Near-Optimal Quantum Algorithms for Multivariate Mean Estimation

Arjan Cornelissen* Yassine Hamoudi† Sofiene Jerbi‡

Last update: October 17, 2024

Abstract

We propose the first near-optimal quantum algorithm for estimating in Euclidean norm the mean of a vector-valued random variable with finite mean and covariance. Our result aims at extending the theory of multivariate sub-Gaussian estimators [LM19a] to the quantum setting. Unlike classically, where any univariate estimator can be turned into a multivariate estimator with at most a logarithmic overhead in the dimension, no similar result can be proved in the quantum setting. Indeed, Heinrich [Hei04] ruled out the existence of a quantum advantage for the mean estimation problem when the sample complexity is smaller than the dimension. Our main result is to show that, outside this low-precision regime, there does exist a quantum estimator that outperforms any classical estimator. More precisely, we prove that the approximation error can be decreased by a factor of about the square root of the ratio between the dimension and the sample complexity. Our approach is substantially more involved than in the univariate setting, where most quantum estimators rely only on phase estimation. We exploit a variety of additional algorithmic techniques such as linear amplitude amplification, the Bernstein-Vazirani algorithm, and quantum singular value transformation. Our analysis is also deeply rooted in proving concentration inequalities for multivariate truncated statistics.

We develop our quantum estimators in two different input models that showed up in the literature before. The first one provides coherent access to the binary representation of the random variable and it encompasses the classical setting. In the second model, the random variable is directly encoded into the phases of quantum registers. This model arises naturally in many quantum algorithms but it is often incomparable to having classical samples. We adapt our techniques to these two settings and we show that the second model is strictly weaker than the other one for solving the mean estimation problem. Finally, we describe several applications of our algorithms, notably in measuring the expectation values of commuting observables and in the field of machine learning.

1 Introduction

Monte Carlo methods are used extensively in various fields of science and engineering, such as statistical physics [BH10], finance [Gla03], or machine learning [AFDJ03]. At the core of these methods is a Monte Carlo process, e.g., a randomized algorithm, whose *expected outcome* is to be estimated via repeated random executions. Quantum computers can speed-up this approach at two different levels [Mon15]. First, novel algorithmic techniques such as Hamiltonian simulation [Fey82] or quantum walks [Sze04] provide faster Monte Carlo simulation processes. Secondly, quantum metrology algorithms (such as phase estimation [Kit95]) give better error rates for computing statistics on these processes. The present paper focuses on this second point through the lens of the *mean estimation problem*. In this problem, the objective is to compute

*QuSoft, University of Amsterdam. arjan.cornelissen@cwi.nl

†Simons Institute for the Theory of Computing, University of California, Berkeley. hamoudi@berkeley.edu

‡Institute for Theoretical Physics, University of Innsbruck. sofiene.jerbi@uibk.ac.at

the closest possible *estimate* $\widetilde{\mu}$ to the mean $\mu = \mathbb{E}[X]$ of a random variable X representing the output of some black-box process. Given the ability to repeat this process n times (the *sample complexity*), one seeks to minimize the error $\|\widetilde{\mu} - \mu\|$ made with high probability.

In the classical setting, a beautiful theory [LM19a] has been developed to solve the mean estimation problem in Euclidean norm. Under the sole assumption that the covariance matrix Σ of X exists, it turns out that the optimal *non-asymptotic* error behaves as if X followed the Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. This motivated the use of the adjective *sub-Gaussian* to qualify the optimal classical estimators. In one dimension, the most well-known sub-Gaussian estimator is arguably the median-of-means [NY83; JVV86; AMS99]. The first computationally efficient sub-Gaussian estimator in high dimension was only found recently by Hopkins [Hop20]. These estimators achieve an optimal error of $\|\widetilde{\mu} - \mu\|_2 \leq O\big(\sqrt{\mathrm{Tr}(\Sigma)/n} + \sqrt{\|\Sigma\| \log(1/\delta)/n}\big)$ with probability $1 - \delta$.

In the quantum setting, the univariate case $X \in \mathbb{R}$ has been studied since the early works on quantum counting [BBHT98]. The celebrated amplitude estimation algorithm [BHMT02] provides a smaller error rate for estimating the mean of any *Bernoulli* random variable compared to the classical estimators. For general univariate distributions, a series of quantum estimators [Gro98; Ter99; AW99; Hei02; WCNA09; BDGT11; Mon15; HM19; Ham21] culminated into a near-optimal algorithm that outperforms any classical estimator. On the other hand, the multivariate case $X \in \mathbb{R}^d$, appearing notably in machine learning applications, remains largely unaddressed by quantum algorithms. Classically, it admits a simple near-optimal approach: the d coordinates of μ can all be estimated simultaneously with d univariate sub-Gaussian estimators run in parallel (i.e., using the same samples from X) with only a logarithmic overhead $\log(d)$ in sample complexity (due to the Hoeffding bound). In the quantum scenario however, this *simultaneous* evaluation of several univariate expectation values is more complicated. Indeed, the quantum algorithms for the univariate case rely on quantum amplitude estimation [BHMT02], which involves as a critical step an encoding of the expectation value in the relative phase of a quantum register. At first sight, it is unclear how a vector of d phases could be encoded simultaneously into d registers without requiring a linear overhead in d. In fact, a lower bound proved by Heinrich [Hei04] rules out the possibility of simply a $\log(d)$ overhead for the quantum multivariate mean estimation problem.

Our paper develops *near-optimal* and *computationally efficient* quantum mean estimators for vector-valued random variables of arbitrary dimension with *binary oracle* access. Unlike in the univariate setting ($d = 1$), where the optimal quantum estimator [Ham21] is strictly more efficient than any classical estimator, we identify two different regimes in higher dimension: (i) if a quantum estimator is limited to accessing the input at most d times (i.e. $n \leq d$) then no advantage can be gained over the classical sub-Gaussian estimators, (ii) if it can access the input at least d times (i.e. $n \geq d$) then the approximation error can be reduced by a near-optimal factor of $\sqrt{d/n}$ compared to classical sub-Gaussian estimators.

We complement this work with new quantum estimators in the weaker *phase oracle* access model, where the information about X are directly encoded into the phases of quantum registers. This model has been considered before [GAW19], albeit not in the context of quantum mean estimation. We adapt some of our techniques to this model and show that here we can even obtain near-optimal estimators with respect to any ℓ_p-norm, with $p \in [1, \infty]$, thereby providing a complete characterization of the query complexities involved in the mean estimation problem. This part of our work shares some overlap with a related paper by a subset of the authors [CJ21] that focused on the probability and phases oracles models for multivariate Monte Carlo estimation.

