\pi + 3 + \frac{1}{2m}}{2^n} + 2\beta\right) \leq 12\pi m\beta.$$

The gate complexity and depth directly follows from Theorem 2.2. The classical computation cost is dominated by Lines 4-6 of Algorithm 3. We can treat all φ_j values as numbers in $[0, 2\pi)$ and define $\varphi'_j \in [0, 2)$ as $\varphi'_j := \frac{\varphi_j}{\pi}$. Then we sort them in time $\mathcal{O}(m\text{poly}(b\log(\frac{m}{\beta})))$. This enables computing each $d'_j := d_j/\pi$ value in time $\mathcal{O}(\text{poly}(b\log(\frac{m}{\beta})))$ using for example binary search. For a numerically stable way of computing w_j we first compute $d''_j := d'_j - \min_{i\in[2m]} d'_i$ and define $w'_j := \exp(-mMd''_j)$ as well as $W' := \sum_{j\in[2m]} w'_j$. It is easy to see that $\frac{w_j}{W} = \frac{w'_j}{W'}$. Then we compute approximations $\tilde{w}_j$ for each j such that $|\tilde{w}_j - w'_j| \leq \mathcal{O}(2^{-n}/m)$ and if $d'_j > 2^{1-b}$, then $\tilde{w}_j \leq w'_j$ and else $\tilde{w}_j \geq w'_j$. Sampling according to $\frac{\tilde{w}_j}{\widetilde{W}}$ ensures that Eq. (6.10) remains valid, as the inaccuracies induce bias towards values j such that $|\varphi_j - \phi| \leq \frac{2\pi}{2^b}$, moreover it also ensures that the total variation distance between the distributions $\frac{\tilde{w}_j}{\widetilde{W}}$ and $\frac{w}{W}$ is of the order 2^{-n}. Tuning the constants and exploiting the fact that the $4m2^{-n}$ total variation distance bound above Eq. (6.12) has some slack ensures that Eq. (6.12) also remains valid and so the entire analysis remains valid even when the approximate $\tilde{w}_j$ are used. Computing all values $d'_j, d''_j, \tilde{w}_j, \widetilde{W}$ and (approximately) sampling according to $\frac{\tilde{w}_j}{\widetilde{W}}$ can be performed in time $\mathcal{O}(\text{poly}(b\log(\frac{m}{\beta})))$, cf. [47]. ∎

We remark that one can similarly show that our suppressed-bias phase estimators give rise to suppressed-bias estimators of $e^{i\phi}$ by using the constructions of Section 6.2, and consequently also allows for the implementation of suppressed-bias probability estimators.

6.5 Unbiased gradient estimation Now we show that our suppressed-bias phase estimation techniques lead to suppressed-bias gradient estimation further improving over Jordan's gradient estimation algorithm, and its variants [18, Lemma 5.1]. Note that if we would use perfect input states and the exact unbiased phase-estimation then we would get a symmetric error distribution. Unfortunately, this no longer holds due to the approximation errors in the input state and the finite bit precision.

To simplify the exposition, let us introduce the notation $B_\infty(g,\varepsilon)$ to denote the (closed) ε-ball around g containing all points x such that $\|x-g\|_\infty \le \varepsilon$.

THEOREM 6.5. (SUPPRESSED-BIAS GRADIENT ESTIMATION) *Let $\varepsilon,\delta \in (0,\frac{1}{6}]$ and $g\in\mathbb{R}^d$ such that $\|g\|_\infty \le \frac{1}{3}$. Let $b := \lceil \log_2(\frac{2}{\varepsilon})\rceil$, $B := 2^b$ and $m := \lceil 8\ln(2d/\delta)\rceil$. If $\left\| |\psi\rangle - \frac{1}{\sqrt{B^d}}\sum_{x\in G_b^d} e^{2\pi i B\langle g,x\rangle}|x\rangle\right\| \le \frac{\delta}{12\pi m}$ and we are given $2m$ copies of $|\psi\rangle$, then we can compute a vector $k\in[-\frac{1}{2},\frac{1}{2}]^d$ such that*

$$\Pr[\|k-g\|_\infty > \varepsilon] \le \delta \tag{6.13}$$

and

$$\|\mathbb{E}[k]-g\|_\infty \le \delta. \tag{6.14}$$

Furthermore, the gate complexity of the procedure is $\mathcal{O}\big(d\log(\frac{d}{\delta})\log(\frac{1}{\varepsilon})\log\big(\frac{d}{\delta}\log(\frac{1}{\varepsilon})\big)\big)$, and the circuit depth is $\mathcal{O}\big(\log(\frac{1}{\varepsilon})\log\big(\frac{d}{\delta}\log(\frac{1}{\varepsilon})\big)\big)$. Finally, there is a random variable $k'\in B_\infty(g,\varepsilon)$ with independent coordinates that is δ-close in total variation distance to k and satisfies $\mathbb{E}[k']\in B_\infty(g,\delta)$.

Proof. We apply Theorem 6.4 with $M=B$, $m=\lceil 8\ln(2d/\delta)\rceil$ and $\beta=\frac{\delta}{12\pi m}$. ∎

Finally, we prove a corollary analogous to Corollary 2.1 which will be the main technical tool in the following Section 7-8.

COROLLARY 6.1. (ALMOST LINEAR BLOCK-HAMILTONIAN TO GRADIENT) *Let $\varepsilon,\delta\in(0,\frac{1}{6}]$, $b := \lceil\log_2(\frac{16}{\varepsilon})\rceil$, $B = 2^b$ and $\beta := \frac{\delta}{96\lceil\ln(6d/\delta)\rceil+12}$. Suppose that we have an a-block-encoding W of a diagonal matrix with diagonal entries $f(x)\in\mathbb{R}$ for $x\in G_b^d$ satisfying $|f(x)-\langle x,g\rangle| \le \frac{\varepsilon\beta}{4\pi}$ for at least a $(1-\beta^2)$ fraction of the points in G_b^d. Then with $\mathcal{O}\Big(\Big(\frac{1}{\varepsilon}+\log(\frac{\log(d)}{\delta})\Big)\log(\frac{d}{\delta})\Big)$ (controlled) uses of W (and its inverse) and $\mathcal{O}\Big(\Big(d\log(\frac{1}{\varepsilon})\log\big(\frac{d}{\delta}\log(\frac{1}{\varepsilon})\big)+a\Big(\frac{1}{\varepsilon}+\log(\frac{\log(d)}{\delta})\Big)\Big)\log(\frac{d}{\delta})\Big)$ other gates with circuit depth $\mathcal{O}\Big(\log(\frac{1}{\varepsilon})\log\big(\frac{d}{\delta}\log(\frac{1}{\varepsilon})\big)+\log(a)\Big(\frac{1}{\varepsilon}+\log(\frac{\log(d)}{\delta})\Big)\Big)$ we can compute a vector $k\in[-4,4]^d$ such that*

$$\Pr[\|k-g\|_\infty > \varepsilon] \le \delta, \tag{6.15}$$

and

$$\|\mathbb{E}[k]-g\|_\infty \le 8\delta. \tag{6.16}$$

Moreover, there is a random variable $k'\in B_\infty(g,\varepsilon)$ with independent coordinates that is δ-close in total variation distance to k and satisfies $\mathbb{E}[k']\in B_\infty(g,8\delta)$.

Proof. We proceed similarly to the proof of Corollary 2.1. The main idea is to apply Theorem 6.5 with preparing the (approximate) initial state via block-Hamiltonian simulation Lemma 2.4. In the proof of Corollary 2.1 it is shown that the assumptions in the statement imply $\|g\|_\infty \le \frac{8}{3}$. Therefore, we will apply Theorem 6.5 to the gradient $\frac{g}{8}$ with precision $\frac{\varepsilon}{8}$. The first step is to prepare a uniform superposition over the grid G_b^d by applying a Hadamard gate to all $d\cdot b$ qubits, that are initially in the $|0\rangle$ state.

First let us assume that we have access to a perfect phase oracle $P := \sum_{x\in G_b^d}|x\rangle\langle x|e^{2\pi i\frac{B}{8}f(x)}$ so that we can prepare the sate $|\psi\rangle = \frac{1}{\sqrt{B^d}}\sum_{x\in G_b^d}|x\rangle e^{2\pi i\frac{B}{8}f(x)}$. We bound the difference from the ideal state $|\phi\rangle$ analogously to

the proof of [18, Lemma 5.1]. Let $S \subseteq G_b^d$ be the set of points for which $|f(x) - \langle x, g\rangle| \leq \frac{\varepsilon\beta}{4\pi}$ holds, then

$$
\begin{aligned}
\||\psi\rangle - |\phi\rangle\|^2 &= \frac{1}{B^d} \sum_{x \in G_b^d} \left| e^{2\pi i \frac{B}{8} f(x)} - e^{2\pi i \frac{B}{8}\langle x,g\rangle} \right|^2 \\
&= \frac{1}{B^d} \sum_{x \in S} \left| e^{2\pi i \frac{B}{8} f(x)} - e^{2\pi i \frac{B}{8}\langle x,g\rangle} \right|^2 + \frac{1}{B^d} \sum_{x \in G_b^d \setminus S} \left| e^{2\pi i \frac{B}{8} f(x)} - e^{2\pi i \frac{B}{8}\langle x,g\rangle} \right|^2 \\
(|e^{iz} - e^{iy}| \leq |z - y|) \qquad &\leq \frac{1}{B^d} \sum_{x \in S} \left| 2\pi \frac{B}{8} f(x) - 2\pi \frac{B}{8} \langle x, g\rangle \right|^2 + \frac{1}{B^d} \sum_{x \in G_b^d \setminus S} 4 \\
&= \frac{1}{B^d} \sum_{x \in S} (2\pi \frac{B}{8})^2 |f(x) - \langle x, g\rangle|^2 + 4 \frac{|G_b^d \setminus S|}{B^d} \\
\text{(by the assumptions of the corollary)} \qquad &\leq \frac{1}{B^d} \sum_{x \in S} 4\beta^2 + 4\beta^2 \\
&\leq 8\beta^2.
\end{aligned}
$$

Finally, we can implement a $(4 - 2\sqrt{2})\beta$-approximation $\widetilde{P}$ of the perfect phase oracle P by applying block-Hamiltonian simulation Lemma 2.4 to W.[3] This lets us preparing an approximate state $|\tilde{\psi}\rangle$ such that $\left\| |\tilde{\psi}\rangle - |\psi\rangle \right\| \leq (4 - 2\sqrt{2})\beta$ and so $\left\| |\tilde{\psi}\rangle - |\phi\rangle \right\| \leq 4\beta$, enabling us to apply Theorem 6.5.

The query complexity follows from the fact that we prepare the state $|\tilde{\psi}\rangle$ a total of $\mathcal{O}\big(\log(\frac{d}{\delta})\big)$ times, each time making $\mathcal{O}\Big(\frac{1}{\varepsilon} + \log(\frac{1}{\beta})\Big) = \mathcal{O}\Big(\frac{1}{\varepsilon} + \log(\frac{\log(d)}{\delta})\Big)$ (controlled) queries to W. The additional gate complexity of preparing $|\tilde{\psi}\rangle$ is $\mathcal{O}(a)$ times the query complexity plus $d \cdot b$ for the Hadamard gates. We get the overall gate complexity by adding the gate cost in Theorem 6.5. ∎

6.6 Application to low depth probability estimation In this section, we sketch a quick application of our results. If we have access to the operation

$$U : |0\rangle \mapsto \sqrt{1-p}|\psi_0\rangle|0\rangle + \sqrt{p}|\psi_1\rangle|1\rangle,$$

then we can estimate p with a depth-t algorithm, by running our version of unbiased probability estimation with $1/M = \mathcal{O}(\log(t)/t)$ and $m = \mathcal{O}(\log(t))$, to obtain an estimate with variance $\widetilde{\mathcal{O}}\big(p(1-p)/t^2 + 1/t^4\big)$. Moreover, we can run this procedure K times in parallel, and take the average of the outcomes. This gives an estimator of p that is still unbiased, and whose variance is

$$\widetilde{\mathcal{O}}\Big(\frac{p(1-p)}{Kt^2} + \frac{1}{Kt^4}\Big).$$

Thus, we obtain a way to estimate p, when we are constrained to using depth-t quantum algorithms, and we can obtain precision ε with high probability if we set $K = \Theta(\max\{p(1-p)/(\varepsilon t)^2, 1/(\varepsilon^2 t^4)\})$.

Now, let $\beta \in (0, 1]$, and suppose the depth that we can use is $t = \Theta(1/\varepsilon^{1-\beta})$. Then, in order to achieve precision ε, we can set $K = \widetilde{\mathcal{O}}\big(\max\{p(1-p)/\varepsilon^{2\beta}, 1/\varepsilon^{4\beta-2}\}\big)$, from which we find that the total number of calls to U becomes $Kt = \widetilde{\mathcal{O}}\big(\max\{p(1-p)/\varepsilon^{1+\beta}, 1/\varepsilon^{3\beta-1}\}\big)$. If we use the crude upper bound $p(1-p) \leq 1$, this reduces to $Kt = \widetilde{\mathcal{O}}\big(1/\varepsilon^{1+\beta}\big)$, and hence we recover the result obtained in [22]. Moreover, we get a slight improvement if we know some small upper bound $q \geq p$ a priori.

7 Second intermezzo: estimating multiple expectation values with a state-preparation oracle

To perform efficient mixed-state tomography we rely on an algorithm to estimate m expectations with few copies of the state. The algorithm is based on constructing the phase oracle for a function whose gradient is the vector of the desired expectation values, similarly to what we did for pure states. The task here is however more complicated, because to ensure that the function is properly normalized we need to bound the weighted

combination of expectation values, where the weights are taken from a hypergrid in $[-\frac{1}{2}, \frac{1}{2}]^m$ (as these are the points used by the gradient algorithm of [18]). This requires some results on random matrices, which we use by translating properties that hold for uniformly random matrices into properties that hold for all but a constant fraction of the points in the hypergrid.

Formally, we assume access to a unitary that prepares a purification of a state $\rho \in \mathbb{C}^{d\times d}$, and its inverse (without requiring them to be controlled). Our goal is to estimate the expectation values $\mathrm{Tr}(E_j\rho)$ of measurement operators E_j for $j = 1, \ldots, m$ up to corresponding errors ε_j, with as few applications of the state-preparation unitary for ρ as possible. We do not apply any gates to the purifying register, other than the state-preparation oracle and its inverse; thus, we do not need to impose any restrictions on how the purification of ρ is constructed. We assume that $\|E_j\| \leq 1$ for all $j = 1, \ldots, m$, which is w.l.o.g. as we can always scale E_j and ε_j down by $\|E_j\|$ to achieve this. Finally, we assume that we are given access to each E_j via a (controlled) block-encoding. Note that this access model is rather generic; for example, if we have an implementation of a POVM for $E_j, I - E_j$, then we can convert this to a block-encoding for E_j via Lemma 2.1.

This task was recently studied in [31], yielding an algorithm that solves the problem using $\mathcal{O}\Big(\sqrt{\sum_{j=1}^m \|E_j\|^2}/\varepsilon\Big)$ applications of the state-preparation unitary and its inverse, in the case where all ε_j are equal to ε. This is $\mathcal{O}(\sqrt{m}/\varepsilon)$ under the assumption $\|E_j\| \leq 1$. The algorithm of [31] is however not optimal in our setting: we want to give an algorithm with a sample complexity that depends on $\sqrt{\|\sum_{j=1}^m E_j^2/\varepsilon_j^2\|}$, because this leads to a saving of a factor d when applied to mixed-state tomography compared to the algorithm of [31], see the details in the next section. For a discussion of other existing approaches to solve the problem of computing expectation values, we refer to the excellent introduction in [31].

7.1 Bounds on uniform matrix series As mentioned above, we first need to prove some properties of uniform random matrices. We do this by adapting a result on Gaussian / Rademacher random matrices given below. Here and in the remainder, for a random matrix Y we define $v(Y) := \left\|\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2\right\|$ as its variance.

Theorem 7.1. (Gaussian & Rademacher matrix series inequality [46, Theorem 4.6.1]) *Let $E_1, \ldots, E_m$ be $d \times d$ Hermitian matrices. Let $\lambda_1, \ldots, \lambda_m$ be drawn from iid standard normal distributions and let $Y = \sum_{j=1}^m \lambda_j E_j$. Then $\mathbb{E}[Y] = 0, v(Y) = \left\|\sum_j E_j^2\right\|$ and*

$$\mathbb{P}[\|Y\| \geq t] \leq 2de^{-\frac{t^2}{2v(Y)}}.$$

The same bounds hold when $\{\lambda_j\}$ is iid uniformly random over $\{-1, 1\}$.

