A Quantum Speed-Up for Approximating the Top Eigenvectors of a Matrix*

Yanlin Chen§ András Gilyén¶ Ronald de Wolf‖

Abstract

Finding a good approximation of the top eigenvector of a given $d \times d$ matrix A is a basic and important computational problem, with many applications. We give two different quantum algorithms that, given query access to the entries of a Hermitian matrix A and assuming a constant eigenvalue gap, output a classical description of a good approximation of the top eigenvector: one algorithm with time complexity $\tilde{\mathcal{O}}(d^{1.75})$ and one with time complexity $d^{1.5+o(1)}$ (the first algorithm has a slightly better dependence on the ℓ_2-error of the approximating vector than the second, and uses different techniques of independent interest). Both of our quantum algorithms provide a polynomial speed-up over the best-possible classical algorithm, which needs $\Omega(d^2)$ queries to entries of A, and hence $\Omega(d^2)$ time. We extend this to a quantum algorithm that outputs a classical description of the subspace spanned by the top-q eigenvectors in time $qd^{1.5+o(1)}$. We also prove a nearly-optimal lower bound of $\tilde{\Omega}(d^{1.5})$ on the quantum query complexity of approximating the top eigenvector.

Our quantum algorithms run a version of the classical power method that is robust to certain benign kinds of errors, where we implement each matrix-vector multiplication with small and well-behaved error on a quantum computer, in different ways for the two algorithms. Our first algorithm estimates the matrix-vector product one entry at a time, using a new "Gaussian phase estimation" procedure. Our second algorithm uses block-encoding techniques to compute the matrix-vector product as a quantum state, from which we obtain a classical description by a new time-efficient unbiased pure-state tomography procedure. This procedure uses an essentially optimal number $\tilde{\mathcal{O}}(d\log(d)/\varepsilon^2)$ of "conditional sample states"; if we have a state-preparation unitary available rather than just copies of the state, then this ε-dependence can be improved further quadratically. Our procedure comes with improved statistical properties and faster runtime compared to earlier pure-state tomography algorithms. We also develop an almost optimal time-efficient process-tomography algorithm for reflections around bounded-rank subspaces, providing the basis for our top-eigensubspace estimation algorithm, and in turn providing a pure-state tomography algorithm that only requires a reflection about the state rather than a state preparation unitary as input.

1 Introduction

Arguably the most important property of a diagonalizable $d \times d$ matrix A is its largest eigenvalue λ_1, with an associated eigenvector v_1. This top eigenvector v_1 can be thought of as the most important "direction" in which the matrix A operates. The ability to efficiently find v_1 is an important tool in many applications, for instance in the PageRank algorithm of Google's search engine, as a starting point for principal component analysis (for clustering or dimensionality-reduction), for Fisher discriminant analysis, or in continuous optimization problems where sometimes the best thing to do is to move the current point in the direction of the top eigenvector of an associated matrix [35, 36].

One way to find the top eigenvector of A is to diagonalize the whole matrix. Theoretically this takes matrix multiplication time: $\mathcal{O}(d^\omega)$ where $\omega \in [2, 2.37\ldots)$ is the still-unknown matrix multiplication exponent [61, 1]. In practice Gaussian elimination (which takes time $\mathcal{O}(d^3)$) is typically faster, unless d is enormous. Diagonalization gives us not only the top eigenvector but a complete orthonormal set of d eigenvectors. However, this is doing too much if we only care about finding the top eigenvector, or the top-q eigenvectors for some $q \ll d$, and better methods exist in this case (see e.g. [51] for a whole book about this).

---

*The full version of the paper can be accessed at https://arxiv.org/abs/2405.14765

§QuSoft, CWI, the Netherlands. yanlin.chen@cwi.nl

¶HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary. Funded by the EU's Horizon 2020 Marie Skłodowska-Curie program QuantOrder-891889 and the QuantERA II project HQCC-101017733 in coordination with the national funding organisation NKFIH. gilyen@renyi.hu

‖QuSoft, CWI and University of Amsterdam, the Netherlands. Partially supported by the Dutch Research Council (NWO) through Gravitation-grant Quantum Software Consortium, 024.003.037. rdewolf@cwi.nl

1.1 The power method for approximating a top eigenvector A quite efficient method for (approximately) finding the top eigenvector is the iterative "power method". This uses simple matrix-vector multiplications instead of any kind of matrix decompositions, and works as follows. We start with a random unit vector w_0 (say with i.i.d. Gaussian entries). This is a linear combination $\sum_{i=1}^{d} \alpha_i v_i$ of the d unit eigenvectors $v_1, \ldots, v_d$ of the Hermitian matrix A, with coefficients of magnitude typically around $1/\sqrt{d}$. Then we apply A to this vector some K times, computing $w_1 = Aw_0$, $w_2 = Aw_1$, etc., up to $w_K = A^K w_0$. This has the effect of multiplying each coefficient α_i with the powered eigenvalue λ_i^K. If there is some "gap" between the first two eigenvalues (say $|\lambda_1| - |\lambda_2| \geqslant \gamma > 0$, or $|\lambda_1/\lambda_2| > 1$), then the relative weight of the coefficient of v_1 will start to dominate all the other coefficients even already for small K, and the renormalization of the final vector $w_K = A^K w_0$ will be close to v_1, up to global phase. Specifically, if A has bounded operator norm and eigenvalue gap γ, then $K = \mathcal{O}\big(\log(d)/\gamma\big)$ iterations suffice to approximate v_1 up to $1/\mathrm{poly}(d)$ ℓ_2-error (see e.g. [27, Section 8.2.1] for details).

The cost of this algorithm is dominated by the K matrix-vector multiplications, each of which costs $\tilde{\mathcal{O}}\big(d^2\big)$ time classically (or $\tilde{\mathcal{O}}(m)$ time if A is sparse, with only m nonzero entries given in some easily-accessible way like lists of nonzero entries for each row and column). Hence if the eigenvalue gap γ is not too small, say constant or at least $1/\mathrm{polylog}(d)$, then the power method takes $\tilde{\mathcal{O}}(d^2)$ time to approximate a top eigenvector.[1] Unsurprisingly, as we show later in this paper, $\Omega(d^2)$ queries to the entries of A are also *necessary* for classical algorithms for this.

1.2 Our results: quantum algorithms Our main results in this paper are faster quantum algorithms for (approximately) finding the top eigenvector of a Hermitian matrix A, and nearly matching lower bounds.[2]

Quantum noisy power method. On a high level we just run the power method to find a good approximation for v_1, with classical representations of w_0 and all intermediate vectors, but we perform each matrix-vector multiplication *approximately* using a quantum computer.[3]

We give two different quantum algorithms for approximate matrix-vector multiplication. Our first algorithm uses that each entry of the vector Aw is an inner product between a row of A and the column-vector w. Because such an inner product is the sum of d numbers, we may hope to approximate it well via some version of amplitude estimation or quantum counting, using roughly $\sqrt{d}$ time per entry and $d^{1.5}$ time for all d entries of Aw together. This approach is easier said than done, because basic quantum-counting subroutines produce small errors in the approximation of each entry, and those errors might add up to a large ℓ_2-error in the d-dimensional vector Aw as a whole. To mitigate this issue we develop a "Gaussian phase estimation" procedure that can estimate one entry of Aw with a complexity that is similar to standard phase estimation, but with well-behaved *sub-Gaussian* error. These well-behaved errors in individual entries typically still add up to a large ℓ_2-error for the vector Aw as a whole. However, with very high probability the error remains small in one or a few fixed directions—including the direction of the unknown top eigenvector. To speed up the computation of each entry, we split the rows into "small" and "large" entries, and handle them separately. This divide-and-conquer approach uses $\tilde{\mathcal{O}}\big(d^{1.75}\big)$ time in total for the d entries of Aw (Theorem 4.4). This is asymptotically worse than our second algorithm (described below), but we still feel it merits inclusion in this paper because it uses an intuitive entry-by-entry approach, it has a slightly better dependence on the precision than our second algorithm, and most importantly our new technique of Gaussian phase estimation may find applications elsewhere.

Our second algorithm is faster and more "holistic". It does not approximate the matrix-vector product Aw entry-by-entry. Instead it implements a block-encoding of the matrix A, uses that to (approximately) produce Aw as a $\log(d)$-qubit state, and then applies a subtle tomography procedure to obtain a classical estimate of the vector Aw with small ℓ_2-error.[4] When used in our version of the power method, this classical vector is then stored

---

[1] The $\tilde{\mathcal{O}}(\cdot)$ notation suppresses polylog factors in d, ε, δ: $\tilde{\mathcal{O}}(T) = \mathcal{O}\big(T\mathrm{polylog}(d/(\varepsilon\delta))\big)$.

[2] If the matrix A is non-Hermitian, then we could instead use the Hermitian matrix $A' = \begin{bmatrix} 0 & A \\ A^\dagger & 0 \end{bmatrix}$ for finding the largest (left or right) singular value of A, replacing the eigenvalue gap with the singular value gap of A.

[3] *Exact* matrix-vector multiplication takes $\Omega(d^2)$ quantum queries to entries of A (and hence does not give speed-up). This is easy to see by taking $A \in \{0, 1/d\}^{d\times d}$ and $w = (\frac{1}{\sqrt{d}}, \ldots, \frac{1}{\sqrt{d}})^T$, because then $d^{1.5}Aw$ gives the number of nonzero entries in A. It is well-known that $\Omega(d^2)$ quantum queries are needed to count this number exactly.

[4] In fact our second algorithm first implements a block-encoding of the rank-1 projector $\Pi = v_1 v_1^\dagger$ using $\tilde{\mathcal{O}}(1/\gamma)$ applications of an approximate block-encoding of the matrix A (by applying QSVT [26, Theorem 31]), and then applies our state tomography algorithm to obtain a classical estimate of the vector Πw, which is proportional to v_1. This increases the spectral gap from γ to $\Theta(1)$ and hence further improves our γ-dependency.

in a QRAM-data structure (see below) that makes it easy to prepare Aw as a quantum state in the next iteration, which applies A again. Our new tomography procedure incorporates ideas from [39] and [5]; it matches the latter's essentially optimal query complexity, but improves upon both their time complexities. Doing tomography at the end of each iteration of the power method is somewhat expensive, but still leaves us with a time complexity of only $d^{1.5+o(1)}$ (Corollary 4.3 with $q = 1$ and sparsity set to $s = d$), which turns out to be a near-optimal quantum speed-up over classical, as our lower bounds (discussed below) imply.

It is worth highlighting our result about preparing the top eigenvector of A as a $\log(d)$-qubit quantum state $|v_1\rangle$, which we have "under the hood" in our second algorithm. This preparation can (with high success probability and small ℓ_2-error) be done in time roughly $(d/\gamma)^{1+o(1)}$ (Corollary 4.4 with sparsity set to $s = d$). This is optimal up to the $o(1)$ in the exponent; maybe surprisingly, for constant gap preparing $|v_1\rangle$ is not significantly more expensive than the easier task of just approximating the top eigen*value* λ_1, for which we prove an $\Omega(d)$ lower bound on the required number of quantum queries to entries of A (Proposition 4.1 with sparsity set to $s = d$). Note that putting some form of quantum state tomography on top of the preparation of $|v_1\rangle$ is *not* enough to achieve the time complexity of our second algorithm, as it would result in a quantum algorithm that produces a classical description of (an approximation of) v_1 in time roughly d^2 rather than roughly $d^{1.5}$.

In both of our algorithms, the vector resulting in each iteration from our approximate matrix-vector multiplication will have small errors compared to the perfect matrix-vector product. In the basic power method we cannot tolerate small errors in adversarial directions: if w_0 has roughly $1/\sqrt{d}$ overlap with the top eigenvector v_1, and we compute Aw_0 with ℓ_2-error $> \lambda_1/\sqrt{d}$, then our approximation to the vector Aw_0 may have *no overlap with v_1 at all anymore*! If this happens, if we lose the initially-small overlap with the top eigenvector, then the power method fails to converge to v_1 even if all later matrix-vector multiplications are implemented perfectly. Fortunately, Hardt and Price [32] have already shown that the power method is robust against errors in the matrix-vector computations if they are sufficiently well-behaved (in particular, entrywise sub-Gaussian errors with moderate variance suffice). Much of the technical effort in our quantum subroutines for matrix-vector multiplication is to ensure that the errors in the resulting vector are indeed sufficiently well-behaved not to break the noisy power method.[5]

Finding the top-q eigenvectors. Going beyond just the top eigenvector, the ability to find the top-q eigenvectors (with $q \ll d$) is crucial for many applications in machine learning and data analysis, such as spectral clustering, principal component analysis, low-rank approximation of A, and dimensionality reduction of d-dimensional data vectors w (e.g., projecting the data vectors onto the span of the top-q eigenvectors of the covariance matrix $A = \mathbb{E}[ww^T]$). In most cases it suffices to (approximately) find the subspace spanned by the top-q eigenvectors of A rather than the individual top-q eigenvectors $v_1, \ldots, v_q$, which is fortunate because distinguishing $v_1, \ldots, v_q$ can be quite expensive if the corresponding eigenvalues $\lambda_1, \ldots, \lambda_q$ are close together.

The noisy power method can also (approximately) find the subspace spanned by the top-q eigenvectors of A, assuming some known gap γ between the qth and $(q+1)$th eigenvalue. Using our knowledge of the gap, we give an algorithm to approximate λ_q (Corollary 4.2). Knowing λ_q and this gap (at least approximately), we then show how a block-encoding of A can be efficiently converted using quantum singular-value transformation [26] into a block-encoding of the projector Π that projects onto the subspace spanned by the top-q eigenvectors. In Section 4.4 we give a new almost optimal process-tomography algorithm for recovering the projector Π, assuming only the ability to apply (controlled) reflections $2\Pi - I$ about the rank-q subspace that we are trying to recover (Theorem 4.5). This algorithm, applied to Π, gives us the subspace corresponding to the top-q eigenvectors of A. For constant eigenvalue gap and desired precision, it uses time $qd^{1.5+o(1)}$. In fact, what we above called our "second algorithm" for finding the top eigenvector is just the special case $q = 1$. In the case where A is s-sparse (meaning each row and column of A has $\leqslant s$ nonzero entries) and we have sparse-query-access to it, the time complexity becomes $q\sqrt{s}d^{1+o(1)}$ (Corollary 4.3; this result implies the claim of the previous sentence by setting

[5]For simplicity we assume here that γ (or some sufficiently good approximation of it) is known to our algorithm. However, because we can efficiently approximate $|\lambda_1|$ (by [6, Lemma 50], one can estimate λ_1 with additive error $\gamma/4$ using $\tilde{\mathcal{O}}(d^{1.5}/\gamma)$ time) and verify whether the output of our algorithm is approximately an eigenvector for this eigenvalue by one approximate matrix-vector computation, we can actually try exponentially decreasing guesses for γ until the algorithm returns an approximate top eigenvector.

It should be noted that our algorithm has polynomial dependence on the precision ε, namely linear in $1/\varepsilon$, which is worse than the $\log(1/\varepsilon)$ dependence of the classical power method with perfect matrix-vector calculations. This is the price we pay for our polynomial speed-up in terms of the dimension. For applications where the precision need not be extremely small, our polynomial dependence on ε would be an acceptable price to pay.

$s = d$). If the pairwise spacing between the first q eigenvalues is at least $\Omega(1/q)$, then we can also (approximately) *find* each of the eigenvectors $v_1, \ldots, v_q$ individually, at the expensive of $\mathrm{poly}(q)$ more time.

As a byproduct of this algorithm we also obtain a qualitatively improved tomography procedure that works assuming the ability to reflect around the state that we want to estimate, but does not need the stronger assumption of being able to prepare that state.

Our computational model. The computational model for our quantum algorithms is that we have query access to the entries of A, which can be, e.g., stored in quantum-accessible classical memory ("quantum read-only memory", a.k.a. QROM). The runtime of a quantum algorithm, or of a classical algorithm with quantum subroutines, is measured by the total number of queries and other elementary operations (one- and two-qubit gates, classical RAM-operations) on the worst-case input.

Our algorithms also use $\tilde{\mathcal{O}}(d)$ bits of quantum-accessible classical-writable memory (QRAM, also sometimes called QCRAM) in order to store classical descriptions of the intermediate d-dimensional approximating vectors. In analogy with classical RAM, a QRAM is a device that stores some m-bit string $z = z_0 \ldots z_{m-1}$ and allows efficient access to the individual bits of z, also to several bits in superposition: a "read" operation corresponds to the unitary O_z that maps $|i, b\rangle \to |i, b \oplus z_i\rangle$, for $i \in [m] - 1$ and $b \in \{0, 1\}$. Note that the QRAM (and QROM) memory content itself is classical throughout the algorithm: it is just a string z, not a superposition of such strings. A "write" operation for the QRAM corresponds to changing a bit of z; this is a purely classical operation, and cannot happen in superposition. If only "read" operations are allowed, this is also sometimes called a QROM ("quantum read-only memory"), and typically classical input is assumed to be given in this form for the type of algorithms we and many others consider. The intuition is that read and write operations should be executable cheaply, for instance in time $\mathcal{O}(\log m)$ if we put the m bits of z on the leaves of a depth-$\log m$ binary tree, and we treat an address i as a bit-by-bit route from the root to the addressed leaf.

It should be noted that QRAM and QROM are controversial notions in some corners of quantum computing. In the case of QROM one could circumvent the controversy by assuming that the entries to our input matrix A are computed for us by a small circuit,[6] but for our use of QRAM this is not an option. The issue is not so much that a circuit for O_z for querying the m-bit string z uses roughly m gates (because the same is true for classical RAM and is not really considered an issue there) but that running such an operation fault-tolerantly on a superposition like $\sum_i \alpha_i |i, 0\rangle$ seems to induce large overheads. Classical RAM is not considered a problematic notion because nowadays it is implemented in very fast and practically error-free hardware without the need to do error-correction. It is conceivable that in the distant future quantum hardware implementations will become so good that one can similarly implement classical RAM on them with similar efficiency. Since one has to allow quantum superposition anyway in order to do anything in quantum computing, we feel that assuming QRAM is conceptually acceptable—especially for a theoretical computer science paper such as this one—though it is clearly not something for the small and noisy quantum computers we have today and in the near future.

We note, however, that in some situations we can actually avoid the use of QRAM altogether. In particular, we can make our second algorithm QRAM-free if we have some way of preparing the quantum state $|w\rangle$ corresponding to the intermediate approximating vectors (instead of using a KP-tree stored in QRAM for state-preparation). Given such a state-preparation procedure, the number of applications of the block-encoding U_Π in Theorem 4.5 remains the same. Accordingly, if we only care about the number of queries to entries of the input matrix (instead of time complexity), then we can get an $\tilde{\mathcal{O}}(d^{1.5+o(1)})$-vs-$\Omega(d^2)$ quantum-classical query-complexity separation for approximating the top eigenvector without using any QRAM, because we can prepare $|w\rangle$ from its classical description using a circuit of $\tilde{\mathcal{O}}(d)$ gates that uses no QRAM and no queries to entries of the input matrix. Also, our state-tomography procedure starting from "conditional samples" does not need any QRAM by itself.

1.3 Our results: lower bounds We also show that our second quantum algorithm for finding the top eigenvector is essentially optimal, by proving an $\tilde{\Omega}(d^{1.5})$ quantum query lower bound for this task. We do this by analyzing a hard instance $A = \frac{1}{d} uu^T + N$, which hides a vector $u \in \{-1, 1\}^d$ using a $d \times d$ matrix N with i.i.d. Gaussian entries of mean 0 and standard deviation $\sim 1/\sqrt{d}$. Note that the value $u_i u_j/d$ has magnitude $1/d$, but is "hidden" in the entry A_{ij} by adding noise to it of much larger magnitude $\sim 1/\sqrt{d}$. One can show that (with high probability) this matrix has a constant eigenvalue gap, and its top eigenvector is close to $u/\sqrt{d}$.

[6] Such an efficient procedure will, however, have to rely on structure present in the data, and unfortunately does not apply in general to problems where one aims to process some real-world data.

First, this hard instance provides the above-mentioned unsurprising[7] $\Omega(d^2)$ query lower bound for *classical* algorithms, as follows. To approximate the top eigenvector (and hence u), an algorithm has to recover most of the d signs u_i of u. Note that the entries of the ith row and column of A are the only entries that depend on u_i. If the algorithm makes T queries overall, then there is an index $\mathbf{i}$ such that the algorithm makes at most $4T/d$ queries to entries in the $\mathbf{i}$th row and column, while still recovering $u_\mathbf{i}$ with good probability. Slightly simplifying, one can think of each of those entries as a sample from either the distribution $N(1/d, 1/d)$ or the distribution $N(-1/d, 1/d)$, with the sign of the mean corresponding to $u_\mathbf{i}$. It is well known that $\Omega(d)$ classical samples are necessary to estimate the mean of the distribution to within $\pm 1/d$ and hence to distinguish between these two distributions. This implies $4T/d = \Omega(d)$, giving us the $T = \Omega(d^2)$ classical query lower bound for approximating the top eigenvector of A (Corollary 5.1).

Second, a similar but more technical argument works to obtain the $\tilde{\Omega}(d^{1.5})$ quantum query lower bound (Corollary 5.2), as follows. A good algorithm recovers most u_i-s with good probability. If it makes T quantum queries overall, then there is an index $\mathbf{i}$ such that the algorithm has at most $4T/d$ "query mass" on the entries of the $\mathbf{i}$th row and column (in expectation over the distribution of A), while still recovering $u_\mathbf{i}$ with good probability. It then remains to show that distinguishing between either the distribution $N(1/d, 1/d)$ or the distribution $N(-1/d, 1/d)$, with the ability to query multiple samples from that distribution in quantum superposition, requires $\tilde{\Omega}(\sqrt{d})$ quantum queries. This we prove by a rather technical modification of the adversary bound of Ambainis [2, 3], adjusted to inputs which are vectors of samples from a continuous distribution, using expectations under a joint distribution μ on pairs of matrices (the two marginal distributions of μ are our hard instance conditioned on $u_\mathbf{i} = 1$ and $u_\mathbf{i} = -1$, respectively) of Hamming distance roughly $\sqrt{d}$ in the $\mathbf{i}$th row and column.

1.4 Related work Our algorithms produce classical descriptions of the top eigenvector(s). This is very different from HHL-style [33] algorithms that return the vectors in the form of a $\log(d)$-qubit state whose vector of amplitudes is (proportional to) the desired vector. Our approach contrasts for instance with algorithms like quantum PCA [44], which can efficiently find the top-q eigenvectors *as quantum states* assuming the ability to prepare A as a mixed quantum state. It also contrasts with the recent work of Seki and Yunoki [57], who show how to apply a given Hermitian matrix many times to a given quantum state $|\psi\rangle$, giving rise to a quantum version of the power method that outputs quantum states. The most closely related quantum algorithm we are aware of is producing classical descriptions of eigenvectors in the special case where A is symmetric and diagonally-dominant (SDD): Apers and de Wolf [9, Claim 7.12] show how to approximately find the top-q eigenvalues and eigenvectors of a dense SDD matrix A in time $\tilde{\mathcal{O}}\big(d^{1.5} + qd\big)$, using their quantum speed-up for Laplacian linear solving.[8]

There is a fair amount of work on finding the largest (or smallest) *eigenvalue* of a given Hamiltonian A, but the setting there is usually different and incomparable to ours: A is viewed as acting on $\log(d)$ qubits, and classically specified as the sum of a small number of terms, each acting non-trivially on only $\mathcal{O}(1)$ of the qubits (this is the canonical QMA-complete problem). There is also work on finding the top eigenvalue of a given matrix in general (not necessarily a local Hamiltonian), for instance Lemma 50 of [6]; some of these works even involve a version of the power method [49]. However, none of these methods readily generalizes to finding a classical description of the top eigenvector itself.

A recent paper by Apers and Gribling [8, Theorem 5.1] also gives a quantum algorithm for approximate matrix-vector multiplication. Their result is incomparable to ours: it uses a different norm to measure the approximation, and it is geared towards the case of tall-and-skinny sparse matrices A; if instead the matrix A is $d \times d$ and dense (the regime we care about in this paper), then their time complexity can be roughly $d^{3.5}$, which is much worse than ours. Their application area is also different from ours: it is to speed up interior-point methods for linear programs where the number of constraints is much larger than the number of variables.

[7] A simpler way to see this lower bound is to consider the problem of distinguishing the all-0 matrix from a matrix that has a 1 in one of the d^2 positions and 0s elsewhere. This is just the d^2-bit OR problem, for which we have an easy and well-known $\Omega(d^2)$ classical query bound. However, the quantum analogue of this approach only gives an $\Omega(d)$ lower bound, since the quantum query complexity of d^2-bit OR is $\Theta(d)$. Therefore we present a more complicated argument for the classical lower bound whose quantum analogue *does* provide an essentially tight bound of $\tilde{\Omega}(d^{1.5})$.

[8] One can actually reduce the general symmetric A to the case of an SDD matrix by defining $A' = A + cI$ for sufficiently large c to make A' SDD, and then renormalizing A' to operator norm $\leqslant 1$. The problem with this reduction is that c could be as big as d and then the eigenvalue gap of the new matrix is much smaller than that of A.

One of our main tools for our upper bounds is a novel, essentially unbiased tomography procedure to estimate a d-dimensional pure state ("essentially unbiased" here means the error vector's expectation is exponentially close to 0) from $\tilde{\mathcal{O}}(d)$ "conditional samples" of the state, see our Section 3.2. There have already been quantum algorithms for essentially unbiased mean estimation for d-dimensional random variables [21, 58], and one might hope to use these to recover our tomography procedure in an easier way. However, their input models are somewhat different from ours and it is not clear how to tweak their algorithms to recover an unbiased state-tomography procedure that works in our input model. In a nutshell, [21, 58] require an oracle that outputs the random vectors in binary, i.e., with each coordinate explicitly written, while our algorithm only requires an exponentially smaller quantum state whose amplitudes represent the vector of interest.

1.5 Future work Here we mention some questions for future work. First, can we improve the d-dependence of our second algorithm from $d^{1.5+o(1)}$ to $d^{1.5}$? Note that the $o(1)$ comes from Low's Hamiltonian simulation result [45]; in his context it is also still an open question whether the $o(1)$ can be removed. Can we also improve our algorithm to use fewer or even no QRAM bits, while retaining the current dependence on d, γ, ε?

Second, our upper bound of roughly $qd^{1.5}$ for finding the subspace spanned by the top-q eigenvectors is essentially optimal for constant q, but it cannot be optimal for large q (i.e., $q = \Omega(d)$) because diagonalization finds all d eigenvectors exactly in time roughly $d^{2.37}$, which is less than $d^{2.5}$. We should try to improve our algorithm for large q.

Third, matrix-vector multiplication is a very basic and common operation in many algorithms. So far there has not been much work on speeding this up quantumly, possibly because easy lower bounds preclude quantum speed-up for *exact* matrix-vector multiplication (Footnote 3). Can we find other applications of our polynomially faster *approximate* matrix-vector multiplication? One such application is computing an approximate matrix-matrix product AB in time roughly $d^{2.5}$, by separately computing AB_i for each of the d columns B_i of B. This would not beat the current-best (but wholly impractical) matrix-multiplication techniques, which take time $d^{2.37\cdots}$, but it would be a very different approach for going beyond the basic $\mathcal{O}\big(d^3\big)$ matrix-multiplication algorithm. A related application is to matrix-product *verification*: we can decide whether AB is close to C in Frobenius norm for given $d \times d$ matrices A, B, C, in quantum time roughly $d^{1.5}$, by combining our approximate matrix-vector computation with Freivalds's algorithm [24]. This should be compared with the quantum algorithm of Buhrman and Špalek [16] that tests if AB is *equal* to C (over an arbitrary field) using $\tilde{\mathcal{O}}\Big(d^{5/3}\Big)$ time.

2 Preliminaries

Throughout the paper, d always denotes the dimension of the ambient space $\mathbb{R}^d$ or $\mathbb{C}^d$, log without a base means the binary logarithm, $\ln = \log_e$ is the natural logarithm, and $\exp(f) = e^f$. We let $[d]$ denote the set $\{1, \ldots, d\}$ and $[d]-1$ denote the set $\{0, \ldots, d-1\}$. We denote by $\|v\|$ the ℓ_2-norm of a vector, and by $\|v\|_p$ its ℓ_p-norm. For $A \in \mathbb{C}^{d \times d'}$, we define the spectral norm $\|A\| = \max\limits_{v \in \mathbb{C}^{d'}} \frac{\|Av\|}{\|v\|}$. For a set S, we define the indicator function $\mathbb{1}_S$ as

$$\mathbb{1}_S(x) = \begin{cases} 1 \text{ if } x \in S, \\ 0 \text{ otherwise.} \end{cases}$$

The *total variation distance* between probability distributions P and Q is defined as $d_{TV}(P,Q) = \sup_A P(A) - Q(A)$, where the supremum is over events A. In particular, for discrete distributions we have $d_{TV}(P,Q) = \frac{1}{2}\sum_x |P(x) - Q(x)|$. We say that two random variables are δ-*close* to each other if the total variation distance between their distributions is at most δ. For two distributions P, Q over the same space, the *relative entropy* $D_{KL}(P \,\|\, Q)$ (also called *Kullback-Leibler divergence* or *KL-divergence*) from P to Q is defined as

$$D_{KL}(P \,\|\, Q) = \int p(x) \cdot \ln \frac{p(x)}{q(x)} \mathrm{d}x = \mathbb{E}_p\left[\ln \frac{p(x)}{q(x)}\right],$$

where $p(x), q(x)$ are the probability density functions (pdf) of P and Q, respectively. In case P is not absolutely continuous with respect to Q we define $D_{KL}(P \,\|\, Q) = \infty$.

For us a projector is always a matrix Π which is idempotent ($\Pi^2 = \Pi$) and Hermitian ($\Pi^\dagger = \Pi$). This is sometimes called an "orthogonal projector" in the literature, but we drop the adjective "orthogonal" in order

to avoid confusion with orthogonality between a pair of projectors. For a subspace S we denote the unique (orthogonal) projector to S by Π_S.

2.1 Computational model and quantum algorithms Our computational model is a classical computer (a classical random-access machine) that can invoke a quantum computer as a subroutine. The input is stored in quantum-readable read-only memory (a QROM), whose bits (or more generally, entries, if one entry is a number with multiple bits) can be queried. Alternatively, we could assume we have some circuit that computes the input bits on demand. The classical computer can also write bits to a quantum-readable classical-writable classical memory (a QRAM). Such a QRAM, storing some m-bit string $z = z_0 \dots z_{m-1}$ admits *quantum queries* (a.k.a. quantum read operations), which correspond to the unitary O_z that maps $|i, b\rangle \to |i, b \oplus z_i\rangle$, where $i \in \{0, \dots, m-1\}$ and $b \in \{0, 1\}$. As already mentioned in the introduction, we think of one QRAM query as relatively "cheap", in the same way that a classical RAM query can be considered "cheap": imagine the m bits of z as sitting on the leaves of a $\log(m)$-depth binary tree, then reading the bit at location i intuitively just means going down the $\log(m)$-length path from the root to the leaf indicated by i.

The classical computer can send a description of a quantum circuit to the quantum computer; the quantum computer runs the circuit (which may include queries to the input bits stored in QROM and to the bits stored by the computer itself in the QRAM), measures the full final state in the computational basis, and returns the measurement outcome to the classical computer. In this model, an algorithm has *time complexity* T if in total it uses at most T elementary classical operations and quantum gates, quantum queries to the input bits stored in QROM, and quantum queries to the QRAM. The *query complexity* of an algorithm only counts the number of queries to the input stored in QROM. We call a (quantum) algorithm *bounded-error* if (for every possible input) it returns a correct output with probability at least $2/3$ (standard methods allow us to change this $2/3$ to any constant in $(1/2, 1)$ by only changing the complexity by a constant factor).