1.1 Contributions

Our main contribution is the design of new quantum mean estimators that achieve the best possible error rates, up to logarithmic factors, in the multivariate setting. We investigate this problem in two different quantum input models. We first consider the *binary oracle* model in

Section 3, which generalizes in a natural way the classical *sample complexity* and is the most frequent setting used in previous work (e.g. [AW99; Hei02; BHH11; BDGT11; Mon15; Ham21]). In this model, the access to a d-dimensional random variable $X : \Omega \to \mathbb{R}^d$ over a probability space $(\Omega, 2^\Omega, \mathbb{P})$ is provided through two unitary operators: one that prepares a superposition over the probability space $U_\mathbb{P} : |0\rangle \mapsto \sum_{\omega\in\Omega} \sqrt{\mathbb{P}(\omega)}|\omega\rangle$, and one that evaluates the random variable over the sample set $\mathcal{B}_X : |\omega\rangle|\vec{0}\rangle \mapsto |\omega\rangle|X(\omega)\rangle$. Note that the mean to be estimated is $\mu = \sum_{\omega\in\Omega} \mathbb{P}(\omega)X(\omega)$. Our first main contribution is to provide an optimal multivariate quantum mean estimator in this setting. Our approach is substantially more involved than in the univariate case [Ham21]. A core primitive developed in our work is a new estimator for random variables bounded in ℓ_2-norm. This can be seen as a multivariate version of the well-known Amplitude Estimation algorithm [BHMT02]. Our techniques are based on the Bernstein-Vazirani algorithm [BV97] (more precisely, its generalization to estimating a linear function over the reals [Jor05]), the quantum singular value transformation framework [GSLW19], and tail inequalities for truncated statistics. We state our first result below with respect to the ℓ_∞-norm to highlight that it will be more natural to use this metric in our algorithms and convert to ℓ_2-norm by standard norm inequalities.

Theorem 3.3 (Informal)**.** *There is a quantum estimator that estimates the mean μ of any d-dimensional random variable X with error $\|\widetilde{\mu} - \mu\|_\infty \leq \frac{\sqrt{L_2}\log(d)}{n}$ and success probability $2/3$, given an upper bound $L_2 \geq \mathbb{E}[\|X\|_2]$ and $\widetilde{O}(n)$ queries to the oracles $U_\mathbb{P}$ and $\mathcal{B}_X$. The error made by this estimator in ℓ_2-norm is $\|\widetilde{\mu} - \mu\|_2 \leq \frac{\sqrt{dL_2}\log(d)}{n}$.*

As an illustration of this result, one can simultaneously estimate the expectation values of d univariate random variables $X_1, \ldots, X_d$ distributed in $[0, 1]$ each with error $\sqrt{d}\log(d)/n$ by doing $\widetilde{O}(n)$ queries. In comparison, running the Amplitude Estimation algorithm on each random variable separately (with $\widetilde{O}(n/d)$ queries) would result in an error of d/n.

Similarly to the Amplitude Estimation algorithm, the above primitive estimator does not always provide an optimal error rate with respect to the trace $\mathrm{Tr}(\Sigma) = \mathbb{E}[\|X\|_2^2] - \|\mathbb{E}[X]\|_2^2$ of the covariance matrix Σ of X. Moreover, it requires the ℓ_2-norm of X to be bounded by 1. We improve upon this result to design the next optimal quantum mean estimator.

Theorem 3.5 (Informal)**.** *There is a quantum estimator in the binary oracle model that estimates the mean μ of any d-dimensional random variable X with error*

$$\|\widetilde{\mu} - \mu\|_2 \leq \begin{cases} \sqrt{\frac{\mathrm{Tr}(\Sigma)}{n}}, & \text{if } n \leq d, \\ \frac{\sqrt{d\,\mathrm{Tr}(\Sigma)}\log(d)}{n}, & \text{if } n > d, \end{cases}$$

and success probability $2/3$, given $\widetilde{O}(n)$ queries to the oracles $U_\mathbb{P}$ and $\mathcal{B}_X$.

This bound is achieved by using any classical sub-Gaussian estimators [LM19a] when $n \leq d$, and a new quantum estimator when $n \geq d$. We show that these two regimes are inevitable since no quantum speed-up is possible when $n \leq d$, whereas our quantum estimator is optimal when $n \geq d$. Our lower bounds are based on the quantum query complexity of approximating a bit string whose entries are determined by parity functions.

Theorems 3.7 and 3.8 (Informal)**.** *For any estimator that uses at most n binary oracle queries, there is a d-dimensional random variable X such that, with probability $2/3$, the error is at least $\|\widetilde{\mu} - \mu\|_2 \geq \Omega\left(\sqrt{\frac{\mathrm{Tr}(\Sigma)}{n}}\right)$ if $n \leq d$, and $\|\widetilde{\mu} - \mu\|_2 \geq \Omega\left(\frac{\sqrt{d\,\mathrm{Tr}(\Sigma)}}{n}\right)$ if $n \geq d$.*

Next, in Section 4, we investigate the mean estimation problem in the *phase oracle* model where the aforementioned unitary $\mathcal{B}_X$ is replaced with phase access $\mathcal{P}_X : |\omega\rangle|j\rangle \mapsto e^{iX(\omega)_j}|\omega\rangle|j\rangle$ to the coordinates of X. This model can be efficiently simulated using a binary oracle but the

converse is generally not true. In fact, even obtaining one classical sample from X using a phase oracle is generally a hard task. On the other hand, as explained in [GAW19], this model arises naturally in the context of variational quantum eigensolvers, QAOA, and quantum auto-encoders for instance. This reason motivates understanding what is the optimal error rate for mean estimation in this weaker setting. Although a phase oracle does not allow to obtain an error depending on the covariance matrix, we manage to adapt some of the techniques developed for binary oracles to arrive at an optimal estimator when X is bounded in ℓ_∞-norm by $\|X\|_\infty \leq 1/4$. Interestingly, our results differ qualitatively from those in the binary oracle setting in two aspects. First, our estimator does not make the same number of queries to the oracles $U_{\mathbb{P}}$ and $\mathcal{P}_X$, and the optimal precision depends in fact differently on these two parameters. Second, in this model, we are actually able to tightly characterize the optimal performance with respect to all ℓ_p-norms, where $p \in [1, \infty]$, up to polylogarithmic factors. We state our results here only with respect to the ℓ_2-norm for ease of exposition and comparison to the binary oracle setting, and we show that these results are nearly optimal in Theorem 4.10.

Theorem 4.5 (Informal)**.** *There is a quantum estimator that estimates the mean μ of any d-dimensional random variable X such that $\|X\|_\infty \leq 1/4$ with error*

$$\|\widetilde{\mu} - \mu\|_2 \leq \begin{cases} \max\left\{\sqrt{\frac{d}{n}}, \frac{d^{3/2}}{n'}\right\} \cdot \log(d), & \text{if } n \leq d, \\ \max\left\{\frac{d}{n}, \frac{d^{3/2}}{n'}\right\} \cdot \log(d), & \text{if } n > d, \end{cases}$$

and success probability $2/3$, given $\widetilde{O}(n)$ queries to the oracle $U_{\mathbb{P}}$ and $\widetilde{O}(n')$ queries to the oracle $\mathcal{P}_X$.

Finally, we conclude this paper by giving some applications of the above results in Section 5. We first explain how our formulation of the multivariate mean estimation problem covers the general task of estimating the expectation values of several mutually commuting observables with respect to a given quantum state (Section 5.1). We then present several applications in the literature, and notably in quantum machine learning (training variational quantum circuits, Boltzmann machines, or reinforcement learning policies), where this problem arises (Section 5.2).

1.2 Proof overview

We give a high-level description of the algorithms developed in Section 3 for addressing the mean estimation problem in the binary oracle model. Similar techniques are employed in Section 4 for the phase oracle model. We simplify the exposition by replacing $\mathrm{Tr}(\Sigma) = \mathbb{E}[\|X - \mu\|_2^2]$ with the second moment $\mathbb{E}[\|X\|_2^2]$, and by taking the failure probability δ to be a small constant. The approximation error $\|\widetilde{\mu} - \mu\|_\infty$ is measured here with respect to the ℓ_∞-norm.