In order to adapt the above result to our setting we invoke a technical statement from [46]:

Proposition 7.1. (Master Bound for a Sum of Independent Random Matrices, [46, Theorem 3.6.1]) *Consider a finite sequence $\{E_j\}$ of independent, random, Hermitian matrices of the same size. Then for all $t \in \mathbb{R}$ we have*

$$\mathbb{P}[\lambda_{max}(\sum_j E_j) \geq t] \leq \inf_{\theta>0} e^{-\theta t}\mathrm{Tr}\left(\exp\left(\sum_j \log \mathbb{E}[e^{\theta E_j}]\right)\right).$$

With the help of this result we prove the following variant of Theorem 7.1 for bounded random variables:

Theorem 7.2. (Bounded Matrix series inequality) *Let $E_1, \ldots, E_m$ be $d \times d$ Hermitian matrices. Let $\lambda_1, \ldots, \lambda_m$ be independent symmetrically distributed random variables supported on $[-1, 1]$ and let $Y = \sum_{j=1}^m \lambda_j E_j$. Then $\mathbb{E}[Y] = 0, v(Y) \leq \left\|\sum_j E_j^2\right\|$ and*

$$\mathbb{P}[\|Y\| \geq t] \leq 2de^{-\frac{t^2}{2v(Y)}}.$$

Proof. We follow the proof of [46, Theorem 4.6.1] and modify it where necessary. First we note that

$$\begin{aligned}
\mathbb{E}[e^{\lambda_j E_j}] &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{\lambda_j^k}{k!} E_j^k\right] \\
\text{(linearity of expectation)} \qquad &= \sum_{k=0}^{\infty} \frac{\mathbb{E}[\lambda_j^k]}{k!} E_j^k \\
(\lambda_j \text{ is symmetrically distributed}) \qquad &= \sum_{q=0}^{\infty} \frac{\mathbb{E}[\lambda_j^{2q}]}{(2q)!} E_j^{2q} \\
(\lambda_j \text{ is bounded}) \qquad &\preceq \sum_{q=0}^{\infty} \frac{1}{(2q)!} E_j^{2q} \\
((2q)! \geq 2^q q!) \qquad &\preceq \sum_{q=0}^{\infty} \frac{1}{q!} (E_j^2/2)^q \\
(7.17) \qquad &= e^{E_j^2/2}.
\end{aligned}$$

Now we show that the above inequality implies that

$$\mathrm{Tr}\left(\exp\left(\sum_j \log(\mathbb{E}[e^{\lambda_j E_j}])\right)\right) \leq \mathrm{Tr}\left(\exp\left(\frac{1}{2}\sum_j E_j^2\right)\right). \tag{7.18}$$

Indeed, we know that [27, Chapter 4.1] the logarithm is operator monotone for positive matrices. Therefore Eq. (7.17) implies that $\log(\mathbb{E}[e^{\lambda_j E_j}]) \preceq \log(e^{E_j^2/2})$, and consequently $\sum_j \log(\mathbb{E}[e^{\lambda_j E_j}]) \preceq \sum_j \log(e^{E_j^2/2}) = \sum_j E_j^2/2$. We conclude by using the fact that the trace of a monotone function is operator monotone [27, Example 3.24], i.e., $A \preceq B$ implies $\mathrm{Tr}(\exp(A)) \leq \mathrm{Tr}(\exp(A))$. We now use this (by absorbing θ into the E_j-s) to get

$$\begin{aligned}
\text{(by Proposition 7.1)} \qquad \mathbb{P}[\lambda_{max}(Y) \geq t] &\leq \inf_{\theta>0} e^{-\theta t} \mathrm{Tr}\left(\exp\left(\sum_j \log \mathbb{E}[e^{\theta\lambda_j E_j}]\right)\right) \\
\text{(by Eq. (7.18))} \qquad &\leq \inf_{\theta>0} e^{-\theta t} \mathrm{Tr}\left(\exp\left(\frac{\theta^2}{2}\sum_j E_j^2\right)\right) \\
&\leq \inf_{\theta>0} e^{-\theta t} d \cdot \left\|\exp\left(\frac{\theta^2}{2}\sum_j E_j^2\right)\right\| \\
&= \inf_{\theta>0} e^{-\theta t} d \cdot \exp\left(\frac{\theta^2}{2}\left\|\sum_j E_j^2\right\|\right) \\
&= d \inf_{\theta>0} \exp\left(-\theta t + \frac{\theta^2}{2} v(Y)\right).
\end{aligned}$$

As the exponential function is monotone increasing, the minimum is attained at the minimum of $-\theta t + \frac{v(Y)}{2}\theta^2$. By differentiating and setting equal to zero we find

$$-t + v(Y)\theta = 0$$

and hence $\theta = \frac{t}{v(Y)}$. Substituting this back we find

$$\mathbb{P}[\lambda_{max}(Y) \geq t] \leq d e^{-\frac{t^2}{2v(t)}}.$$

By symmetry we get the same bound for the smallest eigenvalue and the theorem follows. ∎

7.2 Application to the estimation of multiple expectation values With the tools from the previous section we can tighten the analysis of [31] for the estimation of multiple expectation values. Our running time generalizes the results of [31], and it leads to faster algorithms in some cases that are relevant for tomography.

LEMMA 7.1. *Let $E_1, \ldots, E_m$ be Hermitian matrices with $\|E_j\| \leq 1$, and let $U_E = \sum_{j\in[m]} |j\rangle\langle j| \otimes U_{E_j}$, where U_{E_j} is a a-block-encoding of E_j. Let $\delta > 0$, $\gamma \in \mathbb{R}^m$, $\nu = \|\gamma\|_1$, $\sigma \geq \sqrt{2\left\|\sum_j \gamma_j^2 E_j^2\right\| \ln\left(\frac{2d}{\delta}\right)}$, and $\sigma' := \min\{\nu, \sigma\}$. For any positive integer $b = \mathcal{O}\left(\frac{1}{\varepsilon}\right)$ we can implement a unitary $V = \sum_{x\in G_b^m} |x\rangle\langle x| \otimes V_x$ such that V_x is an $(a+\lceil\log_2(m)\rceil+2)$-block-encoding of a matrix A_x that is ε-close in operator norm to $\frac{1}{\sigma'}\sum x_j\gamma_j E_j$ for at least a $1-\delta$ fraction of points $x \in G_b^m$. This implementation of V uses $\mathcal{O}\left(\frac{\nu}{\sigma'}\log(\frac{\nu}{\sigma'\varepsilon})\right)$ calls to U_E, and $\mathcal{O}\left((a+m)\frac{\nu}{\sigma'}\mathrm{polylog}(\frac{\nu+m}{\sigma'\varepsilon})\right)$ additional two-qubit gates having depth $\mathcal{O}\left(\frac{\nu}{\sigma'}\mathrm{polylog}(\frac{\nu+m}{\sigma'\varepsilon})\right)$.*

Proof. Our goal is to construct a block-encoding of $\frac{1}{\sigma}\sum x_j\gamma_j E_j$. First, we note that this is a valid block-encoding (more precisely, its spectral norm is upper bounded by $\frac{1}{2}$) for at least $1-\delta$ fraction of points $x \in G_b^m$. To see this, we apply Lemma 7.2 to the matrices $\gamma_1 E_1, \gamma_2 E_2, \ldots, \gamma_m E_m$ setting $t = \sqrt{2\left\|\sum_j \gamma_j^2 E_j^2\right\| \ln\left(\frac{2d}{\delta}\right)} \leq \sigma$ and sampling $x \in G_b^m$ uniformly at random to obtain

$$\mathbb{P}_{x\in G_b^m}[\|\sum_j 2x_j\gamma_j E_j\| \geq \sigma] \leq \mathbb{P}_{x\in G_b^m}[\|\sum_j 2x_j\gamma_j E_j\| \geq t] \leq 2de^{-\frac{t^2}{2\left\|\sum_j \gamma_j^2 E_j^2\right\|}} = \delta.$$

Using Lemma 2.2, we first prepare a $(a+\lceil\log_2(m)\rceil+1)$-block-encoding of $\sum_{j=1}^m (x_j\gamma_j/\|\gamma\|_1)E_j$. This requires a single application of U_{E_j}, and one applications of a state-preparation oracle for $\frac{1}{\sqrt{\|\gamma\|_1}}\sum_j \sqrt{x_j\gamma_j}|j\rangle|0\rangle + |\psi\rangle|1\rangle$ (and its inverse), which is trivial to construct with controlled rotations given the binary encoding of $|x\rangle$. We then amplify the block-encoding by a factor $\nu/\sigma = \|\gamma\|_1/\sigma$ using Lemma 2.3: this introduces an overhead equal to the amplification factor. Overall, this requires $\mathcal{O}\left(\left\lceil\frac{\nu}{\sigma}\log(\frac{\nu}{\sigma\varepsilon})\right\rceil\right)$ calls to U_E.

The gate complexity of implementing the state-preparation operation to precision $\mathcal{O}\left(\frac{\varepsilon}{\nu m}\right)$ can be bounded by $\mathcal{O}\left(m\mathrm{polylog}(\frac{\nu+m}{\varepsilon})\right)$, while Lemma 2.3 multiplies this by $\mathcal{O}\left(\left\lceil\frac{\nu}{\sigma}\log(\frac{\nu}{\sigma\varepsilon})\right\rceil\right)$ and additionally introduces $\mathcal{O}\left((a+\log(m)+1)\left\lceil\frac{\nu}{\sigma}\log(\frac{\nu}{\sigma\varepsilon})\right\rceil\right)$ gates proving the gate complexity bound. ∎

THEOREM 7.3. *Let $E_1, \ldots, E_m \in \mathbb{C}^{d\times d}$ be Hermitian matrices with $\|E_j\| \leq 1$, and let $U_E = \sum_{j\in[m]} |j\rangle\langle j| \otimes U_{E_j}$, where U_{E_j} is a (controlled) a_E-block-encoding of E_j. Let $\delta \in (0, \frac{1}{6}]$, $\varepsilon_1, \ldots, \varepsilon_m \in (0, 2]^m$ be error bounds, $\nu = \sum_j \frac{1}{\varepsilon_j}$, $\sigma \geq \max\left\{\sqrt{2\left\|\sum_j E_j^2/\varepsilon_j^2\right\| \ln\left(\frac{2d}{\delta}\right)}, 1\right\}$, and $\sigma' := \min\{\nu, \sigma\}$. Let U_ρ be an a_ρ-qubit state-preparation unitary for a purification of $\rho \in \mathbb{C}^{d\times d}$. There is a quantum algorithm that makes $\mathcal{O}\left(\left(\sigma' + \log(\frac{\log(m)}{\delta})\right)\log(\frac{m}{\delta})\right)$ queries to U_ρ and $U_\rho^\dagger$, and produces estimates $z \in [-1, 1]^m$ such that, with probability at least $1-\delta$,*

$$\forall j \in [m]\colon |\mathrm{Tr}(\rho E_j) - z_j| \leq \varepsilon_j, \tag{7.19}$$

moreover

$$\forall j \in [m]\colon |\mathrm{Tr}(\rho E_j) - \mathbb{E}[z_j]| \leq 16\sigma'\varepsilon_j\delta. \tag{7.20}$$

Furthermore, the quantum algorithm can be implemented by a number of calls to U_E bounded by $\mathcal{O}\left(\nu\log(\frac{\nu\log(m)}{\delta})\log(\frac{m}{\delta}) + \frac{\nu}{\sigma'}\log(\frac{\nu\log(m)}{\delta})\log(\frac{m}{\delta})\log(\frac{\log(m)}{\delta})\right)$, and additional number of two-qubit gates bounded by $\mathcal{O}\left((\sigma' a_\rho + \nu a_E + \nu m)\mathrm{polylog}(\frac{\nu m}{\delta})\right)$ and having depth $\mathcal{O}\left(\nu\mathrm{polylog}(\frac{\nu m}{\delta})\right)$.

Finally, there is a random variable $z' \in \bigtimes_{j\in[m]}[\mathrm{Tr}(\rho E_j) - \varepsilon_j, \mathrm{Tr}(\rho E_j) + \varepsilon_j]$ with independent coordinates that is δ-close in total variation distance to z and also satisfies Eq. (7.20).

Proof. The main idea is to apply Jordan's gradient estimation algorithm to a linear function with derivative vector g such that $g_j = \frac{1}{\sigma'}\mathrm{Tr}\left(\rho\frac{E_j}{\varepsilon_j}\right)$ with accuracy $\varepsilon' := \min\{\frac{1}{\sigma'}, \frac{1}{6}\}$.

Let $b := \lceil \log_2(16/\varepsilon') \rceil$ and let $\beta := \frac{\delta}{96\lceil \ln(6m/\delta)\rceil + 12}$. If $\sigma' < \nu$, then we use Lemma 7.1 in order to construct a unitary $V = \sum_{x \in G_b^m} V_x \otimes |x\rangle\langle x|$ such that V_x is a $c := (a_E + \lceil \log_2(m) \rceil + 2)$-block-encoding of a matrix A_x that is $\frac{\beta}{4\sigma'\pi}$-close in operator norm to $\frac{1}{\sigma'} \sum \frac{x_j E_j}{\varepsilon_j}$ for at least a $1 - \beta^2$ fraction of points $x \in G_b^m$. Otherwise, when $\sigma' = \nu$ then we simply apply the first step in the algorithm of Lemma 7.1, namely Lemma 2.2.

We then define $V_\ell := I_c \otimes U_\rho \otimes I_{G_b^d}$ and $V_r := (V \otimes I_P) \cdot V_\ell$, where I_P acts on the purifying register of U_ρ. Let use the notation $|\rho\rangle_{PS} := U_\rho |0\rangle$. Since by definition $\mathrm{Tr}_P(|\rho\rangle\langle\rho|_{PS}) = \rho$, we have

$$\begin{aligned}
\langle 00x | V_\ell^\dagger V_r | 00y \rangle &= \langle 0 | \langle \rho |_{PS} \langle x | (V \otimes I_P) | 0 \rangle | \rho \rangle_{PS} | y \rangle \\
&= \delta_{xy} \langle 0 | \langle \rho |_{PS} (V_x \otimes I_P) | 0 \rangle | \rho \rangle_{PS} \\
&= \delta_{xy} \langle \rho |_{PS} (A_x \otimes I_P) | \rho \rangle_{PS} \\
&= \delta_{xy} \mathrm{Tr}(\langle \rho |_{PS} (A_x \otimes I_P) | \rho \rangle_{PS}) \\
&= \delta_{xy} \mathrm{Tr}((A_x \otimes I_P) | \rho \rangle\langle \rho |_{PS}) \\
&= \delta_{xy} \mathrm{Tr}(A_x \rho).
\end{aligned}$$

Since $\left\| A_x - \frac{1}{\sigma'} \sum \frac{x_j E_j}{\varepsilon_j} \right\| \leq \frac{\beta}{4\sigma'\pi}$ for at least a $1 - \beta^2$ fraction of points $x \in G_b^m$ we get that $\left| \mathrm{Tr}_P(\rho A_x) - \mathrm{Tr}_P\left(\frac{\rho}{\sigma'} \sum \frac{x_j E_j}{\varepsilon_j} \right) \right| \leq \frac{\beta}{4\sigma'\pi}$ also holds for these points. Thus $W := V_\ell^\dagger (V \otimes I_P) \cdot V_\ell$ is an $(a_\rho + c + 1)$-block-encoding of $f(x)$ that is $\frac{\beta}{4\sigma'\pi}$-close to $\sum_{j \in [m]} \frac{1}{\sigma'} \mathrm{Tr}\left(\rho \frac{x_j E_j}{\varepsilon_j} \right)$ for at least a $1 - \beta^2$ fraction of points $x \in G_b^m$. Moreover, we can get a controlled version of W just by using a controlled implementation of V. Then Eq. (7.19) follows from Corollary 6.1 after multiplying its output coordinate-wise by $\sigma'\varepsilon_j$ and truncating to $[-1, 1]$. Similarly, Eq. (7.20) follows from Corollary 6.1 after incrementing the bias by $8\sigma'\varepsilon_j$ taking into account the truncation error.

The query complexity $\mathcal{O}\left(\left(\sigma' + \log(\frac{\log(m)}{\delta}) \right) \log(\frac{m}{\delta}) \right)$ for U_ρ directly follows from Corollary 6.1. The gate complexity of Corollary 6.1 is $\mathcal{O}\left((m + \sigma'(a_\rho + a_E + 1)) \mathrm{polylog}(\frac{\sigma' m}{\delta}) \right)$, which is supplemented by the complexity of implementing V times the above query complexity. The implementation of V uses $\mathcal{O}\left(\left\lceil \frac{\nu}{\sigma'} \log(\frac{\nu}{\beta}) \right\rceil \right) = \mathcal{O}\left(\frac{\nu}{\sigma'} \log(\frac{\nu \log(m)}{\delta}) \right)$ calls to U_E, and $\mathcal{O}\big((a_E + m) \lceil \frac{\nu}{\sigma'} \rceil \mathrm{polylog}(\frac{\nu + m}{\sigma'\varepsilon}) \big) = \mathcal{O}\big((a_E + m) \frac{\nu}{\sigma'} \mathrm{polylog}(\frac{\nu m}{\delta}) \big)$ additional two-qubit gates having depth $\mathcal{O}\big(\frac{\nu}{\sigma'} \mathrm{polylog}(\frac{\nu m}{\delta}) \big)$. This amounts to a total of $\mathcal{O}\left(\nu \log(\frac{\nu \log(m)}{\delta}) \log(\frac{m}{\delta}) + \frac{\nu}{\sigma'} \log(\frac{\nu \log(m)}{\delta}) \log(\frac{\log(m)}{\delta}) \log(\frac{m}{\delta}) \right)$ calls to U_E. ∎

Note that the assumption $\|E_j\| \leq 1$ is not particularly restrictive, because if $\|E_j\| > 1$ the corresponding block-encoding is subnormalized and we simply need to increase the precision by an amount equal to the subnormalization factor. If all ε_j are equal and we use the assumption $\|E_j\| \leq 1$, we recover the sample complexity $\widetilde{\mathcal{O}}(\sqrt{m}/\varepsilon)$ of the algorithm in [31]. The number of calls to U_E is not directly comparable because we use a different input model: in [31] the algorithm assumes access to to $e^{-i\theta E_j}$ and requires $\widetilde{\mathcal{O}}(\sqrt{m}/\varepsilon)$ calls to each of these operators for $j = 1, \ldots, m$, while we give a version that uses $\widetilde{\mathcal{O}}(m/\varepsilon)$ calls in total to controlled unitaries U_{E_j} block-encoding E_j.

Furthermore, Theorem 7.3 also recovers the query complexity results of the probability distribution estimation problem from [4], by taking $E_j = |j\rangle\langle j|$, for $j \in [d]$, and observing that $\sum_{j=1}^d E_j^2 = I$. Thus, even though [4] and [31] seem to be of different flavor, this result unifies both into a single construction.

For convenience, we state a version of our result only in terms of the number of observables rather than the more involved quantity $\left\| \sum_{j=1}^d E_j^2 \right\|$.