We will represent real numbers in computer memory using a number of bits of precision that is polylogarithmic in d/ε (i.e., $\tilde{\mathcal{O}}(1)$ bits). This ensures that all numbers are represented throughout our algorithms with negligible approximation error and we will ignore those errors later on for ease of presentation. In this paper, we are mainly interested in approximating the top eigenvector of a given matrix A. In the quantum case, we assume entries of A are stored in a QROM, which we can access by means of queries to the oracle $O_A : |ij\rangle|0\rangle \to |ij\rangle|A_{ij}\rangle$. In the special case where A is s-sparse, meaning that each of its rows and columns has at most s-nonzero entries, we additionally assume we can also query the location of the ℓth nonzero entry in the jth column. This is called "sparse-query-access to A", and is a common assumption for quantum algorithms working on sparse matrices (for instance in Hamiltonian simulation). This corresponds to storing the matrix A using an "adjacency list", i.e., the locations and values of the nonzero entries for each row and column in, a QROM.

2.2 Quantum subroutines In this section we describe a few known quantum algorithms that we invoke as subroutines.

THEOREM 2.1. ([26, 62], FIXED-POINT AMPLITUDE AMPLIFICATION) *Let $a, \delta > 0$, U be a unitary that maps $|0\rangle \to |\psi\rangle$ and $R_{\mathcal{A}}, R_{|0\rangle}$ be quantum circuits that reflect through subspaces $\mathcal{A}$ and (the span of) $|0\rangle$ respectively. Suppose $\|\Pi_{\mathcal{A}}|\psi\rangle\| \geqslant a$. There is a quantum algorithm that prepares $|\psi'\rangle$ satisfying $\||\psi'\rangle - \frac{\Pi_{\mathcal{A}}|\psi\rangle}{\|\Pi_{\mathcal{A}}|\psi\rangle\|}\| \leqslant \delta$ using a total number of $\mathcal{O}(\log(1/\delta)/a)$ applications of U, U^{-1}, controlled $R_{\mathcal{A}}, R_{|0\rangle}$ and additional single-qubit gates.*

We also use the following theorem to help us find all marked items in a d-element search space with high probability. The theorem was implicit in [29, 14]. For more details and for a better d, δ-dependency, see [7, Section 3].

THEOREM 2.2. *Let $f : [d] \to \{0, 1\}$ be a function that marks a set of elements $F = \{j \in [d] : f(j) = 1\}$, and $\delta \in (0, 1)$. Suppose we know an upper bound u on the size of F and we have a quantum oracle O_f such that $O_f : |j\rangle|b\rangle \to |j\rangle|b \oplus f(j)\rangle$. Then there exists a quantum algorithm that finds F with probability at least $1 - \delta$, using $\mathcal{O}(\sqrt{du} \cdot \operatorname{poly}\log(d/\delta))$ time.*

THEOREM 2.3. (FOLLOWS FROM SECTION 4 OF [15], AMPLITUDE ESTIMATION) *Let $\delta \in (0, 1)$. Given a natural number M and access to an $(n+1)$-qubit unitary U satisfying*

$$U|0^n\rangle|0\rangle = \sqrt{a}\,|\phi_0\rangle|0\rangle + \sqrt{1-a}\,|\phi_1\rangle|1\rangle,$$

where $|\phi_0\rangle$ and $|\phi_1\rangle$ are arbitrary n-qubit states and $a \in [0,1]$, there exists a quantum algorithm that uses $\mathcal{O}(M\log(1/\delta))$ applications of U and $U^\dagger$ and $\tilde{\mathcal{O}}(M\log(1/\delta))$ elementary gates, and outputs an estimator λ such that, with probability $\geqslant 1-\delta$,

$$|\sqrt{a}-\lambda| \leqslant \frac{1}{M}.$$

2.3 KP-tree and state-preparation Here we introduce a variant of the QRAM data structure developed by Prakash and Kerenidis [52, 38] for efficient state-preparation. We call this data structure a "KP-tree" (or KP_v if we are storing the vector v) in order to credit Kerenidis and Prakash. The variant we use is very similar to the one used in [17].

DEFINITION 2.1. (KP-TREE) *Let $v \in \mathbb{C}^d$. We define a KP-tree KP_v of v as follows:*

- *The root stores the scalar $\|v\|$ and the size of the support $t = |supp(v)|$ of v.*
- *KP_v is a binary tree on $\mathcal{O}(t\log d)$ vertices with depth $\lceil \log d \rceil$.*
- *The number of leaves is t; for each $j \in supp(v)$ there is one corresponding leaf storing v_j.*
- *Each edge of the tree is labeled by a bit; the bits on the edges of the path from the root to the leaf corresponding to the j^{th} entry of v form the binary description of j.*
- *Intermediate nodes store the square root of the sum of their children's squared absolute values.*

For $\ell \leqslant \lceil \log d \rceil$ and $j \in [2^\ell]-1$, we define $KP_v(\ell, j)$ as the scalar value stored in the j^{th} node in the ℓ^{th} layer, i.e., the node that we can reach from the root by the path according to the binary representation of j. If there is no corresponding j^{th} node in the ℓ^{th} layer (that is, we cannot reach a node by the path according to the binary representation of j from the root), then $KP_v(\ell, j)$ is defined as 0. Note that both the numbering of the layer and the numbering of nodes start from 0. When v is the all-0 vector, the corresponding tree consists of a single root node with $t = 0$.

If we have a classical vector $v \in \mathbb{C}^d$, we can build a KP-tree for v using $\tilde{\mathcal{O}}(d)$ time and QRAM bits. Given a KP-tree KP_v, we can efficiently query entries of v and prepare the quantum state $\sum_{j\in[d]-1} \frac{v_j}{\|v\|}|j\rangle$:

THEOREM 2.4. (MODIFIED THEOREM 2.12 OF [17]) *Suppose we have a KP-tree KP_v of $v \in \mathbb{C}^d$, and we can apply a unitary O_{KP_v} that maps $|\ell, k\rangle|0\rangle \to |\ell, k\rangle|KP_v(\ell, k)\rangle$. Then one can implement a unitary U_v that maps $|0\rangle$ to $|\psi_v\rangle = \sum_{j\in[d]-1} \frac{v_j}{\|v\|}|j\rangle$ up to error ε by using $\mathcal{O}(\log d)$ applications of O_{KP_v}, $O^\dagger_{KP_v}$, and $\tilde{\mathcal{O}}(1)$ additional elementary gates.*

Note that if $\|v\| < 1$, then we can similarly prepare the quantum state $|v\rangle = |0\rangle \sum_{j\in[d]-1} v_j|j\rangle + |1\rangle|0\rangle$ using $\tilde{\mathcal{O}}(1)$ time and queries to KP_v.

2.4 Block-encoding and Hamiltonian simulation Block-encoding embeds a scaled version of a (possibly non-unitary) matrix A in the upper-left corner of a bigger unitary matrix U.

DEFINITION 2.2. *Suppose that A is a 2^w-dimensional matrix, $\alpha, \varepsilon > 0$, and $a \in \mathbb{N}$. We call an $(a+w)$-qubit unitary U an (α, a, ε)-block-encoding of A if*

$$\|A - \alpha\langle 0^a| \otimes I_{2^w})U(|0^a\rangle \otimes I_{2^w})\| \leqslant \varepsilon.$$

THEOREM 2.5. ([46]) *Suppose that U is an $(\alpha, a, \varepsilon/|2t|)$-block-encoding of the Hamiltonian H. Then we can implement an ε-precise Hamiltonian-simulation unitary V which is a $(1, a+2, \varepsilon)$-block-encoding of e^{itH}, with $\mathcal{O}(|\alpha t| + \log(1/\varepsilon))$ uses of controlled-U and its inverse, and with $\mathcal{O}(a|\alpha t| + a\log(1/\varepsilon))$ additional elementary gates.*

THEOREM 2.6. ([45, THEOREM 2]) *Let A be a $d \times d$ Hermitian matrix with operator norm $\leqslant 1$, and $t > 0$. Suppose A has sparsity s and we have sparse-query-access to A. Then we can implement a unitary U such that $\|U - \exp(iAt)\| \leqslant \varepsilon$ using $\tilde{\mathcal{O}}\Big(t\sqrt{s}(t\sqrt{s}/\varepsilon)^{o(1)}\Big)$ time and queries.*

If we do not have a sparse oracle or if A is dense, then the time complexity of the above theorem simply becomes $\tilde{\mathcal{O}}((\sqrt{d}t)^{1+o(1)}/\varepsilon^{o(1)})$ by setting $s = d$.

THEOREM 2.7. ([26, COROLLARY 71]) *Let $\varepsilon \in (0, 1/2)$, A be a $d \times d$ Hermitian matrix with operator norm $\leqslant 1/2$, and $U = \exp(iA)$. Then we can implement a $(2/\pi, 2, \varepsilon)$-block-encoding of A, using $\mathcal{O}(\log(1/\varepsilon))$ applications of controlled-U, controlled-U inverse, $\mathcal{O}(\log(1/\varepsilon))$ time, and one auxiliary qubit.*

Combining the above two theorems, we have the following theorem.

THEOREM 2.8. *Let A be a $d \times d$ Hermitian matrix with operator norm $\leqslant 1$. Suppose A has sparsity s and we have sparse-query-access to A. Then we can implement a unitary U which is a $(4/\pi, 2, \varepsilon)$-block-encoding of A with $\tilde{\mathcal{O}}\Big(\sqrt{s}(s/\varepsilon)^{o(1)}\Big)$ time and queries.*

THEOREM 2.9. ([5], LEMMA 6) *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 $U_x : |0\rangle|0^{\otimes a}\rangle \to |0\rangle|\psi_x\rangle + |1\rangle|\tilde{\psi}_x\rangle$ and $V_x : |0\rangle|0^{\otimes a}\rangle \to |0\rangle|\phi_x\rangle + |1\rangle|\tilde{\phi}_x\rangle$ are $(a+1)$-qubit state-preparation unitaries for some (sub-normalized) a-qubit quantum states $|\psi_x\rangle, |\phi_x\rangle$. Then $(I \otimes V^\dagger)(SWAP \otimes I_{2^{a+1}})(I \otimes U)$ is a $(1, a+2, 0)$-block-encoding of the diagonal matrix $\mathrm{diag}(\{\langle\psi_x|\phi_x\rangle\})$, where the SWAP gate acts on the first and second qubits.*

2.5 Singular-value and singular-vector-perturbation bounds We invoke some tight bounds on the perturbation of singular values and singular vectors of a matrix. In the following we order the singular values of a matrix A in decreasing order, such that $i < j \Rightarrow \varsigma_i(A) \geqslant \varsigma_j(A)$.

THEOREM 2.10. (WEYL'S SINGULAR VALUE PERTURBATION BOUND [12, COR. III.2.6, PROBLEM III.6.13]) *Let $A, B \in \mathbb{C}^{n\times m}$ be any matrices, then for all $i \in [n]$ we have*

$$|\varsigma_i(A) - \varsigma_i(B)| \leqslant \|A - B\|.$$

In order to state the following perturbation bound on the singular-value subspaces we define Π_S^X to be the projector onto the subspace spanned by the left-singular vectors of X having singular values in S.

THEOREM 2.11. (WEDIN-DAVIS-KAHAN $\sin(\theta)$ THEOREM [60]) *Let $A, B \in \mathbb{C}^{n\times m}$ be any matrices, and $\alpha, \delta \geqslant 0$, then*[9]

$$\|(I - \Pi_{>\alpha}^A)\Pi_{\geqslant\alpha+\delta}^B\| \leqslant \frac{\|A - B\|}{\delta}.$$

LEMMA 2.1. (OPERATOR NORM EQUIVALENCE TO $\sin(\theta)$ BETWEEN SUBSPACES [12, AFTER EXER. VII.1.11]) *Let $P, Q \in \mathbb{C}^{n\times n}$ be projectors with equal rank, then $\|P - Q\| = \|P(I - Q)\| = \|(I - P)Q\|$.*

2.6 Concentration inequalities Repeated sampling is very important for our quantum tomography algorithms, and here we describe some of the tail bounds we need.

PROPOSITION 2.1. (BENNETT-BERNSTEIN BOUND [13, THEOREM 2.9 & EQN. 2.10]) *Let $X^{(i)} : i \in [n]$ be independent random variables with finite variance such that, for each i, $X^{(i)} \leqslant b$ for some $b > 0$ almost surely (i.e., this event has probability measure 1). Let*

$$S = \sum_{i=1}^n X^{(i)} - \mathbb{E}[X^{(i)}], \qquad v = \sum_{i=1}^n \mathbb{E}[(X^{(i)})^2],$$

[9] This bound is tight for any rank-r projectors A, B, when $\alpha = 0$ and $\delta = 1$ due to Lemma 2.1. Wedin's paper proves the statement for singular-vector subspaces that we use here, and the analogous statement for normal matrices is proven in Bhatia's book [12, Theorem VII.3.1], which actually also implies Theorem 2.11 with a bit of work.

then for any $t > 0$,

$$\Pr[S \geqslant t] \leqslant \exp\left(-\frac{v}{b^2}h\left(\frac{bt}{v}\right)\right) \leqslant \exp\left(-\frac{t^2}{2v + \frac{2}{3}bt}\right),$$

where $h(x) = (1+x)\ln(1+x) - x$.

PROPOSITION 2.2. (CHERNOFF-HOEFFDING BOUND [18], [34, THEOREM 1], [13, SECTION 2.6]) *Let $0 \leqslant X \leqslant 1$ be a bounded random variable and $p := \mathbb{E}[X]$. Suppose we take n i.i.d. samples $X^{(i)}$ of X and denote the normalized outcome by $s = \frac{X^{(1)}+X^{(2)}+\cdots+X^{(n)}}{n}$. Then we have for all $\varepsilon > 0$*

$$\Pr[s \geqslant p + \varepsilon] \leqslant e^{-D_{KL}(p+\varepsilon\|p)n} \leqslant \exp\left(-\frac{\varepsilon^2}{2(p+\varepsilon)}n\right), \tag{2.1}$$

$$\Pr[s \leqslant p - \varepsilon] \leqslant e^{-D_{KL}(p-\varepsilon\|p)n} \leqslant \exp\left(-\frac{\varepsilon^2}{2p}n\right), \tag{2.2}$$

where $D_{KL}(x \parallel p) = x\ln\frac{x}{p} + (1-x)\ln\left(\frac{1-x}{1-p}\right)$ is the Kullback–Leibler divergence between Bernoulli random variables with mean x and p respectively.

Proof. The first inequality is [34, Theorem 1], while (2.2) follows from (2.1) by considering $1 - X^{(i)}$. The rightmost inequalities come from the observation that $\forall x, y \geqslant 0$: $D_{KL}(x \parallel y) \geqslant \frac{(x-y)^2}{2\max\{x,y\}}$. ∎

COROLLARY 2.1. (OKAMOTO-HOEFFDING BOUND) *Let X, s be as in Proposition 2.2, then we have*

$$\Pr[\sqrt{s} \geqslant \sqrt{p} + \varepsilon] \leqslant \exp\Big(-2\varepsilon^2 n\Big),$$
$$\Pr[\sqrt{s} \leqslant \sqrt{p} - \varepsilon] \leqslant \exp\Big(-\varepsilon^2 n\Big).$$

Proof. This directly follows from (2.1)-(2.2) using the observation [50] that

$$D_{KL}(x \parallel p) \geqslant 2\big(\sqrt{x} - \sqrt{p}\big)^2 \qquad \forall\ 0 \leqslant p \leqslant x \leqslant 1,$$
$$D_{KL}(x \parallel p) \geqslant \ \big(\sqrt{p} - \sqrt{x}\big)^2 \qquad \forall\ 0 \leqslant x \leqslant p \leqslant 1.$$

∎

2.7 Matrix concentration inequalities We state some random matrix concentration results for tall matrices (i.e., matrices with more rows than columns), but in our case we mostly apply them to flat matrices (having more columns than rows), thus we effectively apply the statements to $G^\dagger$. We use the following non-asymptotic bounds.

THEOREM 2.12. (WELL-CONDITIONED TALL GAUSSIAN MATRICES [59, THEOREM 5.39 & FOOTNOTE 25]) *There exist absolute constants[10] $c, C \geqslant 1$ such that the following holds for all $N \geqslant n$: if $G \in \mathbb{C}^{N\times n}$ is a random matrix whose matrix elements have i.i.d. real or complex standard normal distribution (in the complex case with independent real and imaginary parts each having centered normal distribution with variance $\frac{1}{2}$), then its smallest singular value $\varsigma_{\min}(G)$ and largest singular value $\varsigma_{\max}(G)$ satisfy, for all $t \geqslant 0$*

$$\Pr[\sqrt{N} - C\sqrt{n} - t < \varsigma_{\min}(G) \leqslant \varsigma_{\max}(G) < \sqrt{N} + C\sqrt{n} + t] > 1 - 2\exp(-t^2/(2c)).$$

[10] In the real case [23, Theorem II.13] gives $c, C = 1$. While [59, Footnote 25] asserts that the statement can be adapted to the complex case, there are no specifics provided, and the proof of [59, Corollary 5.35] is borrowed from [23, Theorem II.13], where the adaptation of the statement to the complex case is presented as an open question. It is tempting to try and adapt the proof of the real case using the observation that $\varsigma_{\min}(G) = \min_{\|u\|=1}\max_{\|v\|=1}\Re\langle u^*, Gv\rangle$ together with the Slepian-Gordon lemma [28, Theorem 1.4], however this approach seems to irrecoverably fail due to a banal issue: while $\||u\rangle|v\rangle - |u'\rangle|v'\rangle\|^2 \leqslant \|u-u'\|^2 + \|v-v'\|^2$ holds for real unit vectors, for complex unit vectors only the weaker $\||u\rangle|v\rangle - |u'\rangle|v'\rangle\|^2 \leqslant 2\|u-u'\|^2 + 2\|v-v'\|^2$ holds in general. Indeed, for any $\alpha \in (0, 2\pi)$ consider the complex numbers $u = 1, v = i\exp(i\alpha), u' = \exp(-i\alpha), v' = i$, then we have $|uv - u'v'|^2/(|u-u'|^2 + |v-v'|^2) = 1 + \cos(\alpha)$.

COROLLARY 2.2. *In the setting of Theorem 2.12, if $N \geqslant 16C^2 n$, we have*

$$\Pr\left[\frac{1}{4}\sqrt{N} < \varsigma_{\min}(G) \leqslant \varsigma_{\max}(G) < \frac{7}{4}\sqrt{N}\right] > 1 - 2\exp(-N/(8c)).$$

Proof. Apply Theorem 2.12 with $t = \sqrt{N}/2$. ∎

The following slightly tighter bound can be used for bounding the norm of individual columns or rows of Gaussian random matrices.

PROPOSITION 2.1. *Let v be an n-dimensional random vector whose coordinates have i.i.d. real or complex standard normal distribution, then $\mathbb{E}[\|v\|] \leqslant \sqrt{n}$ and $\Pr[\|v\| \geqslant \sqrt{n} + t] \leqslant \exp(-\frac{t^2}{2})$ $\forall t \geqslant 0$.*

Proof. We have $\mathbb{E}[\|v\|] \leqslant \sqrt{\mathbb{E}[\|v\|^2]} = \sqrt{n}$ by Jensen's inequality. The function $v \mapsto \|v\|$ is 1-Lipschitz due to the triangle inequality, and hence by the concentration of Lipschitz functions on vectors with the canonical Gaussian measure (Proposition 2.18 & Equation (2.35) of [42]) we have $\Pr[\|v\| \geqslant \sqrt{n} + t] \leqslant \exp(-t^2/2)$.[11] ∎

Next we invoke a concentration bound for the operator norm of a random matrix with independent bounded rows.

THEOREM 2.13. (INDEPENDENT BOUNDED ROWS [59, THEOREM 5.44 & REMARK 5.49]) *There exists an absolute constant $c' > 0$ such that the following holds. Let $G \in \mathbb{C}^{N \times n}$ be a random matrix whose rows G_i are independent, each having mean 0 and a covariance matrix*[12] *with operator norm at most S^2, and almost surely $\|G_i\|_2 \leqslant B$ for all $i \in [N]$. Then for every $t \geqslant 0$, with probability at least $1 - 2n\exp(-c't^2)$ one has*

$$\|G\| \leqslant 2|S|\sqrt{N} + tB. \tag{2.3}$$

In case G is a real-symmetric Gaussian matrix, we have the following non-asymptotic bound.

THEOREM 2.14. (SYMMETRIC GAUSSIAN MATRIX [10, COROLLARY 3.9 WITH $\varepsilon = 0.25$]) *Let $G \in \mathbb{R}^{d \times d}$ be a symmetric matrix with $G_{ij} = b_{ij} \cdot g_{ij}$, where the random variables $\{g_{ij} : i \geqslant j\}$ are i.i.d. $\sim N(0,1)$ and the $\{b_{ij} : i \geqslant j\}$ are arbitrary real scalars. Denote $b_{max} = \max_i \sqrt{\sum_j b_{ij}^2}$ and $b^*_{max} = \max_{ij} |b_{ij}|$. Then for every $t \geqslant 0$,*

$$\Pr\left[\|G\| \geqslant 2.5 \cdot b_{max} + \frac{7.5}{\ln(1.25)} b^*_{max}\sqrt{\ln d} + t\right] \leqslant \exp(-t^2/(4{b^*}^2_{max})).$$

2.8 Bounds on random variables with adaptive dependency structure Here we present some useful bounds on random variables, where the dependency structure follows some martingale-like structure.

The first bound gives an intuitive total variation distance bound for "adaptive" processes. The main idea is to couple two adaptive random processes that are step-wise similar. We use the following folklore [4, p. 1] observation:

THEOREM 2.15. (TOTAL VARIATION DISTANCE AND OPTIMAL COUPLING) *Let X, Y be random variables. Then $d_{TV}(X,Y) \leqslant \varepsilon$ iff there exist ε-coupled random variables $\widetilde{X}$ and $\widetilde{Y}$ (possibly dependent) with the same distribution as X and Y, respectively, such that $\Pr[\widetilde{X} \neq \widetilde{Y}] \leqslant \varepsilon$.*

COROLLARY 2.3. *Let X be a random variable and A an event of the underlying probability space such that $\Pr[A] > 0$. Then for $Y' = X|A$, the conditioned version of X (i.e., $\Pr[Y' \in S] = \Pr[A \,\&\, X \in S]/\Pr[A]$ for all measurable sets S), we have that $d_{TV}(X, Y') \leqslant 1 - \Pr[A]$.*

[11] Actually, in the complex case the upper bound is even stronger: $\exp(-t^2)$.

[12] If $\psi \in \mathbb{C}^n$ is a mean-0 random vector and $C = \mathbb{E}[\psi^\dagger \psi]$ is its covariance matrix, then the covariance matrix of the complex conjugate random variable ψ^* is $\mathbb{E}[\psi^T \psi^*] = C^*$. On the other hand we have $\mathrm{Cov}(\Re(\psi)) + \mathrm{Cov}(\Im(\psi)) = \frac{C + C^*}{2}$, and therefore $\|\mathrm{Cov}(\Re(\psi)) + \mathrm{Cov}(\Im(\psi))\| \leqslant \|C\|$. Thus we can apply [59, Theorem 5.44 & Remark 5.49] separately to the real and imaginary parts of the random vectors G_i, whence the extra (possibly sub-optimal) factor of 2 in (2.3).

Proof. Let Y be a random variable that is independent of A, but its distribution is identical to that of Y', i.e., $\Pr[Y' \in S] = \Pr[Y \in S] = \Pr[A \,\&\, X \in S]/\Pr[A]$ for all measurable sets S. In Theorem 2.15 take $\widetilde{X} := X$, and $\widetilde{Y} := X$ on A and $\widetilde{Y} := Y$ on the complement of A; by construction we have $\Pr[\widetilde{X} \neq \widetilde{Y}] \leqslant 1 - \Pr[A]$. Finally, observe that for all measurable sets S we have

$$\begin{aligned}\Pr[\widetilde{Y} \in S] &= \Pr[A \,\&\, X \in S] + \Pr[\bar{A} \,\&\, Y \in S] = \Pr[A \,\&\, X \in S] + \Pr[\bar{A}] \cdot \Pr[Y \in S] \\ &= \left(1 + \frac{\Pr[\bar{A}]}{\Pr[A]}\right) \Pr[A \,\&\, X \in S] = \frac{\Pr[A \,\&\, X \in S]}{\Pr[A]} = \Pr[X \in S \mid A] = \Pr[Y' \in S].\end{aligned}$$

∎

LEMMA 2.2. (CONDITIONAL TOTAL VARIATION DISTANCE BASED BOUND) *Let $X = (X_1, X_2)$ and $X' = (X'_1, X'_2)$ be two discrete random variables, and let $X_{2x} := X_2|X_1 = x$, $X'_{2x} := X'_2|X'_1 = x$. Suppose that $d_{TV}(X_1, X'_1) \leqslant \varepsilon_1$ and $d_{TV}(X_{2x}, X'_{2x}) \leqslant \varepsilon_2$ for all x such that $\Pr[X_1 = x]\Pr[X'_1 = x] > 0$, then $d_{TV}(X, X') \leqslant \varepsilon_1 + (1 - \varepsilon_1)\varepsilon_2$.*

Proof. Due to Theorem 2.15 we can find ε_1-coupled random variables $\widetilde{X}_1$, $\widetilde{X}'_1$, and similarly ε_2-coupled $\widetilde{X}_{2x}$, $\widetilde{X}'_{2x}$ for all x in the range of X_1, X'_1 respectively. We can assume without loss of generality that $\widetilde{X}_1$ and $\widetilde{X}_{2x}$ are mutually independent for all x in the range of X_1, X'_1 and likewise are $\widetilde{X}_1$ and $\widetilde{X}_{2x}$. We then define $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_{2\widetilde{X}_1})$ and $\widetilde{X}' = (\widetilde{X}'_1, \widetilde{X}'_{2\widetilde{X}'_1})$, so that clearly $\Pr[X = (x_1, x_2)] = \Pr[\widetilde{X} = (x_1, x_2)]$ and $\Pr[X' = (x_1, x_2)] = \Pr[\widetilde{X}' = (x_1, x_2)]$. On the other hand due to the tight coupling of Theorem 2.15 we also get

$$\begin{aligned}d_{TV}(X, X') &\leqslant \Pr[\widetilde{X} \neq \widetilde{X}'] \\ &= \Pr[\widetilde{X}_1 \neq \widetilde{X}'_1] + \sum_x \Pr[\widetilde{X}_1 = \widetilde{X}'_1 = x \,\&\, \widetilde{X}_{2x} \neq \widetilde{X}'_{2x}] \\ &= \Pr[\widetilde{X}_1 \neq \widetilde{X}'_1] + \sum_x \Pr[\widetilde{X}_1 = \widetilde{X}'_1 = x]\Pr[\widetilde{X}_{2x} \neq \widetilde{X}'_{2x}] \\ &\leqslant \varepsilon_1 + \sum_x \Pr[\widetilde{X}_1 = \widetilde{X}'_1 = x]\varepsilon_2 \\ &= \varepsilon_1 + (1 - \varepsilon_1)\varepsilon_2.\end{aligned}$$

∎

The following is essentially a martingale property, which could be stated more generally, but here we prove a simple version for completeness.

LEMMA 2.3. (MARTINGALE-LIKE COVARIANCE SUM) *If $X, Y, Z \in \mathbb{C}^d$ are vector-valued discrete random variables such that $\mathbb{E}[Z|(X, Y)] = 0$ (i.e., $\mathbb{E}[Z|(X = x, Y = y)] = 0$ for all x, y), then $\mathrm{Cov}(X + Z) = \mathrm{Cov}(X) + \mathrm{Cov}(Z)$.*

Proof. It is easy to see that $\mathbb{E}[Z] = 0$, and we can assume without loss of generality that $\mathbb{E}[X] = 0$.

$$\begin{aligned}\mathrm{Cov}(X + Z) &= \mathbb{E}[|X + Z\rangle\langle X + Z|] \\ &= \sum_{\substack{(x,y):\\ \Pr[(X,Y)=(x,y)]>0}} \Pr[(X, Y) = (x, y)]\mathbb{E}[|x + Z\rangle\langle x + Z| \,|(X, Y) = (x, y)] \\ &= \sum_{\substack{(x,y):\\ \Pr[(X,Y)=(x,y)]>0}} \Pr[(X, Y) = (x, y)]\big(|x\rangle\langle x| + \mathbb{E}[|Z\rangle\langle Z| \,|(X, Y) = (x, y)]\big) \\ &= \sum_x \Pr[X = x]|x\rangle\langle x| + \sum_{\substack{(x,y):\\ \Pr[(X,Y)=(x,y)]>0}} \Pr[(X, Y) = (x, y)]\mathbb{E}[|Z\rangle\langle Z| \,|(X, Y) = (x, y)] \\ &= \mathrm{Cov}(X) + \mathrm{Cov}(Z).\end{aligned}$$

∎

2.9 Gaussian, Sub-Gaussian, and discrete Gaussian distributions A random variable X over $\mathbb{R}$ has Gaussian distribution with mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathrm{Var}(X)$, denoted $X \sim N(\mu, \sigma^2)$, if its probability density function is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \text{ for } x \in \mathbb{R}.$$

A well-known property of a Gaussian is that its tail decays rapidly, which can be quantified as:

- If $X \sim N(\mu, \sigma^2)$, then for any $t > 0$, it holds that

$$\Pr[X - \mu > t] \leqslant \frac{1}{\sqrt{2\pi}t} \exp(-t^2/(2\sigma^2)).$$

A random variable X over $\mathbb{R}$ is called τ-sub-Gaussian with parameter $\alpha > 0$, denoted in short by $X \sim \tau\text{-subG}(\alpha^2)$, if its moment-generating function satisfies $\mathbb{E}[\exp(tX)] \leqslant \exp(\tau)\exp(\alpha^2 t^2/2)$ for all $t \in \mathbb{R}$. We list and prove a few useful properties of sub-Gaussian distributions below.

- The tails of $X \sim \tau\text{-subG}(\alpha^2)$ are dominated by a Gaussian with parameter α, i.e.,

$$\Pr[|X| > t] \leqslant 2\exp(\tau)\exp(-t^2/(2\alpha^2)) \quad \forall t > 0. \tag{2.4}$$

- If $X_i \sim \tau\text{-subG}(\alpha_i^2)$ are independent, then for any $a = (a_1, \ldots, a_d)^T \in \mathbb{R}^d$, the weighted sum $\sum_{i\in[d]} a_i X_i$ is $(\sum_{i\in[d]} \tau_i)$-sub-Gaussian with parameter $\tilde{\alpha} = \sqrt{\sum_{i=1}^{d} a_i^2 \alpha_i^2}$.

To prove (2.4), we first use Markov's inequality to obtain that for all $s > 0$

$$\Pr[X \geqslant t] = \Pr[\exp(sX) \geqslant \exp(st)] \leqslant \mathbb{E}[\exp(sX)]/\exp(st) \leqslant \exp(\tau)\exp(\alpha^2 s^2/2 - st).$$

Since the above inequality holds for every $s > 0$, we have $\Pr[X \geqslant t] \leqslant \exp(-t^2/(2\alpha^2))$ because $\min_{s>0}(\alpha^2 s^2/2 - st) = -t^2/(2\alpha^2)$. The same argument applied to $-X$ gives the bound on $\Pr[X \leqslant -t]$.