Bounded multivariate estimator. The main obstacle when trying to generalize most quantum univariate estimators (e.g. [Ter99; Hei02; WCNA09; Mon15; HM19; Ham21]) to the multivariate setting is the absence of an estimator for *bounded* multivariate random variables. In the univariate setting, such an estimator is provided by the well-known Amplitude Estimation algorithm [BHMT02] which, by a well-known trick [Ter99; WCNA09; Mon15], can estimate the mean of any random variable *bounded in* $[-1, 1]$ with an error on the order of $\sqrt{\mathbb{E}[|X|]}/n$. It is worth recalling how this estimator works when X is bounded in $[0, 1]$: the value $\varphi = \arcsin(\sqrt{\mathbb{E}[X]})$ is encoded as the phase of a particular unitary operator and estimated with error $1/n$ using phase estimation [Kit95]. Then, by standard trigonometric identities, $|\widetilde{\varphi} - \varphi| \leq 1/n$ implies that $|\sin^2(\widetilde{\varphi}) - \mathbb{E}[X]| \leq 2\sqrt{\mathbb{E}[X]}/n + 1/n^2$ (the lower-order term $1/n^2$ can be removed by testing if $\mathbb{E}[X] \leq 1/n^2$ and outputting 0 if this is the case). We generalize this idea to higher dimensions in a novel way by considering the *directional mean* function $u \mapsto \langle u, \mathbb{E}[X]\rangle$ where $u \in \mathbb{R}^d$. By using a constant number of queries to X and amplitude-to-phase conversion techniques [GAW19],

one can efficiently approximate the unitary $|u\rangle \mapsto e^{i\langle u,\mathbb{E}[X]\rangle}|u\rangle$ if $|\langle u, X\rangle| \leq 1$ almost surely. We could then estimate the directional mean $\langle u, \mathbb{E}[X]\rangle$ with phase estimation, for sufficiently many values of u, in order to reconstruct an estimate of $\mathbb{E}[X]$. However, this approach would incur a linear cost in the dimension d. Instead, since the directional mean is a *linear function* in u, we can use a variant of the Bernstein-Vazirani algorithm [BV97] to directly recover the entire vector $\mathbb{E}[X]$ (up to a certain precision) with fewer queries. This idea is also at the heart of the quantum gradient estimation algorithms [Jor05; GAW19], however it requires two major improvements for our setting. First, we can only make the assumptions that X is bounded in ℓ_2-norm (i.e. $\|X\|_2 \leq 1$) and u in ℓ_∞-norm (i.e. $\|u\|_\infty \leq 1$). However, these two conditions do not imply that $|\langle u, X\rangle| \leq 1$ as needed by the amplitude-to-phase conversion technique. We overcome this issue by proving tail inequalities for inner products and directional means (Lemma 3.1) showing that they do not exceed 1 with high probability under our assumptions. Hence, by suitable truncations, this gives us a first version of a bounded estimator with error $1/n$. Secondly, we need to incorporate information about X in the error. We cannot reproduce the univariate approach by encoding $\arcsin(\sqrt{|\langle u, \mathbb{E}[X]\rangle|})$ instead of $\langle u, \mathbb{E}[X]\rangle$ into the phase, since it would no longer be a linear function. Instead, we use the quantum singular value transformation framework [GSLW19] to linearly amplify the (squared) amplitude encoding the directional mean $\langle u, \mathbb{E}[X]\rangle$ into $\langle u, \frac{\mathbb{E}[X]}{\mathbb{E}[\|X\|_2]}\rangle$, before applying the amplitude-to-phase conversion technique. Since this amplification step requires $O(1/\sqrt{\mathbb{E}[\|X\|_2]})$ queries, this leaves us with $\widetilde{O}(n\sqrt{\mathbb{E}[\|X\|_2]})$ iterations available for the vector recovering step. Hence, the rescaled mean $\frac{\mathbb{E}[X]}{\mathbb{E}[\|X\|_2]}$ is estimated with error $1/(n\sqrt{\mathbb{E}[\|X\|_2]})$, which translates into the improved error of $\sqrt{\mathbb{E}[\|X\|_2]}/n$ for $\mathbb{E}[X]$ (Theorem 3.3).

Near-Optimal multivariate estimator. We build on the above bounded estimator to remove the assumption on the boundedness of X and decrease the error to $\sqrt{\mathbb{E}[\|X\|_2^2]}/n$. Similarly to the univariate case [Hei02; Mon15; HM19; Ham21], we decompose $X = X_0 + X_1 + X_2 + \dots$ into a sequence of *truncated* random variables $X_j = X\mathbb{1}_{a_{j-1}<\|X\|_2\leq a_j}$ over slices of the ℓ_2-ball, where the values outside the range $(a_{j-1}, a_j]$ are mapped to 0. The truncation levels $0 = a_{-1} < a_0 < a_1 < a_2 < \dots$ are chosen so that the bounded estimator performs well on each X_j individually. In the univariate setting, this sequence followed a geometric progression of ratio 2. Here, we instead choose a_j to be the *quantile value* of order 2^{-j} satisfying $\Pr[\|X\|_2 \geq a_j] \approx 2^{-j}$. This new choice has the advantage that the expected norm of X_j can be *explicitly* bounded as $\mathbb{E}[\|X_j\|_2] \leq 2^{-j-1}$ (Equation (3)), a property needed by our bounded estimator. Moreover, we show that this sequence increases slowly enough so that $a_j \leq 2^{j/2}\sqrt{\mathbb{E}[\|X\|_2^2]}$ (Equation (2)). Consequently, the bounded estimator can estimate *separately* the mean of each X_j with an error of $a_j\sqrt{\mathbb{E}[\|X_j\|_2]}/n \leq 2^{-1/2}\sqrt{\mathbb{E}[\|X\|_2^2]}/n$ (where the a_j factor comes from normalizing X_j to make it fit into the unit ℓ_2-ball). Finally, each quantile a_j can be computed (approximately) in time $O(2^{-j/2})$ using the quantile estimator developed in [Ham21] (Proposition 2.9), and we only need to consider $j \leq O(\log n)$ truncated random variables since the part of X above that threshold does not contribute to a significant portion of the mean (Equation (4)). This leads to the final error of $\sqrt{\mathbb{E}[\|X\|_2^2]}/n$ (Theorem 3.4).

1.3 Related work

There is an extensive literature on classical mean estimators and we refer the reader to [LM19a] for an excellent survey on the optimal *sub-Gaussian estimators* in Euclidean norm. We point out that the *empirical mean* estimator is generally not optimal, and its error is captured by several standard concentration bounds such as the Chebyshev, Chernoff and Bernstein inequalities.

There is a series of quantum *univariate* mean estimators [Gro98; AW99; BDGT11] that get close to the error $1/n$ for random variables distributed in $[0, 1]$ (and success probability $2/3$). The amplitude estimation algorithm [BHMT02; Ter99] leads to a sharper bound of $\sqrt{\mu}/n$.

Nevertheless, the mean μ is always larger than or equal to the variance σ^2 when X is distributed in $[0,1]$. The question of improving the dependence on σ^2 was considered in [Hei02; Mon15; HM19; Ham21], where it is shown that the optimal error is σ/n.