Corollary 7.1. *Let $E_1, \ldots, E_m \in \mathbb{C}^{d \times d}$ be Hermitian matrices with $\|E_j\| \leq 1$, and let $U_E = \sum_{j \in [m]} |j\rangle\langle j| \otimes U_{E_j}$, where U_{E_j} is an a_E-block-encoding of E_j, and let $\delta, \varepsilon \in (0, \frac{1}{6}]$. Let U_ρ be an a_ρ-qubit state-preparation unitary for a purification of $\rho \in \mathbb{C}^{d \times d}$. There is a quantum algorithm that makes $\mathcal{O}\left(\left(\frac{\sqrt{m \log(\frac{d}{\delta})}}{\varepsilon} + \log(\frac{\log(m)}{\delta}) \right) \log(\frac{m}{\delta}) \right)$ queries to U_ρ and $U_\rho^\dagger$, and produces estimates $z \in [-1, 1]^m$ such that, with probability at least $1 - \delta$,*

$$\forall j \in [m] \colon |\mathrm{Tr}(\rho E_j) - z_j| \leq \varepsilon, \tag{7.21}$$

moreover

$$\forall j \in [m]\colon |\mathrm{Tr}(\rho E_j) - \mathbb{E}[z_j]| \leq 16\sqrt{2m\log\left(\frac{2d}{\delta}\right)}\delta. \tag{7.22}$$

Furthermore, the quantum algorithm can be implemented by a number of calls to U_E *bounded by* $\mathcal{O}\big(\frac{m}{\varepsilon}\log(\frac{m}{\varepsilon\delta})\log(\frac{m}{\delta}) + \sqrt{m}\log(\frac{m}{\varepsilon\delta})\log(\frac{m}{\delta})\log(\frac{1}{\delta})\big)$ *and* $\mathcal{O}\left(\left(\frac{\sqrt{m\log(\frac{d}{\delta})}}{\varepsilon}a_\rho + \frac{m}{\varepsilon}a_E + \frac{m^2}{\varepsilon}\right)\mathrm{polylog}(\frac{m}{\varepsilon\delta})\right)$ *additional two-qubit gates while having circuit depth* $\mathcal{O}\big(\frac{m}{\varepsilon}\mathrm{polylog}(\frac{m}{\varepsilon\delta})\big)$.

Finally, there is a random variable $z' \in \bigtimes_{j\in[m]}[\mathrm{Tr}(\rho E_j) - \varepsilon, \mathrm{Tr}(\rho E_j) + \varepsilon]$ *with independent coordinates that is* δ*-close in total variation distance to* z *and also satisfies Eq.* (7.22).

8 Mixed-state tomography

In this section we generalize our pure-state results to mixed states. Throughout this section we use r to denote the rank of the mixed state. As discussed in the introduction, results from the literature on mixed-state tomography consider the case where only copies of the mixed state are available. Gross et al. [23] give an algorithm that uses $\mathcal{O}\big(d^2r^2/\varepsilon^2\big)$ samples; a tighter analysis of their algorithm shows that $\mathcal{O}\big(dr^2/\varepsilon^2\big)$ suffice to get an ε-trace-norm estimate, when measurements are performed on single (i.e., unentangled) copies of the state [25], and [11] shows that this is optimal even for adaptive (but still unentangled) measurements. O'Donnell and Wright [42] and Haah et al. [25] further improve the sample complexity to $\widetilde{\mathcal{O}}\big(dr/\varepsilon^2\big)$, at the cost of requiring joint measurements on many states at once, and with an algorithm that has super-polynomial time complexity. [25] also shows matching lower bounds for both settings, up to polylogarithmic factors, so these complexities are essentially optimal.

We consider a stronger input model where we are given access to a purification of a mixed state. So we assume access to a unitary $U_\rho\colon |00\rangle_{PS} \mapsto |\rho\rangle_{PS}$ such that $\rho = \mathrm{Tr}_P(|\rho\rangle\langle\rho|_{PS})$. Note that there are many possible purifications for the same mixed state.

The simplest idea to use a purification for tomography is to apply our pure-state algorithms directly to the purification, and then post-process by tracing out the unwanted part. The following lemma relates the error in a pure-state estimate to that in the resulting mixed-state estimate.

LEMMA 8.1. *If* $|\rho\rangle, |\tilde{\rho}\rangle \in \mathcal{H}_P \otimes \mathcal{H}_S \simeq \mathbb{C}^{a\cdot d}$*, then* $\frac{1}{2}\|\mathrm{Tr}_P(|\rho\rangle\langle\rho|) - \mathrm{Tr}_P(|\tilde{\rho}\rangle\langle\tilde{\rho}|)\|_1 \leq \sqrt{a\cdot d}\||\rho\rangle - |\tilde{\rho}\rangle\|_\infty$.

Proof. For any matrix M, the relationship $\|\mathrm{Tr}_P(M)\|_1 \leq \|M\|_1$ holds, see, e.g., [43] for a proof. Applying this to our pure state, and using the linearity of (partial) trace, we find:

$$\frac{1}{2}\|\mathrm{Tr}_P(|\rho\rangle\langle\rho|) - \mathrm{Tr}_P(|\tilde{\rho}\rangle\langle\tilde{\rho}|)\|_1 = \frac{1}{2}\|\mathrm{Tr}_P(|\rho\rangle\langle\rho| - |\tilde{\rho}\rangle\langle\tilde{\rho}|)\|_1 \leq \frac{1}{2}\||\rho\rangle\langle\rho| - |\tilde{\rho}\rangle\langle\tilde{\rho}|\|_1.$$

Using the well-known fact that for pure states $\frac{1}{2}\||\rho\rangle\langle\rho| - |\phi\rangle\langle\phi|\|_1 = \sqrt{1 - |\langle\rho|\phi\rangle|^2}$, we can conclude

$$\begin{aligned}
\frac{1}{2}\||\rho\rangle\langle\rho| - |\tilde{\rho}\rangle\langle\tilde{\rho}|\|_1 &= \sqrt{1 - |\langle\rho|\tilde{\rho}\rangle|^2} \\
&= \sqrt{1 + |\langle\rho|\tilde{\rho}\rangle|}\sqrt{1 - |\langle\rho|\tilde{\rho}\rangle|} \\
&\leq \sqrt{2}\sqrt{1 - \Re\langle\rho|\tilde{\rho}\rangle} \\
&= \||\rho\rangle - |\tilde{\rho}\rangle\|_2 \\
&\leq \sqrt{a\cdot d}\||\rho\rangle - |\tilde{\rho}\rangle\|_\infty.
\end{aligned}$$

□

If the purifying register has dimension a then our sampling and phase estimation algorithms would get a sample complexity of $\widetilde{\mathcal{O}}\big(\frac{da}{\varepsilon^2}\big)$ and query complexity of $\widetilde{\mathcal{O}}\big(\frac{da}{\varepsilon}\big)$ respectively to obtain an ε trace-norm estimate. As a can be as small as r, in certain settings this might lead to interesting results, but in general a can be arbitrarily large.

In the rest of this section we describe a tomography algorithm with sample complexity $\widetilde{\mathcal{O}}\big(\frac{dr}{\varepsilon}\big)$, for trace norm error ε, when a unitary (and its inverse) preparing a purification of ρ is available. Note that here we do not even require access to controlled versions of the input unitaries.

8.1 Coordinate-wise unbiased tomography Applying Theorem 7.3 to the set of observables $E_{ij}^{(0)} := \frac{|i\rangle\langle j|+|j\rangle\langle i|}{2}$ and $E_{ij}^{(1)} := \frac{|i\rangle\langle j|-|j\rangle\langle i|}{2\mathbf{i}}$ we get the next result.

THEOREM 8.1. *Let $\varepsilon, \delta \in (0, \frac{1}{6}]$, and let U_ρ be an a-qubit state-preparation unitary for a purification of $\rho \in \mathbb{C}^{d\times d}$. There is a quantum algorithm that makes $\mathcal{O}\left(\left(\frac{\sqrt{d\log(\frac{d}{\delta})}}{\varepsilon} + \log(\frac{\log(d)}{\delta})\right)\log(\frac{d}{\delta})\right)$ queries to U_ρ and $U_\rho^\dagger$, and produces estimates $z \in ([-1,1]\times[-\mathbf{i},\mathbf{i}])^{d^2}$ such that, with probability at least $1-\delta$,*

$$\forall i,j \in [d]\colon |\rho_{ij} - z_{ij}| \le \sqrt{2}\varepsilon, \tag{8.23}$$

moreover

$$\forall i,j \in [d]\colon |\rho_{ij} - \mathbb{E}[z_{ij}]| \le 16\sqrt{2}\sqrt{2d\ln\left(\frac{2d}{\delta}\right)}\delta \le 32\sqrt{d\ln(2d)\delta}. \tag{8.24}$$

Furthermore, the quantum algorithm can be implemented by $\mathcal{O}\left(\frac{\sqrt{d}a+d^3}{\varepsilon}\mathrm{polylog}(\frac{d}{\delta})\right)$ additional two-qubit gates having depth $\mathcal{O}\left(\frac{d}{\varepsilon}\mathrm{polylog}(\frac{d}{\delta})\right)$.

Finally, there is a random variable $z' \in \bigtimes_{i,j\in[d]}([\rho_{ij}-\varepsilon, \rho_{ij}+\varepsilon]\times[\rho_{ij}-\mathbf{i}\varepsilon, \rho_{ij}+\mathbf{i}\varepsilon])$ with independent coordinates that is δ-close in total variation distance to z and also satisfies Eq. (8.24).

Proof. The result follows from Theorem 7.3 by observing that $\sum_{p\in\{0,1\};i,j\in[d]}(E_{ij}^{(p)})^2 = dI$, and a block-encoding of $\sum_{p\in\{0,1\};i,j\in[d]}|pij\rangle\langle pij| \otimes E_{ij}^{(p)}$ can be implemented by $\mathcal{O}(\mathrm{polylog}(d))$ two-qubit gates. We define our estimate as $z_{i,j} := z_{i,j}^{(0)} + \mathbf{i}z_{i,j}^{(1)}$.

The second inequality in Eq. (8.24) follows from the following short computation:

$$\begin{aligned}
\sqrt{\ln\left(\frac{2d}{\delta}\right)}\delta &\le \sqrt{\ln(2d)\delta} \\
&\Updownarrow \\
\ln\left(\frac{2d}{\delta}\right)\delta &\le \ln(2d) \\
&\Uparrow (\delta \le \frac{1}{6}) \\
\ln\left(\frac{1}{\delta}\right)\delta &\le \frac{5}{6}\ln(2d) \\
&\Uparrow (d \ge 1) \\
\ln\left(\frac{1}{\delta}\right)\delta &\le \frac{1}{2}.
\end{aligned}$$

The gate complexities follow by replacing Lemma 7.1 in the proof or Theorem 7.3 by "sparse block-encoding" [19, Lemma 47-48]. This results in reducing the $\Theta(d^2)$ subnormalization factor coming from the generic result of Lemma 7.1 by a factor of d coming from [19, Lemma 47-48]. This improves the gate complexities by about a factor d.[5] ∎

8.2 Matrix norm conversions We now consider the relation between the element-wise max-norm, and the operator norm. To do so we use the following definition and lemma due to [44]:

[5]There is a possibility that using the block-encoding of [38] even an improvement factor by about $d^{1.5}$ is possible, but one needs to be careful with the error bounds, since they are not polylogarithmic in [38].

Definition 8.1. (Subgaussian random variable [44, Definition 2.2]) *A random variable X is* subgaussian *if there exists a $K > 0$, called the subgaussian moment of X, such that*

$$\mathbb{P}(|X| > t) \leq 2e^{-t^2/K^2} \text{ for all } t > 0.$$

Note that a bounded random variable $X \in [-B, B]$ has subgaussian moment $\leq \sqrt{\ln(2)}B$.

Lemma 8.2. (Operator norm of subgaussian matrices [44, Proposition 2.4]) *Let X be an $N \times n$ random matrix whose entries are independent mean zero subgaussian random variables whose subgaussian moments are bounded by K. Then*

$$\mathbb{P}\left(\frac{\|X\|}{K} > C(\sqrt{N} + \sqrt{n}) + t\right) \leq 2e^{-ct^2}, t \geq 0,$$

where C and c denote positive absolute constants.

The above lemma shows that if we can estimate all entries of a state ρ independently and in an unbiased way, then in the conversion to the operator norm error we save an essentially $\sqrt{d}$ factor compared to the worst case: this was one of the main motivation for us to develop unbiased phase estimation in Section 6. We formalize this in the following lemma:

Lemma 8.3. *Let $X \in \mathbb{C}^{d\times d}$ be a matrix, and let $\tilde{X}$ be an ε-approximation of X in the entry-wise* max*-norm, i.e., $|X_{ij} - \tilde{X}_{ij}| \leq \varepsilon$ for each $i, j \in [d]$. Then for the operator norm error we have $\left\|X - \tilde{X}\right\| \leq \varepsilon d$. Also, there are absolute constants $C', c' > 0$ such that, if all entries of $\tilde{X}$ are drawn from independent distributions and $\mathbb{E}[\tilde{X}] = X$, then for every $\tau \geq 1$ we have that $\left\|X - \tilde{X}\right\| \leq C'\sqrt{d}\tau\varepsilon$ with probability at least $1 - 2e^{-c'd\tau^2}$.*

Proof. Let $E = \tilde{X} - X$ be the matrix of errors. For the first statement we note that

$$\|E\| \leq \|E\|_2 \leq d\|E\|_{\max} \leq \varepsilon d,$$

where the first inequality follows from the relation between the operator and the Frobenius norms, and the second follows from a standard norm conversion on the d^2-dimensional vector of entries (we use $\|\cdot\|_{\max}$ for the entry-wise max-norm).

For the second statement we apply Lemma 8.2 to E, noting that each matrix element has subgaussian moment $\leq \sqrt{\ln(2)}\varepsilon$, thus implying that for every $t \geq 0$ we have

$$\mathbb{P}\Big(\|E\| > \sqrt{\ln(2)}\varepsilon\Big(2C\sqrt{d} + t\Big)\Big) \leq 2e^{-ct^2}.$$

We conclude by setting $t := 2C\sqrt{d}\tau$ showing that

$$\mathbb{P}\Big(\|E\| > (2 + 2\tau)\sqrt{\ln(2)}C\sqrt{d}\varepsilon\Big) \leq \mathbb{P}\Big(\|E\| > 4\sqrt{\ln(2)}C\sqrt{d}\tau\varepsilon\Big) \leq 2e^{-4cC^2d\tau^2},$$

so that we can choose $C' := 4\sqrt{\ln(2)}C$ and $c' := 4cC^2$. ∎

When considering estimates of mixed quantum states we mostly consider Schatten q-norms for error bounds (that is, the q-norm of the vector of singular values of the difference between the actual state and our estimate). The most common values for q are $q = \infty$ (operator norm), $q = 2$ (Frobenius norm), and $q = 1$ (trace norm). Using a tiny modification of our norm conversion result, Corollary 3.1, we can obtain the following as a corollary.

Corollary 8.1. *Let $\varepsilon \in (0, 1]$, $1 \leq q$, and let $\rho \in \mathbb{C}^{d\times d}$ be a rank-r quantum state. In order to obtain an ε-Schatten-q-norm estimate of ρ, an η-operator norm estimate suffices, with*

$$\eta = \max\left\{\left(\frac{\varepsilon}{10}\right)^{\frac{1}{1-\frac{1}{q}}}, \frac{\varepsilon}{2(2r)^{\frac{1}{q}}}\right\}.$$

Proof. Let $\tilde{\rho}$ be an η-operator norm estimate of ρ. We can assume without loss of generality that $\tilde{\rho}$ is Hermitian (otherwise take $\frac{\tilde{\rho}+\tilde{\rho}^\dagger}{2}$). First, since $\|\tilde{\rho}-\rho\| \leq \eta$, there must exist a density matrix σ such that $\sigma \succeq 0$, $\|\sigma\|_1 \leq 1$, σ is of rank at most r, and $\|\tilde{\rho}-\sigma\| \leq \eta$, because after all ρ is an example of such a density matrix σ. Let $\tilde{\rho}'$ be any such σ.[6] Then, by the triangle inequality we obtain that $\|\tilde{\rho}'-\rho\| \leq \|\tilde{\rho}'-\tilde{\rho}\| + \|\tilde{\rho}-\rho\| \leq \eta+\eta = 2\eta$.

Thus, we find that ρ and $\tilde{\rho}'$ are both of rank at most r, and therefore by the subadditivity of rank, $\rho-\tilde{\rho}'$ is of rank at most $2r$. This implies by Hölder's inequality that $\|\rho-\tilde{\rho}'\|_q \leq (2r)^{\frac{1}{q}} \cdot 2\eta$.

On the other hand, from the norm conversion lemma, Lemma 3.2, there exist operators $\tilde{\rho}'_{\geq 2\eta}$ and $\rho_{\geq 2\eta}$ such that they are both $\min\{4\eta^{(q-s)/q}, 3r^{1/q}\eta\}$-close to their originals in Schatten-q-norm and both have rank at most $1/(2\eta)$. Then, we obtain by the triangle inequality that

$$\|\tilde{\rho}'-\rho\|_q \leq \left\|\tilde{\rho}'-\tilde{\rho}'_{\geq 2\eta}\right\|_q + \left\|\tilde{\rho}'_{\geq 2\eta}-\rho_{\geq 2\eta}\right\|_q + \|\rho_{\geq 2\eta}-\rho\|_q \leq 2\cdot 4\eta^{\frac{q-1}{q}} + \left(\frac{1}{\eta}\right)^{\frac{1}{q}} 2\eta = 10\eta^{\frac{q-1}{q}}.$$

Combining both results yields $\|\tilde{\rho}'-\rho\|_q \leq \min\{10\eta^{(q-1)/q}, (2r)^{\frac{1}{q}}2\eta\} = \varepsilon$. ∎

8.3 Generic mixed-state tomography We now have all the necessary tools to construct a tomography algorithm with $\widetilde{\mathcal{O}}(dr/\varepsilon)$ sample complexity.