The second property can be easily derived using the independence of the X_i's:

$$\mathbb{E}[\exp(t\sum_{i=1}^{d} a_i X_i)] = \prod_{i=1}^{d} \mathbb{E}[\exp(t a_i X_i)] \leqslant \prod_{i=1}^{d} \exp(\tau_i)\exp(\alpha_i^2 a_i^2 t^2/2) = \exp(\sum_{i=1}^{d} \tau_i)\exp(\sum_{i=1}^{d} \alpha_i^2 a_i^2 t^2/2).$$

Next, we explain what a *discrete* Gaussian is. For any $s > 0$, we define the function $p_s \colon \mathbb{R} \to \mathbb{R}$ as $p_s(x) = \exp(-\pi x^2/s^2)$. When $s = 1$, we simply write $p(x)$. For a countable set S we define $p_s(S) = \sum_{a\in S} p_s(a)$; if $p_s(S) < \infty$, we define $\mathcal{D}_{S,s} = \frac{1}{p_s(S)} p_s$ to be the discrete probability distribution over S such that the probability of drawing $x \in S$ is proportional to $p_s(x)$. We call this the discrete Gaussian distribution over S with parameter s. The following theorem states that such distributions over $\mathbb{Z}$ are sub-Gaussian with parameter s:

THEOREM 2.16. ([47, LEMMA 2.8]) *For any $\lambda, \nu > 0$ and $\tau \in (0, 0.1]$, if $s \geqslant \lambda\sqrt{\log(12/\tau)/\pi}$,*[13] *then $\mathcal{D}_{\lambda\mathbb{Z}+\nu,s}$ is τ-sub-Gaussian with parameter s. Moreover, for every $s > 0$, $\mathcal{D}_{\lambda\mathbb{Z},s}$ is 0-sub-Gaussian with parameter s.*

We also consider truncations of the above infinite discrete Gaussian distributions, and define $\mathcal{D}_{\mathbb{Z},s}^{[-L,R]}$ as $\mathcal{D}_{\mathbb{Z}\cap[-L,R],s}$. For every $\delta \in (0, 1]$, if $L, R \geqslant s\sqrt{2\ln(1/\delta)}$, then by Corollary 2.3 we have $d_{TV}(\mathcal{D}_{\mathbb{Z},s}, \mathcal{D}_{\mathbb{Z},s}^{[-L,R]}) \leqslant 2\delta$, because the tail of $\mathcal{D}_{\mathbb{Z},s}$ can be bounded by 2δ using (2.4) due to Theorem 2.16. We also define $\mathcal{D}_{\mathbb{Z},s}^{\bmod N}$ as a

[13] Here the lattice we consider is the one-dimensional lattice $\lambda\mathbb{Z}$, so the length of the shortest nonzero vector $\lambda_1(\lambda\mathbb{Z})$ is just λ. Also, by using the equation between the smoothing parameter and the length of the shortest vector [48, Lemma 3.3], we have $\eta_\tau(\lambda\mathbb{Z}) \leqslant \sqrt{\log(2 + 2/\tau)/\pi} \cdot \lambda_1(\lambda\mathbb{Z}) \leqslant \lambda\sqrt{\log(3/\tau)/\pi}$. Therefore, if $s \geqslant \lambda\sqrt{\log(3/\tau)/\pi}$, then $s \geqslant \eta_\tau(\lambda\mathbb{Z})$ and hence $\mathcal{D}_{\lambda\mathbb{Z}+\nu,s}$ is $\log((1+\tau)/(1-\tau))$-sub-Gaussian. By the fact $\log((1+\tau)/(1-\tau)) \leqslant 4\tau$ and changing $\tau \to \tau/4$, we get the theorem as stated here.

modular version of the discrete Gaussian distribution $\mathcal{D}_{\mathbb{Z},s}$, which has probability $\mathcal{D}_{\mathbb{Z},s}^{\mathrm{mod}\ N}(k) = \mathcal{D}_{\mathbb{Z},s}(N\cdot\mathbb{Z}+k)$ for every $k \in \{-\lfloor N/2\rfloor, \dots, 0, \dots, \lceil N/2\rceil - 1\}$, and probability 0 for all other $k \in \mathbb{Z}$. Again, by using Theorem 2.16 and (2.4), we have that for every $\delta \in (0,1/2)$, if $N \geqslant 2s\sqrt{2\ln(1/\delta)}$, then $d_{TV}(\mathcal{D}_{\mathbb{Z},s}, \mathcal{D}_{\mathbb{Z},s}^{\mathrm{mod}\ N}) \leqslant 2\delta$. Combining these we get $d_{TV}(\mathcal{D}_{\mathbb{Z},s}^{\mathrm{mod}\ N}, \mathcal{D}_{\mathbb{Z},s}^{[-L,R]}) \leqslant 4\delta$ if $N, L, R \geqslant 2s\sqrt{2\ln(1/\delta)}$. Finally, these arguments can also be applied to $\mathcal{D}_{\mathbb{Z}+c,s}$ with large enough s.

COROLLARY 2.4. *Let $\delta \in (0,1]$. For any $c > 0$ and $\tau \in (0, 0.1]$, if $s \geqslant \sqrt{\log(12/\tau)/\pi}$ and $N, L, R \geqslant 10s\sqrt{2\ln(2/\delta)}$, then $\mathcal{D}_{\mathbb{Z}+c,s}$, $\mathcal{D}_{\mathbb{Z}+c,s}^{[-L,R]}$, and $\mathcal{D}_{\mathbb{Z}+c,s}^{mod\ N}$ are $4\delta\exp(\tau)$-close to each other in total variation distance.*

Using the discussion and corollary above, we can show the truncated discrete Gaussian state is close to the modular discrete Gaussian state when N, s are both reasonably large.

THEOREM 2.17. *Let $\delta \in (0, 0.1]$, $s \geqslant 8\sqrt{2\log(\frac{1}{\delta})}$, $N \geqslant 16s\sqrt{2\ln(\frac{1}{\delta})}$ even number, and $t \in [-\frac{N}{8}, \frac{N}{8}]$. Let $f : \mathbb{R} \to \mathbb{C}$ be an arbitrary phase function such that $|f(x)| = 1$ for every $x \in \mathbb{R}$ and*

$$\begin{aligned}
|G\rangle &= \frac{1}{\sqrt{G}} \sum_{x\in\mathbb{Z}} f(x+t) p_s(x+t)\,|x\rangle,\\
|G^{tr}\rangle &= \frac{1}{\sqrt{G_{tr}}} \sum_{x\in\{-\frac{N}{2}\dots,0,\dots,\frac{N}{2}-1\}} f(x+t) p_s(x+t)\,|x\rangle,\\
|G^{mod}\rangle &= \frac{1}{\sqrt{G_{mod}}} \sum_{x\in\mathbb{Z}} f(x+t) p_s(x+t)\,|(x+\frac{N}{2}\ \ mod\ N) - \frac{N}{2}\rangle,
\end{aligned}$$

where G, G_{mod}, G_{tr} are normalizing factors. Then $|G^{tr}\rangle, |G^{mod}\rangle$, $|G\rangle$ are 9δ-close to each other.[14]

Proof. Let $[\![\pm a]\!]$ denote the set $\{-\lfloor a\rfloor \dots, 0, \dots, \lceil a\rceil - 1\}$. We first show $|G^{tr}\rangle$ is close to $|G\rangle$. Define $|\widetilde{G^{tr}}\rangle = \sqrt{\frac{G_{tr}}{G}}\,|G^{tr}\rangle$, which has ℓ_2-norm $\sqrt{\frac{G_{tr}}{G}} \leqslant 1$. We can see

$$\| \,|\widetilde{G^{tr}}\rangle - |G\rangle\,\|^2 = \frac{1}{G} \sum_{x\in\mathbb{Z}\setminus[\![\pm\frac{N}{2}]\!]} p_s^2(x+t) \leqslant \frac{1}{G} \sum_{x\in\mathbb{Z}\setminus\{x:|x+t|\leqslant N/4\}} p_s^2(x+t) \leqslant \delta^2,$$

where the last equality holds because $p_s^2 = p_{s/\sqrt{2}}$ and $\mathcal{D}_{\mathbb{Z}+t,s/\sqrt{2}}$ is δ^2-sub-Gaussian with parameter $s/\sqrt{2}$ by Theorem 2.16 (note $s/\sqrt{2} \geqslant 8\sqrt{\log(1/\delta)} \geqslant \sqrt{\log(12/\delta^2)/\pi}$) and because of the first property of sub-Gaussians: $\Pr[|x+t| > N/4] \leqslant 2\exp(\delta^2)\exp(-(N/4)^2/(2\cdot s^2/2)) \leqslant \delta^2$. Note that $1 \geqslant \| \,|\widetilde{G^{tr}}\rangle\,\| \geqslant \| \,|G\rangle\,\| - \| \,|\widetilde{G^{tr}}\rangle - |G\rangle\,\| \geqslant 1-\delta$ and hence $\| \,|\widetilde{G^{tr}}\rangle - |G^{tr}\rangle\,\| = \|(\sqrt{\frac{G_{tr}}{G}} - 1)\,|G^{tr}\rangle\,\| \leqslant \delta$. Therefore, we obtain

$$\| \,|G^{tr}\rangle - |G\rangle\,\| \leqslant \| \,|\widetilde{G^{tr}}\rangle - |G^{tr}\rangle\,\| + \| \,|\widetilde{G^{tr}}\rangle - |G\rangle\,\| \leqslant 2\delta.$$

To show $|G^{mod}\rangle$ is close to $|G\rangle$, let us similarly define $|\widetilde{G^{mod}}\rangle = \sqrt{\frac{G_{mod}}{G}}\,|G^{mod}\rangle$. We can see

$$\| \,|\widetilde{G^{mod}}\rangle - |G\rangle\,\|^2 \leqslant \frac{1}{G} \sum_{x\in[\![\pm\frac{N}{2}]\!]} \Big(\sum_{y\in\mathbb{Z}\setminus\{0\}} p_s(x+t+Ny)\Big)^2 + \frac{1}{G} \sum_{x\in\mathbb{Z}\setminus[\![\pm\frac{N}{2}]\!]} \Big(p_s(x+t)\Big)^2.$$

To upper bound the first term in the RHS above, we split the domain of x into three disjoint parts: $D_1 = \{x \in [\![\pm\frac{N}{2}]\!] : |x+t| \leqslant N/4\}$, $D_2 = \{x \in [\![\pm\frac{N}{2}]\!] : N/2 \geqslant |x+t| > N/4\}$, and $D_3 = \{x \in [\![\pm\frac{N}{2}]\!] : |x+t| > N/2\}$.

[14] Equivalently, if $t \in [-5N/8, -3N/8]$ (and the constraints for N, s remain the same), then $|G\rangle$, $|(G^{tr})'\rangle = \frac{1}{\sqrt{G_{tr}}} \sum_{x\in[N]} f(x+t)p_s(x+t)\,|x\rangle$, $|(G^{mod})'\rangle = \frac{1}{\sqrt{G_{mod}}} \sum_{x\in\mathbb{Z}} f(x+t)p_s(x+t)\,|x \bmod N)\rangle$ are 9δ-close to each other.

When $x \in D_1$, we can see $\sum_{y\in\mathbb{Z}\setminus\{0\}} p_s(x+t+Ny) \leqslant p_s(x+t)\cdot 2\sum_{y=1}^{\infty}\delta^{2y} = p_s(x+t)\cdot 2\delta^2/(1-\delta^2)$, because for every $y \in \mathbb{N}\cup\{0\}$ and $x \in D_1$ both

$$p_s(x+t+(y+1)N)/p_s(x+t+yN) = \exp(-\frac{\pi}{s^2}\cdot N(2(x+t)+(2y+1)N)) \leqslant \delta^2 \tag{2.5}$$

and

$$p_s(x+t-(y+1)N)/p_s(x+t-yN) = \exp(-\frac{\pi}{s^2}\cdot N(-2(x+t)+(2y+1)N)) \leqslant \delta^2 \tag{2.6}$$

hold (because $\exp(-\frac{\pi}{s^2}\cdot N(N\pm 2(x+t)) \leqslant \exp(-\frac{\pi}{s^2}\cdot\frac{N^2}{2}) \leqslant \delta^2$ for every $x\in D_1$).

When $x \in D_2$ we use a similar argument, the only difference is that Equations (2.5) and (2.6) now hold for every $y \in \mathbb{N}$ (excluding 0), so

$$\sum_{y\in\mathbb{Z}\setminus\{0\}} p_s(x+t+Ny) \leqslant p_s(x+t)\cdot 2\sum_{y=0}^{\infty}\delta^{2y} = p_s(x+t)\cdot 2/(1-\delta^2).$$

When $x \in D_3$, we can see that either $x+N$ or $x-N$ is $N/2$-close to (but $3N/8$-far from) $-t$, and without loss of generality and for simplicity we can assume $|x+N-(-t)| \in [3N/8, N/2]$. Then using a similar argument as for the $x \in D_2$, we can show that for every $x \in D_3$,

$$\sum_{y\in\mathbb{Z}\setminus\{0\}} p_s(x+t+Ny) \leqslant p_s(x+N+t)\cdot 2/(1-\delta^2).$$

Since $p_s^2 = p_{s/\sqrt{2}}$ and $\mathcal{D}_{\mathbb{Z}+t,s/\sqrt{2}}$ is δ^2-sub-Gaussian with parameter $s/\sqrt{2}$, we have

$$\begin{aligned}
\|\,|\widetilde{G^{mod}}\rangle - |G\rangle\,\|^2 &\leqslant \frac{1}{G}\sum_{x\in D_1\cup D_2\cup D_3}\Big(\sum_{y\in\mathbb{Z}\setminus\{0\}} p_s(x+t+Ny)\Big)^2 + \frac{1}{G}\sum_{x\in\mathbb{Z}\setminus[\![\pm\frac{N}{2}]\!]}\Big(p_s(x+t)\Big)^2\\
&\leqslant \frac{1}{G}\Big((\frac{2\delta^2}{1-\delta^2})^2\sum_{x\in D_1} p_{\frac{s}{\sqrt{2}}}(x+t) + (\frac{2}{1-\delta^2})^2\sum_{x\in D_2} p_{\frac{s}{\sqrt{2}}}(x+t) + (\frac{2}{1-\delta^2})^2\sum_{x\in D_3} p_{\frac{s}{\sqrt{2}}}(x+N+t) + \delta^2 G\Big)\\
&\leqslant \frac{1}{G}\Big(\frac{4\delta^4}{(1-\delta^2)^2}G + \frac{8}{(1-\delta^2)^2}(\delta^2 G) + \delta^2 G\Big) = \frac{4\delta^4+8\delta^2}{1-\delta^2} + \delta^2 \leqslant 10\delta^2,
\end{aligned}$$

where the second equality holds because

$$\begin{aligned}
&\sum_{x\in D_2} p_{s/\sqrt{2}}(x+t) + \sum_{x\in D_3} p_{s/\sqrt{2}}(x+N+t)\\
&\qquad\leqslant \sum_{x\in\mathbb{Z}\setminus\{x:|x+t|\leqslant N/8\}} p_{s/\sqrt{2}}(x+t) \leqslant G\cdot 2\exp(\delta^2)\exp(-(N/8)^2/(2\cdot s^2/2)) \leqslant \delta^2 G.
\end{aligned}$$

Note that $\|\,|\widetilde{G^{mod}}\rangle\,\| \in \|\,|G\rangle\,\| \pm \|\,|\widetilde{G^{mod}}\rangle - |G\rangle\,\|$ (implying $\|\,|\widetilde{G^{mod}}\rangle\,\| \in (1\pm\sqrt{10}\delta)$) and hence $\|\,|\widetilde{G^{mod}}\rangle - |G^{mod}\rangle\,\| = \|(\sqrt{\frac{G_{mod}}{G}}-1)\,|G^{mod}\rangle\,\| \leqslant \sqrt{10}\delta$. As a result, we obtain $\|\,|G^{mod}\rangle - |G\rangle\,\| \leqslant \|\,|\widetilde{G^{mod}}\rangle - |G\rangle\,\| + \|\,|\widetilde{G^{mod}}\rangle - |G^{mod}\rangle\,\| \leqslant 2\sqrt{10}\delta < 7\delta$. By triangle inequality, we obtain $\|\,|G^{mod}\rangle - |G^{tr}\rangle\,\| \leqslant \|\,|G^{mod}\rangle - |G\rangle\,\| + \|\,|G\rangle - |G^{tr}\rangle\,\| \leqslant 9\delta$. ∎

2.10 Fourier transform The Fourier transform $\hat{h}:\mathbb{R}\to\mathbb{C}$ of a function $h:\mathbb{R}\to\mathbb{C}$ is defined as

$$\hat{h}(\omega) = \int_{-\infty}^{\infty} h(x)\exp(-2\pi i x\omega)\,dx.$$

The next facts follow easily from the above definition. If h is defined as $h(\omega) = g(\omega+\nu)$ for some function g and value ν, then we have

$$\hat{g}(\omega) = \hat{h}(\omega)\exp(2\pi i\nu\omega).$$

On the other hand, if $h(x) = g(x)\exp(2\pi i x\nu)$, then

$$\hat{h}(\omega) = \hat{g}(\omega - \nu).$$

Another important fact is that the Fourier transform of p_s is $s \cdot p_{1/s}$ for all $s > 0$. Also, the sum of $p_s(x)$ over $C \cdot \mathbb{Z}$ satisfies the Poisson summation formula [54, Lemma 2.14]:

THEOREM 2.18. *For any scalar $C > 0$ and any Schwartz function $f : \mathbb{R} \to \mathbb{C}$ (i.e., f and each of its derivatives go to 0 faster than every inverse polynomial as the absolute value of the argument goes to infinity),*

$$\sum_{j \in C\cdot\mathbb{Z}} f(j) = C^{-1} \sum_{j \in C^{-1}\cdot\mathbb{Z}} \hat{f}(j).$$

2.11 Area of hyperspherical cap Let $B_d(r)$ denote the d-dimensional ball with radius r. The surface area of $B_d(r)$ is well-known to be $\frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} r^{d-1}$ [43, p. 66], where Γ is the Gamma function. Let $A_d^r(\phi)$ be the surface area of a hyperspherical cap in $B_d(r)$ with spherical angle ϕ. The area of this hyperspherical cap can be calculated by integrating the surface area of a $(d-1)$-dimensional sphere with radius $r\sin\theta$ [43, p. 67]:

$$A_d^r(\phi) = \int_0^\phi 2A_{d-1}^{r\sin\theta}(\pi/2) r \mathrm{d}\theta = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} r^{d-1} \cdot \int_0^\phi \sin^{d-2}\theta \mathrm{d}\theta.$$

We abbreviate $A_d(\phi) := A_d^1(\phi)$ for simplicity.

Following Ravsky's computation in his reply to a question on StackExchange [53], we now use the area of the hyperspherical cap to upper bound the probability of the event that a uniformly random vector u on S^{d-1} only has a small overlap with another (fixed) unit vector v.

THEOREM 2.19. *Let $d \geqslant 3$ be and integer, $v \in \mathbb{R}^d$ be a unit vector, and $a \in [0,1]$. Then we have*

$$\Pr_{u\sim S^{d-1}}\left[|\langle v, u\rangle| < a\right] \leqslant \frac{2}{\sqrt{\pi}} \cdot \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \cdot a.$$

Proof. Let $\phi \in [0, \pi/2]$ such that $\cos\phi = a$. We can see that if $|\langle v, u\rangle| \geqslant a$, then u will be in the hyperspherical cap (whose center is v) with spherical angle ϕ, and the probability that a uniformly-random u lands in that hyperspherical cap is $\frac{A_d(\phi)}{A_d(\pi/2)}$. Therefore,

$$\begin{aligned}
\Pr_{u\sim S^{d-1}}\left[|\langle v, u\rangle| < a\right] =& \frac{A_d(\pi/2) - A_d(\phi)}{A_d(\pi/2)} = \left(\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}\right)^{-1} \cdot \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\right) \int_\phi^{\pi/2} \sin\theta^{d-2}\mathrm{d}\theta \\
\leqslant& \left(\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}\right)^{-1} \cdot \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\right) \int_\phi^{\pi/2} \sin\theta \mathrm{d}\theta = \frac{2\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})} \cdot (-\cos\frac{\pi}{2} + \cos\phi) \\
=& \frac{2}{\sqrt{\pi}} \cdot \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \cdot a.
\end{aligned}$$

∎

By using Legendre's duplication formula $\Gamma(\frac{d}{2})\Gamma(\frac{d-1}{2}) = \frac{\sqrt{\pi}}{2^{d-2}}\Gamma(d-1)$ [56, Chap. 8.21, Eq. 102] and the fact $\Gamma(d) = (d-1)!$, we obtain

$$\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} = \begin{cases} \frac{\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d}{2})\Gamma(\frac{d-1}{2})} = \frac{2^{d-2}((\frac{d}{2}-1)!)^2}{\sqrt{\pi}(d-2)!} = \frac{2^{d-2}}{\sqrt{\pi}} \cdot \binom{d-2}{\frac{d-2}{2}}^{-1}, & \text{if } d \text{ is even,} \\ \frac{\Gamma(\frac{d}{2})\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d-1}{2})^2} = \frac{\sqrt{\pi}(d-2)!}{2^{d-2}(\frac{d-3}{2}!)^2} = \frac{\sqrt{\pi}(d-2)}{2^{d-2}} \cdot \binom{d-3}{\frac{d-3}{2}}, & \text{if } d \text{ is odd.} \end{cases}$$

Plugging the above into Theorem 2.19, we have the following corollary.

COROLLARY 2.5. *Let $d \geqslant 4$ be an integer, $v \in \mathbb{R}^d$ be a unit vector, and $c \geqslant 1$. Then we have*

$$\Pr_{u \sim S^{d-1}}\left[|\langle v, u\rangle| < \frac{1}{c\sqrt{d}}\right] < \frac{1}{c}.$$

Proof. By Theorem 2.19, it suffices to show $\frac{2}{\sqrt{\pi}} \cdot \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \cdot \frac{1}{\sqrt{d}} < 1$ for every $d \geqslant 4$. When d is even, by using Robbins' bound $\frac{4^m}{\sqrt{\pi m}}\exp(-\frac{1}{6m}) \leqslant \binom{2m}{m} \leqslant \frac{4^m}{\sqrt{\pi m}}$ [55, consequence of Eq. 1], we have

$$\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} = \frac{2^{d-2}}{\sqrt{\pi}} \cdot \binom{d-2}{\frac{d-2}{2}}^{-1} \leqslant \sqrt{\frac{d-2}{2}}\exp(\frac{1}{3d-6}) \leqslant \sqrt{\frac{d}{2}}\exp(\frac{1}{6}),$$

implying that $\frac{2}{\sqrt{\pi}} \cdot \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \cdot \frac{1}{\sqrt{d}} \leqslant \sqrt{\frac{2}{\pi}} \cdot \exp(\frac{1}{6}) < 1$. Similarly, when d is odd, we have

$$\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} = \frac{\sqrt{\pi}(d-2)}{2^{d-2}} \cdot \binom{d-3}{\frac{d-3}{2}} \leqslant \frac{d-2}{\sqrt{2(d-3)}},$$

implying that $\frac{2}{\sqrt{\pi}} \cdot \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \cdot \frac{1}{\sqrt{d}} \leqslant \sqrt{\frac{2}{\pi}} \cdot \frac{d-2}{\sqrt{d(d-3)}} < 1$. ∎

3 Time-efficient unbiased pure-state tomography

In this section we design efficient methods for obtaining a good classical description of a pure quantum state (i.e., tomography), by manipulating and measuring multiple copies of that state.

3.1 Pure-state tomography by computational-basis measurements A direct corollary of Corollary 2.1 as observed in [5] is that computational-basis measurements yield a good approximation of the absolute values of the amplitudes of a (sub)normalized quantum state vector.

COROLLARY 3.1. *Suppose that $\varepsilon, \delta \in (0,1]$, $\psi \in \mathbb{C}^d$ has ℓ_2-norm at most 1, and we are given $n \geqslant \frac{1}{\varepsilon^2}\ln\left(\frac{2d}{\delta}\right)$ copies of the pure quantum state $|\varphi\rangle := |\bar{0}\rangle|\psi\rangle + |\bar{0}^\perp\rangle$, where $(|\bar{0}\rangle\langle\bar{0}| \otimes I)|\bar{0}^\perp\rangle = 0$. If we measure each copy in the computational basis and denote by s_i the normalized number (i.e., frequency) of outcomes $|\bar{0}\rangle|i\rangle$ then the vector*

$$\bar{\psi}_i := \sqrt{s_i}$$

with probability at least $1-\delta$ gives an ε-ℓ_∞ approximation of $|\psi|$. Moreover, $\|\bar{\psi}\|_2 \leqslant 1$ with certainty and if $\|\psi\|_2 = 1$, then also $\|\bar{\psi}\|_2 = 1$, and in general $\left|\|\bar{\psi}\|_2 - \|\psi\|_2\right| \geqslant \varepsilon$ holds with probability $\leqslant \frac{\delta}{d}$.

Proof. By Corollary 2.1 we have that

$$\Pr\left[\left|\sqrt{s_i} - |\psi_i|\right| \geqslant \varepsilon\right] \leqslant \frac{\delta}{d},$$

and similarly

$$\Pr\left[\left|\|\bar{\psi}\|_2 - \|\psi\|_2\right| \geqslant \varepsilon\right] = \Pr\left[\left|\sqrt{\sum_{i=0}^{d-1} s_i} - \sqrt{\sum_{i=0}^{d-1}|\psi_i|^2}\right| \geqslant \varepsilon\right] \leqslant \frac{\delta}{d}.$$

Finally, $\|\bar{\psi}\|_2^2 = \sum_{i=0}^{d-1} s_i \leqslant 1$, where the last inequality is an equality if $\|\psi\|_2 = 1$. ∎

3.2 Pure-state tomography using conditional samples Now we show how to produce an unbiased estimator of ψ itself (not just of the magnitudes of its entries) with bounded variance using computational-basis measurements with the help of a reference state $\bar{\psi}$. Our approach is inspired by [39] but improves over their biased estimator by making it unbiased.

LEMMA 3.1. *Suppose that $\psi \in \mathbb{C}^d$ has $\|\psi\|_2 \leqslant 1$, and we are given a copy of the state $|\varphi'\rangle := \Big(|+\rangle\big(|\bar{0}\rangle|\psi\rangle + |\bar{0}^{\perp}\rangle \big) + |-\rangle\big(|\bar{0}\rangle|\bar{\psi}\rangle + |\bar{0}'^{\perp}\rangle \big)\Big)/\sqrt{2}$, where $|\bar{0}\rangle = |0^a\rangle$ for some $a \in \mathbb{N}$, $(|\bar{0}\rangle\langle\bar{0}| \otimes I)\,|\bar{0}^{\perp}\rangle = 0$, and $(|\bar{0}\rangle\langle\bar{0}| \otimes I)\,|\bar{0}'^{\perp}\rangle = 0$. If we measure $|\varphi'\rangle$ in the computational basis and denote by $X \in \{0,1\}^{2d}$ the indicator of the measurement outcomes $|b\rangle\,|\bar{0}\rangle\,|i\rangle$ (this X is a weight-1 Boolean vector indexed by (b,i) where $b \in \{0,1\}$ and $i \in [d]-1$), then the random vector $\psi' \in \mathbb{C}^d$ with coordinates*

$$\psi'_i := \frac{X_{0,i} - X_{1,i}}{|\bar{\psi}_i|}$$

is an unbiased estimator of $\psi_i^{\Re} := \mathrm{Re}\Big(\psi_i \frac{\bar{\psi}_i^*}{|\bar{\psi}_i|}\Big)$, with $\|\psi'\|_2 \leqslant \frac{1}{\min\{|\bar{\psi}_i|\colon i \in [d]-1\}}$ with certainty, and covariance matrix $\mathrm{Cov}(\psi') = \frac{I}{2} + \mathrm{diag}\Big(\frac{|\psi_i|^2}{2|\bar{\psi}_i|^2}\Big) - |\psi^{\Re}\rangle\langle\psi^{\Re}|$.

Proof. The probabilities of getting measurement outcomes $|0\rangle\,|\bar{0}\rangle\,|i\rangle$ and $|1\rangle\,|\bar{0}\rangle\,|i\rangle$ are

$$p_{0,i} := |\langle 0|\langle\bar{0}|\langle i|\,|\varphi'\rangle|^2 = \left|\frac{\psi_i + \bar{\psi}_i}{2}\right|^2 = \frac{|\psi_i|^2 + \psi_i\bar{\psi}_i^* + \psi_i^*\bar{\psi}_i + |\bar{\psi}_i|^2}{4} = \mathbb{E}[X_{0,i}],$$

$$p_{1,i} := |\langle 1|\langle\bar{0}|\langle i|\,|\varphi'\rangle|^2 = \left|\frac{\psi_i - \bar{\psi}_i}{2}\right|^2 = \frac{|\psi_i|^2 - \psi_i\bar{\psi}_i^* - \psi_i^*\bar{\psi}_i + |\bar{\psi}_i|^2}{4} = \mathbb{E}[X_{1,i}],$$

and therefore

$$\mathbb{E}[X_{0,i} - X_{1,i}] = p_{0,i} - p_{1,i} = \frac{\psi_i\bar{\psi}_i^* + \psi_i^*\bar{\psi}_i}{2} = |\bar{\psi}_i|\frac{\psi_i\frac{\bar{\psi}_i^*}{|\bar{\psi}_i|} + \psi_i^*\frac{\bar{\psi}_i}{|\bar{\psi}_i|}}{2} = |\bar{\psi}_i|\psi_i^{\Re}.$$

For the norm bound observe that

$$\|\psi'\|_2 \leqslant \sum_{i=0}^{d-1}|\psi'|_i \leqslant \sum_{i=0}^{d-1}\frac{X_{0,i}+X_{1,i}}{|\bar{\psi}|_i} \leqslant \sum_{i=0}^{d-1}\frac{X_{0,i}+X_{1,i}}{\min\{|\bar{\psi}_i|\colon i \in [d]-1\}} \leqslant \frac{1}{\min\{|\bar{\psi}_i|\colon i \in [d]-1\}}.$$

We can compute the covariance matrix directly as follows

$$\mathrm{Cov}(\psi')_{ij} = \mathbb{E}[\psi'_i\psi'_j] - \mathbb{E}[\psi'_i]\mathbb{E}[\psi'_j] = \delta_{ij}\frac{p_{0,i}+p_{1,i}}{|\bar{\psi}_i|^2} - \psi_i^{\Re}\psi_j^{\Re} = \delta_{ij}\frac{|\psi_i|^2/|\bar{\psi}_i|^2 + 1}{2} - \psi_i^{\Re}\psi_j^{\Re},$$

where the second equality uses that $X_{a,i}X_{b,j} = \delta_{ab}\cdot\delta_{ij}\cdot X_{a,i}$ because X is a weight-1 Boolean vector. ∎

By an analogous argument as in the proof of Lemma 3.1, we can obtain an unbiased estimator of the imaginary parts $\psi_j^{\Im} := \mathrm{Im}\Big(\psi_j\frac{\bar{\psi}_j^*}{|\bar{\psi}_j|}\Big)$ (with the same ℓ_2-norm and covariance matrix gaurantee) by measuring $|\varphi''\rangle := \Big(|+\rangle\big(|\bar{0}\rangle\,|\psi\rangle + |\bar{0}^{\perp}\rangle\big) + i\,|-\rangle\big(|\bar{0}\rangle\,|\bar{\psi}\rangle + |\bar{0}'^{\perp}\rangle\big)\Big)/\sqrt{2}$ in the computational basis.