There are very few works addressing the quantum *multivariate* mean estimation problem. Heinrich [Hei04] proved that the error rate must depend on $1/n$ when the dimension is sufficiently large. Our lower bound in the $n \leq d$ regime (Theorem 3.7) refines this statement by adding a dependence on the covariance matrix. In a recent work, van Apeldoorn [Ape21] proposed a "multidimensional Amplitude Estimation" algorithm. However, it only applies to a restricted set of random variables and the error does not recover that of Amplitude Estimation when $d = 1$. More precisely, the author described a quantum algorithm for estimating with error $1/n$ (in ℓ_∞-norm) a probability vector $p = (p_1, \ldots, p_d)$ given access to a unitary $U : |0\rangle \mapsto \sum_i \sqrt{p_i}|i\rangle$. This is a special case of the multivariate mean estimation problem, where the random variable $X \in \{0,1\}^d$ is equal to the basis vector e_i with probability p_i. Applying our main result (Theorem 3.4) to X decreases the error given in [Ape21] by a factor of $\sqrt{\mathrm{Tr}(\Sigma)} = (1 - \sum_{i\in[d]} p_i^2)^{1/2}$.

Our work shares some similarities with the quantum gradient estimation algorithm of Jordan [Jor05; GAW19], which also uses an extension of the Bernstein-Vazirani algorithm to linear functions over the reals. However, unlike gradient estimation, the mean estimation problem requires combining this technique with further algorithmic steps.

2 Preliminaries

2.1 Notations

Throughout the paper we use the $\widetilde{O}(x)$ and $\widetilde{\Omega}(x)$ notations to hide factors that are polylogarithmic in the argument x. We let $|\vec{0}\rangle$ denote a multiple qubits state $|0\ldots0\rangle$. We use the notation $\mathcal{H}_{\mathrm{aux}}$ when referring to an auxiliary Hilbert space of sufficiently large dimension. We consider the family of ℓ_p-norms, defined as follows.

Definition 2.1 (ℓ_p-NORM)**.** Given $p \in [1, +\infty)$, the *ℓ_p-norm* $\|x\|_p$ of a d-dimensional vector x is defined as $\|x\|_p = (\sum_{i\in[d]} |x_i|^p)^{1/p}$ if $p < \infty$, and $\|x\|_\infty = \max_{i\in[d]} |x_i|$. We also let $\|x\| = \|x\|_2$ denote the ℓ_2-norm, and for a matrix M we set $\|M\|$ to be the induced ℓ_2-norm (or *spectral norm*).

Given $x \in \mathbb{R}^d$ and $0 \leq a < b$, we define the following truncation with respect to the ℓ_2-norm.

$$[\![x]\!]_a^b = \begin{cases} x & \text{if } a < \|x\|_2 \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

We recall the definition of a multivariate random variable. We only consider finite probability spaces for finite encoding reasons. Throughout the paper $d \in \mathbb{N}$ will denote the dimension of the random variable whose mean is to be estimated.

Definition 2.2 (RANDOM VARIABLE)**.** A (finite) *random variable* is a function $X : \Omega \to E$ for some probability space $(\Omega, 2^\Omega, \mathbb{P})$, where Ω is a finite sample set, $\mathbb{P} : \Omega \to [0,1]$ is a probability mass function and $E \subset \mathbb{R}^d$ is the finite support of X. The *covariance matrix* $\Sigma \in \mathbb{R}^{d\times d}$ of X is defined as $\Sigma = \mathbb{E}[XX^\top] - \mathbb{E}[X]\mathbb{E}[X]^\top$.

We say that X is *univariate* if the dimension is $d = 1$, and multivariate otherwise. For any norm $\|.\|$ over $\mathbb{R}^d$, we let $\|X\|$ denote the univariate random variable $\omega \mapsto \|X(\omega)\|$. Finally, we recall the definition of a quantile value (using the complementary cumulative distribution function).

Definition 2.3 (QUANTILE)**.** Given a discrete real-valued random variable X and a real $p \in [0,1]$, the *quantile* of order p is the number $Q(p) = \sup\{x \in \mathbb{R} : \Pr[X \geq x] \geq p\}$.

2.2 Input models

The input to the multivariate mean estimation problem is represented by a random variable X taking values in $\mathbb{R}^d$. In this section, we describe two possible access models for quantum estimators. Before that, we first recall the classical model, which we refer to as a *random experiment*.

Definition 2.4 (Random experiment)**.** Given a random variable X on a probability space $(\Omega, 2^\Omega, \mathbb{P})$, we define a *random experiment* as the process of drawing a sample $\omega \in \Omega$ according to $\mathbb{P}$ and observing the value of $X(\omega) \in \mathbb{R}^d$.

In the quantum setting, we make a distinction between accessing the probability mass function $\mathbb{P}$ and evaluating the function $X : \Omega \to E$. The first operation is provided by means of a *quantum experiment*, defined in the following way.

Definition 2.5 (Quantum experiment)**.** Consider a random variable X on a probability space $(\Omega, 2^\Omega, \mathbb{P})$. Let $\mathcal{H}_\Omega$ be a Hilbert space with basis states $\{|\omega\rangle\}_{\omega\in\Omega}$ and fix a unitary $U_\mathbb{P}$ acting on $\mathcal{H}_\Omega$ such that

$$U_\mathbb{P} : |0\rangle \mapsto \sum_{\omega\in\Omega} \sqrt{\mathbb{P}(\omega)}|\omega\rangle$$

assuming $0 \in \Omega$. We define a *quantum experiment* as the process of applying the unitary $U_\mathbb{P}$ or its inverse $U_\mathbb{P}^{-1}$ on any state in $\mathcal{H}_\Omega$.

We note that $\mathbb{P}$ is sometimes assumed to be the uniform distribution over some large set $\Omega = [N]$ (e.g. [Gro98; NW99; Hei02; BHH11; CFMW10; BDGT11; LW19]). In this case, the access to the unitary $U_\mathbb{P}$ need not be provided as part of the input.

We now describe two different quantum oracles for evaluating X. The first oracle provides a direct access to the value of $X(\omega)$. This model is the most commonly used in previous work on quantum mean estimation (e.g. [Gro98; Ter99; NW99; Hei02; BDGT11; Mon15; HM19]).

Definition 2.6 (Binary oracle)**.** Consider a finite random variable $X : \Omega \to E$ on a probability space $(\Omega, 2^\Omega, \mathbb{P})$. Let $\mathcal{H}_\Omega$ and $\mathcal{H}_E$ be two Hilbert spaces with basis states $\{|\omega\rangle\}_{\omega\in\Omega}$ and $\{|x\rangle\}_{x\in E}$ respectively. We say that a unitary $\mathcal{B}_X$ acting on $\mathcal{H}_\Omega \otimes \mathcal{H}_E$ is a *binary oracle* for X if

$$\mathcal{B}_X : |\omega\rangle|\vec{0}\rangle \mapsto |\omega\rangle|X(\omega)\rangle$$

for all $\omega \in \Omega$, assuming $\vec{0} \in E$.

Observe that one random experiment can be simulated by preparing the state $\sum_{\omega\in\Omega} \sqrt{\mathbb{P}(\omega)}|\omega\rangle$ $|X(\omega)\rangle$ and measuring its last register in the $\{|x\rangle\}_{x\in E}$ basis. This requires using one quantum experiment and one call to the binary oracle.

Our second type of oracle provides individual access to the coordinates of $X(\omega)$, encoded into the phases of a query operator. This model appears naturally in the context of variational eigensolvers, QAOA, and training variational auto-encoders [GAW19], albeit not in relation to the quantum mean estimation problem. This input model can be efficiently simulated using a binary oracle, but the converse is generally not true. In fact, even obtaining one classical sample from X may not be easy to do using a phase oracle.