THEOREM 8.2. *Let $\varepsilon, \delta \in (0, \frac{1}{3}]$, $q \in [1,\infty]$, and let U_ρ be an a-qubit state-preparation unitary for a purification of $\rho \in \mathbb{C}^{d\times d}$. There is a quantum algorithm that makes $\mathcal{O}\left(\frac{d}{\varepsilon} r^{\frac{1}{q}} \log^{\frac{3}{2}}\left(\frac{d}{\delta\varepsilon}\right)\sqrt{\left\lceil \frac{\log(1/\delta)}{d}\right\rceil}\right)$ queries to U_ρ and $U_\rho^\dagger$, and outputs a positive semidefinite $\tilde{\rho}'$ such that with probability at least $1-\delta$ we have $\|\rho-\tilde{\rho}'\|_q \leq \varepsilon$. The quantum algorithm can be implemented by $\mathcal{O}\left(\frac{da+d^{3.5}}{\varepsilon} r^{\frac{1}{q}} \operatorname{polylog}(\frac{d}{\delta\varepsilon})\right)$ additional two-qubit gates having depth $\mathcal{O}\left(\frac{d^{1.5}}{\varepsilon} r^{\frac{1}{q}} \operatorname{polylog}(\frac{d}{\delta\varepsilon})\right)$.*

Proof. First we prove the statement for operator norm by combining Theorem 8.1 and Lemma 8.3. We set $\delta' := \min\{\frac{\delta}{2}, \frac{\varepsilon^2}{2^{11}d^3\ln(2d)}\}$, $\tau := \sqrt{1+\frac{\ln(4/\delta)}{dc'}}$, and $\varepsilon' := \min\left\{\frac{\varepsilon}{4C'\sqrt{d}\tau}, \frac{1}{6}\right\}$ and invoke Theorem 8.1 providing us an estimate $\tilde{\rho} \in \mathbb{C}^{d^2}$ that is $\frac{\delta}{2}$-close in total variation distance to a random variable $\rho' \in \bigtimes_{i,j\in[d]}([\rho_{ij}-\varepsilon, \rho_{ij}+\varepsilon] \times [\rho_{ij}-\mathbf{i}\varepsilon, \rho_{ij}+\mathbf{i}\varepsilon])$. Due to (8.24) we have that $\mathbb{E}[\rho'_{ij}]-\rho_{ij} \leq \frac{\varepsilon}{2d}$ and so by Lemma 8.3 we have $\|\mathbb{E}[\rho']-\rho\| \leq \frac{\varepsilon}{2}$. Using our choice of τ Lemma 8.3 also implies that $\|\mathbb{E}[\rho']-\rho'\| \leq \frac{\varepsilon}{2}$ with probability at least $1-\frac{\delta}{2}$. By the triangle inequality we get that $\|\rho-\rho'\| \leq \varepsilon$ with probability at least $1-\frac{\delta}{2}$. Since ρ' and $\tilde{\rho}$ are $\frac{\delta}{2}$-close in total variation distance this also implies that $\|\rho-\tilde{\rho}\| \leq \varepsilon$ with probability at least $1-\delta$.

As per Theorem 8.1 the algorithm makes $\mathcal{O}\left(\left(\frac{d\sqrt{\log\left(\frac{d}{\delta\varepsilon}\right)\left(1+\frac{\log(1/\delta)}{d}\right)}}{\varepsilon} + \log\left(\frac{d}{\delta\varepsilon}\right)\right)\log\left(\frac{d}{\delta\varepsilon}\right)\right)$ queries to U_ρ and $U_\rho^\dagger$, and can be implemented by $\mathcal{O}\left(\frac{da+d^{3.5}}{\varepsilon}\operatorname{polylog}(\frac{d}{\delta\varepsilon})\right)$ additional two-qubit gates having depth $\mathcal{O}\left(\frac{d^{1.5}}{\varepsilon}\operatorname{polylog}(\frac{d}{\delta\varepsilon})\right)$.

In order to get a positive semidefinite $\tilde{\rho}'$ and to get the results for all Schatten-norms we apply Lemma 8.1 to $\frac{\tilde{\rho}+\tilde{\rho}^\dagger}{2}$ and adjust the value of ε accordingly. ∎

9 Lower bounds

In this section we prove lower bounds for state tomography in several different access models. The first model we consider, in Section 9.1, is the case in which we have access to conditional copies of the state, i.e., we receive states of the form $(|0\rangle|\psi\rangle + |1\rangle|0\rangle)/\sqrt{2}$. From here, we derive matching lower bounds for all the algorithms constructed in Section 4. In the second model, considered in Section 9.2, we assume to have access to a state-preparation unitary and its inverse. The lower bounds derived in this subsection match the complexities of the algorithm constructed in Section 5, up to logarithmic factors. Finally, in Section 9.3, we consider the setting where we access

[6] Removing all negative eigenvalues of $\tilde{\rho}-\eta I$ produces such a matrix $\tilde{\rho}'$. Clearly, $\tilde{\rho}-\rho \preceq \eta I$ and so $\tilde{\rho}-\eta I \preceq \rho$, implying that the n-th largest eigenvalue of ρ majorates that of $\tilde{\rho}-\eta I$ and consequently also that of $\tilde{\rho}'$. This then implies that the rank of $\tilde{\rho}'$ is at most r and that $\|\tilde{\rho}'\|_1 \leq \|\rho\|_1 = 1$; $\|\tilde{\rho}'-\tilde{\rho}\| \leq \eta$ can be shown similarly.

to a unitary constructing a purification of a density matrix that we wish to estimate. We derive lower bounds that match the complexities of the algorithms constructed in Section 8.

Note that van Apeldoorn [4] gives a very similar lower bound result in the pure-state setting where we have access to a state-preparing unitary. However, in the setting of van Apeldoorn the unitary prepares a state of the form $\sum_j \sqrt{p_j}|j\rangle|\phi_j\rangle$, whereas in this paper the state is of the form $\sum_j \sqrt{p_j}|j\rangle$, i.e., without the additional states entangled with $|j\rangle$. Hence our input model is stricter and requires its own lower bound.

The general proof strategy in Section 9.1 and Section 9.2 is very similar – we start by proving a lower bound on estimating probability distributions in the ℓ_1-norm, then use a sequence of reductions to obtain lower bounds on quantum pure-state tomography in any ℓ_q-norm with $q \geq 2$. Since the reductions we use are identical in both cases, we start by presenting it here, and then focus in Section 9.1 and Section 9.2 on proving the lower bound for probability distribution reconstruction in ℓ_1-norm separately for both input models afterwards.

LEMMA 9.1. *Let $0 < \varepsilon \leq 1$, $s > 0$, $q \in [1, \infty]$, and suppose that in order to produce an ε-ℓ_1-estimate of any probability distribution $p \in \Delta_d$ with high probability, one needs to perform at least $\Omega(d/\varepsilon^s)$ queries. Then*

$$\Omega\left(\min\left\{\frac{1}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}, \frac{d^{1-s+\frac{s}{q}}}{\varepsilon^s}\right\}\right)$$

queries are necessary to find a ε-ℓ_q-estimate of p with high probability.

Proof. Suppose that $q > 1$ and $\varepsilon \leq 1/d^{1-\frac{1}{q}}$. Then, by Hölder's inequality,

$$\|\widetilde{p} - p\|_1 \leq d^{1-\frac{1}{q}} \|\widetilde{p} - p\|_q \leq d^{1-\frac{1}{q}} \varepsilon \leq 1,$$

and the number of queries that is required scales as

$$\Omega\left(\frac{d}{\left(d^{1-\frac{1}{q}}\varepsilon\right)^s}\right) = \Omega\left(\frac{d^{1-s+\frac{s}{q}}}{\varepsilon^s}\right).$$

This leaves the case where $q > 1$ and $1/d^{1-\frac{1}{q}} < \varepsilon \leq 1$. This immediately implies that

$$1 \leq \frac{1}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}} < d^{\frac{1-\frac{1}{q}}{1-\frac{1}{q}}} = d.$$

Thus, we can choose

$$d' = \left\lfloor \frac{1}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}} \right\rfloor,$$

and observe that it is an integer between 1 and d. It follows that

$$(d')^{1-\frac{1}{q}} \varepsilon \leq \left(\frac{1}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}\right)^{1-\frac{1}{q}} \cdot \varepsilon = 1.$$

Note that we can embed any d'-dimensional probability distribution $p' \in \Delta_{d'}$ into the first d' coordinates of $p \in \Delta_d$. Moreover, any ε-ℓ_q-estimate $\tilde{p}$ of p naturally leads to an approximation $\tilde{p}'$ to p' by only considering the first d' entries of $\tilde{p}$. Using Hölder's inequality, we find that

$$\|\tilde{p}' - p'\|_1 \leq (d')^{1-\frac{1}{q}} \|\tilde{p}' - p'\|_q \leq (d')^{1-\frac{1}{q}} \|\tilde{p} - p\|_q \leq (d')^{1-\frac{1}{q}} \varepsilon \leq 1,$$

and hence the number of queries in order to find an ε-ℓ_q-estimate of p scales at least as

$$\Omega\left(\frac{d'}{\left((d')^{1-\frac{1}{q}}\varepsilon\right)^s}\right) = \Omega\left(\frac{(d')^{1-s+\frac{s}{q}}}{\varepsilon^s}\right) = \Omega\left(\frac{1}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}\right).$$

Since this expression is indeed smaller than $d^{1/q}/\varepsilon$ precisely when $\varepsilon > 1/d^{1-1/q}$, we find that the lower bound becomes

$$\Omega\left(\min\left\{\frac{1}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}, \frac{d^{1-s+\frac{s}{q}}}{\varepsilon^s}\right\}\right),$$

as claimed. ∎

We know from the first norm-conversion lemma, i.e., Lemma 3.1, that obtaining an estimate of the amplitudes of a quantum state also gives one an estimate of the probability distribution arising from their absolute values squared. Therefore, our reduction from the previous lemma can be extended to give lower bounds on the problem of estimating a quantum state as well.

LEMMA 9.2. *Let $0 < \varepsilon \leq 1$, $q, s \in [1, \infty]$, and suppose that in order to produce an ε-ℓ_1-estimate of any probability distribution $p \in \Delta_d$, defined as $p_j = |\alpha_j|^2$ with $\alpha \in \mathbb{C}^d$, with high probability, one needs to perform at least $\Omega(d/\varepsilon^s)$ queries. Then*

$$\Omega\left(\min\left\{\frac{1}{\varepsilon^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}}, \frac{d^{1-\frac{s}{2}+\frac{s}{q}}}{\varepsilon^s}\right\}\right)$$

queries are necessary to find a ε-ℓ_q-estimate of α with high probability.

Proof. By Lemma 3.1, we know that an ε-ℓ_q-norm estimate of α immediately gives an 4ε-ℓ_r-estimate of p, with $r = 1/(1/q + 1/2)$. Therefore, the result simply follows by substituting r for q into the bounds displayed in Lemma 9.1. ∎

9.1 Lower bound on conditional samples To lower bound the use of conditional samples we will use a prove based on communication complexity. In particular, we give an ensemble of states corresponding to conditional samples, such that an ℓ_1-norm estimate of any of the states would give $\Omega(d)$ bits of information about which state was given, but a copy of the state can only communicate $\widetilde{\mathcal{O}}(\varepsilon^2)$ bits of information.

LEMMA 9.3. *Let $\varepsilon > 0, d \geq 12$ with $d \in \mathbb{N}$. There exists a set of $2^{d/2}$ probability distributions $\{p^{(b)}\}_{b\in\{0,1\}^{d/2}} \in \Delta^d$ indexed by length $d/2$ bit strings, such that for all b and b' where $\left\|p^{(b)} - p^{(b')}\right\|_1 < 2\varepsilon$ we have that b and b' differ on at most $d/6$ bits. Furthermore, let $|\psi_j\rangle = \sum_i \sqrt{p_i^{(j)}}|i\rangle$ and $|\phi_j\rangle = \frac{|0\rangle|\psi_j\rangle+|1\rangle|0\rangle}{\sqrt{2}}$; then, denoting the entropy by S, we have:*

$$S\left(\frac{1}{2^{d/2}}\sum_j |\phi_j\rangle\langle\phi_j|\right) \leq 27\varepsilon^2(1 + \log(d/\varepsilon^2)).$$

Proof. We index the family of probability distributions with bit strings $b \in \{0,1\}^{d/2}$, writing $p^{(b)}$ for the distribution corresponding to string b. Each distribution is over $[d/2] \times \{0,1\}$ and is defined as

$$p_{i,c}^{(b)} = \frac{1 + (-1)^{b_i \oplus c} 6\varepsilon}{d}.$$

In other words, it consists of $d/2$ pairs of entries that correspond to the bits of b, where the bit determines which of the entries in the pair is increased by $6\varepsilon/d$ and which is decreased.

For two bit strings b and b' with Hamming distance $|b \oplus b'|$, the corresponding distributions will be $|b \oplus b'|\frac{12\varepsilon}{d}$ apart in ℓ_1-norm. Hence, if two distributions are less than 2ε apart, then for their bit strings we get $|b \oplus b'| < d/6$, i.e., less than a $1/3$ fraction of the positions differ.

It remains to upper bound the entropy of a uniform mixture of conditional samples. For ease of notation we will write

$$c_\pm = \sqrt{1 \pm 6\varepsilon}$$

and note that $(c_+ + c_-)^2 = 2 + 2c_+c_-$. In our notation, we have:

$$|\phi_b\rangle\langle\phi_b| = \frac{1}{2}\begin{pmatrix} |\psi_b\rangle\langle\psi_b| & |\psi_b\rangle \\ \langle\psi_b| & 1 \end{pmatrix},$$

where for simplicity and without impacting subsequent calculations we have dropped the the $d-1$ all-zero columns on the right and the $d-1$ all-zero rows at the bottom. Let

$$\sigma = \frac{1}{2^{d/2}} \sum_b |\phi_b\rangle\langle\phi_b| = \frac{1}{2^{d/2+1}} \begin{pmatrix} \sum_b |\psi_b\rangle\langle\psi_b| & \sum_b |\psi_b\rangle \\ \sum_b \langle\psi_b| & 1 \end{pmatrix}$$

Considering a single entry of $\sum_b |\psi_b\rangle$, exactly half of the terms will be $\frac{c_+}{\sqrt{d}}$ and half will be $\frac{c_-}{\sqrt{d}}$. So $\frac{1}{2^{d/2+1}} \sum_b |\psi_b\rangle = \frac{c_+ + c_-}{4\sqrt{d}} \vec{1}$ and similar for the row vectors.

We now analyze the term $|\psi_b\rangle\langle\psi_b|$. Consider the 2×2 block of the matrix corresponding to b_i for the rows and b_k for the columns. Depending on the values of those two bits, this block can take four different forms:

1. If $b_i = b_k = 0$ then the block is
$$\frac{1}{d}\begin{pmatrix} c_+^2 & c_+c_- \\ c_+c_- & c_-^2 \end{pmatrix}$$
2. If $b_i = b_k = 1$ then the block is
$$\frac{1}{d}\begin{pmatrix} c_-^2 & c_+c_- \\ c_-c_+ & c_+^2 \end{pmatrix}$$
3. If $b_i = 0, b_k = 1$ then the block is
$$\frac{1}{d}\begin{pmatrix} c_-c_+ & c_+^2 \\ c_-^2 & c_+c_- \end{pmatrix}$$
4. If $b_i = 1, b_k = 0$ then the block is
$$\frac{1}{d}\begin{pmatrix} c_-c_+ & c_-^2 \\ c_+^2 & c_+c_- \end{pmatrix}$$

If $i = k$, i.e., on the diagonal, only (1) and (2) can happen, and by averaging over all possible b (and putting back in the extra factor $1/2$ that appears in the denominator $1/(2^{d/2+1})$ in σ), we get:

$$\frac{1}{2d}\begin{pmatrix} 1 & c_+c_- \\ c_+c_- & 1 \end{pmatrix}.$$

Off-diagonally we average over all 4 possibilities, and get:

$$\frac{1 + c_+c_-}{4d}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Denoting by J the all-ones matrix and by $D := \bigoplus_{i\in[d/2]} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$, the top left block of σ can be written as

$$\frac{1 + c_+c_-}{4d} J + \frac{1 - c_+c_-}{4d} D.$$

So

$$\sigma = \begin{pmatrix} \frac{1+c_+c_-}{4d} J & \frac{c_+ + c_-}{4\sqrt{d}} \vec{1} \\ \frac{c_+ + c_-}{4\sqrt{d}} \vec{1}^\top & 1/2 \end{pmatrix} + \begin{pmatrix} \frac{1-c_+c_-}{4d} D & 0 \\ 0 & 0 \end{pmatrix}.$$

The first term in the above equation is a rank-1 matrix, as it is equal to the outer product of the column vector $(\frac{c_+ + c_-}{2\sqrt{2d}} \vec{1}, 1/\sqrt{2})$ with itself (recall that $(c_+ + c_-)^2/2 = 1 + c_+c_-$). The corresponding eigenvalue is just the norm of this vector, and it is equal to $\frac{3+c_+c_-}{4}$. The second term has $d/2$ nonzero eigenvalues, all equal to $\frac{1-c_+c_-}{2d}$. As $(\frac{c_+ + c_-}{2\sqrt{2d}} \vec{1}, 1/\sqrt{2})$ is in the kernel of the second term, the $d/2+1$ eigenvalues listed above are in fact the eigenvalues of σ. So:

$$\begin{aligned} S(\sigma) &= -\frac{3 + c_+c_-}{4} \log\left(\frac{3 + c_+c_-}{4}\right) - \frac{d}{2}\frac{1 - c_+c_-}{2d} \log\left(\frac{1 - c_+c_-}{2d}\right) \\ &\le \log\left(\frac{4}{3 + c_+c_-}\right) + \frac{36\varepsilon^2}{4} \log\left(\frac{2d}{1 - c_+c_-}\right) \\ &\le \log\left(\frac{1}{1 - 9\varepsilon^2}\right) + 9\varepsilon^2\big(1 + \log(d/18\varepsilon^2)\big) \\ &\le 18\varepsilon^2 + 9\varepsilon^2\big(1 + \log(d/\varepsilon^2)\big) \\ &\le 27\varepsilon^2(1 + \log(d/\varepsilon^2)). \end{aligned}$$

In the chain of inequalities above, we used the fact that

$$1 - 36\varepsilon^2 \leq \sqrt{1 - 36\varepsilon^2} = c_+ c_- \leq 1 - 18\varepsilon^2$$

and hence

$$18\varepsilon^2 \leq 1 - c_+ c_- \leq 36\varepsilon^2,$$

together with the fact that the logarithm is monotonically increasing. ∎

With this entropy bound we are now ready to prove our sample complexity lower bound.