We now give a procedure (the first part of Theorem 3.1) to find an unbiased estimator $\tilde{\psi}$ of ψ that simultaneously has a good bound on the error of the estimator (with overwhelming probability) in some k fixed directions using Lemma 3.1. The second part of Theorem 3.1 shows that the output $\tilde{\psi}$ will be close (in total variation distance) to an "almost ideal" unbiased estimator $\breve{\psi}$ that simultaneously has a good bound on the error of the estimator *with certainty* in some k fixed directions. This will be used later when estimating a matrix-vector product Aw in order to avoid an estimation-error that has too much overlap with k of the eigenvectors of A.

THEOREM 3.1. *Let $\psi \in \mathbb{C}^d$ such that $\|\psi\|_2 \leqslant 1$, $\varepsilon, \delta \in (0,1]$, $\eta \in \mathbb{R}_+$, $k \in \mathbb{N}$, $n \geqslant \frac{4d}{\varepsilon^2}\Big(\frac{4}{3} + \frac{1}{\eta}\Big)\ln\Big(\frac{8k}{\delta}\Big)$. Suppose there exists a "reference state" (not necessarily known to the algorithm) $\bar{\psi} \in \mathbb{C}^d$ such that $|\bar{\psi}_j|^2 \geqslant \max\{\frac{\varepsilon^2}{d}, \eta|\psi_j|^2\}$*

$\forall j \in [d]-1$, *and* $\|\bar{\psi}\| \leqslant 1$. *Given* n *copies of the pure quantum states*

$$|\varphi'\rangle := \Big(|+\rangle\Big(|\bar{0}\rangle|\psi\rangle + |\bar{0}^{\perp}\rangle\Big) + |-\rangle\Big(|\bar{0}\rangle|\bar{\psi}\rangle + |\bar{0}'^{\perp}\rangle\Big)\Big)/\sqrt{2},$$
$$|\varphi''\rangle := \Big(|+\rangle\Big(|\bar{0}\rangle|\psi\rangle + |\bar{0}^{\perp}\rangle\Big) + i\,|-\rangle\Big(|\bar{0}\rangle|\bar{\psi}\rangle + |\bar{0}'^{\perp}\rangle\Big)\Big)/\sqrt{2},$$

where $|\bar{0}\rangle = |0^a\rangle$ *for some* $a \in \mathbb{N}$, $(|\bar{0}\rangle\langle\bar{0}| \otimes I)\,|\bar{0}^{\perp}\rangle = 0$, *and* $(|\bar{0}\rangle\langle\bar{0}| \otimes I)\,|\bar{0}'^{\perp}\rangle = 0$, *if we measure each copy in the computational basis and denote by* $s'_{b,j}$, $s''_{b,j}$ *the normalized number of measurement outcomes* $|b\rangle|\bar{0}\rangle|j\rangle$ *from measuring the states* $|\varphi'\rangle$ *and* $|\varphi''\rangle$ *respectively, then the random vector* $\tilde{\psi} \in \mathbb{C}^d$ *with coordinates*

$$\tilde{\psi}_j := \Big(s'_{0,j} - s'_{1,j} + is''_{0,j} - is''_{1,j}\Big)\frac{\bar{\psi}_j}{|\bar{\psi}_j|^2}$$

is an unbiased estimator of ψ. *Moreover, for every set* $V = \{v^{(j)} : j \in [k]\} \subset \mathbb{C}^d$ *of vectors we have*

$$\Pr\left[\forall v \in V : |\langle\tilde{\psi} - \psi|v\rangle| < \frac{\varepsilon}{\sqrt{d}}\|v\|_2\right] \geqslant 1 - \delta. \tag{3.7}$$

In particular if $k \geqslant d$, *then* $\Pr\Big[\|\tilde{\psi} - \psi\|_2 < \varepsilon\Big] \geqslant 1 - \delta$.

Finally, let $\{v^{(j)} : j \in [k]\}$ *be a fixed set of orthonormal vectors and* Π_k *be the projector to their span. Let* A *be the event that* $\exists j \in [k] : |\langle\tilde{\psi} - \psi|v^{(j)}\rangle| > \frac{\varepsilon}{\sqrt{d}}$, $\bar{A}$ *be the complement of* A, *and* X_ζ *be an independent Bernoulli random variable such that* $\Pr[X_\zeta = 0] = \zeta := \frac{\delta - p}{1-p}$ *for* $p := \Pr[A]$. *Define* $\breve{\psi} \in \mathbb{C}^d$ *as follows*[15]

$$\breve{\psi} = \begin{cases} \tilde{\psi} & \text{on } \bar{A} \cap (X_\zeta = 1) \\ (I - \Pi_k)\tilde{\psi} + \sum\limits_{j\in[k]} |v^{(j)}\rangle\mathbb{E}[\langle v^{(j)}|\tilde{\psi}\rangle\,|A \cup (X_\zeta = 0)] & \text{on } A \cup (X_\zeta = 0). \end{cases}$$

Then $\mathbb{E}[\breve{\psi}] = \psi$, $\Pr[\forall j \in [k] : |\langle\breve{\psi} - \psi|v^{(j)}\rangle| \leqslant \frac{k+3}{k}\frac{\varepsilon}{\sqrt{d}}] = 1$ *(which is why we call* $\breve{\psi}$ *an "almost ideal" unbiased estimator), the total variation distance between* $\tilde{\psi}$, *and* $\breve{\psi}$ *is at most* δ, *and* $\Big\|\mathrm{Cov}(\Pi_k\breve{\psi})\Big\| \leqslant \Big\|\mathrm{Cov}(\Pi_k\tilde{\psi})\Big\| + 25\delta\varepsilon^2\frac{k}{d} \leqslant \left(\frac{1}{4\ln\left(\frac{8k}{\delta}\right)} + 25\delta k\right)\frac{\varepsilon^2}{d}$.

Proof. We prove the first part of Theorem 3.1 first. Let us define the random vectors $\psi', \psi'' \in \mathbb{C}^d$ with coordinates

$$\psi'_j := \frac{X'_{0,j} - X'_{1,j}}{|\bar{\psi}_j|}, \qquad \psi''_j := \frac{X''_{0,j} - X''_{1,j}}{|\bar{\psi}_j|},$$

where $X', X'' \in \{0,1\}^{2d}$ denote the indicator of the measurements outcomes $|b\rangle|\bar{0}\rangle|j\rangle$ for the states $|\varphi'\rangle$ and $|\varphi''\rangle$, respectively. Then by Lemma 3.1 and the discussion after the proof of Lemma 3.1, ψ', ψ'' are unbiased estimators of $\psi_j^{\Re} := \mathrm{Re}\left(\psi_j\frac{\bar{\psi}_j^*}{|\bar{\psi}_j|}\right)$, and $\psi_j^{\Im} := \mathrm{Im}\left(\psi_j\frac{\bar{\psi}_j^*}{|\bar{\psi}_j|}\right)$ respectively, such that $\|\psi'\|_2, \|\psi''\|_2 \leqslant \sqrt{d}/\varepsilon$ with certainty and

$$\mathrm{Cov}(\psi') + \mathbb{E}[\psi']\mathbb{E}[\psi'^T] \preceq \left(\frac{1}{2} + \frac{1}{2\eta}\right)I, \quad \mathrm{Cov}(\psi'') + \mathbb{E}[\psi'']\mathbb{E}[\psi''^T] \preceq \left(\frac{1}{2} + \frac{1}{2\eta}\right)I, \tag{3.8}$$

where for the latter psd inequalities we used that $|\bar{\psi}_j|^2 \geqslant \eta|\psi_j|^2$ and hence $\mathrm{diag}\left(\frac{|\psi_i|^2}{2|\bar{\psi}_i|^2}\right) \preceq \frac{1}{2\eta}I$.

Let $w \in \mathbb{R}^d$, then the random variables $\psi'_w := \langle\psi'|w\rangle, \psi''_w := \langle\psi''|w\rangle$ satisfy $|\psi'_w| \leqslant \|\psi'\|_2\|w\|_2 \leqslant \sqrt{d}\|w\|_2/\varepsilon$, $|\psi''_w| \leqslant \|\psi''\|_2\|w\|_2 \leqslant \sqrt{d}\|w\|_2/\varepsilon$ with certainty. Also, $\mathbb{E}[|\psi'_w|^2]$ and $\mathbb{E}[|\psi''_w|^2]$ are both $\leqslant \left(\frac{1}{2} + \frac{1}{2\eta}\right)\|w\|_2^2$, because for both $\phi = \psi'$ and $\phi = \psi''$, we have

$$\mathbb{E}[|\langle\phi|w\rangle|^2] = \langle w|\,\mathbb{E}[\phi\phi^T]\,|w\rangle = \langle w|\,(\mathrm{Cov}(\phi) + \mathbb{E}[\phi]\mathbb{E}[\phi^T])\,|w\rangle \leqslant \|\mathrm{Cov}(\phi) + \mathbb{E}[\phi]\mathbb{E}[\phi^T]\| \cdot \|w\|_2^2.$$

[15]Here we introduce X_ζ to ensure that $\Pr[A \cup (X_\zeta = 0)]$ exactly equals δ, which is helpful because we use both upper and lower bounds on this probability in the proof.

Let $\Psi', \Psi'' \in \mathbb{R}^d$ be the sum of n i.i.d. copies of ψ', ψ'', respectively, obtained from the measurement outcomes of the n copies of $|\varphi'\rangle$ and $|\varphi''\rangle$, so that

$$\tilde{\psi}_j = (\Psi'_j + i\Psi''_j)\frac{\bar{\psi}_j}{n|\bar{\psi}_j|}. \tag{3.9}$$

Let us analogously define $\Psi'_w := \langle \Psi'|w\rangle$, and $\Psi''_w := \langle \Psi''|w\rangle$. Then clearly $\mathbb{E}[\Psi'_w] = n\langle \psi^{\Re}|w\rangle$, and $\mathbb{E}[\Psi''_w] = n\langle \psi^{\Im}|w\rangle$. For all $\tau \geqslant 1$, the Bennett-Bernstein tail bound (Proposition 2.1) implies

$$\begin{aligned}\Pr\left[|\Psi'_w - n\langle \psi^{\Re}|w\rangle| \geqslant \tau\frac{\varepsilon n\|w\|_2}{2\sqrt{d}}\right] &\leqslant 2\exp\left(-\frac{\tau^2\varepsilon^2 n^2\|w\|_2^2/(4d)}{\left(1+\frac{1}{\eta}\right)\|w\|_2^2 n + \frac{\tau}{3}\|w\|_2^2 n}\right)\\ &\leqslant 2\exp\left(-\tau\frac{\varepsilon^2 n/(4d)}{\frac{4}{3}+\frac{1}{\eta}}\right) \leqslant 2\left(\frac{\delta}{8k}\right)^{\tau},\end{aligned} \tag{3.10}$$

and similarly $\Pr[|\Psi''_w - n\langle \psi^{\Im}|w\rangle| \geqslant \tau\varepsilon n\|w\|_2/\sqrt{4d}] \leqslant 2\left(\frac{\delta}{8k}\right)^{\tau}$.

Let $\tilde{v}_j := v_j\frac{\bar{\psi}_j^*}{|\bar{\psi}_j|}$, $\tilde{v}_j^{\Re} := \mathrm{Re}(\tilde{v}_j)$, $\tilde{v}_j^{\Im} := \mathrm{Im}(\tilde{v}_j)$, and observe that for any $v \in V$

$$\begin{aligned}\langle\tilde{\psi}|v\rangle - \langle\psi|v\rangle &= \frac{1}{n}\langle\Psi' + i\Psi''|\tilde{v}\rangle - \langle\psi^{\Re} + i\psi^{\Im}|\tilde{v}\rangle\\ &= \Bigg(\underbrace{\frac{\langle\Psi'|\tilde{v}^{\Re}\rangle}{n} - \langle\psi^{\Re}|\tilde{v}^{\Re}\rangle}_{:=a\|\tilde{v}^{\Re}\|_2} \underbrace{- \frac{\langle\Psi''|\tilde{v}^{\Im}\rangle}{n} + \langle\psi^{\Im}|\tilde{v}^{\Im}\rangle}_{:=b\|\tilde{v}^{\Im}\|_2}\Bigg) + i\Bigg(\underbrace{\frac{\langle\Psi'|\tilde{v}^{\Im}\rangle}{n} - \langle\psi^{\Re}|\tilde{v}^{\Im}\rangle}_{:=c\|\tilde{v}^{\Re}\|_2} \underbrace{- \frac{\langle\Psi''|\tilde{v}^{\Re}\rangle}{n} + \langle\psi^{\Im}|\tilde{v}^{\Re}\rangle}_{:=d\|\tilde{v}^{\Im}\|_2}\Bigg),\end{aligned}$$

so

$$\begin{aligned}|\langle\tilde{\psi}|v\rangle - \langle\psi|v\rangle| &\leqslant \sqrt{2}\max\left\{\left|\mathrm{Re}(\langle\tilde{\psi}|v\rangle - \langle\psi|v\rangle)\right|, \left|\mathrm{Im}(\langle\tilde{\psi}|v\rangle - \langle\psi|v\rangle)\right|\right\}\\ &= \sqrt{2}\max\left\{\left|a\|\tilde{v}^{\Re}\|_2 + b\|\tilde{v}^{\Im}\|_2\right|, \left|c\|\tilde{v}^{\Re}\|_2 + d\|\tilde{v}^{\Im}\|_2\right|\right\}\\ &\leqslant \sqrt{2}\max\{|a|,|b|,|c|,|d|\}(\|\tilde{v}^{\Re}\|_2 + \|\tilde{v}^{\Im}\|_2)\\ &\leqslant 2\max\{|a|,|b|,|c|,|d|\}\|v\|_2,\end{aligned}$$

where the last step uses that $|\tilde{v}_j| = |v_j|$ for all j, and Cauchy-Schwarz. Using Eq. (3.10) four times with different choices of w , and the union bound over 4 events, we have $\max\{|a|,|b|,|c|,|d|\} < \frac{\tau\varepsilon}{2\sqrt{d}}$ except with probability $\leqslant 8\left(\frac{\delta}{8k}\right)^{\tau}$. Eq. (3.7) now follows by choosing $\tau = 1$ and taking the union bound over all k vectors $v \in V$. If $k \geqslant d$ then we can apply the statement for a set V containing the computational basis, and then $\|\tilde{\psi} - \psi\|_\infty \leqslant \frac{\varepsilon}{\sqrt{d}}$ implies $\|\tilde{\psi} - \psi\|_2 \leqslant \varepsilon$ by Cauchy-Schwarz.

Now we prove the second part of Theorem 3.1, where V is an orthonormal set (the "Finally" part). Note that the previous paragraph already proved that for every $v \in \mathbb{C}^d$ and for all $\tau \geqslant 1$,

$$\Pr[|\langle v|\tilde{\psi} - \psi\rangle| > \tau\frac{\varepsilon}{\sqrt{d}}\|v\|_2] \leqslant 8\left(\frac{\delta}{8k}\right)^{\tau}. \tag{3.11}$$

Defining $\mathring{\varepsilon} := \varepsilon/\sqrt{d}$ and $\mathring{\psi}^j := \langle v^{(j)}|\tilde{\psi} - \psi\rangle$, we bound

$$\mathbb{E}\Big[|\mathring{\psi}^j|\mathbb{1}_{(\mathring{\varepsilon},\infty)}(|\mathring{\psi}^j|)\Big] = \sum_{\ell=0}^{\infty} \mathbb{E}\Big[|\mathring{\psi}^j|\mathbb{1}_{(2^\ell\mathring{\varepsilon},2^{\ell+1}\mathring{\varepsilon}]}(|\mathring{\psi}^j|)\Big] \leqslant \sum_{\ell=0}^{\infty} 2^{\ell+1}\mathring{\varepsilon}\Pr\Big[|\mathring{\psi}^j| \in (2^\ell\mathring{\varepsilon}, 2^{\ell+1}\mathring{\varepsilon}]\Big]$$

$$\left(\Pr[|\mathring{\psi}^j| > \tau\mathring{\varepsilon}] \leqslant 8\Big(\tfrac{\delta}{8k}\Big)^\tau\right) \quad \leqslant \sum_{\ell=0}^{\infty} 2^{\ell+1}\mathring{\varepsilon}\Pr\Big[|\mathring{\psi}^j| > 2^\ell\mathring{\varepsilon}\Big] \leqslant \sum_{\ell=0}^{\infty} 2^{\ell+1}\mathring{\varepsilon}\,8\Big(\frac{\delta}{8k}\Big)^{2^\ell}$$

(since $\delta/k \leqslant 1$, and $2^\ell \geqslant \ell+1$)

$$\leqslant \sum_{\ell=0}^{\infty} 2^{\ell+1}\mathring{\varepsilon}\,8\Big(\frac{\delta}{8k}\Big)^{\ell+1}$$

(since $\delta/k \leqslant 1$)

$$= 2\frac{\delta\varepsilon}{k\sqrt{d}}\sum_{\ell=0}^{\infty}\Big(\frac{\delta}{4k}\Big)^\ell \leqslant 2\frac{\delta\varepsilon}{k\sqrt{d}}\sum_{\ell=0}^{\infty}\Big(\frac{1}{4}\Big)^\ell$$

$$< \frac{3\delta}{k}\frac{\varepsilon}{\sqrt{d}}. \tag{3.12}$$

We have already proven Eq. (3.7), implying that $p = \Pr[A] \leqslant \delta$. Observe that

$$\mathbb{E}[\langle v^{(j)}|\tilde{\psi}\rangle|A \cup (X_\zeta = 0)] = \langle v^{(j)}|\psi\rangle + \mathbb{E}[\mathring{\psi}^j|A \cup (X_\zeta = 0)],$$

and hence

$$\mathbb{E}[\langle v^{(j)}|\tilde{\psi} - \psi\rangle|A \cup (X_\zeta = 0)] = \mathbb{E}[\mathring{\psi}^j|A \cup (X_\zeta = 0)].$$

Since $\Pr[A \cup (X_\zeta = 0)] = 1 - \Pr[\bar{A} \cap (X_\zeta = 1)] = \delta$, using (3.12) we have

$$|\mathbb{E}[\mathring{\psi}^j|A \cup (X_\zeta = 0)]| \leqslant \mathbb{E}[|\mathring{\psi}^j| \mid A \cup (X_\zeta = 0)] = \frac{\mathbb{E}[|\mathring{\psi}^j|\mathbb{1}_{A\cup(X_\zeta=0)}]}{\delta} \leqslant \frac{\mathring{\varepsilon}\delta + \frac{3\delta}{k}\frac{\varepsilon}{\sqrt{d}}}{\delta} = \left(1 + \frac{3}{k}\right)\frac{\varepsilon}{\sqrt{d}}. \tag{3.13}$$

Since we modified $\tilde{\psi}$ on an event of probability δ to get $\check{\psi}$, the total variation distance between the distributions of the random variables $\check{\psi}$ and $\tilde{\psi}$ is at most δ. The boundedness of $\check{\psi}$ is by construction and the unbiasedness is inherited from that of $\tilde{\psi}$, as follows: abbreviating the event $\bar{A} \cap (X_\zeta = 1)$ to B, we have

$$\begin{aligned}
\mathbb{E}[\check{\psi}] &= \Pr[B]\cdot\mathbb{E}[\check{\psi} \mid B] + \Pr[\bar{B}]\cdot\mathbb{E}[\check{\psi} \mid \bar{B}] \\
&= \Pr[B]\cdot\mathbb{E}[\tilde{\psi} \mid B] + \Pr[\bar{B}]\cdot\mathbb{E}[(I - \Pi_k)\tilde{\psi} + \sum_{j\in[k]} |v^{(j)}\rangle\mathbb{E}[\langle v^{(j)}|\tilde{\psi}\rangle] \mid \bar{B}] \\
&= \Pr[B]\cdot\mathbb{E}[\tilde{\psi} \mid B] + \Pr[\bar{B}]\cdot\mathbb{E}[(I - \Pi_k)\tilde{\psi} + \sum_{j\in[k]} |v^{(j)}\rangle\langle v^{(j)}|\cdot\tilde{\psi} \mid \bar{B}] \\
&= \Pr[B]\cdot\mathbb{E}[\tilde{\psi} \mid B] + \Pr[\bar{B}]\cdot\mathbb{E}[(I - \Pi_k)\tilde{\psi} + \Pi_k\tilde{\psi} \mid \bar{B}] \\
&= \Pr[B]\cdot\mathbb{E}[\tilde{\psi} \mid B] + \Pr[\bar{B}]\cdot\mathbb{E}[\tilde{\psi} \mid \bar{B}] = \mathbb{E}[\tilde{\psi}] = \psi.
\end{aligned}$$

Finally, defining $\bar{\varepsilon} := \varepsilon\sqrt{\frac{k}{d}}$ and $\mathring{\psi} := \Pi_k(\tilde{\psi} - \psi) = \sum_{j\in[k]} \mathring{\psi}^j v^{(j)}$ we bound

$$\begin{aligned}
\left\|\mathbb{E}\Big[|\mathring{\psi}\rangle\langle\mathring{\psi}|\mathbb{1}_{(\bar{\varepsilon},\infty)}(\|\mathring{\psi}\|_2)\Big]\right\| &\leqslant \mathbb{E}\Big[\|\mathring{\psi}\|_2^2\mathbb{1}_{(\bar{\varepsilon},\infty)}(\|\mathring{\psi}\|_2)\Big] = \sum_{\ell=0}^{\infty}\mathbb{E}\Big[\|\mathring{\psi}\|_2^2\mathbb{1}_{(2^\ell\bar{\varepsilon},2^{\ell+1}\bar{\varepsilon}]}(\|\mathring{\psi}\|_2)\Big] \\
&\leqslant \sum_{\ell=0}^{\infty} 4^{\ell+1}\bar{\varepsilon}^2\Pr\Big[\|\mathring{\psi}\|_2 \in (2^\ell\bar{\varepsilon},2^{\ell+1}\bar{\varepsilon}]\Big] \leqslant \sum_{\ell=0}^{\infty} 4^{\ell+1}\bar{\varepsilon}^2\Pr\Big[\|\mathring{\psi}\|_2 > 2^\ell\bar{\varepsilon}\Big] \\
\left(\Pr[\|\mathring{\psi}\|_2 > \tau\bar{\varepsilon}] \leqslant 8k\Big(\tfrac{\delta}{8k}\Big)^\tau\right) &\leqslant \sum_{\ell=0}^{\infty} 4^{\ell+1}\bar{\varepsilon}^2 8k\left(\frac{\delta}{8k}\right)^{2^\ell} \\
\text{(since } \delta/k \leqslant 1\text{, and } 2^\ell \geqslant \ell+1) &\leqslant \sum_{\ell=0}^{\infty} 4^{\ell+1}\bar{\varepsilon}^2 8k\left(\frac{\delta}{8k}\right)^{\ell+1} \\
\text{(since } \delta/k \leqslant 1) &= 4\delta\varepsilon^2\frac{k}{d}\sum_{\ell=0}^{\infty}\left(\frac{\delta}{2k}\right)^\ell \leqslant 4\delta\varepsilon^2\frac{k}{d}\sum_{\ell=0}^{\infty}\left(\frac{1}{2}\right)^\ell \\
(3.14) &= 8\delta\varepsilon^2\frac{k}{d}.
\end{aligned}$$

This then implies that

$$\begin{aligned}
\left\|\mathrm{Cov}(\Pi_k\tilde{\psi}) - \mathrm{Cov}(\Pi_k\breve{\psi})\right\| &= \left\|\mathbb{E}\Big[|\mathring{\psi}\rangle\langle\mathring{\psi}| - \Pi_k|\breve{\psi}-\psi\rangle\langle\breve{\psi}-\psi|\Pi_k\Big]\right\| \\
&= \left\|\mathbb{E}\Big[\Big(|\mathring{\psi}\rangle\langle\mathring{\psi}| - \Pi_k|\breve{\psi}-\psi\rangle\langle\breve{\psi}-\psi|\Pi_k\Big)\mathbb{1}_{A\cup(X_\zeta=0)}\Big]\right\| \\
(3.15) &\leqslant \left\|\mathbb{E}\Big[|\mathring{\psi}\rangle\langle\mathring{\psi}|\mathbb{1}_{A\cup(X_\zeta=0)}\Big]\right\| + \left\|\mathbb{E}\Big[\Pi_k|\breve{\psi}-\psi\rangle\langle\breve{\psi}-\psi|\Pi_k\cdot\mathbb{1}_{A\cup(X_\zeta=0)}\Big]\right\|,
\end{aligned}$$

where the second equality is because $\tilde{\psi} = \breve{\psi}$ on the complement of the event $A \cup (X_\zeta = 0)$. Using the definition of $\breve{\psi}$, and the fact that $\Pi_k(I - \Pi_k) = 0$, we can see that $\Pi_k\breve{\psi}$ conditioned on $A \cup (X_\zeta = 0)$ is actually a fixed vector $\mathbb{E}[\Pi_k\tilde{\psi} \mid A \cup (X_\zeta = 0)]$, not a random variable anymore. We now have

$$\left\|\mathbb{E}\Big[\Pi_k|\breve{\psi}-\psi\rangle\langle\breve{\psi}-\psi|\Pi_k\cdot\mathbb{1}_{A\cup(X_\zeta=0)}\Big]\right\| = \Pr[A\cup(X_\zeta=0)]\cdot\left\|\mathbb{E}\Big[\Pi_k\,|\tilde{\psi}-\psi\rangle \mid A\cup(X_\zeta=0)\Big]\right\|_2^2.$$

Continuing with Eq. (3.15), we have

$$\begin{aligned}
\left\|\mathrm{Cov}(\Pi_k\tilde{\psi}) - \mathrm{Cov}(\Pi_k\breve{\psi})\right\| &\leqslant \left\|\mathbb{E}\Big[|\mathring{\psi}\rangle\langle\mathring{\psi}|\mathbb{1}_{A\cup(X_\zeta=0)}\Big]\right\| + \delta\|\mathbb{E}[\mathring{\psi} \mid A\cup(X_\zeta=0)]\|_2^2 \\
\text{(by (3.13) and } (1+3/k\leqslant 4)) &\leqslant \mathbb{E}\Big[\|\mathring{\psi}\|_2^2\mathbb{1}_{A\cup(X_\zeta=0)}\Big] + 16\delta\varepsilon^2\frac{k}{d} \\
&= \mathbb{E}\left[\|\mathring{\psi}\|_2^2\mathbb{1}_{A\cup(X_\zeta=0)}\Big(\mathbb{1}_{[0,\bar{\varepsilon}]}(\|\mathring{\psi}\|_2) + \mathbb{1}_{(\bar{\varepsilon},\infty)}(\|\mathring{\psi}\|_2)\Big)\right] + 16\delta\varepsilon^2\frac{k}{d} \\
\text{(by (3.14) and } \Pr[A\cup(X_\zeta=0)] = \delta) &\leqslant \delta\bar{\varepsilon}^2 + 8\delta\varepsilon^2\frac{k}{d} + 16\delta\varepsilon^2\frac{k}{d} = 25\delta\varepsilon^2\frac{k}{d}.
\end{aligned}$$

We obtain

$$\left\|\mathrm{Cov}(\Pi_k\breve{\psi})\right\| \leqslant \left\|\mathrm{Cov}(\Pi_k\tilde{\psi}) - \mathrm{Cov}(\Pi_k\breve{\psi})\right\| + \left\|\mathrm{Cov}(\Pi_k\tilde{\psi})\right\| \leqslant 25\delta\varepsilon^2\frac{k}{d} + \frac{\varepsilon^2}{4d\ln\left(\frac{8k}{\delta}\right)}$$

because $\left\|\mathrm{Cov}(\Pi_k\tilde{\psi})\right\| \leqslant \left\|\mathrm{Cov}(\tilde{\psi})\right\| = \frac{1}{n}\left\|\mathrm{Cov}(\psi') + \mathrm{Cov}(\psi'')\right\|$, and the matrix inside the latter norm can be upper bounded by $2(\frac{1}{2} + \frac{1}{2\eta})I$ using Eq. (3.8). ∎

If we have n conditional samples $|\varphi\rangle := (|0\rangle\left(|\bar{0}\rangle|\psi\rangle + |\bar{0}^{\perp}\rangle\right) + |1\rangle|\bar{0}\rangle|0\rangle)/\sqrt{2}$, then we can first use Corollary 3.1 to produce (with success probability $\geqslant 1-\frac{\delta}{2}$) a $\frac{1}{\sqrt{d}}$-ℓ_∞ approximation ψ' of the vector $|\psi|$ of the magnitudes of entries, which has $\|\psi'\|_2 \leqslant 1$. Setting $\bar{\psi}_i := \frac{|\psi'_i| + \frac{1}{\sqrt{d}}}{2}$, and building a KP-tree for $\bar{\psi}$ to be able to efficiently prepare a state that is coordinate-wise $\frac{1}{4\sqrt{d}}$-close to $|\bar{\psi}\rangle/\|\bar{\psi}\|_2$, we can transform the conditional copies $|\varphi\rangle$ to the form required by Theorem 3.1 using $\mathcal{O}\Big(n\log^2(d)\Big)$ classical operations, ordinary quantum gates and QRAM calls. Since $\eta = \Omega(1)$, we get a time-efficient unbiased tomography algorithm using $\mathcal{O}\left(\frac{d}{\varepsilon^2}\ln\left(\frac{2d}{\delta}\right)\right)$ conditional samples.

3.3 Improved pure-state tomography using state-preparation oracles If we have a state-preparation oracle available, rather than copies of the state, then the precision-dependence can be quadratically improved using iterative refinement [25]:

COROLLARY 3.2. *Let $\psi \in \mathbb{C}^d$ such that $\|\psi\| \leqslant 1$, and $\varepsilon, \delta \in (0, \frac{1}{2}]$. Suppose we have access to a controlled unitary U (and its inverse) that prepares the state $U|0^{\otimes a'}\rangle = |0^{\otimes a}\rangle|\psi\rangle + |0^{\otimes a}{}^{\perp}\rangle$, where $a', a = \mathcal{O}\big(\mathrm{poly}\log(d/(\delta\varepsilon))\big)$. There is a quantum algorithm that outputs a random vector $\tilde{\psi} \in \mathbb{C}^d$ such that, for every set $V = \{v^{(1)}, v^{(2)}, \ldots, v^{(k)}\}$ of unit vectors, with probability at least $1-\delta$, $|\langle\psi - \tilde{\psi}|v\rangle| \leqslant \varepsilon/\sqrt{d}$ for all $v \in V$, using $\mathcal{O}(\frac{d}{\varepsilon}\mathrm{poly}\log(kd/(\varepsilon\delta)))$ applications of controlled U, $U^\dagger$, two-qubit quantum gates, read-outs of a QRAM of size $\mathcal{O}(d \cdot \mathrm{poly}\log(kd/(\varepsilon\delta)))$, and classical computation.*

If $k = d$ and V is an orthonormal set, then $\tilde{\psi}$ is δ-close in total variation distance to an "almost ideal" discrete random variable $\breve{\psi} \in \mathbb{C}^d$ such that $\mathbb{E}[\breve{\psi}] = \psi$, $\Pr[\forall v \in V\colon |\langle\breve{\psi} - \psi|v\rangle| \leqslant \frac{\varepsilon}{\sqrt{d}}] = 1$, and $\|\mathrm{Cov}(\breve{\psi})\| \leqslant \frac{\varepsilon^2}{d}$.

Proof. The idea is to use the tomography algorithm of [25] to get an estimator ψ' with ℓ_2-error ε, with success probability $\geqslant 1-\frac{\delta}{4}$, using $\mathcal{O}\Big(\frac{d}{\varepsilon}\log(d/\delta)\Big)$ queries in time $\mathcal{O}(\frac{d}{\varepsilon}\log(1/\delta)\cdot\mathrm{poly}\log(d/\varepsilon))$. In case of failure we set $\breve{\psi} = \psi$.