Definition 2.7 (Phase oracle)**.** Consider a finite random variable $X : \Omega \to E$ on a probability space $(\Omega, 2^\Omega, \mathbb{P})$. Let $\mathcal{H}_\Omega$ be a Hilbert space with basis states $\{|\omega\rangle\}_{\omega\in\Omega}$. We say that a unitary $\mathcal{P}_X$ acting on $\mathcal{H}_\Omega \otimes \mathbb{C}^d$ is a *phase oracle* for X if

$$\mathcal{P}_X : |\omega\rangle|j\rangle \mapsto e^{iX(\omega)_j}|\omega\rangle|j\rangle$$

for all $\omega \in \Omega$ and $j \in [d]$.

2.3 Algorithmic tools

We first recall the optimal classical error bound for estimating the mean of a multivariate random variable with respect to the Euclidean norm.

Proposition 2.8 (Classical sub-Gaussian estimators, [LM19a]). *Let X be a d-dimensional random variable with mean μ and covariance matrix Σ. Given $\delta \in (0,1)$ and $n \geq \log(1/\delta)$, the* sub-Gaussian estimators *outputs a mean estimate $\widetilde{\mu}$ such that*

$$\|\widetilde{\mu} - \mu\|_2 \leq \sqrt{\frac{\mathrm{Tr}(\Sigma)}{n}} + \sqrt{\frac{\|\Sigma\| \log(1/\delta)}{n}}$$

with probability at least $1-\delta$, by using $O(n)$ random experiments.

We now present four quantum subroutines used in our work. We first need an algorithm introduced in [DH96; NW99] and generalized in [Ham21] for estimating the quantiles (Definition 2.3) of a univariate random variable quadratically faster than it is possible classically.

Proposition 2.9 (Quantile estimator, [Ham21]). *Let X be a univariate random variable. Given two reals $p, \delta \in (0,1)$, the* quantile estimation *algorithm* $\mathsf{Quantile}(X, p, \delta)$ *returns an approximate quantile $\widetilde{Q}$ that satisfies*

$$Q(p) \leq \widetilde{Q} \leq Q(cp)$$

with probability at least $1-\delta$ for some universal constant $c \in (0,1)$. The algorithm uses $O\left(\frac{\log(1/\delta)}{\sqrt{p}}\right)$ quantum experiments and binary oracle queries to X.

Next, we will use a variant of amplitude amplification [BHMT02] that provides a precise linear amplification of the amplitude.

Proposition 2.10 (Linear amplitude amplification, Theorem 6.10 in [Low17] or Lemma 11 in [GL20]). *Let V be a unitary operator and let Π be a projection operator acting on the same Hilbert space. Given two reals $t \geq 1$ and $\epsilon \in (0,1)$ there is a unitary operator $V_{t,\epsilon}$ that can be implemented with $O(t \log(1/\epsilon))$ applications of V, $V^\dagger$ and $I - 2\Pi$, and such that*

$$\big|\|\Pi V_{t,\epsilon}|\vec{0}\rangle\| - t\|\Pi V|\vec{0}\rangle\|\big| \leq \epsilon \quad \text{if} \quad t\|\Pi V|\vec{0}\rangle\| \leq 1/2.$$

Finally, the next two results provide efficient algorithms for converting between phase and amplitude encodings.

Lemma 2.11 (Amplitude-to-Phase conversion, Corollary 4.1 in [GAW19]). *Let V be a unitary operator acting on some Hilbert space $\mathcal{H}_U \otimes \mathcal{H}$ such that*

$$V : |u\rangle|\vec{0}\rangle \mapsto |u\rangle(\sqrt{1-p_u}|\psi_u^0\rangle|0\rangle + \sqrt{p_u}|\psi_u^1\rangle|1\rangle)$$

where $\{|u\rangle\}_{u\in U}$ is the standard basis of $\mathcal{H}_U$, $p_u \in (0,1)$ and $|\psi_u^0\rangle$, $|\psi_u^1\rangle$ are some arbitrary unit states. Then, given two reals $t \geq 0$ and $\epsilon \in (0,1)$, there is a unitary operator $\mathcal{P}_{t,\epsilon}$ acting on $\mathcal{H}_U \otimes \mathcal{H} \otimes \mathcal{H}_{\mathrm{aux}}$ that can be implemented with $O(t + \log(1/\epsilon))$ applications of V and $V^\dagger$, such that

$$\mathcal{P}_{t,\epsilon} : |u\rangle|\vec{0}\rangle \mapsto |u\rangle|\varphi_u\rangle \quad \text{where} \quad \||\varphi_u\rangle - e^{itp_u}|\vec{0}\rangle\| \leq \epsilon,$$

for all $u \in U$.

Lemma 2.12 (Phase-to-Amplitude conversion, Lemma 16 in [GAW17]). *Let $\mathcal{P}$ be a unitary operator acting on some Hilbert space $\mathcal{H}_U$ such that*

$$\mathcal{P} : |u\rangle \mapsto e^{ip_u}|u\rangle$$

where $\{|u\rangle\}_{u\in U}$ is the standard basis of $\mathcal{H}_U$ and $p_u \in [\delta, 1-\delta]$ for some $\delta \in (0, 1/2)$. Then, given a real $\epsilon \in (0,1)$, there is a unitary operator $V_{\epsilon,\delta}$ acting on $\mathcal{H}_U \otimes \mathcal{H}_{\mathrm{aux}} \otimes \mathbb{C}^2$ that can be implemented with $O(\log(1/\epsilon)/\delta)$ applications of $\mathcal{P}$ and $\mathcal{P}^\dagger$, such that

$$V_{\epsilon,\delta} : |u\rangle|\vec{0}\rangle|0\rangle \mapsto |u\rangle\big(\sqrt{p'_u}|\vec{0}\rangle|0\rangle + \sqrt{1-p'_u}|\psi\rangle|1\rangle\big) \quad \text{where} \quad \big|\sqrt{p'_u} - \sqrt{p_u}\big| \leq \epsilon,$$

for all $u \in U$ and some state $|\psi\rangle$.

3 Mean estimation with binary oracles

3.1 Bounded multivariate estimator

In this section, we generalize the univariate bounded estimator [Ter99; WCNA09; Mon15] derived from Amplitude Estimation [BHMT02] to the multivariate setting $X \in \mathbb{R}^d$. Our main ingredient is the construction of an approximate phase oracle for the directional mean $\langle u, \mathbb{E}[X]\rangle$, where the vectors $u \in \mathbb{R}^d$ are selected from the grid of points,

$$G = \left\{ \frac{j}{m} - \frac{1}{2} + \frac{1}{2m} : j \in \{0, \dots, m-1\} \right\}^d \subset (-1/2, 1/2)^d$$

with m being defined in step 2 of Algorithm 1.

We let $u \sim G$ denote a vector obtained according to the uniform distribution over G. We also define $\mathcal{H}_G$ to be the Hilbert space whose standard basis is indexed by the elements of G. Our algorithm requires encoding the inner product $\langle u, X\rangle$ into an amplitude. However, this quantity can be as large as $\sqrt{d}$ assuming $\|X\|_2 \leq 1$. The next crucial result shows that it is in fact much smaller than $\sqrt{d}$ for most values of u.