PROPOSITION 9.1. *Let $p \in \Delta^d$ be a probability distribution, and let $|\psi\rangle = \sum_j \sqrt{p_j}|j\rangle$. Then $\tilde{\Omega}(d/\varepsilon^2)$ copies of*

$$\frac{|0\rangle|\psi\rangle + |1\rangle|0\rangle}{\sqrt{2}}$$

are required to learn p up to ℓ_1-norm error ε.

Proof. We consider a communication scenario where Alice picks a $b \in \{0,1\}^{d/2}$ and encodes this in k copies of $|\phi_b\rangle$ from Lemma 9.3. She sends these copies to Bob. If Bob can estimate p_b up to ε-ℓ_1-norm using k copies then, by rounding to the closest distribution $p_{\tilde{b}}$, he can learn a $\tilde{b}$ that agrees with b on at least a $2/3$ fraction of the bits, and hence he has learned $\Omega(d)$ bits of information about Alice's string. By Holevo's Theorem we know that the maximum amount of information that can be communicated by an ensemble of pure states is upper bounded by its entropy. As the entropy of k copies of a state is equal to k times the entropy of a single state, we have

$$\Omega(d) \leq kS\left(\frac{1}{2^{d/2}}\sum_j |\phi_j\rangle\langle\phi_j|\right) \leq k36\varepsilon^2(1 + \log(d/\varepsilon^2))$$

and hence $k = \Omega(\frac{d}{\varepsilon^2 \log(d/\varepsilon^2)})$ copies are needed for an ℓ_1-norm estimate. ∎

We can now apply our norm conversion lemmas to obtain the following theorem.

THEOREM 9.1. *Let $|\psi\rangle = \sum_j \alpha_j |j\rangle$. Then*

$$\tilde{\Omega}\left(\min\left\{\left(\frac{3}{\varepsilon}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{q}}}, \frac{d^{\frac{2}{q}}}{\varepsilon}\right\}\right)$$

copies of

$$\frac{|0\rangle|\psi\rangle + |1\rangle|0\rangle}{\sqrt{2}}$$

are required to learn α up to ℓ_q-norm error ε.

Proof. We combine the ℓ_1-reduction Lemma 9.2 with $s = 2$, with Proposition 9.1. ∎

9.2 Lower bound with inverse state preparation We start with showing that if we want to obtain an ℓ_1-estimate of a probability distribution $p \in \Delta_d$, we need at least $\Omega(d/\varepsilon)$ queries to the operation that prepares it $U : |0\rangle \mapsto \sum_{j=1}^d \sqrt{p_j}|j\rangle$. We do this by reducing the problem to the problem of recovering a constant fraction of the bits in a bit string, which is known to have a lower bound on of $\Omega(d/\varepsilon)$.

LEMMA 9.4. *Let $0 < \varepsilon < 1/16$, $d \in \mathbb{N}$, $p \in \Delta^d$ a probability distribution, and let U be a unitary that prepares $\sum_j \sqrt{p_j}|j\rangle$. Then $\Omega\left(\frac{d}{\varepsilon}\right)$ applications of U and its inverse are necessary to find a ε-ℓ_1-estimate of p with high probability.*

Proof. Let $x \in \{0,1\}^d$ be a bit string, and suppose that we have controlled access to x by means of a fractional phase oracle, i.e., we can access a controlled oracle that acts on $\mathbb{C}^d$ as

$$O_x : |j\rangle \mapsto e^{4\pi i \varepsilon x_j}|j\rangle.$$

Recovering more than three quarters of the bits of x with high probability is known to require $\Omega(d/\varepsilon)$ queries to O_x.[7]

Now, we construct a specific probability distribution p, dependent on x, whose corresponding quantum state can be constructed using only one call to O_x, and that allows for recovering at least 3/4 of the bit string if it is estimated up to ε in ℓ_1-norm. To that end, suppose that we start in the state

$$\frac{1}{\sqrt{2d}} \sum_{j=1}^{d} |j\rangle \Big(e^{-\frac{\pi i}{4}}|0\rangle + e^{\frac{\pi i}{4}}|1\rangle \Big).$$

Now, we can apply $O_x^\dagger$ to the first register if the last qubit is in state $|0\rangle$, and O_x if the last qubit is in state $|1\rangle$. This results in the state

$$\frac{1}{\sqrt{2d}} \sum_{j=1}^{d} |j\rangle \Big(e^{-\frac{\pi i}{4} - 4\pi i \varepsilon x_j}|0\rangle + e^{\frac{\pi i}{4} + 4\pi i \varepsilon x_j}|1\rangle \Big).$$

Next, after applying a Hadamard gate to the final qubit, we obtain the state

$$\frac{1}{\sqrt{d}} \sum_{j=1}^{d} |j\rangle \Big(\cos\Big(\frac{\pi}{4} + 4\pi\varepsilon x_j\Big)|0\rangle - i \sin\Big(\frac{\pi}{4} + 4\pi\varepsilon x_j\Big)|1\rangle \Big),$$

which after applying an S-gate to the final qubit is turned into the state

$$|\psi\rangle = \sum_{j\in[d]} \sum_{b\in\{0,1\}} \sqrt{p_{j,b}}|j\rangle|b\rangle, \qquad \text{where} \qquad p_{j,b} = \begin{cases} \frac{\cos^2\left(\frac{\pi}{4} + 4\pi\varepsilon x_j\right)}{d}, & \text{if } b = 0, \\ \frac{\sin^2\left(\frac{\pi}{4} + 4\pi\varepsilon x_j\right)}{d}, & \text{if } b = 1. \end{cases}$$

Since $\varepsilon < 1/16$, we have

$$\Big|\cos^2\Big(\frac{\pi}{4}\Big) - \cos^2\Big(\frac{\pi}{4} + 4\pi\varepsilon\Big)\Big| = \Big|\sin^2\Big(\frac{\pi}{4}\Big) - \sin^2\Big(\frac{\pi}{4} + 4\pi\varepsilon\Big)\Big| = \frac{1}{2}\sin(8\pi\varepsilon) > 8\varepsilon.$$

Now suppose that we can find some estimate $\tilde{p}$ such that $\|\tilde{p} - p\|_1 \le \varepsilon$. Then, define the bit string $\tilde{x} \in \{0,1\}^d$ as:

$$\tilde{x}_j = \begin{cases} 1, & \text{if } \tilde{p}_{j,0} < \frac{1}{d}\big(\frac{1}{2} - 4\varepsilon\big), \\ 0, & \text{otherwise.} \end{cases}$$

It follows that $4\varepsilon|\tilde{x}_j - x_j|/d \le |\tilde{p}_{j,0} - p_{j,0}|$, and so the number of bits of $\tilde{x}$ that differ from those in x is at most $d/4$. Hence, finding an ε-ℓ_1-norm estimate of a $2d$-dimensional probability distribution must take at least $\Omega(d/\varepsilon)$ calls to a state-preparation oracle as well. ∎

It now remains to apply our ℓ_1-reduction to complete the lower bound for general ℓ_q-norms.

THEOREM 9.2. *Let $|\psi\rangle = \sum_{j\in[d]} \alpha_j |j\rangle$ be a quantum state with and let U be a unitary that prepares $|\psi\rangle$. Then*

$$\Omega\left(\min\left\{ \frac{1}{\varepsilon^{\frac{1}{2} - \frac{1}{q}}}, \frac{d^{\frac{1}{q} + \frac{1}{2}}}{\varepsilon} \right\} \right)$$

applications of U and its inverse are necessary to find an ε-ℓ_q-estimate of $|\alpha|$ for $q \in [2, \infty]$.

Proof. This follows from combining the ℓ_1-reduction Lemma 9.2 with $s = 1$, and Lemma 9.1. ∎

[7] Proving this is done in two steps – first one proves that this takes at least $\Omega(d)$ calls to a regular phase oracle to x, which can be easily shown using an information theoretic argument stemming from [14]. Next, this can be combined with Appendix B from [35] and the general adversary bound for relations from [7], to get to the desired bound of $\Omega(d/\varepsilon)$. We can also reduce the problem to recovering the bit string exactly, and then directly apply the phase adversary bound from [3, Ch. 6].

9.3 Lower bounds on mixed-state tomography In this section, we prove tight lower bounds on the problem of mixed state tomography. More specifically, we prove that the algorithm outlined in Section 8 is essentially optimal in all Schatten-q-norms. The optimality in all other Schatten norms follows directly from it. The results obtained in this section are summarized in Corollary 9.1.

The core idea is to take a bit string $b \in \{0,1\}^{dr}$, and hide it in a density matrix $\rho_b \in \mathbb{C}^{d\times d}$, with rank at most r, such that its purification can be constructed with exactly one fractional phase query O_b^ε. Next, we show that if we learn a classical description $\widetilde{\rho}$ of ρ_b such that $\|\widetilde{\rho} - \rho_b\|_1 \le \varepsilon$, we narrow down the size of the set of bit strings $b' \in \{0,1\}^{dr}$ that satisfy $\|\widetilde{\rho} - \rho_{b'}\|_1 \le \varepsilon$ to $O(2^{-cdr})$, for some constant $c > 0$. Finally, we show that this must require $\Omega(dr/\varepsilon)$ queries to O_b^ε, and hence the state tomography algorithm must make at least this number of queries to the state preparation unitary too.

We start by defining the density matrices $\rho_b \in \mathbb{C}^{d\times d}$, in which we hide the bit string $b \in \{0,1\}^{dr}$, in the following definition.

Definition 9.1. *Let $\varepsilon \in [0,1]$, $d \in \mathbb{N}$, $r \in [d]$, and $U^{(0)}, \dots, U^{(r-1)} \in \mathbb{C}^{d\times d}$ unitaries to be fixed later. Let $b \in \{0,1\}^{dr}$ a bit string of length dr. We write $b = (b^{(0)}, \dots, b^{(r-1)})$, where $b^{(j)}$ is the jth block of size d. Let*

$$|\psi_b\rangle = \frac{1}{\sqrt{dr}} \sum_{j=0}^{r-1} \sum_{k=0}^{d-1} \sum_{c\in\{-1,1\}} \sqrt{\frac{1}{2} + \frac{1}{2} c\varepsilon (-1)^{b_k^{(j)}}}\, |c\rangle U^{(j)} |k\rangle |j\rangle,$$

and let ρ_b be the density matrix found by tracing out the last register of $|\psi_b\rangle\langle\psi_b|$.

We can already derive some interesting properties of the density matrices ρ_b defined above, without fixing the unitaries $U^{(j)}$, as we discuss in Lemma 9.5 (We choose the unitaries in Definition 9.2.)

Lemma 9.5. *Let ε, d, r and $U^{(0)}, \dots, U^{(r-1)}$ as in Definition 9.1. Let $b, \overline{b} \in \{0,1\}^{dr}$. Let $\delta = (-1)^b - (-1)^{\overline{b}} \in \{-2,0,2\}^{dr}$. Let*

$$X = \begin{bmatrix} U^{(0)}\delta^{(0)} & \cdots & U^{(r-1)}\delta^{(r-1)} \end{bmatrix}, \qquad \text{and} \qquad Y = \begin{bmatrix} U^{(0)}\mathbb{1} & \cdots & U^{(r-1)}\mathbb{1} \end{bmatrix},$$

Then

$$\left\|\rho_b - \rho_{\overline{b}}\right\|_1 \ge \frac{\varepsilon}{rd} \left\|XY^\dagger\right\|_1 - 4\varepsilon^2.$$

Proof. By taking the partial trace, we obtain that

$$\rho_b = \frac{1}{dr} \sum_{j=0}^{r-1} \sum_{k,k'=0}^{d-1} \sum_{c,c'\in\{-1,1\}} \sqrt{\frac{1}{2} + \frac{1}{2} c\varepsilon (-1)^{b_k^{(j)}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2} c'\varepsilon (-1)^{b_{k'}^{(j)}}}\, |c\rangle\langle c'| \otimes U^{(j)} |k\rangle\langle k'| \left(U^{(j)}\right)^\dagger.$$

Next, we approximate both the square roots by their tangents around $1/2$, i.e., we write $\sqrt{1/2 + x} \approx (1 + x/2)/\sqrt{2}$. We denote the resulting matrix by $\overline{\rho}_b$, and we can express it directly as

$$\overline{\rho}_b = \frac{1}{2dr} \sum_{j=0}^{r-1} \sum_{k,k'=0}^{d-1} \sum_{c,c'\in\{-1,1\}} \left(1 + \frac{1}{2} c\varepsilon (-1)^{b_k^{(j)}} + \frac{1}{2} c'\varepsilon (-1)^{b_{k'}^{(j)}}\right) |c\rangle\langle c'| \otimes U^{(j)} |k\rangle\langle k'| \left(U^{(j)}\right)^\dagger.$$

Next, we characterize the error introduced by this linearization step. To that end, observe that if $c(-1)^{b_k^{(j)}}$ and $c'(-1)^{b_{k'}^{(j)}}$ are equal, then both expressions under the square root are equal, and hence their product becomes exactly the product of the linearizations. Thus, the only entries in which the matrices ρ_b and $\overline{\rho}_b$ differ are those for which $c(-1)^{b_k^{(j)}}$ and $c'(-1)^{b_{k'}^{(j)}}$ differ. This allows us to write the difference between ρ_b and $\overline{\rho}_b$ succinctly, as

$$\rho_b - \overline{\rho}_b = \frac{1 - \sqrt{1-\varepsilon^2}}{2dr} \sum_{j=0}^{r-1} \sum_{c,c'\in\{-1,1\}} \sum_{\substack{k,k'=0 \\ c(-1)^{b_k^{(j)}} \neq c'(-1)^{b_{k'}^{(j)}}}}^{d-1} |c\rangle\langle c'| \otimes U^{(j)} |k\rangle\langle k'| \left(U^{(j)}\right)^\dagger.$$

Thus, by taking the trace norm, applying the triangle inequality, removing the unitary transformations on the last register, and using $1-\sqrt{1-\varepsilon^2} \le \varepsilon^2$, we obtain that

$$\|\rho_b - \overline{\rho}_b\|_1 \le \frac{\varepsilon^2}{2dr}\sum_{j=0}^{r-1}\left[\left\|\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\otimes \sum_{\substack{k,k'=0\\ (-1)^{b_k^{(j)}}\neq(-1)^{b_{k'}^{(j)}}}}^{d-1} |k\rangle\langle k'|\right\|_1 + \left\|\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\otimes \sum_{\substack{k,k'=0\\ (-1)^{b_k^{(j)}}=(-1)^{b_{k'}^{(j)}}}}^{d-1} |k\rangle\langle k'|\right\|_1\right].$$

Since the 2×2 matrices both have 2 eigenvalues that are both of modulus 1, we can factor them out at the expense of a factor of 2. Then, after reordering rows and columns such that all k's with $b_k^{(j)}=1$ are at the upper-left corner, we obtain that

$$\|\rho_b - \overline{\rho}_b\|_1 \le \frac{\varepsilon^2}{dr}\sum_{j=0}^{r-1}\left[\left\|\begin{array}{cc} & \begin{array}{cc} b_k^{(j)}=1 & b_k^{(j)}=0\end{array}\\ \begin{array}{c} b_{k'}^{(j)}=1\\ b_{k'}^{(j)}=0\end{array} & \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\end{array}\right\|_1 + \left\|\begin{array}{cc} & \begin{array}{cc} b_k^{(j)}=1 & b_k^{(j)}=0\end{array}\\ \begin{array}{c} b_{k'}^{(j)}=1\\ b_{k'}^{(j)}=0\end{array} & \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\end{array}\right\|_1\right].$$

The trace norm of the left term is d, since the two diagonal blocks have the single eigenvalues $|b^{(j)}|$ and $d-|b^{(j)}|$. On the right-hand side, the two eigenvalues are $\pm\sqrt{|b^{(j)}|(d-|b^{(j)}|)}$, and so the trace norm is $2\sqrt{|b^{(j)}|(d-|b^{(j)}|)} \le 2\sqrt{d^2/4} = d$. Plugging these bounds into the above equation yields $\|\rho_b - \overline{\rho}_b\|_1 \le 2\varepsilon^2$. Thus, we have

$$\left\|\overline{\rho}_b - \overline{\rho}_{\overline{b}}\right\|_1 \le \|\rho_b - \overline{\rho}_b\|_1 + \left\|\rho_b - \rho_{\overline{b}}\right\|_1 + \left\|\rho_{\overline{b}} - \overline{\rho}_{\overline{b}}\right\|_1 \le \left\|\rho_b - \rho_{\overline{b}}\right\|_1 + 4\varepsilon^2,$$

and hence it suffices to prove that $\left\|\overline{\rho}_b - \overline{\rho}_{\overline{b}}\right\|_1 = \varepsilon\left\|XY^\dagger\right\|_1/(dr)$. To that end, observe that we can write the difference as

$$\overline{\rho}_b - \overline{\rho}_{\overline{b}} = \frac{\varepsilon}{4dr}\sum_{j=0}^{r-1}\sum_{k,k'=0}^{d-1}\sum_{c,c'\in\{-1,1\}}\left(c\delta_k^{(j)} + c'\delta_{k'}^{(j)}\right)|c\rangle\langle c'|\otimes U^{(j)}|k\rangle\langle k'|\left(U^{(j)}\right)^\dagger.$$