We first build the KP-tree for ψ' in QRAM. We can now prepare a state $|0\rangle|\psi'\rangle + |1\rangle|.\rangle$, and thus also the state $|00\rangle|(\psi-\psi')/2\rangle + |1\rangle|..\rangle$, and using linearized amplitude amplification [26, Theorem 30], we can also prepare a subnormalized state ϕ such that $\|\phi - (\psi-\psi')/(2\varepsilon)\| \leqslant \frac{\delta}{16\sqrt{d}}$ with $\mathcal{O}\big(\log(d/\delta)/\varepsilon\big)$ (controlled) uses of U and $U^\dagger$.

As discussed at the start of this subsection, by Corollary 3.1 using $d\ln(\frac{6d}{\delta})$ copies of $|\phi\rangle$ we can output a vector $\boldsymbol{\mu} \in [0,1]^d$ such that with probability at least $1-\frac{\delta}{4}$, $|\boldsymbol{\mu}_j - |\phi_j|| \leqslant \frac{1}{\sqrt{d}}$ for every $j \in [d]$. (In case of failure we once again set $\breve{\psi} = \psi$.) Upon success, the vector $\boldsymbol{\mu'} := \frac{1}{2}\boldsymbol{\mu} + \frac{1}{2\sqrt{d}}\mathbf{1}_d$ where $\mathbf{1}_d$ is the d-dimensional all-1 vector, satisfies $|\boldsymbol{\mu}'_j|^2 \geqslant \frac{1}{4}\max\{|\phi_j|^2, \frac{1}{d}\}$ for every $j \in [d]$. Also, by using $\tilde{\mathcal{O}}(d)$ time and QRAM bits, we can construct a KP-tree $\mathrm{KP}_{\boldsymbol{\mu'}}$ for $\boldsymbol{\mu'}$. Thus, by using one query to $\mathrm{KP}_{\boldsymbol{\mu'}}$ and $\tilde{\mathcal{O}}(1)$ time, we can prepare a state $|\bar{0}\rangle|\boldsymbol{\mu'}\rangle + |\bar{0}'^{\perp}\rangle$, where $|\boldsymbol{\mu'}\rangle = \sum_{j\in[d]} \boldsymbol{\mu'}_j|j\rangle$.

By Theorem 3.1 we can output an unbiased estimator $\tilde{\phi}$ of ϕ such that $\Pr[\forall v \in V\colon |\langle\tilde{\phi} - \phi|v\rangle| \leqslant \frac{1}{32\sqrt{d}}] \geqslant 1-\frac{\delta}{4}$. Defining $\tilde{\psi} := \psi' + 2\varepsilon\tilde{\phi}$ we then have $\Pr[\forall v \in V\colon |\langle\tilde{\psi} - \psi|v\rangle| \leqslant \frac{\varepsilon}{8\sqrt{d}}] \geqslant 1-\frac{\delta}{4}$ since $\|\phi - (\psi-\psi')/(2\varepsilon)\| \leqslant \frac{\delta}{16\sqrt{d}}$. If V is an orthonormal basis, then furthermore $\tilde{\phi}$ is $\delta/4$-close to an "ideal" (though not error-free) unbiased estimator ϕ' of ϕ such that $\Pr[\forall v \in V\colon |\langle\phi' - \phi|v\rangle| \leqslant \frac{1}{8\sqrt{d}}] = 1$. Since $\|\phi - (\psi-\psi')/(2\varepsilon)\| \leqslant \frac{\delta}{16\sqrt{d}}$ there is another discrete-valued estimator $\breve{\phi}$ within total variation distance $\frac{\delta}{4}$ to ϕ' that satisfies $\mathbb{E}[\breve{\phi}] = (\psi-\psi')/(2\varepsilon)$ and $\Pr[\|\breve{\phi} - \phi'\| > \frac{1}{4\sqrt{d}}] = 0$ in turn implying $\Pr[\forall v \in V\colon |\langle\breve{\phi} - (\psi-\psi')/(2\varepsilon)|v\rangle| \leqslant \frac{1}{2\sqrt{d}}] = 1$. We then set $\breve{\psi} := \psi' + 2\varepsilon\breve{\phi}$ (in case no failure happened). We have $\|\mathrm{Cov}(\breve{\phi})\| \leqslant 2\|\mathrm{Cov}(\breve{\phi} - \phi')\| + 2\|\mathrm{Cov}(\phi')\|$, because for all vectors a, b, the matrix $2aa^\dagger + 2bb^\dagger - (a+b)(a+b)^\dagger = (a-b)(a-b)^\dagger$ is psd. Therefore the claimed properties of $\tilde{\psi}, \breve{\psi}$ follow from those of $\tilde{\phi}, \phi'$ as guaranteed by Theorem 3.1, assuming $\delta \leqslant \frac{1}{k}$ without loss of generality. ∎

4 Quantum noisy power method

In this section we introduce quantum algorithms for approximating the top eigenvector or top-q eigenvectors. For simplicity, we assume the input matrix A is real and Hermitian, and has operator norm $\|A\| \leqslant 1$ (which implies all entries are in $[-1,1]$). Since A is Hermitian, its eigenvalues $\lambda_1, \ldots, \lambda_d$ are real, and we assume them to be

ordered in descending order according to their absolute value.[16] Since the entries of A are real, there is always an associated orthonormal basis of *real* eigenvectors $v_1, \ldots, v_d$. For simplicity and without loss of generality, when we mention the qth eigenvector of A, we mean v_q in this basis. The goal of the algorithms in this section is to find a unit vector w which has large overlap with v_1 in the sense that $|\langle w, v_1\rangle| \geqslant 1 - \varepsilon^2/2$. Note that this is equivalent to finding a unit vector w satisfying that either $\|w - v_1\|_2$ or $\|w + v_1\|_2$ is small (at most ε), and hence we say w approximates v_1 with small ℓ_2-error.

4.1 Classical noisy power method for approximating the top eigenvector For the sake of completeness and pedagogy, we start with the noisy power method of Hardt and Price [32], given in Algorithm 1 below. Like the usual power method, it works by starting with a random vector and applying A some K times to it; the resulting vector will converge to the top eigenvector after relatively small K, assuming some gap between the first and second eigenvalues of A. We include a short proof explaining how the noisy power method can approximate the top eigenvector of A even if there is a small noise vector G_k in the kth matrix-vector computation that does not have too much overlap with v_1.

input : a Hermitian matrix $A \in [-1, 1]^{d\times d}$ with operator norm $\|A\| \leqslant 1$;
Let w_0 be a unit vector randomly chosen from S^{d-1};
for $k \leftarrow 0$ **to** $K - 1$ **do**
  $y_k = Aw_k + G_k$;
  $w_{k+1} = y_k/\|y_k\|_2$;
end
output: w_K;

Algorithm 1: Noisy power method (NPM) for approximating the top eigenvector of A

THEOREM 4.1. *Let A be a $d \times d$ Hermitian matrix with top eigenvector v_1, first and second eigenvalues λ_1 and λ_2, and $\gamma = |\lambda_1| - |\lambda_2|$. Let $\varepsilon \in (0, 0.5)$ and $K = \frac{10|\lambda_1|}{\gamma}\log(20d/\varepsilon)$ (for larger K the theorem still holds). Suppose $|\langle G_k, v_1\rangle| \leqslant \gamma/(50\sqrt{d})$ and $\|G_k\|_2 \leqslant \varepsilon\gamma/5$ for all $k \in [K] - 1$. Then the unit vector w_K in Algorithm 1 satisfies $|\langle w_K, v_1\rangle| \geqslant 1 - \varepsilon^2/2$ with probability $\geqslant 0.9$.*

Proof. Let $w_0 = \sum_{i\in[d]} \alpha_i^{(0)} v_i$. Because w_0 is chosen uniformly at random over the unit sphere, by Corollary 2.5 (without loss of generality assuming $d \geqslant 4$), with probability $\geqslant 0.9$, we have $|\alpha_1^{(0)}| \geqslant 1/(10\sqrt{d})$ and hence we assume $|\alpha_1^{(0)}| \geqslant 1/(10\sqrt{d})$ below for simplicity. Suppose $w_k = \sum_{i\in[d]} \alpha_i^{(k)} v_i$. We define the tangent angle of w_k as $\tan\theta_k = \frac{\sin\theta_k}{\cos\theta_k} = \frac{\sqrt{\sum_{i=2}^d (\alpha_i^{(k)})^2}}{|\alpha_1^{(k)}|}$, and hence $\tan\theta_0 \leqslant 10\sqrt{d}$. It suffices to show that $\tan\theta_K \leqslant \varepsilon/2$, because that implies $|\langle w_K, v_1\rangle| = \cos\theta_K \geqslant 1 - \varepsilon^2/4$.

Since $w_k = \sum_{i\in[d]} \alpha_i^{(k)} v_i$, we have $Aw_k = \sum_{i\in[d]} \alpha_i^{(k)} \lambda_i v_i$. Also because G_k satisfies $|\langle G_k, v_1\rangle| \leqslant \gamma/(50\sqrt{d})$

[16] Sometimes the eigenvalues are ordered $1 \geqslant \lambda_1 \geqslant \cdots \geqslant \lambda_d \geqslant -1$ according to their value (instead of their absolute value). To find the v_1 associated with λ_1 in this situation, one can just let $A' = A/3 + 2I/3$. Then the eigenvectors of A and A' are the same, and the eigenvalues of A' now are all between 1/3 and 1 (and hence one can use Algorithm 1 to find v_1). This trick can also be used to find v_d by simply considering $A'' = -A/3 + 2I/3$. Note that the gap between the top and the second eigenvalues of A' might be different from the gap between the top and the second eigenvalues of A''.

and $\|G_k\|_2 \leqslant \varepsilon\gamma/50$, we can give an upper bound for $\tan\theta_{k+1}$ as follows:

$$\begin{aligned}
\tan\theta_{k+1} &\leqslant \frac{\sqrt{\sum_{i=2}^d(\lambda_i^2\alpha_i^{(k)})^2}+\|G_k\|_2}{|\lambda_1|\cdot|\alpha_1^{(k)}|-|\langle G_k,v_1\rangle|} \leqslant \frac{|\lambda_2|\sqrt{\sum_{i=2}^d(\alpha_i^{(k)})^2}+\varepsilon\gamma/5}{|\lambda_1|\cdot|\alpha_1^{(k)}|-\gamma/(50\sqrt{d})} \leqslant \frac{1}{\cos\theta_k}\cdot\frac{|\lambda_1|\sin\theta_k+\varepsilon\gamma/5}{|\lambda_1|-\gamma/5} \\
&= \frac{1}{\cos\theta_k}\cdot\frac{|\lambda_1|\sin\theta_k+\varepsilon\gamma/5}{|\lambda_2|+4\gamma/5} = \frac{\sin\theta_k}{\cos\theta_k}\cdot\frac{|\lambda_1|}{|\lambda_2|+4\gamma/5}+\frac{1}{\cos\theta_k}\cdot\frac{\varepsilon\gamma/5}{|\lambda_2|+4\gamma/5} \\
&\leqslant \tan\theta_k\cdot\frac{|\lambda_1|}{|\lambda_2|+4\gamma/5}+(1+\tan\theta_k)\cdot\frac{\varepsilon\gamma/5}{|\lambda_2|+4\gamma/5} \\
&= \big(1-\frac{\gamma/5}{|\lambda_2|+\gamma/5}\big)\frac{|\lambda_2|+\varepsilon\gamma/5}{|\lambda_2|+3\gamma/5}\tan\theta_k+\frac{\gamma/5}{|\lambda_2|+4\gamma/5}\varepsilon \leqslant \max\{\varepsilon,\frac{|\lambda_2|+\varepsilon\gamma/5}{|\lambda_2|+3\gamma/5}\tan\theta_k\}.
\end{aligned}$$

Note that $\frac{|\lambda_2|+\varepsilon\gamma/5}{|\lambda_2|+3\gamma/5} \leqslant \max\{\varepsilon,\frac{|\lambda_2|}{|\lambda_2|+2\gamma/5}\}$ because the left-hand side is a weighted mean of the components on the right ($\frac{|\lambda_2|+\varepsilon\gamma/5}{|\lambda_2|+3\gamma/5} = \varepsilon\cdot\frac{\gamma/5}{|\lambda_2|+3\gamma/5}+\frac{|\lambda_2|}{|\lambda_2|+2\gamma/5}\cdot\frac{|\lambda_2|+2\gamma/5}{|\lambda_2|+3\gamma/5}$). Also, $\frac{|\lambda_2|}{|\lambda_2|+2\gamma/5} \leqslant (\frac{|\lambda_2|}{|\lambda_2|+5\gamma/5})^{2/5} = (\frac{|\lambda_2|}{|\lambda_1|})^{2/5}$, so we have $\tan\theta_{k+1} \leqslant \max\{\varepsilon,\tan\theta_k\cdot\max\{\varepsilon,(|\frac{\lambda_2}{\lambda_1}|)^{2/5}\}\}$. By letting $K = \frac{10|\lambda_1|}{\gamma}\log(20d/\varepsilon)$, we obtain $\tan\theta_K \leqslant \varepsilon/2$, which concludes the proof. ∎

4.2 Quantum Gaussian phase estimator Before we explain our quantum noisy power method, we introduce another tool which we call the "quantum Gaussian phase estimator". Its aim is to do phase estimation with (approximately) Gaussian error on the estimate. The high-level idea of this estimator is to replace the initial uniform superposition in the algorithm of phase estimation [40, 20] by a discrete Gaussian quantum state, with standard deviation s; then the distribution of the error $\tilde{a}-a$ between the amplitude a and the estimator $\tilde{a}$ produced by the quantum Gaussian amplitude estimator, is also a discrete Gaussian distribution, now with standard deviation $1/s$.[17] Since the latter distribution is sub-Gaussian with parameter $1/s$, with probability at least $1-\delta$ the output is at most $\sqrt{\log(1/\delta)}/s$ away from a. Recall that p_s is the pdf for the Gaussian with standard deviation s, defined in Section 2.9.

THEOREM 4.2. *Let $\delta\in(0,0.1]$, $s \geqslant 20\sqrt{2\log(1/\delta)}$, $a\in[0,1]$, $N = 200\cdot\lceil s\sqrt{\log(100/\delta)}\rceil$, U be a unitary, $|\psi\rangle$ be an eigenvector of U such that $U|\psi\rangle = \exp(\pi i a/4)|\psi\rangle$. There exists a quantum algorithm that for every such U and a, given one copy of $|\psi\rangle$, outputs an estimator $\tilde{a}$ satisfying that $a-\tilde{a}$ distributes δ-close to $\mathcal{D}_{\frac{8}{N}\cdot\mathbb{Z}-\frac{8}{N}\nu,\frac{8}{\sqrt{2}s}}$ for some $\nu\in[0,1)$, using $\mathcal{O}(s\cdot\mathrm{polylog}(s/\delta))$ applications of controlled-U, controlled-U^{-1} and $\tilde{\mathcal{O}}(s\cdot\mathrm{polylog}(s/\delta))$ time.*

Proof. We first explain the algorithm of our quantum Gaussian phase estimator. Let $|\tilde{p_s}\rangle = \frac{1}{\sqrt{\tilde{G}}}\sum_{z\in\{-N/2,\ldots,N/2-1\}} p_s(z)|z\rangle$, where $\tilde{G}$ is a normalizing constant. Let U_z be a unitary that maps $|z\rangle|\psi\rangle \to |z\rangle U^z|\psi\rangle$ for $z\in\{-N/2,\ldots,N/2-1\}$, and U_π be a unitary that maps $|z\rangle\to(-1)^z|z\rangle$ (this U_π is basically a Z-gate on the least-significant bit of z).

The algorithm is as follows. We first prepare the state $|\tilde{p}_s\rangle|w'\rangle$. We then apply U_z to this state, apply U_π, apply QFT_N^{-1} on the first register, and then measure the first register in the computational basis, divide the outcome value by $N/8$, subtract 4 from it, and output this value.

We first explain the time complexity of the above algorithm. To prepare $|\tilde{p_s}\rangle$, it suffices to compute all values of $p_s(z)$ for $z\in\{-N/2,\ldots,N/2-1\}$,[18] and computing all those values $(p_s(-N/2),\ldots,p_s(N/2-1))$ and constructing a KP-tree with those values stored in its leaves takes $\mathcal{O}(N\cdot\mathrm{poly}\log N)$ time. To construct the unitary U_z, it suffices to use $\mathcal{O}(N\mathrm{poly}\log N)$ applications of controlled-U, controlled-U^{-1} and time. Also, QFT_N can be implemented using $\mathcal{O}(\log^2 N)$ elementary gates. As a result, the total cost here is $\mathcal{O}(N\log N\cdot\log(1/\delta)+\log^2 N) = \mathcal{O}(s\cdot\mathrm{polylog}(s/\delta))$ time and $\mathcal{O}(N\cdot\mathrm{poly}\log N) = \mathcal{O}(s\cdot\mathrm{polylog}(s/\delta))$ applications of controlled-U, controlled-U^{-1}.

[17] There have been other variations of quantum phase estimation with non-uniform initial state to improve the statistics of the outcome, for instance in the context of Hamiltonian simulation [19], but to the best of our knowledge ours is the first application with a discrete *Gaussian* initial state.

[18] Once we have those values, we can do the controlled-rotation tricks similar to how the KP-tree routine produces the quantum state. Since we only need to prepare $|\tilde{p}_s\rangle$ once, it is fine for us to prepare the state using $N\log N$ time. This procedure does not require the use of QRAM. See the discussion above Theorem 2.13 in [17].

Now we show the correctness of the above algorithm. Applying $U_\pi U_z$ to $|\tilde{p}_s\rangle|w'\rangle$ gives the state $\frac{1}{\sqrt{G'}} \sum_{z\in\{-N/2,\ldots,N/2-1\}} p_s(z)(-1)^z \exp(\pi i a z/4)\,|z\rangle\,|w'\rangle$. If we discard the second register, which is in tensor product with the rest of the state, then the remaining state is also 9δ-close to

$$|\Psi\rangle = \frac{1}{\sqrt{G}} \sum_{z\in\mathbb{Z}} p_s(z) \exp(2\pi i(a/8+1/2)z)\,|(z+N/2 \bmod N) - N/2\rangle$$

because $s \geqslant 8\sqrt{2\log(1/\delta)}$, $N > 16s\sqrt{2\ln(1/\delta)}$, and by Theorem 2.17, where G is a normalizing constant. Therefore, the distribution of the outcome of the quantum Gaussian phase estimator is 9δ-close to the distribution obtained by measuring the following state (using $a' = a/8 + 1/2$)

$$\begin{aligned}
\mathrm{QFT}_N^{-1}|\Psi\rangle =& \frac{1}{\sqrt{NG}} \sum_{y\in[N]} \sum_{z\in\mathbb{Z}} p_s(z) \exp(2\pi i a' z) \exp(-2\pi i y((z+N/2 \bmod N) - N/2)/N)\,|y\rangle \\
=& \frac{1}{\sqrt{NG}} \sum_{y\in[N]} \sum_{z\in\mathbb{Z}} p_s(z) \exp(2\pi i z(Na'-y)/N)\,|y\rangle && (e^{-2\pi i y N z/N} = 1) \\
=& \frac{1}{\sqrt{NG}} \sum_{y\in[N]} \Big\{ \sum_{z\in\frac{1}{N}\cdot\mathbb{Z}} p_{s/N}(z) \exp(2\pi i z(Na'-y)) \Big\}\,|y\rangle && (z \leftarrow z/N) \\
=& \frac{1}{\sqrt{NG}} \sum_{y\in[N]} \Big\{ N \cdot \sum_{x\in N\cdot\mathbb{Z}} \widehat{p_{s/N} \exp\left(2\pi i z(Na'-y)\right)}(x) \Big\}\,|y\rangle && \text{(by Theorem 2.18)} \\
=& \frac{1}{\sqrt{NG}} \sum_{y\in[N]} \Big\{ \frac{s}{N} \cdot N \cdot \sum_{x\in N\cdot\mathbb{Z}} p_{N/s}(x - Na' + y) \Big\}\,|y\rangle && (\widehat{p_{s/N}} = \frac{s}{N} \cdot p_{N/s}) \\
=& \frac{s}{\sqrt{NG}} \sum_{y\in[N]} p_{N/s}(N\cdot\mathbb{Z} - Na' + y)\,|y\rangle \\
=& \frac{s}{\sqrt{NG}} \sum_{y\in\mathbb{Z}} p_{N/s}(-Na' + y)\,|y \mod N\rangle.
\end{aligned}$$

By Theorem 2.17 again, because $a' \in [1/2, 5/8]$, $N/s \geqslant 8\sqrt{2\log(1/\delta)}$, and $N \geqslant 16(N/s)\sqrt{2\ln(1/\delta)}$, we know $\mathrm{QFT}_N^{-1}|\Psi\rangle$ is 9δ-close to $\frac{1}{\sqrt{G''}} \sum_{y\in[N]} p_{N/s}(-Na'+y)\,|y\rangle$ where G'' is a normalizing constant. Therefore, the probability distribution of $y - Na'$ (letting y be the measurement outcome) is 9δ-close to $\mathcal{D}^{[-N/2,N/2-1]}_{\mathbb{Z}-Na',\frac{N}{\sqrt{2}s}}$. Moreover, since $\sqrt{\log(12/\delta)/\pi} \leqslant \frac{N}{\sqrt{2}s}$ and $10\frac{N}{\sqrt{2}s}\sqrt{2\ln(1/\delta)} \leqslant N/2$, by Corollary 2.4 we know $y - Na'$ is also $9\delta + 4\delta\exp(\delta)$-close to $\mathcal{D}_{\mathbb{Z}-Na',\frac{N}{\sqrt{2}s}} = \mathcal{D}_{\mathbb{Z}-\nu,\frac{N}{\sqrt{2}s}}$ for some $\nu \in [0,1)$, implying that the distribution of $8y/N - 4 - a$ is $9\delta + 4\delta\exp(\delta)$-close to $\mathcal{D}_{\frac{8}{N}\cdot\mathbb{Z}-\frac{8}{N}\nu,\frac{8}{\sqrt{2}s}}$. As a result, the output of the algorithm in the second paragraph is $9\delta + 9\delta + 4\delta\exp(\delta)$-close to $\mathcal{D}_{\frac{8}{N}\cdot\mathbb{Z}-\frac{8}{N}\nu,\frac{8}{\sqrt{2}s}}$. Rescaling δ by a multiplicative constant, we finish the proof. ∎

Using Theorem 2.16, since $\frac{4\sqrt{2}}{s} \geqslant 8\sqrt{\log(12/\delta)/\pi}/N$ by the choice of N, we can see $\mathcal{D}_{\frac{8}{N}\cdot\mathbb{Z}-8\nu,\frac{4\sqrt{2}}{s}}$ is δ-sub-Gaussian with parameter $\frac{4\sqrt{2}}{s}$. By letting $s = \frac{4\sqrt{2}}{\varepsilon}$, we have the following corollary.

COROLLARY 4.1. (SUB-GAUSSIAN PHASE ESTIMATOR, SUBGPE(U,ε,τ)) *Let $\varepsilon,\tau \in (0, 0.1]$, $a \in [0,1]$, U be a unitary, $|\psi\rangle$ be an eigenvector of U such that $U|\psi\rangle = \exp(\pi i a/4)|\psi\rangle$. There exists a quantum algorithm that, given one copy of $|\psi\rangle$, outputs an estimator $\tilde{a}$ satisfying that $a - \tilde{a}$ is τ-close to τ-sub-Gaussian with parameter ε using $\tilde{\mathcal{O}}(\mathrm{poly}\log(1/\tau)/\varepsilon)$ applications of controlled-U, controlled-U^{-1} and elementary gates.*

4.3 Quantum noisy power method using Gaussian phase estimator In this subsection we combine the noisy power method and the quantum Gaussian phase estimator (introduced in the previous subsection) to get a quantum version of the noisy power method. It approximates the top eigenvector of a given matrix A with additive ℓ_2-error ε in $\tilde{\mathcal{O}}(d^{1.75}/(\gamma^2\varepsilon))$ time, which is a factor $d^{0.25}$ faster in its d-dependence than the best-possible classical algorithm (see Section 5.2 for the $\Omega(d^2)$ classical lower bound).

We first prove the following theorem, which helps us estimate an individual entry of a matrix-vector product Aw (the vector u would be one of the rows of A).

THEOREM 4.3. (INNER PRODUCT ESTIMATOR, IPE$(\tau, \delta, \varepsilon)$) *Let $\tau, \delta, \varepsilon \in (0, 0.1]$, and $u, w \in B_2^d$ s.t. $\|w\|_2 = 1$. Suppose we can access a KP-tree KP_w of w and have quantum query access to entries of u by a unitary O_u. There is a quantum algorithm that with probability at least $1-\delta$, outputs an estimator $\tilde{\mu}$ satisfying that $\tilde{\mu} - \langle u, w\rangle$ is τ-close to τ-subG(ε^2), using $\tilde{\mathcal{O}}(d^{0.75}\text{poly}\log(d/\delta) + d^{0.25}\text{poly}\log(1/\tau)/\varepsilon)$ time.*

Proof. In this proof, we index entries of vectors starting from 0. Let $I_1 = [-d^{-0.25}, d^{-0.25}]$ and $I_2 = [-1, 1]\backslash I_1$. Let $u = u_1 + u_2$, where $(u_1)_j = u_j \mathbb{1}_{I_1}(u_j)$ and $(u_2)_j = u_j \mathbb{1}_{I_2}(u_j)$ for every $j \in [d] - 1$. Informally, u_1 is the vector with smallish entries, and u_2 is the vector with largish entries. We separately estimate $\langle u_1, w\rangle$ and $\langle u_2, w\rangle$.

Finding u_2 and computing $\langle u_2, w\rangle$. The number of nonzero entries of u_2 is at most $\sqrt{d}$ because $\|u\|_2 \leqslant 1$. We first find (with success probability at least $1-\delta/2$) all the nonzero entries of u_2 in $\mathcal{O}(\sqrt{d \cdot \sqrt{d}} \cdot \text{poly}\log(d/\delta)) = \mathcal{O}(d^{0.75} \cdot \text{poly}\log(d/\delta))$ time using Theorem 2.2. Since we can query the entries of w through its KP-tree, we can now compute $\langle u_2, w\rangle$ in time $\tilde{\mathcal{O}}(\sqrt{d})$.

Estimating $\langle u_1, w\rangle$. Our goal below is first to show how to prepare (in time $\tilde{\mathcal{O}}(1)$) a superposition corresponding to the vector $d^{-0.25}u_1$, using the fact that all entries of u_1 are small; and then to use this to estimate $\langle u_1, w\rangle$ with Gaussian error in time $\tilde{\mathcal{O}}(d^{0.25}/\varepsilon)$.

We can implement O_{u_1} using 2 queries to O_u and $\tilde{\mathcal{O}}(1)$ elementary gates: query O_u, and then apply O_u^{-1} conditional on the magnitude of the value being $> d^{-0.25}$ to set the value back to 0 for entries that are in the support of u_2 rather than u_1. Let CR be a controlled rotation such that for every $a \in [-d^{-0.25}, d^{-0.25}]$

$$CR\,|a\rangle\,|0\rangle = |a\rangle\,(a \cdot d^{0.25}\,|0\rangle + \sqrt{1 - a^2 \cdot d^{0.5}}\,|1\rangle).$$

This can be implemented up to negligibly small error by $\tilde{\mathcal{O}}(1)$ elementary gates. Using one application each of O_{u_1}, $O_{u_1}^{-1}$, and CR, and $\tilde{\mathcal{O}}(1)$ elementary gates, we can map

$$\begin{aligned}
|0^{\otimes \log d}\rangle\,|0\rangle\,|0\rangle &\xrightarrow{H^{\otimes \log d}\otimes I} \sum_{j\in[d]-1} d^{-0.5}\,|j\rangle\,|0\rangle\,|0\rangle \xrightarrow{O_{u_1}\otimes I} \sum_{j\in[d]-1} d^{-0.5}\,|j\rangle\,|(u_1)_j\rangle\,|0\rangle \\
&\xrightarrow{I_d\otimes CR} \sum_{j\in[d]-1} (d^{-0.25}(u_1)_j\,|j\rangle\,|(u_1)_j\rangle\,|0\rangle + (\sqrt{d^{-1} - (u_1)_j^2 \cdot d^{-0.5}}\,|j\rangle\,|(u_1)_j\rangle\,|1\rangle) \\
&\xrightarrow{O_{u_1}^{-1}\otimes I} \sum_{j\in[d]-1} (d^{-0.25}(u_1)_j\,|j\rangle\,|0\rangle\,|0\rangle + (\sqrt{d^{-1} - (u_1)_j^2 \cdot d^{-0.5}}\,|j\rangle\,|0\rangle\,|1\rangle).
\end{aligned}$$

Swapping the second register to the front of the state, we showed how to implement the state-preparation unitary $U_{d^{-0.25}u_1}$ that maps

$$U_{d^{-0.25}u_1} : |0\rangle\,|0^{\otimes \log d}\rangle \to |0\rangle \sum_{j\in[d]-1} (d^{-0.25}(u_1)_j)\,|j\rangle + |1\rangle\,|\Phi\rangle,$$

for some arbitrary unnormalized state $|\Phi\rangle$.

Now we show how to estimate $\langle u_1, w\rangle$. Since we have a KP-tree of w, we can implement the state-preparation unitary U_w that maps $|0^{\otimes \log d}\rangle \to \sum_{j\in[d]-1} w_j\,|j\rangle$ using $\tilde{\mathcal{O}}(1)$ time by Theorem 2.4. Let $W = I \otimes U_w \otimes Z$ and $V = U_{d^{-0.25}u_1} \otimes I$. Using Theorem 2.9 (with x ranging over 2 cases), we can implement a $(1, \log d + 2, 0)$ block-encoding of $\text{diag}(\{d^{-0.25}\langle u_1, w\rangle, -d^{-0.25}\langle u_1, w\rangle\})$ in $\tilde{\mathcal{O}}(1)$ time. In order to be able to use our Gaussian phase estimator, we want to convert $\langle u_1, w\rangle$ into an eigenphase. To that end, by Theorem 2.5, we implement a unitary U_{exp} which is a $(1, \log d + 2, 0)$-block-encoding of $W = \exp(i\pi(\langle u_1, w\rangle/4)Z)$ using $\tilde{\mathcal{O}}(d^{0.25})$ time. Since $|0\rangle$ is an eigenvector of W with eigenvalue $\exp(i\pi\langle u_1, w\rangle/4)$, our Gaussian phase estimator (Corollary 4.1) can output (with success probability $\geqslant 1 - \delta/2$) an estimator $\tilde{\eta}$ such that $\tilde{\eta} - \langle u_1, w\rangle$ is τ-close to τ-subG(ε^2) using $\tilde{\mathcal{O}}(\text{poly}\log(1/\tau)/\varepsilon)$ applications of controlled-U_{exp}, controlled-U_{exp}^{-1}, and time.

Because $\langle u, w\rangle = \langle u_1, w\rangle + \langle u_2, w\rangle$, with probability at least $1 - \delta$ we obtain an estimator $\tilde{\mu} = \tilde{\eta} + \langle u_2, w\rangle$ such that $\tilde{\mu} - \langle u, w\rangle = \tilde{\eta} - \langle u_1, w\rangle$ is τ-close to τ-subG(ε^2). ∎

Note that the two terms in the time complexity of the above theorem are roughly equal if ε is a small constant times $1/\sqrt{d}$. Such a small error-per-coordinate translates into a small overall ℓ_2-error for a d-dimensional vector.