Lemma 3.1. *Let $\alpha > 0$. For any vector $x \in \mathbb{R}^d$ and any random variable X over $\mathbb{R}^d$ we have,*

$$\Pr_{u\sim G}\big[\alpha|\langle u, x\rangle| \geq \|x\|_2\big] \leq 2e^{-2/\alpha^2} \quad \text{and} \quad \Pr_{u\sim G}\big[\alpha\mathbb{E}[|\langle u, X\rangle|] \geq \mathbb{E}[\|X\|_2]\big] \leq \alpha/2.$$

Proof. We use that the coordinates of a uniformly random vector $u \in G$ are independent centered random variables bounded in $(-1/2, 1/2)$. The first result is obtained using Hoeffding's inequality, $\Pr_u[\alpha|\langle u, x\rangle| \geq \|x\|_2] \leq 2\exp\left(\frac{-2\|x\|_2^2}{\sum_{j=1}^d |\alpha x_j|^2}\right) = 2e^{-2/\alpha^2}$, since $\mathbb{E}_u[\langle u, x\rangle] = 0$ for all $x \in \mathbb{R}^d$. For the second result, we have by Markov's inequality that $\Pr_{u\sim G}\big[\alpha\mathbb{E}[|\langle u, X\rangle|] \geq \mathbb{E}[\|X\|_2]\big] \leq \frac{\mathbb{E}_u[\alpha\mathbb{E}[|\langle u,X\rangle|]]}{\mathbb{E}[\|X\|_2]} = \frac{\alpha\mathbb{E}[\mathbb{E}_u[|\langle u,X\rangle|]]}{\mathbb{E}[\|X\|_2]}$. By Cauchy-Schwarz inequality, $\mathbb{E}_u[|\langle u, x\rangle|] \leq \sqrt{\mathbb{E}_u[\langle u, x\rangle^2]} = \sqrt{\sum_{j=1}^d \mathbb{E}_{u_j}[(u_j x_j)^2]} \leq \|x\|_2/2$ for all $x \in \mathbb{R}^d$. Thus, $\Pr_{u\sim G}[\alpha\mathbb{E}[|\langle u, X\rangle|] \geq \mathbb{E}[\|X\|_2]] \leq \alpha/2$. □

Using the above lemma, one can encode (for most values of u) the truncated directional mean $\mathbb{E}[[\![\alpha\langle u, X\rangle]\!]_0^1]$ into an amplitude and apply oracle conversion techniques to approximate the phase oracle $|u\rangle \mapsto e^{i\mathbb{E}[[\![\alpha\langle u,X\rangle]\!]_0^1]}|u\rangle$ with accuracy ϵ at cost $O(\log(1/\epsilon))$. The cost of applying m times this oracle is then $O(m\log(1/\epsilon))$. We describe a more subtle algorithm where the latter complexity becomes $\widetilde{O}(m\sqrt{L_2}\log^2(1/\epsilon))$ given an upper-bound $L_2 \geq \mathbb{E}[\|X\|_2]$. The dependence on $\mathbb{E}[\|X\|_2]$ generalizes the dependence on $\mathbb{E}[|X|]$ provided by the univariate bounded estimator [Ter99; BHMT02; WCNA09; Mon15].

Proposition 3.2 (Directional mean oracle)**.** *Let X be a d-dimensional bounded random variable such that $\|X\|_2 \leq 1$. Given four reals $L_2 \in (0, 1]$, $m \geq 1/L_2$, $\alpha, \epsilon \in (0, 1)$ such that $\mathbb{E}[\|X\|_2] \leq L_2$, there exists a unitary operator $\widetilde{\mathcal{P}}_{X,L_2,m,\alpha,\epsilon} : |u\rangle|\vec{0}\rangle \mapsto |u\rangle|\varphi_u\rangle$ acting on $\mathcal{H}_G \otimes \mathcal{H}_{\text{aux}}$ that can be implemented using $\widetilde{O}\big(m\sqrt{L_2}\log^2(1/\epsilon)\big)$ quantum experiments and binary oracle queries to X, and such that*

$$\Big\| |\varphi_u\rangle - e^{im\mathbb{E}\left[[\![\alpha\langle u,X\rangle]\!]_0^1\right]}|\vec{0}\rangle \Big\| \leq \epsilon$$

for a fraction at least $1 - \alpha/2$ of all $u \in G$.

Proof. Fix u and consider the random variable X_+ defined over the same probability space as X such that $X_+(\omega) = X(\omega)$ when $\alpha\langle u, X(\omega)\rangle > 0$ and $X_+(\omega) = 0$ otherwise. Similarly, define X_- such that $X_-(\omega) = X(\omega)$ if $\alpha\langle u, X(\omega)\rangle < 0$ and $X_-(\omega) = 0$ otherwise. Since $\mathbb{E}\big[[\![\alpha\langle u, X\rangle]\!]_0^1\big] = \mathbb{E}\big[[\![\alpha\langle u, X_+\rangle]\!]_0^1\big] + \mathbb{E}\big[[\![\alpha\langle u, X_-\rangle]\!]_0^1\big]$ it is sufficient to explain how to construct a unitary $\mathcal{P}_+ : |u\rangle|\vec{0}\rangle \mapsto |u\rangle|\varphi_{+,u}\rangle$ such that $\big\||\varphi_{+,u}\rangle - e^{im\mathbb{E}\left[[\![\alpha\langle u,X_+\rangle]\!]_0^1\right]}|\vec{0}\rangle\big\| \leq \epsilon/2$ when $\alpha\mathbb{E}[|\langle u, X\rangle|] \leq L_2$. One can construct $\mathcal{P}_-$ that encodes $\mathbb{E}\big[[\![\alpha\langle u, X_-\rangle]\!]_0^1\big]$ using a similar approach. The proposition then

follows by taking the product $\widetilde{\mathcal{P}}_{X,L_2,m,\alpha,\epsilon} = \mathcal{P}_+\mathcal{P}_-$ and noting that $\Pr_u[\alpha\mathbb{E}[|\langle u, X\rangle|] \geq L_2] \leq \alpha/2$ by the second part of Lemma 3.1 since $\mathbb{E}[\|X\|_2] \leq L_2$.

There are three steps in the construction of $\mathcal{P}_+$. First, we construct a unitary V_+ acting on $\mathcal{H}_G \otimes \mathcal{H}_\Omega \otimes \mathcal{H}_E \otimes \mathbb{C}^2$ as follows,

$$\begin{aligned} V_+ : |u\rangle|\vec{0}\rangle|0\rangle &\mapsto \sum_{\omega\in\Omega} \sqrt{\mathbb{P}(\omega)}|u\rangle|\omega, X(\omega)\rangle|0\rangle \\ &\mapsto \sum_{\omega\in\Omega} \sqrt{\mathbb{P}(\omega)}|u\rangle|\omega, X(\omega)\rangle\Big(\sqrt{1-[\![\alpha\langle u, X_+(\omega)\rangle]\!]_0^1}|0\rangle + \sqrt{[\![\alpha\langle u, X_+(\omega)\rangle]\!]_0^1}|1\rangle\Big) \\ &= \sqrt{1-\mathbb{E}[[\![\alpha\langle u, X_+\rangle]\!]_0^1]}|u\rangle|\psi_u^0\rangle|0\rangle + \sqrt{\mathbb{E}[[\![\alpha\langle u, X_+\rangle]\!]_0^1]}|u\rangle|\psi_u^1\rangle|1\rangle \end{aligned}$$