Next, some of the indices can be decoupled from one another, which results in

$$\begin{aligned}\overline{\rho}_b - \overline{\rho}_{\overline{b}} = \frac{\varepsilon}{4dr}\Bigg[&\sum_{j=0}^{r-1}\sum_{c\in\{-1,1\}} c|c\rangle \sum_{c'\in\{-1,1\}}\langle c'|\otimes U^{(j)}\sum_{k=0}^{d-1}\delta_k^{(j)}|k\rangle\sum_{k'=0}^{d-1}\langle k'|\left(U^{(j)}\right)^\dagger\\ &+\sum_{j=0}^{r-1}\sum_{c\in\{-1,1\}}|c\rangle\sum_{c'\in\{-1,1\}} c'|c'\rangle\otimes U^{(j)}\sum_{k=0}^{d-1}|k\rangle\sum_{k'=0}^{d-1}\delta_{k'}^{(j)}\langle k'|\left(U^{(j)}\right)^\dagger\Bigg],\end{aligned}$$

Now, we can let

$$V = \begin{bmatrix}X\\ -X\end{bmatrix}, \qquad \text{and} \qquad W = \begin{bmatrix}Y\\ Y\end{bmatrix},$$

from which we can directly observe that the expression for $\overline{\rho}_b - \overline{\rho}_{\overline{b}}$ simplifies to

$$\overline{\rho}_b - \overline{\rho}_{\overline{b}} = \frac{\varepsilon}{4rd}\left[VW^\dagger + WV^\dagger\right].$$

To evaluate the trace norm, we observe that

$$\left\|\overline{\rho}_b - \overline{\rho}_{\overline{b}}\right\|_1 = \mathrm{Tr}\left[\sqrt{\left(\rho_b - \rho_{\overline{b}}\right)^2}\right] = \frac{\varepsilon}{4dr}\mathrm{Tr}\left[\sqrt{(VW^\dagger + WV^\dagger)^2}\right].$$

By direct calculation, we find that $V^\dagger W = W^\dagger V = 0$, and so when we expand the square, two of the terms cancel, and we end up with

$$\left\|\overline{\rho}_b - \overline{\rho}_{\overline{b}}\right\|_1 = \frac{\varepsilon}{4rd}\mathrm{Tr}\left[\sqrt{VW^\dagger WV^\dagger + WV^\dagger VW^\dagger}\right].$$

Again, since $V^\dagger W = W^\dagger V = 0$, we find that the two terms underneath the square root are only acting non-trivially on mutually orthogonal subspaces. Therefore, the square root of the sum is the sum of the square roots, and we end up with

$$\left\|\overline{\rho}_b - \overline{\rho}_{\overline{b}}\right\|_1 = \frac{\varepsilon}{4rd}\left[\mathrm{Tr}\left[\sqrt{VW^\dagger WV^\dagger}\right] + \mathrm{Tr}\left[\sqrt{WV^\dagger VW^\dagger}\right]\right].$$

We rewrite the term on the left within parentheses as

$$\mathrm{Tr}\left[\sqrt{VW^\dagger WV^\dagger}\right] = \mathrm{Tr}\left[\sqrt{\begin{bmatrix} 2XY^\dagger YX^\dagger & -2XY^\dagger YX^\dagger \\ -2XY^\dagger YX^\dagger & 2XY^\dagger YX^\dagger \end{bmatrix}}\right] = \mathrm{Tr}\left[\sqrt{\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} \otimes XY^\dagger YX^\dagger}\right].$$

The (non-zero) spectrum of the expression inside the square root is $4\sigma(XY^\dagger YX^\dagger)$, and hence the trace becomes $2\mathrm{Tr}[\sqrt{XY^\dagger YX^\dagger}] = 2\left\|XY^\dagger\right\|_1$. Analogously, the term on the right yields $2\left\|YX^\dagger\right\|$, and since $\left\|A^\dagger\right\|_1 = \|A\|_1$, for any matrix A, we obtain the expression from the statement of the lemma. ∎

It is worth noting that the matrix $XY^\dagger$ from the previous lemma statement is in general not Hermitian. Therefore, it is important to stress that the trace norm has to be interpreted as the sum of the singular values, rather than the sum of eigenvalues.

Next, we fix the unitary matrices $U^{(j)}$, for $j \in [r]$.

Definition 9.2. *Let $d, r \in \mathbb{N}$, and let $j \in [r]$. Define $U^{(j)} \in \mathbb{C}^{d\times d}$ as*

$$U^{(j)}_{k\ell} = \frac{1}{\sqrt{d}}\omega_d^{(j-k)\ell}.$$

The benefit of this particular choice of unitaries $U^{(j)}$ is that they further simplify the expressions that appear in Lemma 9.5. The details are presented in the lemma below.

Lemma 9.6. *Let ε, d and r as in Definition 9.1. Then, for all $j \in [r]$, $U^{(j)}$ is unitary,*

$$Y = \sqrt{d}\begin{bmatrix} I_r \\ 0 \end{bmatrix},$$

and we have $\Delta \in \mathbb{C}^{d\times r}$ such that

$$\left\|\rho_b - \rho_{\overline{b}}\right\|_1 \geq \frac{\varepsilon}{r\sqrt{d}}\|\Delta\|_1 - 2\varepsilon^2, \qquad \textit{with} \qquad \Delta_{kj} = \omega_d^{jk}\delta_k^{(j)}.$$

Proof. Let $j \in [r]$, and $\ell, \ell' \in [d]$. Then,

$$\left[\left(U^{(j)}\right)^\dagger U^{(j)}\right]_{\ell,\ell'} = \sum_{k=0}^{d-1}\overline{U}^{(j)}_{k\ell}U^{(j)}_{k\ell'} = \frac{1}{d}\sum_{k=0}^{d-1}\omega_d^{-(j-k)\ell}\omega_d^{(j-k)\ell'} = \frac{1}{d}\sum_{k=0}^{d-1}\omega_d^{(\ell'-\ell)k} = 1_{\ell=\ell'},$$

and thus indeed $\left(U^{(j)}\right)^\dagger U^{(j)} = I$. Next, let $j \in [r]$ and $k \in [d]$. We observe that

$$Y_{jk} = \left(U^{(j)}\mathbb{1}\right)_k = \sum_{\ell=0}^{d-1}U^{(j)}_{k\ell} = \frac{1}{\sqrt{d}}\sum_{\ell=0}^{d-1}\omega_d^{(j-k)\ell} = \sqrt{d}1_{j=k},$$

as claimed. Finally, for the trace norm, observe that it suffices to show that $\|X\|_1 = \|\Delta\|_1$. To that end, for all $j \in [r]$ and $k \in [d]$,

$$\left[\left(U^{(0)}\right)^\dagger X\right]_{kj} = \left[\left(U^{(0)}\right)^\dagger U^{(j)}\delta^{(j)}\right]_k = \frac{1}{d}\sum_{\ell,\ell'=0}^{d-1}\omega_d^{k\ell}\omega_d^{(j-\ell)\ell'}\delta^{(j)}_{\ell'} = \frac{1}{d}\sum_{\ell'=0}^{d-1}\left[\sum_{\ell=0}^{d-1}\omega_d^{(k-\ell')\ell}\right]\omega_d^{j\ell'}\delta^{(j)}_{\ell'}.$$

The inner summation vanishes if $k \neq \ell'$, and evaluates to d when $k = \ell'$. Therefore, the whole expression simplifies to $\omega_d^{jk}\delta_k^{(j)} = \Delta_{kj}$. Thus, $\Delta = \left(U^{(0)}\right)^\dagger X$, and since $U^{(0)}$ is unitary, X and Δ have the same singular values, and hence the same trace norm as well. This completes the proof. ∎

Next, we choose two bit strings $b, \overline{b} \in \{0,1\}^{dr}$ uniformly at random, and we analyze the probability of the trace norm of the difference $\rho_b - \rho_{\overline{b}}$ being small. On a high level, the smaller this probability is, the fewer $\rho_{\overline{b}}$s are close to ρ_b chosen uniformly at random.

LEMMA 9.7. *Let $\varepsilon \in [0, 1/128]$, $d \in \mathbb{N}$ and $r \in [d]$. Let $b, \overline{b} \in \{0,1\}^{dr}$ uniformly at random. Then, there exist constants $c, d_0 > 0$ such that for all $d > d_0$,*

$$\mathbb{P}\Big[\big\|\rho_b - \rho_{\overline{b}}\big\|_1 \leq \frac{\varepsilon}{64}\Big] \leq e^{-crd}.$$

Proof. By the previous lemma, we know that if $\|\Delta\|_1 \geq r\sqrt{d}/32$, then $\big\|\rho_b - \rho_{\overline{b}}\big\|_1 \geq \varepsilon/32 - 2\varepsilon^2 \geq \varepsilon/32 - \varepsilon/64 = \varepsilon/64$. Thus, it suffices to prove that there exist constants $c, d_0 > 0$ such that for all $d > d_0$,

$$\mathbb{P}\left[\|\Delta\|_1 \leq \frac{r\sqrt{d}}{32}\right] \leq e^{-crd}.$$

Next, let $A \in \mathbb{C}^{d\times d}$ be such that $\|A\| \leq 1$, and let $\Delta = U\Sigma V$ be the singular value decomposition of Δ, with the singular values $\sigma_1, \ldots, \sigma_r$ (where we allow the σs to be 0 if the rank turns out to be strictly smaller than r). For technical reasons, we pad the matrices Σ and V with zeros, such that we have $U, \Sigma, V \in \mathbb{C}^{d\times d}$. Let $\mathbf{u}_j$ and $\mathbf{v}_j$ denote the jth columns of U and V, respectively. Then,

$$|\mathrm{Tr}[A\Delta]| = |\mathrm{Tr}[AU\Sigma V]| = |\mathrm{Tr}[VAU\Sigma]| \leq \sum_{j=1}^{r}\Big|[VAU]_{jj}\Big|\sigma_j = \sum_{j=1}^{r}\Big|\mathbf{v}_j^\dagger A\mathbf{u}_j\Big|\sigma_j \leq \sum_{j=1}^{r}\sigma_j = \|\Delta\|_1,$$

where we used that $|\mathbf{v}_j^\dagger A\mathbf{u}_j| \leq \|\mathbf{v}_j\| \cdot \|A\| \cdot \|\mathbf{u}_j\| \leq 1$. Thus, it is sufficient to find some matrix $A \in \mathbb{C}^{d\times d}$ such that $\|A\| \leq 1$, which is allowed to depend on Δ, and prove that there exists a constant $c > 0$ such that

$$\mathbb{P}\left[|\mathrm{Tr}[A\Delta]| \leq \frac{r\sqrt{d}}{32}\right] \leq e^{-crd}.$$

To that end, for any vector $\mathbf{v}$, we let the matrix $A_{\mathbf{v}}$ be the operation that reflects through the subspace $\mathrm{Span}\{\mathbf{v}\}^\perp$, i.e.,

$$A_{\mathbf{v}} = I - 2\frac{\mathbf{v}\mathbf{v}^\dagger}{\|\mathbf{v}\|^2}.$$

Since $A_{\mathbf{v}}$ is a reflection, $A_{\mathbf{v}}$ is a unitary matrix and in particular we have $\|A_{\mathbf{v}}\| \leq 1$. Additionally, suppose if we have $\mathbf{v}$ and $\mathbf{w}$ with $\|\mathbf{v}\| = \|\mathbf{w}\| = 1$ and $\mathrm{Im}(\mathbf{v}^\dagger\mathbf{w}) = 0$, then since $(\mathbf{v}-\mathbf{w})^\dagger(\mathbf{v}+\mathbf{w}) = \|\mathbf{v}\|^2 - \|\mathbf{w}\|^2 = 0$, we obtain

$$A_{\mathbf{v}-\mathbf{w}}\mathbf{v} = A_{\mathbf{v}-\mathbf{w}} \cdot \frac{1}{2}[\mathbf{v} - \mathbf{w} + \mathbf{v} + \mathbf{w}] = \frac{1}{2}[-\mathbf{v} + \mathbf{w} + \mathbf{v} + \mathbf{w}] = \mathbf{w}.$$

Now, we let $\mathbf{y}_1 = \mathbf{x}_1$, and for all $j \in [r]$ recursively define

$$\mathbf{z}_j = \begin{cases} \frac{\mathbf{y}_j}{\|\mathbf{y}_j\|} \cdot \frac{\overline{(y_j)_1}}{|(y_j)_1|}, & \text{if } (y_j)_1 \neq 0, \\ \frac{\mathbf{y}_j}{\|\mathbf{y}_j\|}, & \text{if } \mathbf{y}_j \neq \mathbf{0}, \\ \mathbf{e}_1, & \text{if } \mathbf{y}_j = \mathbf{0}, \end{cases} \qquad \varphi_j = \begin{cases} \frac{\overline{(y_j)_1}}{|(y_j)_1|} & \text{if } (y_j)_1 \neq 0, \\ 1, & \text{if } (y_j)_1 = 0, \end{cases} \qquad A_j = \left[\begin{array}{c|c} I_{j-1} & 0 \\ \hline 0 & \varphi_j A_{\mathbf{z}_j - \mathbf{e}_1} \end{array}\right],$$

and furthermore

$$A_j \cdots A_1\mathbf{x}_{j+1} = \left[\begin{array}{c} \mathbf{w}_{j+1} \in \mathbb{C}^j \\ \hline \mathbf{y}_{j+1} \in \mathbb{C}^{d-j} \end{array}\right], \qquad \text{and} \qquad A = A_r \cdots A_1.$$

From the construction, it is clear that all A_js are unitary, and so $\|A\| \leq 1$. Additionally,

$$A\mathbf{x}_j = A_r \cdots A_1\mathbf{x}_j = A_r \cdots A_j\left[\begin{array}{c} \mathbf{w}_j \\ \hline \mathbf{y}_j \end{array}\right] = A_r \cdots A_{j+1}\left[\begin{array}{c} \mathbf{w}_j \\ \hline \varphi_j A_{\mathbf{z}_j - \mathbf{e}_1}\mathbf{y}_j \end{array}\right].$$

We can directly observe that $\mathrm{Im}(\mathbf{e}_1^\dagger \mathbf{z}_j) = 0$, for all $j \in [r]$. If $(y_j)_1 \neq 0$, then we find $\varphi_j A_{\mathbf{z}_1-\mathbf{e}_1}\mathbf{y}_j = |(y_j)_1|\|\mathbf{y}_j\|\varphi_j A_{\mathbf{z}_1-\mathbf{e}_1}\mathbf{z}_1/\overline{(y_j)_1} = \|\mathbf{y}_1\|\mathbf{e}_1$. Similarly, if $(y_j)_1 = 0$ but $\mathbf{y}_j \neq \mathbf{0}$, then $\varphi_j A_{\mathbf{z}_1-\mathbf{e}_1}\mathbf{y}_j = \|\mathbf{y}_1\|A_{\mathbf{z}_1-\mathbf{e}_1}\mathbf{z}_1 = \|\mathbf{y}_1\|\mathbf{e}_1$. Finally, if $\mathbf{y}_j = \mathbf{0}$, then $\varphi_j A_{\mathbf{z}_1-\mathbf{e}_1}\mathbf{y}_j = \mathbf{0} = \|\mathbf{y}_j\|\mathbf{e}_1$. Thus, in all cases we find

$$A\mathbf{x}_j = A_r \cdots A_{j+1}\left[\frac{\mathbf{w}_j}{\|\mathbf{y}_j\|\mathbf{e}_1}\right] = \left[\frac{\mathbf{w}_j}{\|\mathbf{y}_j\|\mathbf{e}_1}\right].$$

Moreover, observe that $A_j \cdots A_1$ is unitary and it only depends on the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_j$. Therefore, $\mathbf{y}_{j+1}$ is a projection of $\mathbf{x}_{j+1}$ onto an $(d-j)$-dimensional subspace, which we denote by S_{j+1}, and we observe that S_{j+1} does not depend on the vector $\mathbf{x}_{j+1}$ itself. Finally, we have

$$\|\Delta\|_1 \geq |\mathrm{Tr}[A\Delta]| = \sum_{j=1}^{r}\|\mathbf{y}_j\| \geq \sum_{j=1}^{\min\{r,\lfloor d/2\rfloor\}}\|\mathbf{y}_j\|.$$

Now, if there are at least $r/4$ js in $1, \ldots, \min\{r, \lfloor d/2\rfloor\}$ for which $\|\mathbf{y}_j\| > \sqrt{d}/8$, then this would imply that $\|\Delta\|_1 > r\sqrt{d}/32$. Thus, by the contrapositive, if $\|\Delta\|_1 \leq r\sqrt{d}/32$, then there must be at least $\min\{r, \lfloor d/2\rfloor\} - r/4 + 1 \geq r/4$ js for which $\|\mathbf{y}_j\| \leq \sqrt{d}/8$. Let us label all these js $j_1 < \cdots j_k \leq d/2$, where we now know that $k \geq r/4$. Then,

$$\mathbb{P}\left[\bigwedge_{\ell=1}^{k}\|\mathbf{y}_{j_\ell}\| \leq \frac{\sqrt{d}}{8}\right] = \prod_{\ell=1}^{k}\mathbb{P}\left[\|\mathbf{y}_{j_\ell}\| \leq \frac{\sqrt{d}}{8}\middle|\bigwedge_{m=1}^{\ell-1}\|\mathbf{y}_{j_m}\| \leq \frac{\sqrt{d}}{8}\right].$$

Now, recall that $\mathbf{y}_{j_\ell} = \Pi_{S_{j_\ell}}\mathbf{x}_{j_\ell}$. Since the entries of $\mathbf{x}_{j_\ell}$ are independent and sub-Gaussian with constant $K = 2$, we know from [45], Corollary 3.1, combined with the complexification techniques outlined in the last paragraph of said document, that there exists a constant $c' > 0$ such that

$$\mathbb{P}\left[\left\|\Pi_{S_{j_\ell}}\mathbf{x}_{j_\ell}\right\| \leq \frac{\sqrt{d}}{8}\right] \leq \mathbb{P}\left[\left|\left\|\Pi_{S_{j_\ell}}\mathbf{x}_{j_\ell}\right\| - \sqrt{d-j_\ell}\right| \leq \left|\sqrt{d-j_\ell} - \frac{\sqrt{d}}{8}\right|\right] \leq 2e^{-c'd}.$$

Moreover, since this bound holds for all subspaces S_{j_ℓ} of dimension $d - j_\ell$, we obtain

$$\mathbb{P}\left[\|\Delta\|_1 \leq \frac{r\sqrt{d}}{32}\right] \leq \prod_{\ell=1}^{k}\mathbb{P}\left[\left\|\Pi_{S_{j_\ell}}\mathbf{x}_{j_\ell}\right\| \leq \frac{\sqrt{d}}{8}\middle|\bigwedge_{m=1}^{\ell-1}\|\mathbf{y}_{j_m}\| \leq \frac{\sqrt{d}}{8}\right] \leq \left(2e^{-c'd}\right)^k.$$

Thus, whenever $d \geq \ln(2)/c' =: d_0$, we can choose $c = (c' - \ln(2)/d)/4$. Then, $c > 0$, and

$$\mathbb{P}\left[\|\Delta\|_1 \leq \frac{r\sqrt{d}}{32}\right] \leq \left(2e^{-c'd}\right)^k = \left(2e^{-\left(4c+\frac{\ln(2)}{d}\right)d}\right)^k \leq e^{-4cdk} \leq e^{-cdr},$$

where in the last step, we used that $k \geq r/4$. This completes the proof. ∎

Now, we are able to finish the proof. The proof strategy followed from here onward is very similar to those presented in [13], Section 5.