Accordingly, such a setting of ε is what we use in our quantum noisy power method for estimating the top eigenvector.

input : a Hermitian matrix $A \in [-1,1]^{d\times d}$ with operator norm at most 1;
Let $\gamma = |\lambda_1(A)| - |\lambda_2(A)|$; $\varepsilon \in (0,1)$; $K = \frac{10|\lambda_1|}{\gamma}\log(20d/\varepsilon)$;
Let $\delta = 1/(1000Kd)$; $\tau = \delta/(1000Kd^2)$; $\zeta = \frac{\varepsilon\gamma}{100d^{0.5}\sqrt{\log(1000Kd/\delta)}}$;
Let w_0 be a unit vector randomly chosen from S^{d-1};
for $k \leftarrow 0$ **to** $K-1$ **do**
  Prepare a KP-tree for w_k;
  For every $j \in [d]$, compute an estimator $(y_k)_j$ of $\langle A_j, w_k\rangle$ using IPE(τ,δ,ζ) of Theorem 4.3 (A_j is the jth row of A);
  $w_{k+1} = y_k/\|y_k\|_2$;
end
output: w_K;

Algorithm 2: Quantum noisy power method using Gaussian phase estimator

Theorem 4.4. (Quantum noisy power method using Gaussian phase estimator, Algorithm 2) *Let $A \in [-1,1]^{d\times d}$ be a symmetric matrix with operator norm at most 1, first and second eigenvalues $\lambda_1(A)$ and $\lambda_2(A)$, $\gamma = |\lambda_1(A)| - |\lambda_2(A)|$, $v_1 = v_1(A)$ be the top eigenvector of A, and $\varepsilon \in (0,1]$. Suppose we have quantum query access to entries of A. There exists a quantum algorithm (namely Algorithm 2) that with probability at least 0.89, outputs a d-dimensional vector w such that $|\langle w, v_1\rangle| \geqslant 1 - \varepsilon^2/2$, using $\tilde{\mathcal{O}}(d^{1.75}/(\gamma^2\varepsilon))$ time and $\tilde{\mathcal{O}}(d)$ QRAM bits.*

Proof. Each iteration of Algorithm 2 uses $\tilde{\mathcal{O}}(d)$ time and QRAM bits to build the KP-tree for w_k, and for every $j \in [d]$, we use $\tilde{\mathcal{O}}(d^{0.75}\mathrm{poly}\log(d/\delta) + d^{0.25}\mathrm{poly}\log(1/\tau)/\zeta)$ time and $\tilde{\mathcal{O}}(\sqrt{d})$ QRAM bits for estimating $\langle A_j, w_\ell\rangle$ by IPE(τ,δ,ζ) in Theorem 4.3. Hence the total number of elementary gates we used and queries to entries of A is $\tilde{\mathcal{O}}(d\cdot(d^{0.75}+d^{0.25}/\zeta)\cdot K) = \tilde{\mathcal{O}}(d^{1.75}/(\gamma^2\varepsilon))$.

Now we are ready to show the correctness of Algorithm 2. By Theorem 4.1 and the union bound, it suffices to show that for each $k \in [K]-1$, both $\|y_k - Aw_k\|_2 \leqslant \gamma\varepsilon/5$ and $|\langle y_k - Aw_k, v_1\rangle| \leqslant \gamma/(50\sqrt{d})$ hold with probability $\geqslant 1-1/(100K)$. Fix k. By Theorem 4.3, we know that for every $j \in [d]$, with probability at least $1-\delta$, $(y_k - Aw_k)_j$ is τ-close to τ-subG(ζ^2) for every $j \in [d]$. Let $e = y_k - Aw_k$. There are three different kinds of bad events, whose probabilities we now analyze. Firstly, we can see that for every $j \in [d]$, with probability at least $1-\delta$,

$$\Pr\left[|e_j| > \frac{\gamma\varepsilon}{5\sqrt{d}}\right] \leqslant 2\exp(\tau)\cdot\exp\left(-\frac{(\frac{\gamma\varepsilon}{5\sqrt{d}})^2}{2\zeta^2}\right) + \tau = 2\exp(\tau - 200\log(1000Kd/\delta)) + \tau \leqslant \frac{\delta}{100Kd}.$$

Therefore, with probability at least $1 - d\delta - \delta/(100K)$, $|e_j| \leqslant \frac{\gamma\varepsilon}{5\sqrt{d}}$ for every $j \in [d]$, implying that $\|e\|_2 \leqslant \frac{\gamma\varepsilon}{5}$. Secondly, by the properties of sub-Gaussians from Section 2.9, we know that with probability at least $1-d\delta$, $\langle e, v_1\rangle$ is $d\tau$-close to $d\tau$-subG$(\sum_{j\in[d]}(v_1)_j^2\zeta^2)$. Thirdly, since v_1 is a unit vector, we have that (with probability at least $1-d\delta$) $\langle e, v_1\rangle$ is $d\tau$-close to $d\tau$-subG(ζ^2) and hence

$$\begin{aligned}\Pr\left[|\langle e, v_1\rangle| > \frac{\gamma}{50\sqrt{d}}\right] &\leqslant 2\exp(d\tau)\cdot\exp\left(-\frac{(\frac{\gamma}{50\sqrt{d}})^2}{2\zeta^2}\right) + d\tau \\ &= 2\exp(d\tau)\cdot\exp\left(-2\frac{\log(1000Kd/\delta)}{\varepsilon^2}\right) + d\tau \leqslant \frac{\delta}{100Kd}.\end{aligned}$$

As a result, by the union bound over the three kinds of error probabilities, for each $k \in [K]\cup\{0\}$, with probability $1 - 2d\delta - \delta/(50K) \geqslant 1 - 1/(100K)$, we have both $\|y_k - Aw_k\|_2 \leqslant \gamma\varepsilon/5$ and $|\langle y_k - Aw_k, v_1\rangle| \leqslant \gamma/(50\sqrt{d})$. This proves correctness of Algorithm 2. ∎

4.4 Almost optimal process-tomography of "low-rank" reflections In this subsection we describe an essentially optimal algorithm for the "tomography" of projectors Π of rank at most q (or of the corresponding unitary reflection $2\Pi - I$).[19] We will use this in the next subsection to approximate the eigensubspace spanned by the top-q eigenvectors. For generality, we will from here on allow our matrices to have complex entries, not just real entries like in the earlier subsections.

Our algorithm is inspired by the noisy power method and has query complexity $\tilde{\mathcal{O}}\big(dq/\varepsilon\big)$ and time complexity $\tilde{\mathcal{O}}\big(dq/\varepsilon + dq^2\big)$ (using QRAM). When $q \ll d$ this gives a better complexity than the optimal unitary process-tomography algorithm of Haah, Kothari, O'Donnell, and Tang [31]. Also in the special case when $q = 1$ this gives a qualitative improvement over prior pure-state tomography algorithms [39, 5] which required a state-preparation unitary, while for us it suffices to have a reflection about the state, which is a strictly weaker input model.[20] Surprisingly, it turns out that this weaker input essentially does not affect the query and time complexity.

Observe that if we have two projectors Π, Π' of rank r, then

$$\|\Pi - \Pi'\|_1/(2r) \leqslant \|\Pi - \Pi'\| \leqslant \|\Pi - \Pi'\|_1/2. \tag{4.16}$$

This implies that the ε-precise estimation of a rank-1 projector Π (i.e., the density matrix corresponding to a pure state) is equivalent to ε-precise approximation of the quantum state Π. Since the complexity of pure-state tomography using state-preparation unitaries is known to be $\widetilde{\Theta}(d/\varepsilon)$ [5], it follows that our algorithm is optimal up to log factors in the $q = 1$ case.

Similarly, the query complexity of our algorithm is optimal for $\varepsilon = \frac{1}{6}$ up to log factors, as can be seen by an information-theoretic argument using ε-nets, since by Lemma 4.1 there is an ensemble of $\exp(\Omega(dq))$ rank-q' projectors for $q' = \min(q, \lfloor \frac{d}{2} \rfloor)$ such that each pair of distinct projectors is more than $\frac{1}{3}$ apart in operator norm. We conjecture that the query complexity of the task is actually $\widetilde{\Omega}(dq/\varepsilon)$, meaning that our algorithm has essentially optimal query complexity for all $\varepsilon \in (0, \frac{1}{6}]$.

Now we briefly explain our algorithm for finding an orthonormal basis of the image of Π, assuming for ease of exposition that its rank is exactly q. We want to find a $d \times q$ isometry W such that $\|WW^\dagger - \Pi\|$ is small. Our algorithm (Algorithm 3 below) can be seen as a variant of the noisy power method: We start with generating m Gaussian vectors $g_1, \ldots, g_m$ ($m = \widetilde{\Theta}(q)$ will be slightly bigger than q), with i.i.d. (complex) normal entries each having standard deviation $\sim 1/\sqrt{d}$. Subsequently, we repeat the following process $K \approx \log \sqrt{d/m}$ times: we estimate Πg_i for every $i \in [m]$ using our quantum state tomography algorithm (Theorem 3.1 and Corollary 3.2), and multiply the outcome by 2, resulting in m new vectors (these factors of 2 allows our analysis to treat all the errors together in one geometric series later). Finally, let V be the $d \times m$ matrix whose columns are the versions of those m vectors after the last iteration. The algorithm classically computes the singular value decomposition (SVD) of this V, and outputs those left-singular vectors $w_1, w_2, \ldots, w_q$ that have singular value greater than the threshold of $\frac{1}{14}$.

Note that the span of $g_1, \ldots, g_m$ includes the q-dimensional image of Π almost surely. Hence if no error occurs in the tomography step, then $V = 2^K \Pi[g_1, \ldots, g_m]$, and the image of V is the image of Π (almost surely). Thus, if $\sum_{i=1}^{q} \varsigma_i w_i u_i^\dagger$ is the SVD of V, then $\sum_{i \in [q]} w_i w_i^\dagger = \Pi$. That is, we get the desired output W by rounding the singular values of V appropriately: the first q singular values are rounded to 1, and the others are rounded to 0. To show the stability of this approach it remains to show that the singular values $\varsigma_1, \ldots, \varsigma_q$ of the final V in the non-error-free case are still large (≈ 1), the other $d - q$ singular values are still essentially 0, and the error incurred by the quantum state tomography is small. Our analysis relies on the concentration of the singular values of sufficiently random matrices, see Section 2.7.

To bound the effect of errors induced by tomography we combine the operator norm bound on random matrices of Theorem 2.13 with our unbiased tomography algorithm (Theorem 3.1). The key observation is that Corollary 3.2 gives an estimator $\tilde{\psi}$ that is δ-close in total variation distance to an "ideal" (though not error-free) estimator $\check{\psi}$ that satisfies $\mathbb{E}[\check{\psi}] = \psi$, whose covariance matrix has operator norm at most $S^2 \leqslant \frac{\varepsilon^2}{d}$, and

[19] Having access to a controlled reflection $2\Pi - I$ is equivalent up to constant factors to having access to controlled $U_\Pi^{\pm 1}$ (i.e., controlled-U_Π and its inverse) for a block-encoding U_Π of the projector Π, as follows from the QSVT framework [26].

[20] Indeed, we can implement a reflection about a state $|\psi\rangle$ by a state-preparation unitary and its inverse as in amplitude amplification. However, if we only have access to a reflection about an unknown classical basis state $|i\rangle$ for $i \in [d]$, then we need to use this reflection $\Omega(\sqrt{d})$ times to find i (because of the optimality of Grover search) showing that the reflection input is substantially weaker than the state-preparation-unitary input.

$\|\Pi\check{\psi} - \Pi\psi\|_2 \leqslant \varepsilon\sqrt{\frac{q}{d}}$ with certainty. For the sake of analysis we can assume that we work with $\check{\psi}$, because we only notice the difference between $\check{\psi}$ and $\tilde{\psi}$ with probability at most δ, which can be made negligibly small at a logarithmic cost in Corollary 3.2. Finally, we use perturbation bounds on singular values and vectors from Section 2.5.

Input : Dimension d, maximum failure probability $\delta' \in (0, \frac{1}{2}]$, target precision $\varepsilon \in (0, \frac{1}{2}]$, $\tilde{\mathcal{O}}\left(\frac{\varepsilon\delta'}{dq}\right)$-approximate block-encoding U_Π of a projector Π of rank $\leqslant q$

Output: ε-approximate orthonormal basis W of the image of Π, i.e., $\left\|WW^\dagger - \Pi\right\| \leqslant \varepsilon$

Init: $m := \min\left(\left\lceil \max\left(16C^2q, 8c\ln\left(\frac{10}{\delta'}\right)\right)\right\rceil, d\right)$, $K := \left\lceil \log_2\left(\sqrt{\frac{d}{m}} + 1\right)\right\rceil$, $\varepsilon' := \frac{\varepsilon}{65+98\sqrt{\ln\left(\frac{10d}{\delta'}\right)/c'}}$

`// the constants` c, C `come from Corollary 2.2 and in the real case` [10] $c, C = 1$

`// the constant` c' `comes from Theorem 2.13`

1.) **if** $m = d$ **then** set $g_j = |j\rangle/4$, **else** generate m random vectors g_j with i.i.d. complex (or real if $\Pi \in \mathbb{R}^{d\times d}$) standard normal entries multiplied by $\eta := \frac{4\cdot 2^{-K}}{7\sqrt{m}}$ `// note` $\eta < \frac{4}{7\sqrt{d}}$

2.) **for** $j = 1$ **to** m **do**

  Set $g_j^{(0)} = g_j$; **if** $\|g_j\| > 2$ **then ABORT**

  for $k = 0$ **to** $K - 1$ **do**

    Classically compute $\|g_j^{(k)}\|$ and store $g_j^{(k)}/\|g_j^{(k)}\|$ in a KP-tree

    (we can now unitarily prepare $|\psi\rangle = \Pi g_j^{(k)}/\|g_j^{(k)}\|$ using this KP-tree and U_Π)

    Obtain $y_j^{(k+1)}$ via $\varepsilon'/\|g_j^{(k)}\|$-precise tomography (Corollary 3.2) on $|\psi\rangle$ setting $\delta \leftarrow \delta'/(5mK)$

    `// query complexity is` $\tilde{\mathcal{O}}(d/\varepsilon)$

    Set $g_j^{(k+1)} \leftarrow 2\|g_j^{(k)}\|y_j^{(k+1)}$; **if** $\|g_j^{(k+1)}\| > 2$ **then ABORT**

  endfor

endfor

3.) Output the left-singular vectors of $V = [g_1^{(K)}, \ldots, g_m^{(K)}]$ with singular value above $\frac{1}{14}$.

`// Classical complexity is` $\tilde{\mathcal{O}}(dq^2)$ `via direct diagonalization of` $V^\dagger V$

`// The output is correct with probability at least` $1 - \delta'$

`// The total` U_Π`-query and quantum gate complexity is` $\tilde{\mathcal{O}}(dq/\varepsilon)$

Algorithm 3: Time-efficient approximation of the top-q eigensubspace

We say that U_Π is an ε-approximate block-encoding of Π if $\|U_\Pi - U\| \leqslant \varepsilon$ for some U satisfying $\Pi = (\langle 0^a| \otimes I)U(|0^a\rangle \otimes I)$.[21]

THEOREM 4.5. (CORRECTNESS OF ALGORITHM 3) *Let $\Pi \in \mathbb{C}^{d\times d}$ be an orthonormal projector of rank at most q, given via an $\tilde{\mathcal{O}}\left(\frac{\varepsilon\delta'}{dq}\right)$-approximate block-encoding U_Π. Algorithm 3 outputs an isometry W such that, with probability at least $1 - \delta'$, $\|WW^\dagger - \Pi\| \leqslant \varepsilon$, using $\tilde{\mathcal{O}}\left(\frac{dq}{\varepsilon}\right)$ controlled U_Π, $U_\Pi^\dagger$, two-qubit quantum gates, read-outs of a QRAM of size $\tilde{\mathcal{O}}(d)$, and $\tilde{\mathcal{O}}(dq^2)$ classical computation.*

Proof. First consider the case when there is no error in tomography and in the implementation of Π. Then we end up with $g_j^{(K)} = 2^K\Pi g_j^{(0)}$, and the corresponding matrix $V_{ideal} = 2^K[\Pi g_1^{(0)}, \ldots, \Pi g_m^{(0)}]$ has almost surely rank(Π)

[21] The condition $\|(\langle 0^a| \otimes I)U_\Pi(|0^a\rangle \otimes I) - \Pi\| \leqslant \varepsilon'$ appears similar, however is in some sense quadratically weaker. Consider, e.g., $\Pi = 1$, $a = 1$, and $U_\Pi = \begin{pmatrix} \cos(x) & -\sin(x) \\ \sin(x) & \cos(x) \end{pmatrix}$, then $1 - \cos(x) = x^2/2 + \mathcal{O}(x^4)$, but for any unitary $U = |0\rangle\langle 0| + z|1\rangle\langle 1|$ we have $\|U_\Pi - U\| \geqslant |\sin(x)| = |x| + \mathcal{O}(|x|^3)$. Nevertheless, because Π is a projector, we can remedy this in general by converting U_Π to an (approximate) block-encoding of $\Pi/2$ via linear combination of unitaries, and then applying quantum singular value transformation with the polynomial $-T_3(x) = 3x - 4x^3$; the resulting unitary is then indeed $\mathcal{O}(\varepsilon')$-close to a perfect block-encoding of Π, see for example the proof of [26, Lemma 23].

nonzero singular values with associated left-singular vectors lying in the image of Π. The $m = d$ case is trivial. If $m < d$, then due to Corollary 2.2 all corresponding singular values are in $(\frac{1}{7}, 1)$ with probability at least $1 - \frac{\delta'}{5}$.[22] By Proposition 2.1 (and a union bound over all $j \in [m]$) we know that with probability at least $1 - \frac{\delta'}{10}$ we have for all $j \in [m]$ that[23] $\|g_j\| < \eta(\sqrt{d} + \sqrt{2\ln(10m/\delta')}) < \frac{6}{7}$. Thereby with probability at least $1 - \frac{3\delta'}{10}$ Algorithm 3 does not abort at the intialization of $g_j^{(0)}$, and the left-singular vectors of V_{ideal} with singular value at least $1/7$ form the columns of the desired matrix W such that $WW^\dagger = \Pi$; in the remainder of the proof we assume this is the case.

Second, we consider what happens when the tomography has error, but we can implement Π exactly. Let $\tilde{e}_j^{(k,\ell)} := g_j^{(k)} - 2^{k-\ell}\Pi g_j^{(\ell)}$ be the aggregate tomography error that occurred from the ℓ-th iteration to the k-th iteration where $k \in [K]$, $\ell \in [k] - 1$, and observe that $\tilde{e}_j^{(k,\ell)} = \tilde{e}_j^{(k,k-1)} + \sum_{i=\ell+1}^{k-1} 2^{k-i} \cdot \Pi\tilde{e}_j^{(i,i-1)}$. For each iteration, we do the tomography with precision $\varepsilon'/\|g_j^{(k)}\|$ on $\Pi g_j^{(k)}/\|g_j^{(k)}\|$ via Corollary 3.2, guaranteeing that the random variable $\tilde{e}_j^{(k,k-1)}$ is δ-close in total variation distance to an "ideal" random variable $e_j^{(k,k-1)}$ such that $\|e_j^{(k,k-1)}\| \leqslant \varepsilon'$ and $\|\Pi e_j^{(k,k-1)}\| \leqslant \varepsilon'\sqrt{q/d}$ almost surely, and $\mathbb{E}[e_j^{(k,k-1)}] = 0$, $\|\mathrm{Cov}(e_j^{(k,k-1)})\| \leqslant \frac{\varepsilon'}{d^2}$ (to see this, choose V to be an orthonormal basis whose first q elements span the image of Π). We define analogously $e_j^{(k,\ell)} := e_j^{(k,k-1)} + \sum_{i=\ell+1}^{k-1} 2^{k-i} \cdot \Pi e_j^{(i,i-1)}$. Note that the distribution of $\tilde{e}_j^{(k,k-1)}$ (and $e_j^{(k,k-1)}$) can depend on $\tilde{e}_j^{(i,i-1)}$ (and $e_{j'}^{(i,i-1)}$, respectively) only if $j = j'$ and $i \leqslant k$. Let $\tilde{\boldsymbol{e}}_j := (\tilde{e}_j^{(1,0)}, \tilde{e}_j^{(2,1)}, \ldots, \tilde{e}_j^{(K,K-1)})$, and define $\boldsymbol{e}_j$ analogously. We can apply Lemma 2.2 recursively to show that $d_{TV}(\tilde{\boldsymbol{e}}_j, \boldsymbol{e}_j) \leqslant K\delta$. Since the $\tilde{\boldsymbol{e}}_j$ are independent from each other, we can assume without loss of generality that so are the $\boldsymbol{e}_j$. Once again by Lemma 2.2 we get that when comparing the two sequences of m random variables, we have $d_{TV}((\tilde{\boldsymbol{e}}_j : j \in m), (\boldsymbol{e}_j : j \in m)) \leqslant mK\delta = \frac{\delta'}{5}$. From now on we replace $(\tilde{\boldsymbol{e}}_j : j \in m)$ by $(\boldsymbol{e}_j : j \in m)$ throughout the analysis, which can therefore hide an additional failure probability of at most $\frac{\delta'}{5}$.

By the triangle inequality we have that

$$\begin{aligned}\|e_j^{(k,0)}\| &\leqslant \|e_j^{(k,k-1)}\| + \sum_{i=1}^{k-1} 2^{k-i}\|\Pi e_j^{(i,i-1)}\| \leqslant \varepsilon' + \sum_{i=1}^{k-1} 2^{k-i}\varepsilon'\sqrt{q/d} \\ &\leqslant (1 + 2^k\sqrt{q/d})\varepsilon' < (1 + 2(\sqrt{d/m} + 1)\sqrt{q/d})\varepsilon' \leqslant 5\varepsilon'\end{aligned}$$

for every $k \in [K]$. This also implies that for every $k \in [K]$ we have $\|g_j^{(k)}\| \leqslant \|2^k\Pi g_j^{(0)}\| + \|e_j^{(k,0)}\| \leqslant \|2^K\Pi g_j^{(0)}\| + 5\varepsilon' \leqslant 1 + 5\varepsilon' \leqslant 2$, and therefore Algorithm 3 also does not abort in the for-loop.

Let us define $E_{tomo} := [e_1^{(K,0)}, \ldots, e_m^{(K,0)}] = V - V_{ideal}$ as the matrix of accumulated tomography errors. We can apply Lemma 2.3 recursively with $X = \Pi e_j^{(k-1,0)}, Y = (I - \Pi)e_j^{(k-1,0)}, Z = e_j^{(k,k-1)}$ to show that for all $k \in [K]$ we have

$$\begin{aligned}\|\mathrm{Cov}(e_j^{(k,0)})\| &\leqslant \|\mathrm{Cov}(e_j^{(k,k-1)})\| + \sum_{i=1}^{k-1} 4^{k-i}\|\mathrm{Cov}(\Pi e_j^{(i,i-1)})\| \\ &\leqslant \|\mathrm{Cov}(e_j^{(k,k-1)})\| + \sum_{i=1}^{k-1} 4^{k-i}\|\mathrm{Cov}(e_j^{(i,i-1)})\| \\ &\leqslant \sum_{i=1}^{k} 4^{k-i}\frac{\varepsilon'^2}{d} \leqslant \frac{4^K\varepsilon'^2}{3d} \leqslant \frac{4(\sqrt{d/m} + 1)^2\varepsilon'^2}{3d} \leqslant \frac{16\varepsilon'^2}{3m}.\end{aligned}$$

[22] The matrix $\frac{1}{\eta}[g_1^{(0)}, \ldots, g_m^{(0)}]$ is a $d \times m$ random matrix with i.i.d. (complex) standard normal entries. After multiplying by Π this effectively (up to a rotation) becomes a $q \times m$ random matrix with i.i.d. (complex) standard normal entries. We apply Corollary 2.2 (with $N = m$) to the latter matrix, obtaining the interval $[\frac{1}{4}\sqrt{m}, \frac{7}{4}\sqrt{m}]$ for its singular values. Multiplying by $\eta 2^K = 4/(7\sqrt{m})$ we get the interval $[\frac{1}{7}, 1]$ for the singular values of V_{ideal}.

[23] We have $\eta(\sqrt{d} + \sqrt{2\ln(10m/\delta')}) \leqslant \eta\sqrt{d} + \frac{2}{7\sqrt{m}}\sqrt{2\ln(10m/\delta')} < \frac{4}{7} + \frac{2}{7}\sqrt{\frac{2\ln(10m/\delta')}{m}} = \frac{4}{7} + \frac{2}{7}\sqrt{\frac{2\ln(m) + 2\ln(10/\delta')}{m}} < \frac{6}{7}$, because $m \geqslant 8\ln(10/\delta')$ and $\frac{2\ln(x)}{x}$ takes its maximum at $x = e$, where it is less than $\frac{3}{4}$.

Applying Theorem 2.13 for $t = \sqrt{\ln(10d/\delta')/c'}$ gives that with probability at least $1 - \frac{\delta'}{5}$ we have

$$\|E_{tomo}\| \leqslant \frac{8\varepsilon'}{\sqrt{3}} + 7\varepsilon'\sqrt{\ln(10d/\delta')/c'} \leqslant \frac{\varepsilon}{14}. \tag{4.17}$$

Since $\|E_{tomo}\| \leqslant \frac{\varepsilon}{14}$, using the notation of Theorem 2.11 we have that $\Pi = \Pi_{>0}^{V_{ideal}}$ and the rank of $\Pi_{>\frac{1}{14}}^{V}$ is $\operatorname{rank}(\Pi) \leqslant q$ due to Weyl's bound (Theorem 2.10). Therefore, by Theorem 2.11 and Lemma 2.1 we have $\|\Pi - \Pi_{>\frac{1}{14}}^{V}\| \leqslant 14\|V - V_{ideal}\| = 14\|E_{tomo}\| \leqslant \varepsilon$ as desired.

Finally, let us analyze the effect of implementation errors in U_Π. We perform tomography mK times via Corollary 3.2, each time using $T = \mathcal{O}(\frac{d}{\varepsilon}\operatorname{poly}\log(d/(\varepsilon\delta))$ applications of $U_\Pi^{\pm 1}$, therefore the induced total variation distance[24] in the output distribution is at most $TmK \cdot \tilde{\mathcal{O}}\left(\frac{\varepsilon\delta'}{dq}\right) \leqslant \frac{\delta'}{10}$. Preparing a KP-tree that allows a similar precision for the preparation of $g_j^{(k)}/\|g_j^{(k)}\|$ likewise induces at most an additional $\frac{\delta'}{10}$ total variation distance, implying that our algorithm outputs a sufficiently precise answer with probability at least $1 - \delta'$ when all approximations are considered. The quantum gate complexity comes entirely from Corollary 3.2, which is gate efficient, while the final computation requires computing the SVD of a $d \times m$ matrix, which can be performed in $\tilde{\mathcal{O}}(dm^2) = \tilde{\mathcal{O}}(dq^2)$ classical time. ∎

Using basic quantum information theory, one can see that recovering the q-dimensional subspace with small constant error $\varepsilon = 1/6$ gains us $\Omega(dq)$ bits of information about the subspace, and hence requires $\tilde{\Omega}(dq)$ quantum queries. This shows that the query complexity of our previous algorithm is essentially optimal in its dq-dependence. We first show that there exists a large set of q-dimensional projectors that are all far apart from each other. This might be a standard result, but we did not find a reference in the literature so provide our own proof.

LEMMA 4.1. *Let $q \leqslant d/2$. There exists a set S of q-dimensional subspaces of $\mathbb{R}^d$ of size $\exp(\Omega(dq))$ such that for any distinct $s, r \in S$ we have $\|\Pi_s - \Pi_r\| > \frac{1}{3}$, where Π_t denotes the projector to the subspace t.*

Proof. We can assume without loss of generality that $d \geqslant 128$[3].

First let us assume that $q \leqslant d/64$; we show the existence of such a set S via the probabilistic method, by showing that for any set S of subspaces, if $|S| < \exp(qd/32 - 1)$, then with non-zero probability a Haar-random q-dimensional subspace r satisfies $\|\Pi_s - \Pi_r\| > \frac{1}{3}$ for every $s \in S$ (thereby we can take $S \leftarrow S \cup \{r\}$). We sample r as follows: generate a random matrix $R \in \mathbb{R}^{d\times q}$ with i.i.d. standard normal entries, and accept R only if $\varsigma_{\min}(R) \geqslant \frac{3}{4}\sqrt{d}$ (i.e., take a sample conditioned on this happening – we know by Theorem 2.12 that this happens with probability $\geqslant \frac{1}{e}$).

Upon acceptance we compute a singular value decomposition $R = U\Sigma V^\dagger$ and define $\Pi = UU^\dagger$ as the projector corresponding to the subspace. Since $\Pi = UU^\dagger$ is an orthogonal projection to the image of $R = U\Sigma V^\dagger$ we have $\Pi r_i = r_i$ for all columns r_i of R and thus

$$\begin{aligned}
\|\Pi_s r_i\| &\geqslant \|\Pi r_i\| - \|(\Pi - \Pi_s)r_i\| \\
&= \|r_i\| - \|(\Pi - \Pi_s)r_i\| \\
&\geqslant (1 - \|\Pi - \Pi_s\|)\|r_i\| \\
&\geqslant (1 - \|\Pi - \Pi_s\|)\varsigma_{\min}(R) \\
&\geqslant \frac{3}{4}\sqrt{d}(1 - \|\Pi - \Pi_s\|).
\end{aligned}$$

Hence $\|\Pi - \Pi_s\| \leqslant \frac{1}{3}$ is only possible if $\|\Pi_s r_i\| \geqslant \sqrt{d}/2$ for all columns of R. Without the conditioning (on $\varsigma_{\min}(R) \geqslant \frac{3}{4}\sqrt{d}$, the condition we accept R), $\Pi_s r_i$ is effectively a random q-dimensional vector with i.i.d. standard normal entries. Since $\sqrt{d}/2 - \sqrt{q} \geqslant \sqrt{d}/4$, by Proposition 2.1 for any $s \in S$ we have $\Pr[\|\Pi_s r_i\| \geqslant \sqrt{d}/2] \leqslant \exp(-d/32)$, and due to the independece of the columns the probability that this happens for all $i \in [q]$ is less than $\exp(1 - qd/32)$ even after conditioning. Taking the union bound over all $s \in S$ we can conclude that with non-zero probability $\|\Pi - \Pi_s\| > \frac{1}{3}$ for all $s \in S$.

[24] With more careful tracking of error spreading in the estimated vectors it might be possible to show that it suffices to have access to a block-encoding satisfying the weaker condition $\|(\langle 0^a| \otimes I)U_\Pi(|0^a\rangle \otimes I) - \Pi\| \leqslant \frac{\varepsilon'}{\sqrt{dm}}$.

The statement for $q = \Omega(d)$ follows from [30, Lemma 8]. Alternatively, if $q > d/64$, we can set $q' := \lceil q/128 \rceil$, $d' := d - (q - q')$ so that $q' \leqslant d'/64$. Then from a large set S' of q'-dimensional subspaces of $\mathbb{R}^{d'}$ satisfying $\|\Pi_{s'} - \Pi_{r'}\| > \frac{1}{3}$ for any distinct $s', r' \in S'$ we construct $S := \{s \colon \Pi_s = \Pi_{s'} \oplus I_{q-q'} \text{ for some } s' \in S'\}$ so that also $\|\Pi_s - \Pi_r\| > \frac{1}{3}$ for any distinct $s, r \in S$. ∎

4.5 Time-efficiently approximating the subspace spanned by top-q eigenvectors In this subsection, we give a quantum algorithm to "approximate the top-q eigenvectors" in a strong sense using $qd^{1.5+o(1)}$ time (and $q\sqrt{s}d^{1+o(1)}$ time if the matrix is s-sparse). In particular, when $q = 1$, this algorithm outputs a vector that approximates the top eigenvector using $d^{1.5+o(1)}$ time. This is what we referred to as our "second algorithm" in Section 1.2.