where the first step uses one quantum experiment and one binary oracle query to X, the second step performs a sequence of controlled rotations, and in the last line $|\psi_u^0\rangle, |\psi_u^1\rangle$ are some irrelevant unit states. Secondly, if $L_2 < 1/4$, we apply the Linear Amplitude Amplification algorithm of Proposition 2.10 on V_+ with $t = 1/(2\sqrt{L_2})$ and accuracy $\epsilon/(32mL_2)$, which gives a unitary $W_+ : |u\rangle|\vec{0}\rangle|0\rangle \mapsto \sqrt{1-p_u}|u\rangle|\psi_u^0\rangle|0\rangle + \sqrt{p_u}|u\rangle|\psi_u^1\rangle|1\rangle$ where, for each $u \in G$,

$$\left|\sqrt{p_u} - \sqrt{\frac{\mathbb{E}[[\![\alpha\langle u, X_+\rangle]\!]_0^1]}{4L_2}}\right| \leq \frac{\epsilon}{32mL_2}, \qquad \text{if } \mathbb{E}[[\![\alpha\langle u, X_+\rangle]\!]_0^1] \leq L_2,$$

using $O(\log(mL_2/\epsilon)/\sqrt{L_2})$ applications of V_+ and $V_+^\dagger$. If $L_2 < 1/4$ we directly take $W_+ = V_+$. Thirdly, we define $L_2' = \min(L_2, 1/4)$ and apply the phase conversion algorithm of Lemma 2.11 on W_+ with $t = 4mL_2'$ and accuracy $\epsilon/4$. We obtain a unitary $\mathcal{P}_+ : |u\rangle|\vec{0}\rangle \mapsto |u\rangle|\varphi_{+,u}\rangle$ such that, by the triangle inequality, the state $|\varphi_{+,u}\rangle$ satisfies $\big\||\varphi_{+,u}\rangle - e^{im\mathbb{E}[[\![\alpha\langle u,X_+\rangle]\!]_0^1]}|\vec{0}\rangle\big\| \leq \big\||\varphi_{+,u}\rangle - e^{i4mL_2'p_u}|\vec{0}\rangle\big\| + \big|e^{i4mL_2'p_u} - e^{im\mathbb{E}[[\![\alpha\langle u,X_+\rangle]\!]_0^1]}\big| \leq \epsilon/4 + \big|4mL_2'p_u - m\mathbb{E}[[\![\alpha\langle u, X_+\rangle]\!]_0^1]\big| = \epsilon/4 + 4mL_2'\Big|\sqrt{p_u} - \sqrt{\frac{\mathbb{E}[[\![\alpha\langle u,X_+\rangle]\!]_0^1]}{4L_2'}}\Big| \cdot \Big|\sqrt{p_u} + \sqrt{\frac{\mathbb{E}[[\![\alpha\langle u,X_+\rangle]\!]_0^1]}{4L_2'}}\Big|$. Thus, for each $u \in G$,

$$\Big\||\varphi_{+,u}\rangle - e^{im\mathbb{E}[[\![\alpha\langle u,X_+\rangle]\!]_0^1]}|\vec{0}\rangle\Big\| \leq \epsilon/2, \qquad \text{if } \mathbb{E}[[\![\alpha\langle u, X_+\rangle]\!]_0^1] \leq L_2,$$

and $\mathcal{P}_+$ uses $O(mL_2+\log(1/\epsilon))$ applications of W_+ and $W_+^\dagger$. Overall, we used $\widetilde{O}(m\sqrt{L_2}\log^2(1/\epsilon))$ quantum experiments and binary oracle queries to X to implement one application of $\mathcal{P}_+$. Moreover, the condition $\mathbb{E}[[\![\alpha\langle u, X_+\rangle]\!]_0^1] \leq L_2$ is implied by $\alpha\mathbb{E}[|\langle u, X\rangle|] \leq L_2$. □

We finally describe the algorithm that estimates the mean of any bounded random variable (Algorithm 1). Our approach relies on applying the quantum Fourier transform over G to the above directional mean oracle in a similar manner as in previous work (e.g. [BV97; Jor05; GAW19]). The results of Lemma 3.1 play again a central role in the analysis.

Theorem 3.3 (Bounded multivariate estimator)**.** *Let X be a d-dimensional bounded random variable such that $\|X\|_2 \leq 1$. Given three reals $L_2 \in (0,1]$, $\delta \in (0,1)$ and $n \geq 1$ such that $\mathbb{E}[\|X\|_2] \leq L_2$, the* bounded multivariate estimator $\mathsf{QBounded}_d(X, L_2, n, \delta)$ *(Algorithm 1) outputs a mean estimate $\widetilde{\mu}$ of $\mu = \mathbb{E}[X]$ such that*

$$\|\widetilde{\mu} - \mu\|_\infty \leq \frac{\sqrt{L_2}\log(d/\delta)}{n}$$

with probability at least $1-\delta$. It uses $\widetilde{O}(n)$ quantum experiments and binary oracle queries to X.

Proof. If $n \leq \frac{\log(d/\delta)}{\sqrt{L_2}}$ then by choosing $\widetilde{\mu} = 0$ at step 1 we directly have $\|\widetilde{\mu} - \mu\|_\infty \leq \mathbb{E}[\|X\|_2] \leq \frac{\sqrt{L_2}\log(d/\delta)}{n}$. Thus, we can suppose from now on that $n > \frac{\log(d/\delta)}{\sqrt{L_2}}$ (in particular, $m \geq 1/L_2$).

1. If $n \le \frac{\log(d/\delta)}{\sqrt{L_2}}$ then output $\widetilde{\mu} = 0$.
2. Set $\alpha = \frac{1}{\sqrt{\log(400\pi n\sqrt{d})}}$ and $m = 2^{\left\lceil \log\left(\frac{8\pi}{\alpha}\cdot\frac{n}{\sqrt{L_2}\log(d/\delta)}\right)\right\rceil}$.
3. For $k = 1, \ldots, \lceil 18\log(d/\delta)\rceil$:
   (a) Compute the uniform superposition $|G\rangle := \frac{1}{m^{d/2}}\sum_{u\in G}|u\rangle$ over G.
   (b) Compute the state $|\psi\rangle := \widetilde{\mathcal{P}}_{X,L_2,m,\alpha,\epsilon}|G\rangle|\vec{0}\rangle \in \mathcal{H}_G \otimes \mathcal{H}_{\text{aux}}$, where $\widetilde{\mathcal{P}}_{X,L_2,m,\alpha,\epsilon}$ is the directional mean oracle constructed in Proposition 3.2 with $\epsilon = 1/25$.
   (c) Compute the state $|\phi\rangle := (\mathrm{QFT}_G^{-1} \otimes \mathbb{I}_{\text{aux}})|\psi\rangle$ where the unitary $\mathrm{QFT}_G : |u\rangle \mapsto \frac{1}{m^{d/2}}\sum_{v\in G} e^{2i\pi m\langle u,v\rangle}|v\rangle$ is the quantum Fourier transform over G.
   (d) Measure the $\mathcal{H}_G$ register of $|\phi\rangle$ in the computational basis and let $\widetilde{v}^{(k)} \in G$ denote the obtained result. Set $\widetilde{\mu}^{(k)} = \frac{2\pi}{\alpha}\widetilde{v}^{(k)}$.
4. Output the coordinate-wise median $\widetilde{\mu} = \text{median}(\widetilde{\mu}^{(1)}, \ldots, \widetilde{\mu}^{(\lceil 18\log(d/\delta)\rceil)})$.

Algorithm 1: Bounded multivariate estimator, $\mathsf{QBounded}_d(X, L_2, n, \delta)$.