THEOREM 9.3. *Let $\varepsilon \in [0, 1/128]$, $d \in \mathbb{N}$, and $r \in [d]$. Suppose that we have a Q-query quantum algorithm that given access to an (inverse) state-preparation unitary for a purification of a $d \times d$ density matrix ρ of rank at most r, outputs an approximation $\widetilde{\rho}$ such that $\|\widetilde{\rho} - \rho\|_1 \leq \varepsilon/128$, with probability at least $2/3$. Then $Q = \Omega(dr/\varepsilon)$.*

Proof. Let G be a bipartite graph with $2 \cdot 2^{rd}$ nodes, labeled by the bit strings $b \in \{0,1\}^{rd}$ on one side, and $\overline{b} \in \{0,1\}^{rd}$ on the other. Let there be an edge between b and $\overline{b}$, if $\left\|\rho_b - \rho_{\overline{b}}\right\|_1 \leq \varepsilon/64$. From the previous lemma, we obtain that there exists a constants $c, d_0 > 0$ such that whenever $d \geq d_0$, the number of edges of m in G satisfies

$$m \leq 2^{2rd} \cdot e^{-crd}.$$

We abbreviate $f = e^{-crd}$, and observe that

$$\sum_{b\in\{0,1\}^{rd}} \deg(b) = m \leq f \cdot 2^{2rd},$$

where $\deg(b)$ denotes the degree of b in G. Next, let $B = \{b \in \{0,1\}^{rd} : \deg(b) \geq 2^{rd}\sqrt{f}\}$, i.e., the set of nodes on the left side that have high degree. Then, by an argument that is usually referred to as the pigeonhole principle, we obtain that $|B| \leq 2^{rd}\sqrt{f}$.

Let $b \in \{0,1\}^{rd} \setminus B$, and suppose that we have access to b through the phase oracle

$$O_{b,\varepsilon'} : |j\rangle \mapsto e^{i\varepsilon b_j}|j\rangle.$$

We now use our Q-query mixed-state tomography algorithm to construct a new algorithm that recovers b with some very low probability.

The first step is to implement the unitary U_b that maps

$$U_b : |0\rangle \mapsto \frac{1}{\sqrt{r}}\sum_{j=0}^{r-1} |\psi_b^{(j)}\rangle|j\rangle.$$

Using the same construction as in Lemma 9.4, we can construct a circuit implementing this unitary U_b with K calls to $O_{b,\varepsilon}$, where $K = \Theta(1)$. Next, since this unitary U_b prepares a purification of ρ_b, we can use Q queries to it to obtain an estimate $\widetilde{\rho}$, such that $\|\widetilde{\rho} - \rho_b\|_1 \leq \varepsilon/128$, with probability at least $2/3$.

Next, suppose that $\overline{b} \in \{0,1\}^{rd}$ satisfies $\left\|\widetilde{\rho} - \rho_{\overline{b}}\right\|_1 \leq \varepsilon/128$. Then, by the triangle inequality, we have that $\left\|\rho_b - \rho_{\overline{b}}\right\|_1 \leq \|\rho_b - \widetilde{\rho}\|_1 + \left\|\widetilde{\rho} - \rho_{\overline{b}}\right\|_1 \leq \varepsilon/64$, and hence we find that b and $\overline{b}$ are neighbors in G. Since we chose b to be in $\{0,1\}^{rd}\setminus B$, we know that $\deg(b) \leq 2^{rd}\sqrt{f}$, and hence there are at most $2^{rd}\sqrt{f}$ choices for $\overline{b}$, among which is b itself. Thus, if we uniformly choose one such $\overline{b}$, it will be equal to b with probability at least $2/3 \cdot 2^{-rd} f^{-1/2}$.

The procedure above uses KQ queries to $O_{b,\varepsilon'}$, and recovers b with probability at least $2/3 \cdot 2^{-rd} f^{-1/2}$. It is known that if we can solve this problem with KQ queries to the fractional phase oracle $O_{b,\varepsilon'}$, we can also solve it with at most $K'KQ$ queries to the regular phase oracle O_b, with $K' = \Theta(\varepsilon)$.[8] According to [14], Equation 4, this implies that

$$2^{rd} - |B| \leq \frac{3}{2}\cdot 2^{rd}\sqrt{f}\cdot \sum_{k=0}^{K'KQ}\binom{rd}{k} \leq \frac{3}{2}\cdot 2^{rd}\sqrt{f}\cdot 2^{rdH\left(\frac{K'KQ}{rd}\right)},$$

where $H(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function, and the rightmost inequality can be found in several text books, e.g., [16], Lemma 16.19.

We can now plug everything into the above equation. Since $|B| \leq 2^{rd}\sqrt{f} \leq 2^{rd}e^{-1/2} \leq 2^{rd}/\sqrt{2}$, and hence we write

$$2^{rd+\log\left(1-\frac{1}{\sqrt{2}}\right)} \leq 2^{\log(3)-1+rd-crd\log(e)+rdH\left(\frac{K'KQ}{dr}\right)}.$$

Comparing the exponents we obtain

$$\log(3) - 1 - \log\left(1 - \frac{1}{\sqrt{2}}\right) + rd\left(-c\log(e) + H\left(\frac{K'KQ}{dr}\right)\right) \geq 0,$$

and thus $H(\frac{K'KQ}{dr}) = \Omega(1)$. Since the binary entropy function is monotonously increasing from 0 to 1 in the interval $[0, 1/2]$, we find that $K'KQ = \Omega(dr)$, and hence $Q = \Omega(rd/\varepsilon)$. ∎

We can now derive the lower bounds for all other Schatten norms too.

COROLLARY 9.1. *Let $\varepsilon \in [0, 1/128]$, $d \in \mathbb{N}$, $r \in [d]$ and $q \in [1, \infty]$. Suppose that we have a quantum algorithm that estimates a density matrix ρ up to precision ε in Schatten-q-norm, using Q (inverse) queries to a unitary preparing its purification. Then,*

$$Q = \Omega\left(\min\left\{\frac{dr^{\frac{1}{q}}}{\varepsilon}, \frac{d}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}\right\}\right).$$

[8] See the footnote in Lemma 9.4 for more details.

Proof. First, suppose that $\varepsilon \leq 1/(256r^{1-1/q})$. Then, using standard norm conversion, we can obtain a $r^{1-1/q}\varepsilon$-precise approximation of ρ in trace norm in Q queries. Since $r^{1-1/q}\varepsilon \leq 1/128$, using Theorem 9.3 we find that

$$Q = \Omega\left(\frac{dr}{r^{1-1/q}\varepsilon}\right) = \Omega\left(\frac{dr^{\frac{1}{q}}}{\varepsilon}\right).$$

On the other hand, if $1/(128r^{1-1/q}) < \varepsilon \leq 1/128$, then we can choose an integer $1 \leq r' < r$, such that $1/(256(r')^{1-1/q}) < \varepsilon \leq 1/(128(r')^{1-1/q})$. By the previous argument for the case $\varepsilon \leq 1/(256r^{1-1/q})$, we can now use Q queries to obtain a 2ε-precise Schatten-q-approximation of any density matrix of rank at most r', using our algorithm for density matrices of rank r. We already know that this takes $\Omega(d(r')^{1/q}/\varepsilon)$ queries, and since $r' = \Theta(1/\varepsilon^{1/(1-1/q)})$, we obtain that

$$Q = \Omega\left(\frac{d(r')^{\frac{1}{q}}}{\varepsilon}\right) = \Omega\left(\frac{d}{\varepsilon^{1+\frac{1/q}{1-1/q}}}\right) = \Omega\left(\frac{d}{\varepsilon^{\frac{1}{1-\frac{1}{q}}}}\right).$$

Finally, observe that $d/\varepsilon^{1/(1-1/q)} < dr^{1/q}/\varepsilon$ is equivalent to $\varepsilon^{1/(q-1)} > 1/r^{1/q}$, and hence to $\varepsilon > 1/r^{1-1/q}$, so indeed the minimum picks out the right branch of the lower bound. This completes the proof. ∎

Note that this exactly matches the complexity that we obtained in Lemma 8.1. Thus, up to polylogarithmic factors, we have completely characterized the query complexity of state tomography in this model.

10 Open problems

We end the paper with a discussion on some open questions.

State preparation without an inverse. In Section 4.2 we consider tomography using conditional samples, and in Section 9.1 we show that our upper bounds are optimal up to log factors. Conditional samples are directly inspired by controlled usage of a state-preparation unitary, without access to the inverse of this unitary. Such a state-preparation unitary is at least as powerful as conditional samples, and at most as powerful as state preparation with the inverse as well.

Even in the two dimensional case of standard amplitude estimation, the best upper bound of $\widetilde{\mathcal{O}}(1/\varepsilon^2)$ comes from conditional samples, while the best lower bound of $\Omega(1/\varepsilon)$ also holds when the inverse is allowed. Hence the question of finding a quantum algorithm that requires $o(1/\varepsilon^2)$ applications of a controlled state-preparation unitary to perform amplitude estimation, and that does not require access to the inverse of this unitary. We conjecture that the answer is negative, but we are not aware of any lower bound techniques that differentiate between normal and inverse usage of an input oracle.

Vector estimate conversions. The two lemmas in Section 3 still leave some open questions. While Lemma 3.1 gives the relation between amplitude and probability estimates in general, it is unclear whether a similar relation holds for amplitudes of a purification and the associated density matrix. Lemma 8.1 gives a relation between the ℓ_2-norm for amplitudes and the Schatten-1-norm (tracer norm) for the density matrix, does a similar relation hold for the ℓ_q-norm and Schatten-$\frac{1}{1/2+1/q}$-norm?

Similarly, Lemma 3.2 shows how to obtain a ℓ_q-norm estimate of a ℓ_s-normalized vector using an ℓ_∞-norm estimate. It is still unclear whether an ℓ_p norm estimate can be used in a similar manner to obtain an ℓ_q-norm estimate, when $p > q > s$.

Simple sample-based estimates for mixed states in other norms. All single-copy sampling methods for pure-state tomography that we are aware of estimate the state directly in a Schatten q-norm, and then convert to the trace norm. In order to find the initial estimate, a set of random measurements is performed, and an optimization problem is solved to find a $\tilde{\rho}$ that matches best with the measurement statistics. Could a very efficient estimate in the max-norm possibly lead to a simpler algorithm? In Appendix A we show how a probability distribution can be constructed that is proportional to the d^2 elements in the density operator, so an ℓ_2-norm approximation of this distribution gives a Frobenius norm estimate of ρ.

An alternative approach could be to use a procedure inspired by shadow tomography to estimate all the E_{ij} and obtain a max-norm estimate with $\widetilde{\mathcal{O}}(1/\varepsilon^2)$ samples. If these estimates can be made symmetric and unbiased, then this would imply an operator norm estimate with $\widetilde{\mathcal{O}}(d/\varepsilon^2)$ samples, a Frobenius norm estimate with $\widetilde{\mathcal{O}}(dr/\varepsilon^2)$ samples, and a trace norm estimate with $\widetilde{\mathcal{O}}(dr^2/\varepsilon^2)$ samples. This would matching the optimal

bound by [23, 25] for single copy measurements. There is some hope for this, as recent shadow tomography results [30] require only $\widetilde{\mathcal{O}}(1/\varepsilon^2)$ copies when the Frobenius norm of the measurements is constant. Furthermore, these methods are rather simple, and there is no post processing needed, unlike the result by [23, 25] that requires the solution of a convex optimization problem. The main problem to overcome is that the outputs from shadow tomography might not be independent.

Time complexity of expectation value estimation. When we apply expectation value estimation to mixed-state tomography, we give a tailored implementation of the block-encoding of $\sum_i \lambda_i E_i$ in order to avoid a large subnormalization. In general however the block-encoding is sub-normalized by $\sum_i |\lambda_i| \|E_i\|$ due to the usage of the LCU-lemma. The pre-amplification of this block encoding then requires a number of iterations which scales with $N = \sum_i \|E_i\|$.

On the other hand, the set of operators of the form $E_i / \left\| \sum_j E_j \right\|$ could be turned into a POVM measurement, as their sum has operator norm at most 1. Hence, by estimating all expectation values of this POVM with precision $\varepsilon / \left\| \sum_j E_j \right\|$ by simply measuring, we would be able to learn all original expectation values with precision ε, and a sample complexity dependent on $\left\| \sum_j E_j \right\|$ (as opposed to $\sum_j \|E_j\|$. Can our techniques be improved to also depend on $\left\| \sum_j E_j \right\|$? Or, more generally, is there a version of the LCU-lemma and pre-amplification that achieves this time complexity? Low [38] uses a technique that might be related to this in order to improve sparse block-encodings for matrices with bounded norm, and a general answer might improve the method by Low slightly.

Acknowledgements

We are grateful to Srinivasan Arunachalam, Ronald de Wolf and anonymous referees for useful discussions and comments. Joran van Apeldoorn is supported by the Dutch Research Council (NWO/OCW), as part of QSC (024.003.037) and by QuantumDelta NL. András Gilyén acknowledges funding provided by the EU's Horizon 2020 Marie Skłodowska-Curie program 891889-QuantOrder. Giacomo Nannicini was partially supported by the Army Research Office under grant number W911NF-20-1-0014.

References

[1] S. Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, June 2018.

[2] A. Acharya, S. Saha, and A. M. Sengupta. Informationally complete povm-based shadow tomography, 2021. arXiv: `2105.05992`

[3] J. van Apeldoorn. *A Quantum View on Convex Optimization*. PhD thesis, Universiteit van Amsterdam, 2020.

[4] J. van Apeldoorn. Quantum probability oracles & multidimensional amplitude estimation. In *Proceedings of the 16th Conference on the Theory of Quantum Computation, Communication, and Cryptography (TQC)*, pages 9:1–9:11, 2021.

[5] J. van Apeldoorn and A. Gilyén. Improvements in quantum SDP-solving with applications. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 99:1–99:15, 2019. arXiv: `1804.05058`

[6] A. Barenco, A. Ekert, K.-A. Suominen, and P. Törmä. Approximate quantum Fourier transform and decoherence. *Physical Review A*, 54:139–146, 1996. arXiv: `quant-ph/9601018`

[7] A. Belovs. Variations on quantum adversary. arXiv: `1504.06943`, 2015.

[8] E. Bernstein and U. Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. Earlier version in STOC'93.

[9] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *Contemporary Mathematics Series*, pages 53–74. AMS, 2002. arXiv: `quant-ph/0005055`

[10] C. L. Canonne. A short note on learning discrete distributions. *arXiv preprint arXiv:2002.11457*, 2020.

[11] S. Chen, B. Huang, J. Li, A. Liu, and M. Sellke. Tight bounds for state tomography with incoherent measurements, 2022. arXiv: `2206.05265`

[12] R. Cleve and J. Watrous. Fast parallel circuits for the quantum Fourier transform. In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 526–536, 2000. arXiv: `quant-ph/0006004`

[13] A. Cornelissen and S. Jerbi. Quantum algorithms for multivariate monte carlo estimation, 2021.

[14] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Bound on the number of functions that can be distinguished with k quantum queries. *Physical Review A*, 60:4331–4333, 11 1999. arXiv: `quant-ph/9901012`

[15] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012.

[16] J. Flum and M. Grohe. *Parameterized Complexity Theory*. Springer Berlin, Heidelberg, 2006.

[17] A. Gilyén. *Quantum Singular Value Transformation & Its Algorithmic Applications*. PhD thesis, University of Amsterdam, 2019.

[18] A. Gilyén, S. Arunachalam, and N. Wiebe. Optimizing quantum optimization algorithms via faster quantum gradient computation. In *Proceedings of the 30th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1425–1444, 2019. arXiv: `1711.00465`

[19] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics [full version], 2018. arXiv: `1806.01838`

[20] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st ACM Symposium on the Theory of Computing (STOC)*, pages 193–204, 2019. arXiv: `1806.01838`

[21] V. Giovannetti, S. Lloyd, and L. Maccone. Architectures for a quantum random access memory. *Phys. Rev. A*, 78:052310, Nov 2008.