Consider the following situation: for $q \in [d]$, suppose we only know there is a significant eigenvalue gap between the qth eigenvalue λ_q and the $(q+1)$th eigenvalue λ_{q+1}. Is there a way we can learn the subspace spanned by the top-q eigenvectors? Here we consider the subspace instead of the top-q eigenvectors directly, because there might be degeneracy among $\lambda_1, \ldots, \lambda_q$, in which case the set of the top-q eigenvectors is not uniquely defined.

We first estimate the magnitude of λ_q (with additive error $\gamma/100$) using the following theorem.

THEOREM 4.6. *Let $\delta \in (0,1)$, $q < d$, $A \in \mathbb{C}^{d\times d}$ be a Hermitian matrix with operator norm at most 1, $v_1, \ldots, v_d$ be an orthonormal basis of eigenvectors of A, and corresponding eigenvalues $\lambda_1, \ldots, \lambda_d$ such that $|\lambda_1| \geqslant \cdots \geqslant |\lambda_d|$, where we know the gap $\gamma = |\lambda_q| - |\lambda_{q+1}|$. Suppose $U_A = \exp(\pi i A)$. There is a quantum algorithm that with probability at least $1-\delta$, estimates $|\lambda_q|$ with additive error $\gamma/100$, using $\mathcal{O}\Big(\frac{\sqrt{qd}}{\gamma}\log(\frac{\log(1/\gamma)}{\delta})\log(\frac{d}{\delta})\log(\frac{1}{\gamma})\Big)$ controlled applications of $U_A^{\pm}$ and $\tilde{\mathcal{O}}\Big(\frac{\sqrt{qd}}{\gamma}\log(\frac{\log(1/\gamma)}{\delta})\log(\frac{d}{\delta})\log(\frac{1}{\gamma})\Big)$ time.*

Proof. Let $\delta' = \delta/(10d)$, $A = \sum_{i\in[d]} \lambda_i |v_i\rangle\langle v_i|$, and $T = 2^{\lceil \log(200\log(1/\delta')/\gamma)\rceil+2}$. Unitary $W = \sum_{t=0}^{T-1} |t\rangle\langle t| \otimes \exp(\pi i t A)$ does Hamiltonian simulation according to A on the second register, for an amount of time specified in the first register. Observe that $\frac{1}{\sqrt{d}}\sum_{i\in[d]} |i\rangle|i\rangle = \frac{1}{\sqrt{d}}\sum_{i\in[d]} |v_i\rangle|v_i^*\rangle$ because of the invariance of maximally entangled states under unitaries of the form $U \otimes U^{\dagger}$. Hence we can apply phase estimation with precision $\gamma/200$ and failure probability δ' to the quantum state $\frac{1}{\sqrt{d}}\sum_{i\in[d]} |v_i\rangle|v_i^*\rangle|0\rangle$ using W, to obtain the state

$$\frac{1}{\sqrt{d}}\sum_{i\in[d]} |v_i\rangle|v_i^*\rangle|L_i\rangle, \tag{4.18}$$

where the state $|L_i\rangle$ contains a superposition over different estimates $\tilde{\lambda_i}$ of λ_i. For each $i \in [d]$, if we were to measure $|L_i\rangle$ in the computational basis, then with probability at least $1-\delta'$ we get an outcome $\tilde{\lambda_i}$ such that $|\lambda_i - \tilde{\lambda_i}| \leqslant \gamma/200$.[25] Let $\mu \in [0,1]$, and R_μ be a unitary that marks whether a number's absolute value is $< \mu$, i.e., for every $a \in [-1,1]$

$$R_\mu |a\rangle|0\rangle = \begin{cases} |a\rangle|0\rangle, & \text{if } |a| \geqslant \mu \\ |a\rangle|1\rangle, & \text{otherwise.} \end{cases}$$

This unitary can be implemented up to negligibly small error by $\tilde{\mathcal{O}}(1)$ elementary gates. Applying R_μ on the last register of the state of Eq. (4.18) and an additional $|0\rangle$, we obtain $\sqrt{p_\mu}|\phi_0\rangle|0\rangle + \sqrt{1-p_\mu}|\phi_1\rangle|1\rangle$ for some $|\phi_0\rangle$ and $|\phi_1\rangle$, where p_μ is the probability of outcome 0 if we were to measure the last qubit. Note that if $\mu \geqslant |\lambda_q| + \gamma/150 > |\lambda_q| + \gamma/200$, then $p_\mu \leqslant (q-1)/d + (d-q+1)\delta'/d$, where the first term on the right-hand side is the maximal contribution (to the probability p_μ of getting outcome 0 for the last qubit) coming from $|L_i\rangle$ with $i \leqslant q-1$ and the second term is the maximal contribution coming from $|L_i\rangle$ with $i > q-1$. On the other hand, if $\mu \leqslant |\lambda_q| - \gamma/150 < |\lambda_q| - \gamma/200$, then $p_\mu \geqslant (q/d)\cdot(1-\delta')$, which is the minimal contribution coming from $|L_i\rangle$

[25] There's a small technical issue here: the unitary $e^{\pi i A}$ (to which we apply phase estimation) has phases ranging between $-\pi$ and π because the λ_j range between -1 and 1, and phase estimation treats $-\pi$ and π the same. However, we can easily fix that by applying phase estimation to the unitary $e^{\pi i A/2}$, whose phases range between $-\pi/2$ and $\pi/2$.

with $i \leqslant q$. The difference between the square-roots of these two values is therefore

$$\sqrt{\frac{q}{d}\cdot(1-\delta')}-\sqrt{\frac{q-1}{d}+\frac{d-q+1}{d}\delta'} \geqslant \sqrt{\frac{q}{d}-\delta'}-\sqrt{\frac{q-1}{d}+\delta'} = \frac{\frac{1}{d}-2\delta'}{\sqrt{\frac{q}{d}-\delta'}+\sqrt{\frac{q-1}{d}+\delta'}}, \tag{4.19}$$

where the last equality is because $a-b=(a^2-b^2)/(a+b)$. Because $\delta'=\delta/(10d)\in(0,1/(10d))$, both terms in the denominator are $\leqslant\sqrt{q/d}$, and hence the right-hand side of Eq. (4.19) is at least $2/(5\sqrt{qd})$. To estimate $|\lambda_q|$ with additive error $\gamma/100$, it therefore suffices to do binary search over the values of μ (with precision $\gamma/300$, that is, binary search over $\mu\in\{0,\gamma/300,2\gamma/300,\dots,1\}$), in each iteration estimating $\sqrt{p_\mu}$ to within $\pm 1/(5\sqrt{qd})$. We can implement the unitary that maps $|0\rangle|0\rangle\to\sqrt{p_\mu}\,|\phi_0\rangle|0\rangle+\sqrt{1-p_\mu}\,|\phi_1\rangle|1\rangle$ using one application of W and $\tilde{\mathcal{O}}(1)$ time. Let $\delta''=\delta/(10\log(300/\gamma))>0$. By Theorem 2.3 with additive error $\eta=1/(5\sqrt{qd})=\Theta(1/\sqrt{qd})$ and with failure probability δ'', one iteration of the binary search succeeds with probability at least $1-\delta''$ and uses $\mathcal{O}(\log(1/\delta'')/\eta)=\mathcal{O}\big(\sqrt{qd}\log(\frac{\log(1/\gamma)}{\delta})\big)$ applications of W and $W^\dagger$, and $\tilde{\mathcal{O}}\big(\sqrt{qd}\log(\frac{\log(1/\gamma)}{\delta})\big)$ time. Therefore by the union bound, with probability at least $1-(\lceil\log(300/\gamma)\rceil+1)\cdot\delta''\geqslant 1-\delta$, all iterations of the binary search give a sufficiently good estimate of the value $\sqrt{p_\mu}$ of that iteration, so the binary search gives us an estimate of $|\lambda_q|$ to within $\pm(\gamma/150+\gamma/300)=\pm\gamma/100$.

Since we use $O(\log(1/\gamma))$ iterations of binary search, we use $\mathcal{O}\Big(\sqrt{qd}\log(\frac{\log(1/\gamma)}{\delta})\log(\frac{1}{\gamma})\Big)$ applications of W and $W^\dagger$ and $\tilde{\mathcal{O}}\Big(\sqrt{qd}\log(\frac{\log(1/\gamma)}{\delta})\log(\frac{1}{\gamma})\Big)$ time for the whole binary search. We can implement W and $W^\dagger$ using $\mathcal{O}(T)=\mathcal{O}(\frac{\log(d/\delta)}{\gamma})$ controlled applications of $U_A^\pm$ and $\tilde{\mathcal{O}}(\frac{\log(d/\delta)}{\gamma})$ time. Thus we obtain a good estimate of $|\lambda_q|$ with probability $\geqslant 1-\delta$, using $\mathcal{O}\Big(\frac{\sqrt{qd}}{\gamma}\log(\frac{\log(1/\gamma)}{\delta})\log(\frac{d}{\delta})\log(\frac{1}{\gamma})\Big)$ controlled applications of $U_A^\pm$ and $\tilde{\mathcal{O}}\Big(\frac{\sqrt{qd}}{\gamma}\log(\frac{\log(1/\gamma)}{\delta})\log(\frac{d}{\delta})\log(\frac{1}{\gamma})\Big)$ time. ∎

The above theorem assumes perfect access to $\exp(\pi iA)=U_A$ for doing phase estimation. If we only assume we have sparse-query-access to A, then by Theorem 2.6 we can implement a unitary $\widetilde{U}_A$ such that $\|\widetilde{U}_A-\exp(\pi iA)\|\leqslant\varepsilon$ using $\tilde{\mathcal{O}}(s^{0.5+o(1)}/\varepsilon^{o(1)})$ time and queries.

Since the procedure in Theorem 4.6 makes use of $D=\mathcal{O}\Big(\frac{\sqrt{qd}}{\gamma}\log(\frac{\log(1/\gamma)}{\delta})\log(\frac{d}{\delta})\log(\frac{1}{\gamma})\Big)$ controlled applications of $U_A^\pm$, if we replace U_A with $\widetilde{U}_A$, the algorithm still outputs the desired answer with success probability at least $1-\delta-D\varepsilon$.[26] By plugging in the time complexity for constructing $\widetilde{U}_A$ with $\varepsilon=\Theta\Big(\frac{\delta\gamma}{\sqrt{qd}\log(\log(1/\gamma)/\delta)\log(d/\delta)\log(1/\gamma)}\Big)$, with the constant in the $\Theta(\cdot)$ chosen such that $T\varepsilon\leqslant\delta$ (and rescaling δ by factor of 2), we immediately have the following corollary.

COROLLARY 4.2. *Let $q<d$ and $A\in\mathbb{C}^{d\times d}$ be a Hermitian matrix with operator norm at most 1, $v_1,\dots,v_d$ be an orthonormal basis of eigenvectors of A, and eigenvalues $\lambda_1,\dots,\lambda_d$ such that $|\lambda_1|\geqslant\cdots\geqslant|\lambda_d|$, where we know the gap $\gamma=|\lambda_q|-|\lambda_{q+1}|$, and $\delta\in(0,1)$. Suppose A has sparsity s and we have sparse-query-access to A. There is a quantum algorithm that with success probability at least $1-\delta$, estimates $|\lambda_q|$ with additive error $\gamma/100$, using $\tilde{\mathcal{O}}\Big(\frac{1}{\delta^{o(1)}}(\frac{\sqrt{dqs}}{\gamma})^{1+o(1)}\Big)$ queries and time.*

The following proposition shows that every bounded-error quantum algorithm needs $\Omega(\sqrt{ds})$ sparse-access queries to estimate the top eigenvalue of an s-sparse matrix with constant additive error. This implies the above corollary is near-optimal when $q=1$.

PROPOSITION 4.1. *Let $A\in\mathbb{C}^{d\times d}$ be a Hermitian matrix with operator norm at most 3. Suppose A has sparsity s and we have sparse-query-access to A. Every bounded-error quantum algorithm that estimates the top eigenvalue of A with additive error 0.1 uses $\Omega(\sqrt{ds})$ queries.*

[26] Here we use the fact that if two unitaries are ε-close in operator norm, and they are applied to the same quantum state, then the resulting two states are ε-close in Euclidean norm, and the two probability distributions obtained by measuring the resulting two states in the computational basis are ε-close in total variation distance.

Proof. For simplicity and without loss of generality we assume A has sparsity $2s+1$ and $d \geqslant 2s+1$ is a multiple of s. The idea is to encode an $s(d-s)$-bit string into a $2s+1$-sparse $d \times d$ matrix. Given a string $X \in \{0,1\}^{s(d-s)} \coloneqq X^{(1)}X^{(2)}\dots X^{(d/s)-1}$ with Hamming weight either 0 or 1, where $X^{(k)}$ is an s^2-bit Boolean string for each $k \in [d/s-1]$. For every $k \in [d/s-1]$, define $Y^{(k)} \in \{0,1\}^{s\times s}$ as $(Y^{(k)})_{ij} = X^{(k)}_{s\cdot i+j}$. Let A be defined by $d/s \cdot d/s = d^2/s^2$ many $s \times s$ square matrices such that for $i \geqslant j$

$$A_{ij} = \begin{cases} I_s & \text{if } i = j \\ Y^{(i)} + 2^{-d} \cdot J_s & \text{if } i = j+1 \\ 0_s & \text{otherwise,} \end{cases}$$

where I_s is the $s \times s$ identity matrix, J_s is the $s \times s$ all-1 matrix, and 0_s is the $s \times s$ all-0 matrix; and for $i < j$, $A_{ij} = A^T_{ji}$. One can easily see that A has sparsity $2s+1$, and because the Hamming weight $\mathrm{Ham}(X)$ of X is at most 1, the operator norm of A is at most $2 + 2s \cdot 2^{-d} \leqslant 2 + d \cdot 2^{-d} \leqslant 3$. Note that given access to the oracle O_X that maps $|i\rangle|0\rangle \to |i\rangle|X_i\rangle$ for every $i \in [s(d-s)]$, one can construct an oracle that allows us to make sparse-query-access to A using 2 applications of $O_X^{\pm}$.

Observe that if $\mathrm{Ham}(X)$=0, then the operator norm of A is at most $1+2s\cdot 2^{-d} \leqslant 1+d\cdot 2^{-d} \leqslant 1+1/(e\cdot\ln 2) < 1.6$, while if $\mathrm{Ham}(X)$=1, then the operator norm of A is at least 2. Therefore, if there exists a T-query quantum algorithm $\mathcal{A}$ that estimates the top eigenvector of A with additive error 0.1, then $\mathcal{A}$ can also be used to decide whether the $s(d-s)$-bit string X has Hamming weight 0 or 1 using $2T$ queries to O_X. By the well-known quantum query lower bound for search, every bounded-error quantum algorithm needs $\Omega(\sqrt{ds})$ queries for this, implying $T = \Omega(\sqrt{ds})$. ∎

Note that a $\sqrt{ds}$ lower bound for estimating the top eigenvalue implies a $\sqrt{ds}$ lower bound for approximating the quantum state $|v_1\rangle$ of the top eigenvector: once we have $|v_1\rangle$, we can apply the quantum phase estimation algorithm with precision ε to estimate λ_1 with additive error ε using $\tilde{\mathcal{O}}\big(1/\varepsilon\big)$ controlled applications of $\exp(\pi i A)$ and $\exp(-\pi i A)$. Combining this with Hamiltonian simulation (Theorem 2.6) and our lower for approximating λ_1 (Proposition 4.1), we conclude that every bounded-error quantum algorithm that outputs a state at ℓ_2-distance $\leqslant 0.05$ from $|v_1\rangle$, needs $\Omega(\sqrt{ds})$ queries.

Once we know $|\lambda_q|$, we can apply [26, Theorem 31] to implement a block-encoding of U_Π using $\tilde{\mathcal{O}}\big(1/\gamma\big)$ applications of a block-encoding of A, where $\Pi = \sum_{i\in[q]} v_i v_i^\dagger$. Combining the above argument with Theorem 2.8, we directly get the following corollary of Theorem 4.5:

Corollary 4.3. *Let $q < d$ and $A \in \mathbb{C}^{d\times d}$ be a Hermitian matrix with operator norm at most 1, $v_1, \dots, v_d$ be an orthonormal eigenbasis of A with respective eigenvalues $\lambda_1, \dots, \lambda_d$ such that $|\lambda_1| \geqslant \cdots \geqslant |\lambda_d|$, where we know the gap $\gamma = |\lambda_q| - |\lambda_{q+1}|$. Let $\varepsilon, \delta \in (0,1)$ and $\Pi = \sum_{i\in[q]} v_i v_i^\dagger$. Suppose A has sparsity s and we have sparse-query-access to A. There exists a quantum algorithm that outputs a $d \times q$ matrix W with orthonormal columns such that, with probability at least $1-\delta$, $\|WW^\dagger - \Pi\| \leqslant \varepsilon$, using $\tilde{\mathcal{O}}\Big(\Big(\frac{d\sqrt{sq}}{\gamma\varepsilon}\Big)^{1+o(1)} + \frac{1}{\delta^{o(1)}}\Big(\frac{\sqrt{dqs}}{\gamma}\Big)^{1+o(1)} + dq^2\Big)$ time and $\tilde{\mathcal{O}}(d)$ QRAM bits.*

For the case of dense matrix A, we can set $s = d$ to get time complexity roughly $qd^{1.5}$. The special case $q = 1$ gives our main result for approximating the top eigenvector (with additive ℓ_2-error ε)[27] in time $\tilde{\mathcal{O}}\Big(\big(d^{1.5}/(\gamma\varepsilon)\big)^{1+o(1)}\Big)$. The ε-dependency is slightly worse than the algorithm in Section 4.3, while both the d-dependency and γ-dependency are significantly better (for d, the power is $1.5 + o(1)$ instead of 1.75; for γ, the power is $1 + o(1)$ instead of 2). In Section 5 we show that its d-dependence to be essentially optimal. However, the complexity with respect to q is sub-optimal for q close to d, because one can diagonalize the entire matrix A classical in matrix-multiplication time $\mathcal{O}(d^\omega)$.

We also remark that one can (approximately) prepare the quantum state $|v_1\rangle$ of the top eigenvector of the matrix A (with success probability $\geqslant 0.9$) in time $\sim (\frac{\sqrt{ds}}{\gamma})^{1+o(1)}$ without using QRAM: first use Corollary 4.2 to

[27] Note that for every unit $w, v \in \mathbb{C}^d$ it holds that $\|ww^\dagger - vv^\dagger\| = 2\sqrt{1-|\langle w, v\rangle|^2}$, thus $\varepsilon \geqslant \|ww^\dagger - vv^\dagger\|$ implies $|\langle w, v\rangle| \geqslant \sqrt{1-\varepsilon^2/4} \geqslant 1-\varepsilon^2/4 \geqslant 1-\varepsilon^2/2$.

estimate λ_1 (with additive error $\gamma/100$ and with success probability $\geqslant 0.99$) and then implement an ϵ'-approximate block-encoding of projector $|v_1\rangle\langle v_1|$ in time $\tilde{\mathcal{O}}\Big(\frac{\sqrt{s}(s/\epsilon')^{o(1)}}{\gamma}\Big)$ by Theorem 2.8 and [26, Theorem 31]. After that, we generate a vector g with i.i.d. (complex) standard normal entries, prepare the corresponding normalized quantum state $|g\rangle$ (with probability $\geqslant 0.99 - \exp(-d/2)$ it has overlap $\geqslant 1/(100\sqrt{2d})$ with v_1)[28] and apply the block-encoding to project it down, obtaining a state that looks like $\alpha\,|0\rangle\,|v_1\rangle + \sqrt{1-|\alpha|^2}\,|1\rangle\,|v_1^{\perp}\rangle$ where $|\alpha| = \Omega(1/\sqrt{d})$. By Theorem 2.1, $\mathcal{O}\Big(\sqrt{d}\log(1/\varepsilon)\Big)$ amplitude amplification rounds then suffice to prepare $|v_1\rangle$ (up to ℓ_2-norm error $\varepsilon/2$), and the total cost is therefore $\tilde{\mathcal{O}}\Big(\frac{1}{\varepsilon^{o(1)}}(\frac{\sqrt{ds}}{\gamma})^{1+o(1)}\Big)$ by choosing $\varepsilon' = \Theta(\varepsilon/(\sqrt{d}\log(1/\varepsilon)))$ such that $T\varepsilon' \leqslant \varepsilon/2$, where T is number of amplitude amplification rounds. This algorithm is near-optimal in its d-dependence and s-dependence because of the argument after the proof of Proposition 4.1.

COROLLARY 4.4. *Let $A \in \mathbb{C}^{d\times d}$ be a Hermitian matrix with operator norm at most 1, $v_1, \ldots, v_d$ be an orthonormal eigenbasis of A with respective eigenvalues $\lambda_1, \ldots, \lambda_d$ such that $|\lambda_1| \geqslant \cdots \geqslant |\lambda_d|$, where we know the gap $\gamma = |\lambda_1| - |\lambda_2|$. Let $\varepsilon \in (0,1)$. Suppose A has sparsity s and we have sparse-query-access to A. There exists a quantum algorithm that with probability at least $0.98 - \exp(-d/2)$, outputs $|\tilde{v_1}\rangle$ that approximates $|v_1\rangle$ with ℓ_2-error ε using $\tilde{\mathcal{O}}\Big(\frac{1}{\varepsilon^{o(1)}}(\frac{\sqrt{ds}}{\gamma})^{1+o(1)}\Big)$ time.*

5 Lower bounds for approximating the top eigenvector

In this section we prove essentially tight classical and quantum query lower bounds for approximating the top eigenvector of a matrix whose entries we can query.

5.1 The hard instance for the lower bound Consider the following case, which is the "hard instance" for which we prove the lower bounds. Let $u \in \{-1,1\}^d$ be a vector, and define symmetric random matrix $A = \frac{1}{d}uu^T + N$ where the entries of N are i.i.d. $N_{ij} \sim N(0, \frac{1}{4\cdot 10^6 d})$ for all $1 \leqslant i \leqslant j \leqslant d$ (and $N_{ij} = N_{ji}$ if $i > j$); the goal is to recover most (say, 99%) of the entries of the vector u. In this problem, the information about the u_i-s is hidden in the matrix A: the entry A_{ij} is clearly a sample from $N(\frac{u_iu_j}{d}, \frac{1}{4\cdot 10^6 d})$. Hence to learn entries of u, intuitively we should be able to distinguish the distribution $N(\frac{1}{d}, \frac{1}{4\cdot 10^6 d})$ from the distribution $N(-\frac{1}{d}, \frac{1}{4\cdot 10^6 d})$. In the classical case (where querying an entry of A_{ij} is the same as obtaining one sample fom the distribution) it requires roughly $\Omega(d)$ queries to the entries of the ith row and column to learn one u_i, even if all u_j with $j \neq i$ are already known. In the quantum case, it requires $\Omega(\sqrt{d})$ queries. Intuitively, learning $0.99d$ of the u_i-s should then require roughly d times more queries, so $\Omega(d^2)$ and $\Omega(d^{1.5})$ classical and quantum queries in total, respectively. We show in the next two subsections that this is indeed the case.

First we show that (with high probability) A has a large eigenvalue gap. Note that $A = \frac{1}{d}uu^T + \frac{1}{2000\sqrt{d}}G$ where $G_{ij} \sim N(0,1)$ for $1 \leqslant i \leqslant j \leqslant d$ and $G_{ij} = G_{ji}$ for the lower-triangular elements. Since G itself is symmetric and its entries are i.i.d. $\sim N(0,1)$, by Theorem 2.14 with $t = 0.4\sqrt{d}$ (here $b_{max} = \sqrt{d}$ and $b^*_{max} = 1$) we have that the operator norm of G is upper bounded by $2.5\sqrt{d} + (0.4 + o(1))\sqrt{d} \leqslant 3\sqrt{d}$ with probability at least $1 - \exp(-0.04d)$; below we assume this is indeed the case. Therefore, by triangle inequality, the top eigenvalue of A is upper bounded by $1 + \frac{3}{2000}$ and lower bounded by $\|A\frac{u}{\sqrt{d}}\|_2 \geqslant 1 - \frac{3}{2000}$, implying that there is a unit top eigenvector $v_1 = v_1(A)$ of A that has inner product nearly 1 with the unit vector $\frac{u}{\sqrt{d}}$:

$$1 - \frac{3}{2000} \leqslant \|Av_1\|_2 \leqslant \|\frac{1}{d}uu^Tv_1\|_2 + \|\frac{1}{2000\sqrt{d}}Gv_1\|_2 \leqslant |\frac{1}{\sqrt{d}}u^Tv_1| + \frac{3}{2000},$$

hence $|\langle v_1, \frac{u}{\sqrt{d}}\rangle| \geqslant 1 - \frac{3}{1000}$. We may assume without loss of generality that the eigenvector v_1 has been chosen such that $\langle v_1, \frac{u}{\sqrt{d}}\rangle$ is positive, so we can ignore the absolute value sign. The signs of $v_1(A)$ have to agree with the signs of u in at least 99.4% of the d entries, because each entry where the signs are different contributes at least

[28] Since g has i.i.d. (complex) standard normal entries, $\langle v_1|g\rangle$ has (complex) standard normal distribution. Since by Proposition 2.1, with probability $\geqslant 1 - \exp(-d/2)$, $\|g\| \leqslant \sqrt{2d}$, and since the pdf of the standard normal distribution is $p(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$), we know with probability $\geqslant 1 - \exp(-d/2) - 0.01\cdot 2\cdot\frac{1}{\sqrt{2\pi}} \geqslant 0.99 - \exp(-d/2)$, $|\langle v_1|g\rangle|/\|g\| \geqslant 1/(100\sqrt{2d})$.

$1/d$ to the squared distance between $u/\sqrt{d}$ and $v_1(A)$:

$$\frac{1}{d}\#\{j \in [d] \mid u_j \cdot (v_1(A))_j \leqslant 0\} \leqslant \|\frac{u}{\sqrt{d}} - v_1(A)\|_2^2 = 2 - 2\langle \frac{u}{\sqrt{d}}, v_1(A)\rangle \leqslant \frac{3}{500}.$$

Moreover, the second eigenvalue $\lambda_2(A)$ of A is at most 0.08:

$$\lambda_2(A) = \max_{w:\|w\|_2=1, w\perp v_1} \|Aw\|_2 \leqslant \frac{3}{2000} + \max_{w:\|w\|_2=1, w\perp v_1} \|\frac{uu^T w}{d}\|_2 \leqslant \frac{3}{2000} + \frac{\sqrt{3\cdot 1997}}{997} < 0.08,$$

where the last inequality holds because $\langle v_1, \frac{u}{\sqrt{d}}\rangle \geqslant 1 - \frac{3}{1000}$.[29] Hence there is a constant gap between the top and the second eigenvalue of A.

If we have an algorithm that outputs a vector $\tilde{u}$ satisfying $\|\tilde{u} - v_1(A)\|_2 \leqslant \frac{1}{1000}$, then we can use the signs of $\tilde{u}$ to learn 99% of the u_i-s. Hence a (classical or quantum) query lower bound for recovering (most of) u is also a query lower bound for approximating the top eigenvector of our hard instance.

5.2 A classical lower bound We first show that every classical algorithm that recovers 99% of the u_i-s needs $\Omega(d^2)$ queries.

THEOREM 5.1. *Let $u \in \{-1,1\}^d$ be a vector, $e_{11}, e_{12}, \ldots, e_{1d}, e_{22}, \ldots, e_{dd}$ be $d(d+1)/2$ independent samples drawn from $N(0, \frac{1}{4\cdot 10^6 d})$, and $A \in \mathbb{R}^{d\times d}$ be the matrix defined by*

$$A_{ij} = \begin{cases} \frac{1}{d}u_i u_j + e_{ij}, & \text{if } 1 \leqslant i \leqslant j \leqslant d, \\ A_{ji}, & \text{otherwise.} \end{cases}$$

Suppose we have query access to entries of A. Every bounded-error classical algorithm that computes a $\tilde{u} \in \{-1,1\}^d$ at Hamming distance $\leqslant d/100$ from u, uses $\Omega(d^2)$ queries.

Proof. Suppose there exists a T-query bounded-error classical algorithm $\mathcal{A}$ to compute such a $\tilde{u}$, with worst-case error probability $\leqslant 1/20$. Note that the only entries that depend on u_i are in the ith column and row of A. Let random variable T_i be the number of queries that $\mathcal{A}$ makes in the ith column and row (here the randomness comes from the input distribution and from the internal randomness of $\mathcal{A}$). Because every query is counted at most twice among the T_i-s (and only once if the query is to a diagonal entry of A), we have $\mathbb{E}[\sum_{i\in[d]} T_i] \leqslant 2T$. Define index i as "good" if $\Pr[u_i = \tilde{u}_i] \geqslant 0.8$, and let I_G be the set of good indices. Since $\mathcal{A}$ has error probability at most 1/20, we can bound the expected Hamming distance by

$$\mathbb{E}[\mathrm{Ham}(u,\tilde{u})] \leqslant \Pr[\mathrm{Ham}(u,\tilde{u}) \leqslant \frac{d}{100}] \cdot \frac{d}{100} + \Pr[\mathrm{Ham}(u,\tilde{u}) > \frac{d}{100}] \cdot d \leqslant 1 \cdot \frac{d}{100} + \frac{1}{20} \cdot d \leqslant \frac{d}{10}.$$

By plugging in the definition of I_G, we obtain

$$\frac{d}{10} \geqslant \mathbb{E}[\mathrm{Ham}(u,\tilde{u})] = \sum_{i\in[d]} \Pr[u_i \neq \tilde{u}_i] \geqslant \sum_{i\in[d]\setminus I_G} \Pr[u_i \neq \tilde{u}_i] \geqslant \sum_{i\in[d]\setminus I_G} \frac{1}{5} = \frac{d - |I_G|}{5},$$

implying that $|I_G| \geqslant d/2$. Because $\mathbb{E}[\sum_{i\in I_G} T_i] \leqslant \mathbb{E}[\sum_{i\in[d]} T_i] \leqslant 2T$, by averaging there exists an index $\mathbf{i} \in I_G$ such that $\mathbb{E}[T_i] \leqslant 4T/d$. This implies that there is a classical algorithm $\mathcal{A}'$ that recovers $u_{\mathbf{i}}$ with probability at least 0.8 using an expected number of at most $4T/d$ queries to entries in the $\mathbf{i}$th row and column of A.

Now suppose we want to distinguish $u_{\mathbf{i}} = 1$ from $u_{\mathbf{i}} = -1$ using samples from $N(\frac{u_{\mathbf{i}}}{d}, \frac{1}{4\cdot 10^6 d})$. We can use $\mathcal{A}'$ for this task, as follows. Generate $u_1, \ldots, u_{\mathbf{i}-1}, u_{\mathbf{i}+1}, \ldots, u_d$ uniformly at random from ± 1, and generate e'_{ij} from

[29] Let $v_1 = \alpha_1 \frac{u}{\sqrt{d}} + \beta_1 w_1$ and $v_2 = \alpha_2 \frac{u}{\sqrt{d}} + \beta_2 w_2$ for some unit vectors $w_1, w_2 \perp \frac{u}{\sqrt{d}}$ and for some $\alpha_1, \alpha_2, \beta_1, \beta_2 \in [-1,1]$ satisfying $\alpha_1^2 + \beta_1^2 = \alpha_2^2 + \beta_2^2 = 1$. Since $v_1 \perp v_2$, we have $\langle v_1, v_2\rangle = \alpha_1\alpha_2 + \beta_1\beta_2\langle w_1, w_2\rangle = 0$, implying $|\alpha_2| = \frac{|\beta_1\beta_2|}{|\alpha_1|}|\langle w_1, w_2\rangle| \leqslant \frac{|\beta_1|}{|\alpha_1|} = \frac{\sqrt{1-\alpha_1^2}}{|\alpha_1|} \leqslant \frac{\sqrt{3\cdot 1997}}{997}$.

$N(0, \frac{1}{4\cdot 10^6 d})$ for all $1 \leqslant i \leqslant j \leqslant d$. Then we define a $d \times d$ matrix A' as

$$A'_{ij} = \begin{cases} \frac{1}{d}u_i u_j + e'_{ij}, & \text{if } 1 \leqslant i \leqslant j \leqslant d, \mathbf{i} \notin \{i, j\} \\ u_j \cdot \text{sample from } N(\frac{u_\mathbf{i}}{d}, \frac{1}{4\cdot 10^6 d}) & \text{if } i = \mathbf{i} \text{ and } i < j \\ u_i \cdot \text{sample from } N(\frac{u_\mathbf{i}}{d}, \frac{1}{4\cdot 10^6 d}) & \text{if } j = \mathbf{i} \text{ and } i < j \\ A'_{ji}, & \text{otherwise.} \end{cases}$$

Note that for every $i, j \in [d]$, A'_{ij} is a sample drawn from $N(\frac{u_i u_j}{d}, \frac{1}{4\cdot 10^6 d})$. Given that our algorithm knows the values it generated itself (in particular, all u_j with $j \neq \mathbf{i}$), it can implement one query to an entry in the $\mathbf{i}$th row or column of A' by at most one new sample from $N(\frac{u_\mathbf{i}}{d}, \frac{1}{4\cdot 10^6 d})$, and queries to other entries of A' do not cost additional samples.

By running $\mathcal{A}'$ on the input matrix A', we have $\Pr[u_\mathbf{i} = \tilde{u}_\mathbf{i}] \geqslant 0.8$. Hence by using an expected number of $4T/d$ samples drawn from $N(\frac{u_\mathbf{i}}{d}, \frac{1}{4\cdot 10^6 d})$, we can distinguish $u_\mathbf{i} = 1$ from $u_\mathbf{i} = -1$ with probability $\geqslant 0.8$. Then by Markov's inequality, worst-case $40T/d$ samples suffice to distinguish $u_\mathbf{i} = 1$ from $u_\mathbf{i} = -1$ with probability $\geqslant 0.7$.

The KL-divergence between $N(\frac{1}{d}, \frac{1}{4\cdot 10^6 d})$ and $N(-\frac{1}{d}, \frac{1}{4\cdot 10^6 d})$ is $\mathbb{E}_{x\sim N(\frac{1}{d}, \frac{1}{4\cdot 10^6 d})}[8 \cdot 10^6 x] = \frac{8\cdot 10^6}{d}$. As a result, by using the well-known Pinsker's inequality ($d_{TV}(P, Q) \leqslant \sqrt{\frac{1}{2}D_{KL}(P \,\|\, Q)}$ [41, 22, 37]) and the fact that $D_{KL}(P^{\otimes t} \,\|\, Q^{\otimes t}) = t \cdot D_{KL}(P \,\|\, Q)$ for any distributions P, Q and natural number t, we obtain

$$\begin{aligned} \Omega(1) &\leqslant d_{TV}\Big(N(\frac{1}{d}, \frac{1}{4\cdot 10^6 d})^{\otimes \frac{40T}{d}}, N(-\frac{1}{d}, \frac{1}{4\cdot 10^6 d})^{\otimes \frac{40T}{d}}\Big) \\ \text{(by Pinsker's inequality)} \qquad &\leqslant \sqrt{\frac{1}{2}D_{KL}\Big(N(\frac{1}{d}, \frac{1}{4\cdot 10^6 d})^{\otimes \frac{40T}{d}} \,\Big\|\, N(-\frac{1}{d}, \frac{1}{4\cdot 10^6 d})^{\otimes \frac{40T}{d}}\Big)} \\ &= \sqrt{\frac{1}{2}\cdot\frac{40T}{d}\cdot D_{KL}\Big(N(\frac{1}{d}, \frac{1}{4\cdot 10^6 d}) \,\Big\|\, N(-\frac{1}{d}, \frac{1}{4\cdot 10^6 d})\Big)} = \mathcal{O}\Big(\sqrt{T/d^2}\Big), \end{aligned}$$

implying $T = \Omega(d^2)$. ∎

By the discussion in Section 5.1, we therefore obtain the following corollary.

COROLLARY 5.1. *Let A be a $d \times d$ symmetric matrix with $\|A\| = \mathcal{O}(1)$ and an $\Omega(1)$ gap between its top and second eigenvalues. Suppose we have query access to the entries of A. Every classical algorithm that with probability at least $\geqslant 99/100$, approximates the top eigenvector of A with ℓ_2-error at most $\frac{1}{1000}$ uses $\Omega(d^2)$ queries.*

This query lower bound is tight up to the constant factor, since we can compute the top eigenvector exactly using d^2 queries: just query every entry of A and diagonalize the now fully known matrix A (without any further queries) to find the top eigenvector exactly.

5.3 A quantum lower bound Now we move to the quantum case, still using the same hard instance. Our proof uses similar ideas as the hybrid method [11] and adversary method [2, 3] for quantum query lower bounds, but adjusted to continuous random variables instead of input bits.

THEOREM 5.2. *Let $u \in \{-1, 1\}^d$ be a uniformly random vector, $e_{11}, e_{12}, \ldots, e_{1d}, e_{22}, \ldots, e_{dd}$ be $d(d+1)/2$ independent samples drawn from $N(0, \frac{1}{4\cdot 10^6 d})$, and $A \in \mathbb{R}^{d\times d}$ be the random matrix defined by*

$$A_{ij} = \begin{cases} \frac{1}{d}u_i u_j + e_{ij}, & \text{if } 1 \leqslant i \leqslant j \leqslant d, \\ A_{ji}, & \text{otherwise.} \end{cases}$$

Suppose we have quantum query access to entries of A. Every bounded-error quantum algorithm that computes $\tilde{u} \in \{-1, 1\}^d$ at Hamming distance $\leqslant d/100$ from u, uses $\Omega(d^{1.5}/\sqrt{\log d})$ queries.

Proof. Let ν denote the input distribution given in the theorem statement, and $X \sim \nu$ be a random $d \times d$ input matrix according to that distribution (with instantiations of random variable X denoted by lower-case x). Let O_x denote the query oracle to input matrix x. Suppose there exists a T-query quantum algorithm

$\mathcal{A} = U_T O_x U_{T-1} \cdots U_1 O_x U_0$, alternating queries and input-independent unitaries on some fixed initial state (say, all-0), to compute such a $\tilde{u}$ with error probability $\leqslant 1/20$, probability taken over both ν and the internal randomness of $\mathcal{A}$ caused by the measurement of its final state. Our goal is to lower bound T.

For $t \in \{0, \ldots, T-1\}$, let $|\psi_x^t\rangle = \sum_{i,j \in [d]} \alpha_{xij}^t |i,j\rangle |\phi_{xij}^t\rangle$ be the quantum state of algorithm $\mathcal{A}$ just before its $(t+1)$st query on input matrix x. Let $|\psi_x^T\rangle$ be the final state on input x. We define the query mass on $(i,j) \in [d] \times [d]$ on input x as $p_{ijx} = \sum_{t \in [T]} |\alpha_{xij}^t|^2$, and define the query mass on i on input x as $p_{ix} = \sum_{t \in [T], j \in [d]} |\alpha_{xij}^t|^2 + |\alpha_{xji}^t|^2$. Note that $\sum_{i,j} |\alpha_{xij}^t|^2 \leqslant 1$ for all x and t. Every $|\alpha_{xij}^t|^2$ is counted once in p_{ix} and once in p_{jx} if $i \neq j$, and counted only in p_{ix} if $i = j$. Hence we have $\sum_{i \in [d]} p_{ix} \leqslant 2T$ for every x. Define $T_i = \mathbb{E}_{x \sim \nu}[p_{ix}]$ as the expected query mass on the ith row and column of the input matrix, then $\sum_{i \in [d]} T_i \leqslant 2T$.

Call index $i \in [d]$ "good" if $\Pr[u_i = \tilde{u}_i] \geqslant 0.8$, where the probability is taken over ν and the algorithm's internal randomness. Let I_G be the set of good indices. Since $\mathcal{A}$ has error probability at most $1/20$, we have

$$\mathbb{E}_x[\mathrm{Ham}(u,\tilde{u})] \leqslant \Pr[\mathrm{Ham}(u,\tilde{u}) \leqslant \frac{d}{100}] \cdot \frac{d}{100} + \Pr[\mathrm{Ham}(u,\tilde{u}) > \frac{d}{100}] \cdot d \leqslant 1 \cdot \frac{d}{100} + \frac{1}{20} \cdot d \leqslant \frac{d}{10}.$$

Using linearity of expectation and the definition of I_G, we have

$$\frac{d}{10} \geqslant \mathbb{E}_x[\mathrm{Ham}(u,\tilde{u})] = \sum_{i \in [d]} \Pr[u_i \neq \tilde{u}_i] \geqslant \sum_{i \in [d] \setminus I_G} \Pr[u_i \neq \tilde{u}_i] \geqslant \frac{d - |I_G|}{5},$$

which implies $|I_G| \geqslant d/2$. Since $\sum_{i \in I_G} T_i \leqslant \sum_{i \in [d]} T_i \leqslant 2T$, by averaging there exists an index $\mathbf{i} \in I_G$ such that $T_{\mathbf{i}} \leqslant 4T/d$. We fix this $\mathbf{i}$ for the rest of the proof. Note that because $\mathbf{i}$ has $\Pr_\nu[u_{\mathbf{i}} = \tilde{u}_{\mathbf{i}}] \geqslant 0.8$, we also have $\Pr_{\nu_+}[u_{\mathbf{i}} = \tilde{u}_{\mathbf{i}}] \geqslant 0.6$ and $\Pr_{\nu_-}[u_{\mathbf{i}} = \tilde{u}_{\mathbf{i}}] \geqslant 0.6$, where the distributions ν_+ and ν_- are ν conditioned on $u_{\mathbf{i}} = 1$ and $u_{\mathbf{i}} = -1$, respectively.

We now define an (adversarial) joint distribution μ on (X,Y)-pairs of input matrices, such that the marginal distribution of X is ν_+ and the marginal distribution of Y is ν_-. First sample a matrix x (with associated $u \in \{-1,1\}^d$ with $u_{\mathbf{i}} = 1$) according to ν_+. We want to probabilistically modify this into a matrix y by changing only a small number of entries, and only in the $\mathbf{i}$th row and column of x. Let f and g be the pdf of $N(\frac{1}{d}, \frac{1}{4 \cdot 10^6 d})$ and $N(-\frac{1}{d}, \frac{1}{4 \cdot 10^6 d})$, respectively. Consider an entry $x_{\mathbf{i}j}$ in the $\mathbf{i}$th row of x, with $j \neq \mathbf{i}$. Conditioned on the particular u we sampled, its pdf was f if $u_j = 1$ and g if $u_j = -1$. If the pdf of $x_{\mathbf{i}j}$ was f, then obtain $y_{\mathbf{i}j}$ from $x_{\mathbf{i}j}$ as follows: if $x_{\mathbf{i}j} > 0$, then negate it with probability $\frac{f(x_{\mathbf{i}j}) - g(x_{\mathbf{i}j})}{f(x_{\mathbf{i}j})}$, else leave it unchanged.

Claim: If $x_{\mathbf{i}j} \sim N(\frac{1}{d}, \frac{1}{4 \cdot 10^6 d})$, then $y_{\mathbf{i}j} \sim N(-\frac{1}{d}, \frac{1}{4 \cdot 10^6 d})$.
Proof: Let h be the pdf of $y_{\mathbf{i}j}$. For a value $z > 0$, we have $h(z) = f(z) - f(z) \cdot \frac{f(z) - g(z)}{f(z)} = g(z)$.
For $z \leqslant 0$ we have $h(z) = f(z) + f(-z) \cdot \frac{f(-z) - g(-z)}{f(-z)} = f(z) + f(-z) - g(-z) = f(-z) = g(z)$. ■

If the pdf of $x_{\mathbf{i}j}$ was g instead of f, then we do something analogous: if $x_{\mathbf{i}j} < 0$, then negate it with probability $\frac{g(x_{\mathbf{i}j}) - f(x_{\mathbf{i}j})}{g(x_{\mathbf{i}j})}$. This gives the analogous claim: the pdf of $y_{\mathbf{i}j}$ is then f.

Let matrix y be obtained by applying this probabilistic process to all entries in the $\mathbf{i}$th row of x, and changing the entries in the $\mathbf{i}$th column to equal the new $\mathbf{i}$th row (since the resulting y needs to be a symmetric matrix). Outside of the $\mathbf{i}$th row and column, x and y are equal. Let μ be the resulting joint distribution on (x,y) pairs. An important property of this distribution that we use below, is that the $d \times d$ matrices x and y typically only differ in roughly $\sqrt{d}$ entries, because the probability with which $x_{\mathbf{i}j}$ is modified (=negated) is $\mathcal{O}(1/\sqrt{d})$. The marginal distribution of Y is ν_-, because the change we made in the X-distribution corresponds exactly to changing $u_{\mathbf{i}}$ from 1 to -1. We could equivalently have defined μ by first sampling $Y \sim \nu_-$, and then choosing x_{ij} by an analogous negating procedure on y_{ij}.

We now use the general template of the adversary method [2] together with our distribution μ to lower bound the total number T of queries that $\mathcal{A}$ makes. Define a progress measure $S_t = \mathbb{E}_{xy \sim \mu}[\langle \psi_x^t | \psi_y^t \rangle]$. As usual in the adversary method, this measure is large at the start of the algorithm and becomes small at the end: $S_0 = 1$ because $\langle \psi_x^0 | \psi_y^0 \rangle = 1$ for all x, y (since the initial state is fixed, independent of the input); and $S_T \leqslant 1 - \Omega(1)$ because for $(x,y) \sim \mu$, our algorithm outputs 1 with probability at least 0.6 on x and outputs -1 with probability

at least 0.6 on y, meaning that $\langle\psi_x^T|\psi_y^T\rangle$ is typically bounded below 1. Let $\Delta_t = |S_{t+1} - S_t|$ be the change in the progress measure due to the $(t+1)$st query. We can upper bound that change as follows:

$$\begin{aligned}
\Delta_t &= |\mathbb{E}_{xy\sim\mu}[\langle\psi_x^{t+1}|\psi_y^{t+1}\rangle - \langle\psi_x^t|\psi_y^t\rangle]| \\
&= |\mathbb{E}_{xy\sim\mu}[\langle\psi_x^t|\big(O_x^\dagger O_y - I\big)|\psi_y^t\rangle]| \\
&= |\mathbb{E}_{xy\sim\mu}\left[\sum_{i,j\in[d]} \alpha_{xij}^t \langle i,j|\langle\phi_{xij}^t| \sum_{i,j:x_{ij}\neq y_{ij}} \alpha_{yij}^t |i,j\rangle|\phi_{yij}'^t\rangle\right]| \\
&\leqslant \mathbb{E}_{xy\sim\mu}\left[\sum_{i,j:x_{ij}\neq y_{ij}} |\alpha_{xij}^t|\cdot|\alpha_{yij}^t|\right] \\
&= \mathbb{E}_{xy\sim\mu}\left[\sum_{j:x_{\mathbf{i}j}\neq y_{\mathbf{i}j}} |\alpha_{x\mathbf{i}j}^t|\cdot|\alpha_{y\mathbf{i}j}^t| + \sum_{j:x_{j\mathbf{i}}\neq y_{j\mathbf{i}}} |\alpha_{xj\mathbf{i}}^t|\cdot|\alpha_{yj\mathbf{i}}^t|\right] \\
&\leqslant \frac{1}{2}\mathbb{E}_{xy\sim\mu}\left[\sum_{j:x_{\mathbf{i}j}\neq y_{\mathbf{i}j}} \Big(|\alpha_{x\mathbf{i}j}^t|^2 + |\alpha_{y\mathbf{i}j}^t|^2\Big) + \sum_{j:x_{j\mathbf{i}}\neq y_{j\mathbf{i}}} \Big(|\alpha_{xj\mathbf{i}}^t|^2 + |\alpha_{yj\mathbf{i}}^t|^2\Big)\right],
\end{aligned}$$

where we use that $|\psi_x^{t+1}\rangle = U_{t+1}O_x|\psi_x^t\rangle$ and $|\psi_y^{t+1}\rangle = U_{t+1}O_y|\psi_y^t\rangle$, that $O_x^\dagger O_y : |i,j,b\rangle \to |i,j\rangle|b - x_{ij} + y_{ij}\rangle$, that x and y only differ in the $\mathbf{i}$th row and column, and the AM-GM inequality $(ab \leqslant (a^2+b^2)/2)$ in the last step.

Now observe that

$$\begin{aligned}
\mathbb{E}_{xy\sim\mu}\left[\sum_{j:x_{\mathbf{i}j}\neq y_{\mathbf{i}j}} |\alpha_{x\mathbf{i}j}^t|^2\right] &= \sum_{j\in[d]} \mathbb{E}_x[\Pr_{y\sim\mu|x}[x_{\mathbf{i}j} \neq y_{\mathbf{i}j}]\cdot|\alpha_{x\mathbf{i}j}^t|^2] \\
&= \sum_j \int_0^\infty \frac{f(x_{\mathbf{i}j}) - g(x_{\mathbf{i}j})}{f(x_{\mathbf{i}j})}\cdot|\alpha_{x\mathbf{i}j}^t|^2\cdot f(x_{\mathbf{i}j})\mathrm{d}x_{\mathbf{i}j} \\
&= \sum_j \Big(\int_0^{\frac{10\sqrt{\log d}}{\sqrt{d}}} \big(1 - \frac{g(x_{\mathbf{i}j})}{f(x_{\mathbf{i}j})}\big)\cdot|\alpha_{x\mathbf{i}j}^t|^2\cdot f(x_{\mathbf{i}j})\mathrm{d}x_{\mathbf{i}j} \\
&\quad + \int_{\frac{10\sqrt{\log d}}{\sqrt{d}}}^\infty \big(f(x_{\mathbf{i}j}) - g(x_{\mathbf{i}j})\big)\cdot|\alpha_{x\mathbf{i}j}^t|^2\mathrm{d}x_{\mathbf{i}j}\Big) \\
&\leqslant \sum_j \Big(\max_{z\in[0,\frac{10\sqrt{\log d}}{\sqrt{d}}]} |1-\exp(-8\cdot10^6 z)|\cdot\mathbb{E}_x[|\alpha_{x\mathbf{i}j}^t|^2] \\
&\quad + \int_{\frac{10\sqrt{\log d}}{\sqrt{d}}}^\infty \big(f(x_{\mathbf{i}j}) - g(x_{\mathbf{i}j})\big)\mathrm{d}x_{\mathbf{i}j}\Big) \\
&\leqslant \sum_j \Big(\frac{8\cdot10^7\sqrt{\log d}}{\sqrt{d}}\cdot\mathbb{E}_{x\sim\nu_+}[|\alpha_{x\mathbf{i}j}^t|^2] + 2d^{-100}\Big),
\end{aligned}$$

where the first part of the first inequality holds because $\frac{g(x_{\mathbf{i}j})}{f(x_{\mathbf{i}j})} = \exp(-8\cdot10^6 x_{\mathbf{i}j})$, the first part of the second inequality holds because for every z, $1-\exp(-z) \leqslant z$, and the second part of the second inequality holds because both f and g are Gaussians with variance $\frac{1}{4\cdot10^6 d}$ and $\frac{10\sqrt{\log d}}{\sqrt{d}} - \frac{1}{d} \geqslant 10\cdot\frac{1}{2000\sqrt{d}}$.

We can similarly upper bound $\sum_{t\in[T]} \mathbb{E}_{xy\sim\mu}[\sum_{j:x_{j\mathbf{i}}\neq y_{j\mathbf{i}}} |\alpha_{xj\mathbf{i}}^t|^2]$, $\sum_{t\in[T]} \mathbb{E}_{xy\sim\mu}[\sum_{j:x_{\mathbf{i}j}\neq y_{\mathbf{i}j}} |\alpha_{y\mathbf{i}j}^t|^2]$,

and $\sum_{t\in[T]} \mathbb{E}_{xy\sim\mu}[\sum_{j:x_{j\mathbf{i}}\neq y_{j\mathbf{i}}} |\alpha^t_{yj\mathbf{i}}|^2]$. Now we have

$$\begin{aligned}
\Omega(1) &\leqslant |S_0 - S_T| \\
&\leqslant \sum_{t\in[T]-1} \Delta_t \\
&\leqslant \frac{1}{2} \sum_{t\in[T]-1} \mathbb{E}_{xy\sim\mu} \left[\sum_{j:x_{\mathbf{i}j}\neq y_{\mathbf{i}j}} \left(|\alpha^t_{x\mathbf{i}j}|^2 + |\alpha^t_{y\mathbf{i}j}|^2 \right) + \sum_{j:x_{j\mathbf{i}}\neq y_{j\mathbf{i}}} \left(|\alpha^t_{x\mathbf{i}j}|^2 + |\alpha^t_{y\mathbf{i}j}|^2 \right) \right] \\
&\leqslant \sum_{t\in[T]-1, j\in[d]} \left(\frac{4\cdot 10^7 \sqrt{\log d}}{\sqrt{d}} \left(\mathbb{E}_{x\sim\nu_+}[|\alpha^t_{x\mathbf{i}j}|^2 + |\alpha^t_{x\mathbf{i}j}|^2] + \mathbb{E}_{y\sim\nu_-}[|\alpha^t_{y\mathbf{i}j}|^2 + |\alpha^t_{yj\mathbf{i}}|^2] \right) + 4d^{-100} \right) \\
&\leqslant \frac{4\cdot 10^7 \sqrt{\log d}}{\sqrt{d}} \left(\mathbb{E}_{x\sim\nu_+}[p_{\mathbf{i}x}] + \mathbb{E}_{y\sim\nu_-}[p_{\mathbf{i}y}] \right) + 4d^{-97} \\
&\leqslant \frac{8\cdot 10^7 \sqrt{\log d}}{\sqrt{d}} \mathbb{E}_{x\sim\nu}[p_{\mathbf{i}x}] + 4d^{-97} \\
&\leqslant \frac{8\cdot 10^7 \sqrt{\log d}}{\sqrt{d}} \cdot \frac{4T}{d} + 4d^{-97},
\end{aligned}$$

where the sixth inequality uses that $\nu_+ + \nu_- = 2\nu$, and that j ranges over d values and t ranges over T values (and $T \leqslant d^2$ without loss of generality). This implies $T = \Omega(d^{1.5}/\sqrt{\log d})$. ∎

Again invoking the discussion in Section 5.1, we obtain the following corollary which shows that our second algorithm is close to optimal.

COROLLARY 5.2. *Let A be a $d \times d$ symmetric matrix with $\|A\| = \mathcal{O}(1)$ and an $\Omega(1)$ gap between its top and second eigenvalues. Suppose we have quantum query access to the entries of A. Every quantum algorithm that with probability at least $\geqslant 99/100$, approximates the top eigenvector of A with ℓ_2-error at most $\frac{1}{1000}$ uses $\Omega(d^{1.5}/\sqrt{\log d})$ queries.*

References

[1] Josh Alman, Ran Duan, Virginia Vassilevska Williams, Yinzhan Xu, Zixuan Xu, and Renfei Zhou. More asymmetry yields faster matrix multiplication, 2024. arXiv: 2404.16349 1

[2] Andris Ambainis. Quantum lower bounds by quantum arguments. *Journal of Computer and System Sciences*, 64(4):750–767, 2002. 5, 38, 39

[3] Andris Ambainis. Polynomial degree vs. quantum query complexity. *Journal of Computer and System Sciences*, 72(2):220–238, 2006. 5, 38

[4] Omer Angel and Yinon Spinka. Pairwise optimal coupling of multiple random variables, 2019. arXiv: 1903.00632 11

[5] Joran van Apeldoorn, Arjan Cornelissen, András Gilyén, and Giacomo Nannicini. Quantum tomography using state-preparation unitaries. In *Proceedings of the 36th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1265–1318, 2023. arXiv: 2207.08800 3, 9, 17, 29

[6] Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Quantum SDP-solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020. Earlier version in FOCS'17. arXiv: 1705.01843 3, 5

[7] Joran van Apeldoorn, Sander Gribling, and Harold Nieuwboer. Basic quantum subroutines: finding multiple marked elements and summing numbers. To appear in Quantum. arXiv: 2302.10244, 2023. 7

[8] Simon Apers and Sander Gribling. Quantum speedups for linear programming via interior point methods. arXiv: 2311.03215, 2023. 5

[9] Simon Apers and Ronald de Wolf. Quantum speedup for graph sparsification, cut approximation, and Laplacian solving. *SIAM Journal on Computing*, 51(6):1703–1742, 2022. 5

[10] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016. 11

[11] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh V. Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997. 38

[12] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, 1997. 9

[13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. 9, 10

[14] Michel Boyer, Gilles Brassard, Peter Høyer, and Alain Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46(4–5):493–505, 1998. arXiv: `quant-ph/9605034` 7

[15] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain 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` 7

[16] Harry Buhrman and Robert Špalek. Quantum verification of matrix products. In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 880–889, 2006. arXiv: `quant-ph/0409035` 6

[17] Yanlin Chen and Ronald de Wolf. Quantum algorithms and lower bounds for linear regression with norm constraints. In *Proceedings of 50th International Colloquium on Automata, Languages, and Programming*, volume 261 of *LIPIcs*, pages 38:1–38:21, 2023. 8, 25

[18] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493 – 507, 1952. 10

[19] Andrew M. Childs. On the relationship between continuous- and discrete-time quantum walk. *Communications in Mathematical Physics*, 294(2):581–603, 2009. 25

[20] Richard Cleve, Artur Ekert, Chiara Macchiavello, and Michele Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society A*, 454(1969):339–354, 1998. arXiv: `quant-ph/9708016` 25

[21] Arjan Cornelissen, Yassine Hamoudi, and Sofiene Jerbi. Near-optimal quantum algorithms for multivariate mean estimation. In *Proceedings of the 54st Annual ACM SIGACT Symposium on Theory of Computing*, pages 33–43. ACM, 2022. arXiv:2111.09787. 6

[22] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967. 38

[23] Kenneth R. Davidson and Stanisław J. Szarek. Local operator theory, random matrices and Banach spaces. In William B. Johnson and Joram Lindenstrauss, editors, *Handbook of the Geometry of Banach Spaces, Vol. I*, volume I, pages 317–366, North-Holland, Amsterdam, 2001. Elsevier Science. 10

[24] Rusins Freivalds. Probabilistic machines can use less running time. In *Proceedings of 7th IFIP Congress*, pages 839–842, 1977. 6

[25] András Gilyén. Iterative refinement for improved tomography and quantum linear equation solving. Poster presented at QIP'24, manuscript forthcoming, 2023. 23

[26] András Gilyén, Yuan Su, Guang Hao Low, and Nathan 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. Full version in arXiv: `1806.01838` 2, 3, 7, 9, 23, 29, 30, 35, 36

[27] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, volume 3 of *Johns Hopkins Studies in Mathematical Sciences*. Johns Hopkins University Press, fourth edition, 2013. 2

[28] Yehoram Gordon. Some inequalities for Gaussian processes and applications. *Israel Journal of Mathematics*, 50(4):265–289, 1985. 10

[29] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th ACM Symposium on the Theory of Computing (STOC)*, pages 212–219, 1996. arXiv: `quant-ph/9605043` 7

[30] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. arXiv: `1508.01797` 33

[31] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. arXiv: `2302.14066`, 2023. 29

[32] Moritz Hardt and Eric Price. The noisy power method: A meta algorithm with applications. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 27, 2014. 3, 24

[33] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009. arXiv: `0811.3171` 5

[34] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 10

[35] Ian Jolliffe. *Principal Components in Regression Analysis*. Springer New York, second edition, 2002. 1

[36] Ravindran Kannan and Santosh Vempala. Spectral algorithms. *Foundations and Trends in Theoretical Computer Science*, 4:57–288, 2009. 1

[37] Johannes H.B. Kemperman. On the optimum rate of transmitting information. In M. Behara, K. Krickeberg, and J. Wolfowitz, editors, *Probability and Information Theory*, volume 89 of *Lecture Notes in Mathematics*, pages 126–169. Springer Berlin Heidelberg, 1969. 38

[38] Iordanis Kerenidis and Anupam Prakash. Quantum recommendation systems. In *Proceedings of the 8th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 49:1–49:21, 2017. arXiv: 1603.08675 8

[39] Iordanis Kerenidis and Anupam Prakash. A quantum interior point method for LPs and SDPs. *ACM Transactions on Quantum Computing*, 1(1), 2020. arXiv: 1808.09266 3, 17, 29

[40] Alexei Yu. Kitaev. Quantum measurements and the Abelian stabilizer problem. arXiv: quant-ph/9511026, 1995. 25

[41] Solomon Kullback. A lower bound for discrimination information in terms of variation. *IEEE Transactions on Information Theory*, 13(1):126–127, 1967. Correction, volume 16(5), p. 652, 1970. 38

[42] Michel Ledoux. *The Concentration of Measure Phenomenon*. Mathematical surveys and monographs. American Mathematical Society, 2001. 11

[43] Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, pages 66–70, 2011. 16

[44] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10:631–633, 2014. arXiv: 1307.0401 5

[45] Guang Hao 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 6, 9

[46] Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by qubitization. *Quantum*, 3:163, 2019. arXiv: 1610.06546 8

[47] Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In *Proceedings of the Advances in Cryptology 31st Annual International Conference on the Theory and Applications of Cryptographic Techniques*, volume 7237, pages 700–718, 2012. 13

[48] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on Gaussian measures. *SIAM Journal on Computing*, 37(1):267–302, 2007. 13

[49] Nhat A. Nghiem and Tzu-Chieh Wei. Quantum algorithm for estimating largest eigenvalues. *Physics Letters A*, 488(12913857), 2023. 5

[50] Masashi Okamoto. Some inequalities relating to the partial sum of binomial probabilities. *Annals of the Institute of Statistical Mathematics*, 10(1):29–35, 1959. 10

[51] Beresford N. Parlett. *The Symmetric Eigenvalue Problem*. Number 20 in Classics in Applied Mathematics. SIAM, 1987. 1

[52] Anupam Prakash. *Quantum algorithms for linear algebra and machine learning*. PhD thesis, University of California, Berkeley, 2014. 8

[53] Alex Ravsky. Reply on StackExchange. StackExchange, 2021. 16

[54] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):1–40, 2009. 16

[55] Herbert Robbins. A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, 1955. 17

[56] Walter Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1976. 16

[57] Kazuhiro Seki and Seiji Yunoki. Quantum power method by a superposition of time-evolved states. *PRX Quantum*, 2:010333, 2021. 5

[58] Aaron Sidford and Chenyi Zhang. Quantum speedups for stochastic optimization. In *Proceedings of NeurIPS*, 2023. arXiv:2308.01582. 6

[59] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Yonina C. Eldar and Gitta Kutyniok, editors, *Compressed Sensing: Theory and Applications*, pages 210–268. Cambridge University Press, 2012. arXiv: 1011.3027 10, 11

[60] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972. 9

[61] Virginia Vassilevska Williams, Yinzhan Xu, Zixuan Xu, and Renfei Zhou. New bounds for matrix multiplication: from alpha to omega. In *Proceedings of the 2024 Symposium on Discrete Algorithms*, pages 3792–3835, 2024. 1

[62] Theodore J. Yoder, Guang Hao Low, and Isaac L. Chuang. Fixed-point quantum search with an optimal number of queries. *Physical Review Letters*, 113(21):210501, 2014. arXiv: 1409.3305 7