[22] T. Giurgica-Tiron, I. Kerenidis, F. Labib, A. Prakash, and W. J. Zeng. Low depth algorithms for quantum amplitude estimation. *CoRR*, abs/2012.03348, 2020.

[23] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105(15), 2010. arXiv: `0909.3304`

[24] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020.

[25] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. arXiv: `1508.01797`

[26] A. W. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009. arXiv: `0811.3171`

[27] F. Hiai and D. Petz. *Introduction to Matrix Analysis and Applications*. Universitext. Springer, 2014.

[28] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[29] H.-Y. Hu, R. LaRose, Y.-Z. You, E. Rieffel, and Z. Wang. Logical shadow tomography: Efficient estimation of error-mitigated observables, 2022. arXiv: `2203.07263`

[30] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, June 2020.

[31] W. J. Huggins, K. Wan, J. McClean, T. E. O'Brien, N. Wiebe, and R. Babbush. Nearly optimal quantum algorithm for estimating multiple expectation values, 2021. arXiv: `2111.09283`

[32] S. P. Jordan. Fast quantum algorithm for numerical gradient estimation. *Physical Review Letters*, 95(5):050501, 2005. arXiv: `quant-ph/0405146`

[33] I. Kerenidis and A. Prakash. A quantum interior point method for LPs and SDPs. *ACM Transactions on Quantum Computing*, 1(1), 2020. arXiv: `1808.09266`

[34] R. Kueng, H. Rauhut, and U. Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

[35] T. Lee, R. Mittal, B. W. Reichardt, R. Špalek, and M. Szegedy. Quantum query complexity of state conversion. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 344–353, 2011. arXiv: `1011.3020`

[36] N. Linden and R. de Wolf. Average-case verification of the Quantum Fourier Transform enables worst-case phase estimation. arXiv: `2109.10215`, 2021.

[37] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

[38] G. H. Low. Hamiltonian simulation with nearly optimal dependence on spectral norm. In *Proceedings of the 51st ACM Symposium on the Theory of Computing (STOC)*, pages 491–502, 2019. arXiv: `1807.03967`

[39] G. H. Low and I. L. Chuang. Hamiltonian simulation by uniform spectral amplification. arXiv: `1707.05391`, 2017.

[40] G. H. Low and I. L. Chuang. Hamiltonian simulation by qubitization. *Quantum*, 3:163, 2019. arXiv: `1610.06546`

[41] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

[42] R. O'Donnell and J. Wright. Efficient quantum tomography. In *Proceedings of the 48th ACM Symposium on the Theory of Computing (STOC)*, pages 899–912, 2016. arXiv: `1508.01907`

[43] A. E. Rastegin. Relations for certain symmetric norms and anti-norms before and after partial trace. *Journal of Statistical Physics*, 148(6):1040–1053, 2012. arXiv: `1202.3853`

[44] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians (ICM)*, volume 3, pages 1576–1602, 2010. arXiv: 1003.2990
[45] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:1 – 9, 2013. arXiv: 1306.2872
[46] J. A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015. arXiv: arXiv:1501.01571
[47] Wikipedia. Computational complexity of mathematical operations. 2022. Accessed July 14, 2022.
[48] R. de Wolf. Quantum computing: Lecture notes, 2019. arXiv: 1907.09415
[49] H. Yuen. An improved sample complexity lower bound for quantum state tomography, 2022. arXiv: 2206.11185

A Direct mixed-state tomography using copies

We show how to perform mixed-state tomography with $\widetilde{\mathcal{O}}\big(rd^2/\varepsilon\big)$ copies of the state and a small amount of quantum power. Note that the well-known algorithm consisting of performing measurements in random bases already achieves this sample complexity, see the discussion in [23, 15, 34]; this is optimal for unentangled, non-adaptive algorithms [25]. Thus, the algorithm presented here does not improve over the known upper bounds. We discuss it anyway for the following reasons: first, the algorithm is easy to analyze (in our opinion, at least as simple, if not simpler, than [24]); second, the algorithm uses very similar techniques to Section 4 for pure states; third, it is fast to implement, because it only requires a few quantum gates and simple classical postprocessing.

Proposition A.1. *Let $\rho = \sum_{k=1}^r p_k|\psi_k\rangle\langle\psi_k|$ for some orthonormal $|\psi_k\rangle = \sum_{j\in[d]} \alpha_j^{(k)}|j\rangle$. There is a quantum algorithm that, given $\widetilde{\mathcal{O}}\big(rd^2/\varepsilon^2\big)$ copies of ρ and the ability to perform unitary operations on them, outputs $\tilde{\rho}$ such that $\|\rho - \tilde{\rho}\|_1 \leq \varepsilon$ with probability at least $2/3$. The algorithm is non-adaptive and does not require entangled measurements between copies of ρ.*

Proof. Recall that ρ is a $d \times d$ matrix with entries:

$$\rho_{u,v} = \sum_{k=1}^r p_k \alpha_u^{(k)} (\alpha_v^{(k)})^\dagger.$$

To avoid cumbersome equations, it is easier to analyze the algorithm by working with a purification $|\rho\rangle = \sum_{k=1}^r \sqrt{p_k}|\psi_k\rangle_A|\phi_k\rangle_B$ of ρ, where $|\phi_k\rangle$ are orthonormal; note that we never act on the purifying register, and the purification is solely for convenience. Add one fresh qubit in state $|0\rangle$ to the system; suppose it is the first. For $h \in [d]$, apply a Hadamard on the first qubit, followed by the unitary $|0\rangle\langle 0| \otimes I_A \otimes I_B + |1\rangle\langle 1| \otimes \sum_{j\in[d]} |(j-h) \bmod d\rangle\langle j| \otimes I_B$, and finally another Hadamard on the first qubit. In the following, for brevity we write $j+h$ instead of $(j+h) \mod d$: we use this notation only to index basis elements, so the context should avoid any ambiguity. The larger system is now described by the following pure state:

$$\frac{1}{2}|0\rangle \sum_{k=1}^r \sqrt{p_k}\left(\sum_{j\in[d]} (\alpha_j^{(k)} + \alpha_{j+h}^{(k)})|j\rangle|\phi_k\rangle\right) + \frac{1}{2}|1\rangle \sum_{k=1}^r \sqrt{p_k}\left(\sum_{j\in[d]} (\alpha_j^{(k)} - \alpha_{j+h}^{(k)})|j\rangle|\phi_k\rangle\right).$$

Next, we trace out the purifying register B, and compute the probability of finding the first qubit in state $|0\rangle$ and system A in state $|j\rangle$:

$$\begin{aligned}\frac{1}{4}\sum_{k=1}^r p_k(\alpha_j^{(k)} + \alpha_{j+h}^{(k)})(\alpha_j^{(k)} + \alpha_{j+h}^{(k)})^\dagger &= \frac{1}{4}\sum_{k=1}^r p_k\Big(|\alpha_j^{(k)}|^2 + 2\Re(\alpha_j^{(k)}(\alpha_{j+h}^{(k)})^\dagger) + |\alpha_{j+h}^{(k)}|^2\Big) \\ &= \frac{1}{4}(\rho_{j,j} + 2\Re(\rho_{j,j+h}) + \rho_{j+h,j+h}) = q_{0j}^{(h)}.\end{aligned}$$

Similarly, the probability of finding the first qubit in state $|1\rangle$ and system A in state $|j\rangle$ is:

$$\frac{1}{4}(\rho_{j,j} - 2\Re(\rho_{j,j+h}) + \rho_{j+h,j+h}) = q_{1j}^{(h)}.$$

By definition the vector $q^{(h)}$ represents a discrete probability distribution. We can obtain an ℓ_2-norm estimate $\tilde{q}^{(h)}$ of $q^{(h)}$ with error $\bar{\varepsilon}$ and with high probability taking $\widetilde{\mathcal{O}}\big(1/\bar{\varepsilon}^2\big)$ samples, see e.g., [10]. Note that for $h = 0$,

this immediately yields an estimate $(\tilde{\rho}_{0,0}, \ldots, \tilde{\rho}_{d-1,d-1})$ of the diagonal of ρ with ℓ_2-norm error at most $\bar{\varepsilon}$. For $h \in [d] \setminus \{0\}$, we can then compute an estimate $\tilde{\rho}_{j,j+h}$ for the real part of $\rho_{j,j+h}$ as $2(\tilde{q}_{0j}^{(h)} - \frac{1}{2}\tilde{\rho}_{j,j} - \frac{1}{2}\tilde{\rho}_{j+h,j+h})$. For convenience, let us call v the vector with entries $\rho_{j,j}$ for $j \in [d]$, $v^{(h)}$ the vector with entries $\rho_{j+h,j+h}$, and similarly for $\tilde{v}$ and $\tilde{v}^{(h)}$. The total ℓ_2-norm squared error for a set of d of these off-diagonal elements can be bounded as follows:

$$\sum_{j\in[d]} (\tilde{\rho}_{j,j+h} - \Re(\rho_{j,j+h}))^2 = 2\sum_{j\in[d]} \left((\tilde{q}_{0j}^{(h)} - \frac{1}{2}\tilde{\rho}_{j,j} - \frac{1}{2}\tilde{\rho}_{j+h,j+h}) - (q_{0j}^{(h)} - \frac{1}{2}\rho_{j,j} - \frac{1}{2}\rho_{j+h,j+h}) \right)^2 =$$

$$2\left\| (\tilde{q}_0^{(h)} - \frac{1}{2}\tilde{v} - \frac{1}{2}\tilde{v}^{(h)}) - (q_0^{(h)} - \frac{1}{2}v - \frac{1}{2}v^{(h)}) \right\|^2 \leq 2\Big(\left\| \tilde{q}_0^{(h)} - q_0^{(h)} \right\|^2 + \frac{1}{4}\|\tilde{v} - v\|^2 + \frac{1}{4}\left\| \tilde{v}^{(h)} - v^{(h)} \right\|^2 +$$

$$\frac{1}{2}\left\| \tilde{q}_0^{(h)} - q_0^{(h)} \right\| \|\tilde{v} - v\| + \frac{1}{2}\left\| \tilde{q}_0^{(h)} - q_0^{(h)} \right\| \left\| \tilde{v}^{(h)} - v^{(h)} \right\| + \frac{1}{4}\|\tilde{v} - v\| \left\| \tilde{v}^{(h)} - v^{(h)} \right\| \Big) \leq 6\bar{\varepsilon}^2,$$

where we use Cauchy-Schwarz plus the fact that $\left\| \tilde{q}_0^{(h)} - q_0^{(h)} \right\|$, $\|\tilde{v} - v\|$ and $\left\| \tilde{v}^{(h)} - v^{(h)} \right\|$ are all $\leq \bar{\varepsilon}$. This implies that we can get an $\mathcal{O}(\bar{\varepsilon})$-$\ell_2$-estimate of the real part of d elements of ρ with $\widetilde{\mathcal{O}}(1/\bar{\varepsilon}^2)$ samples. A similar approach, with the addition of a phase gate to multiply all coefficients by i, allows us to retrieve the imaginary part with the same complexity.

The above algorithm is repeated d times, for $h \in [d]$. Combining these d estimates of d coefficients each, setting $\bar{\varepsilon} = \varepsilon/\sqrt{d}$ and using a union bound for the probability of success, we obtain $\tilde{\rho}$ such that $\|\tilde{\rho} - \rho\|_F \leq \varepsilon$ taking $\widetilde{\mathcal{O}}(d^2/\varepsilon^2)$ samples. To convert from Frobenius norm to trace norm, using the fact that there are at most r nonzero eigenvalues by assumption, we need to decrease the error $\bar{\varepsilon}$ by a further factor $\sqrt{r}$. Then, this yields a trace-norm estimate of ρ with $\widetilde{\mathcal{O}}(rd^2/\bar{\varepsilon}^2)$ samples. ∎

B Implementing a QRAM

In this appendix we prove our claim that a d-qubit QRAM can be implemented with $\mathcal{O}(d)$ gates in $\mathcal{O}(\log(d))$ depth. Although QRAM implementations have been discussed at length in the literature, e.g. [21] and follow-up works, these discussions focus on the number of "activated" gates. While physically relevant in order to argue about error rates, from a complexity point of view there is no difference between an activated or non-activated gate.

We expect that the results below appear in the literature, but we were unable to locate them and hence proof them for completeness. If the reader is aware of earlier works with the same results, we would be grateful if the could inform us so that we can update this section to give proper attribution.

LEMMA B.1. *Let d be a power of 2. There is a unitary, called indexed-CNOT-out (stylized iCNOTo), acting on $\log(d) + 1 + d$ qubits plus $2d - 3$ ancillary qubits that can be implemented using $2d - 2 - 2\log(d)$ CNOT gates and $4d - 4$ Toffoli and X gates in $10\log(d)$ depth, and acts as follows on computational basis states*

$$iCNOTo|i\rangle|b\rangle|q_1\rangle \ldots |q_d\rangle = |i\rangle|b \oplus q_i\rangle|q_1\rangle \ldots |q_d\rangle.$$

There is also a unitary, called indexed-CNOT-in (stylized iCNOTi), acting on the same amount of qubits, that can be implemented in the same depth and number of gates, acting as

$$iCNOTi|i\rangle|b\rangle|q_1\rangle \ldots |q_d\rangle = |i\rangle|b\rangle|q_1\rangle \ldots |q_{i-1}\rangle|q_i \oplus b\rangle|q_{i+1}\rangle \ldots |q_d\rangle.$$

Proof. We first note that a FANOUT gate acting (for $a \in \{0, 1\}$) as

$$\text{FANOUT}|a\rangle|0^k\rangle = |a^k\rangle$$

can be build using $k - 1$ CNOT gates in depth $\log(k)$.

We will implement the ICNOTo gate as a tournament bracket. In the first step, if i is even then we first copy over all q_j for even j to a fresh layer of $d/2$ qubits. If i is odd then we do this for the odd j. The information whether i is even or odd is contained in its least significant bit, which, using a FANOUT to $d/2$ can be distributed

to $d/2$ fresh qubits in depth $\log(d) - 1$. Now, conditioned on the kth of these parity qubits either q_{2k} or q_{2k+1} is put in a fresh qubits, using 2 Toffoli gates and 2 X gates in depth 4.

We then do exactly the same circuit for the next layer, as if we were implementing a iCNOTo on $d/2$ qubits. After $\log(d)$ levels we end up with a (fixed) register in the state $|q_i\rangle$, and we can CNOT this value with $|b\rangle$. In fact, we can use $|b\rangle$ as the target for the final level, instead of a fresh qubit. After this we can uncompute all intermediate values using the same depth and gate count.

For the depth, note that all FANOUT gates can be performed in parallel. The deepest has depth $\log(d)$. The tournament bracket has depth 4 per layer, and $\log(d)$ layers. Including the uncompute the total depth is $10\log(d)$.

As for the ancillary qubits, there are $d-1$ parity bits used, one for each decision in the tournament bracket. There are $d-2$ intermediate bits used in the tournament, as we use b for the final result. Hence the circuit uses $2d-3$ ancillary qubits.

The CNOT count of all the fan outs is $\sum_{i=1}^{\log(d)} \left(\frac{d}{2^i} - 1\right) = d - 1 - \log(d)$. The tournament requires 2 Toffoli gates per decision, of which there are $d-1$, so the Toffoli count of this part is $2d-2$ (and the X count is the same). The total, including uncomputation becomes $2d - 2 - 2\log(d)$ CNOT gates, and $4d-4$ Toffoli and X gates.

The iCNOTi gate is implemented in almost the same way, but now b is distributed from the top of the tournament to the leave corresponding to q_i. ∎

There are two types of indexed SWAP that we may build. The first type has a fixed qubit that can be swaped with the ith qubit controlled on i. The second, most general indexed SWAP is controlled by both an i and j register and swaps the two. In the body of the paper we do not make this disintion, as there complexities are of the same order, but as the constant differ we will do so here.

LEMMA B.2. *Let d be a power of 2. There is a unitary, called single-indexed-SWAP (stylized iSWAP[9]), acting on $\log(d) + 1 + d$ qubits plus $2d-3$ ancillary qubits that can be implemented using $2d - 2 - 2\log(d)$ CNOT gates and $12d - 12$ Toffoli and X gates in $26\log(d)$ depth, and acts as follows on computational basis states*

$$iSWAP|i\rangle|b\rangle|q_1\rangle\dots|q_d\rangle = |i\rangle|q_i\rangle|q_1\rangle\dots|q_{i-1}\rangle|b\rangle|q_{i+1}\rangle\dots|q_d\rangle.$$

There is also a unitary, called double-indexed-SWAP (stylized iiSWAP), acting on $2\log(d) + d$ qubits plus $4d-5$ ancillary qubits, that can be implemented using $4d - 4 - 4\log(d)$ CNOT gates and $36d - 36$ Toffoli and X gates in $74\log(d)$ depth, and acts as follows on computational basis states

$$iiSWAP|i\rangle|j\rangle|q_1\rangle\dots|q_d\rangle = |i\rangle|j\rangle|q_1\rangle\dots|q_{i-1}\rangle|q_j\rangle|q_{i+1}\rangle\dots|q_{j-1}\rangle|q_i\rangle|q_{j+1}\rangle\dots|q_d\rangle.$$

Proof. For the iSWAP implementation we note that the SWAP gate can be implemented using 3 CNOT gates. In particular we can use two calls to iCNOTo and a single call to $iCNOTi$. Note that we can reuse the parity information bits and do not need to repeat the FANOUT.

For the iiSWAP, we note that we can perform a doubly indexed CNOT, i.e., a CNOT from qubit i to qubit j, by first retrieving the bit in the ith position with a iCNOTo, then running iCNOTi with index j, and then erasing the recovered bit with another call to iCNOTo. We can reuse the b bit for this. Again, 3 of these doubly indexed CNOTs are sufficient to implement a iiSWAP. We can again reuse the parity bits without redoing the FANOUT, but we have to implement these bits for both i and j. The stated counts follow. ∎

[9]Note that this is not related to the iSWAP gate that applies the phase i if qubits are swapped, sometimes discussed in the literature.