Let $|\psi'\rangle, |\psi''\rangle \in \mathcal{H}_G \otimes \mathcal{H}_{\text{aux}}$ be the two unit states defined as follows,

$$|\psi'\rangle = \frac{1}{m^{d/2}}\sum_{u\in G} e^{im\mathbb{E}\left[[\![\alpha\langle u,X\rangle]\!]_0^1\right]}|u\rangle|\vec{0}\rangle \quad \text{and} \quad |\psi''\rangle = \frac{1}{m^{d/2}}\sum_{u\in G} e^{im\alpha\langle u,\mathbb{E}[X]\rangle}|u\rangle|\vec{0}\rangle.$$

We first show that the state $|\psi\rangle$ at step 3b satisfies $\||\psi\rangle - |\psi''\rangle\| \le 1/12$. On one hand, by Proposition 3.2, we have $\||\psi\rangle - |\psi'\rangle\|^2 = \frac{1}{m^d}\sum_u \|\widetilde{\mathcal{P}}_{X,L_2,m,\alpha,\epsilon}(|u\rangle|\vec{0}\rangle) - e^{im\mathbb{E}\left[[\![\alpha\langle u,X\rangle]\!]_0^1\right]}|u\rangle|\vec{0}\rangle\|^2 \le \epsilon^2 + \alpha$. On the other hand, by using the inequality $\sin^2(x) \le |x|$, we have $\||\psi'\rangle - |\psi''\rangle\|^2 = \frac{4}{m^d}\sum_{u\in G}\sin^2\left(\frac{m}{2}\left(\mathbb{E}\left[[\![\alpha\langle u,X\rangle]\!]_0^1\right] - \alpha\langle u,\mathbb{E}[X]\rangle\right)\right) \le \frac{2m}{m^d}\sum_{u\in G}\left|\mathbb{E}\left[[\![\alpha\langle u,X\rangle]\!]_0^1\right] - \alpha\langle u,\mathbb{E}[X]\rangle\right| \le \frac{2m}{m^d}\mathbb{E}\left[\sum_u\left|[\![\alpha\langle u,X\rangle]\!]_0^1 - \alpha\langle u,X\rangle\right|\right] \le 2m\alpha\sqrt{d}\mathbb{E}[\Pr_u[\alpha|\langle u,X\rangle| > 1]]$ where the last step uses $|\langle u,X\rangle| \le \sqrt{d}$. By the first part of Lemma 3.1, we have $\Pr_u\left[\alpha|\langle u,X\rangle| \ge 1\right] \le 2e^{-2/\alpha^2}$ since $\|X\|_2 \le 1$. Thus, $\||\psi'\rangle - |\psi''\rangle\|^2 \le 4m\alpha\sqrt{d}e^{-2/\alpha^2}$. By the triangle inequality, $\||\psi\rangle - |\psi''\rangle\| \le \sqrt{\epsilon^2+\alpha} + 2\sqrt{m\alpha}d^{1/4}e^{-1/\alpha^2} \le 1/12$.

We now analyse steps 3c and 3d of the algorithm assuming $|\psi\rangle$ is replaced with $|\psi''\rangle$. Let $\widetilde{v} \in G$ denote the vector obtained by measuring the first register of $(\mathrm{QFT}_G^{-1} \otimes \mathbb{I}_{\text{aux}})|\psi''\rangle$ in the computational basis. Since the phases satisfy $\|\alpha\mathbb{E}[X]\|_\infty \le 2\pi/3$, we can apply the analysis of phase estimation (e.g. [GAW19, Lemma 5.1]) to conclude that $\left|\widetilde{v}_j - \frac{\alpha}{2\pi}\mathbb{E}[X]_j\right| \le 4/m$ with probability at least $5/6$ for each $j \in [d]$. By replacing $|\psi\rangle$ with $|\psi''\rangle$, we achieve the same result with probability at least $5/6 - 2\||\psi\rangle - |\psi''\rangle\| \ge 2/3$. Finally, by the Chernoff bound, $\left\|\widetilde{\mu} - \mathbb{E}[X]\right\|_\infty \le 8\pi/(\alpha m) \le \sqrt{L_2}\log(d/\delta)/n$ with probability at least $1-\delta$.

The total number of quantum experiments and binary oracle queries to X is $O(\log(d/\delta)\cdot m\sqrt{L_2}\log^2(1/\epsilon)) = O(n/\alpha) = \widetilde{O}(n)$. □

3.2 Near-optimal multivariate estimator

We describe in Algorithm 2 our main quantum algorithm for estimating the mean of a d-dimensional random variable in the binary oracle model. Our approach relies on applying the bounded multivariate estimator, developed in the previous section, to a sequence of carefully chosen truncated random variables.

Theorem 3.4 (Near-optimal multivariate estimator). *Let X be a d-dimensional random variable with mean μ and covariance matrix Σ. Given two reals $\delta \in (0,1)$ and $n \ge \log(d/\delta)$, the*

1. Set $k = \lceil 2\log\left(\frac{2\sqrt{2}n}{\log(d/\delta)}\right)\rceil$ and $n' = n \cdot \frac{(k+1)4\log(5kd/\delta)}{\sqrt{c}\log(d/\delta)}$ where c is the constant mentioned in Theorem 2.9.
2. Run any classical sub-Gaussian estimator (Proposition 2.8) on X using $O(\log(1/\delta))$ samples to compute a mean estimate $\eta \in \mathbb{R}^d$ such that $\Pr\left[\|\eta - \mu\|_2 > \sqrt{\mathrm{Tr}(\Sigma)}\right] \leq \delta/2$.
3. Define the random variable $Y = X - \eta$.
4. For $j = 0, \ldots, k$:
   (a) Compute an estimate a_j of the quantile of order 2^{-j} of $\|Y\|_2$ by using the quantile estimator $\mathsf{Quantile}(\|Y\|_2, 2^{-j}, \delta/(5k))$.
   (b) Define the bounded random variable $Y_j = \frac{1}{a_j}[\![Y]\!]_{a_{j-1}}^{a_j}$ (where $a_{-1} = 0$). If $a_{j-1} = a_j$ then set $\widetilde{\mu}_j = 0$, else compute an estimate $\widetilde{\mu}_j$ of $\mathbb{E}[Y_j]$ by using the bounded multivariate estimator $\mathsf{QBounded}_d(Y_j, 2^{-(j-1)}, n', \delta/(5k))$.
5. Output $\widetilde{\mu} = \eta + \sum_{j=0}^{k} a_j \widetilde{\mu}_j$.

Algorithm 2: Near-optimal multivariate estimator, $\mathsf{QEstimator}_d(X, n, \delta)$.

quantum multivariate estimator $\mathsf{QEstimator}_d(X, n, \delta)$ *(Algorithm 2) outputs a mean estimate* $\widetilde{\mu}$ *such that*

$$\|\widetilde{\mu} - \mu\|_\infty \leq \frac{\sqrt{\mathrm{Tr}(\Sigma)}\log(d/\delta)}{n}$$

with probability at least $1 - \delta$*. It uses* $\widetilde{O}(n)$ *quantum experiments and binary oracle queries to* X*.*

Proof. The main part of the proof is to show that the mean estimate $\widetilde{\mu}_Y = \sum_{j=0}^{k} a_j \widetilde{\mu}_j$ of $\mu_Y = \mathbb{E}[Y]$ satisfies

$$\|\widetilde{\mu}_Y - \mu_Y\|_\infty \leq \frac{\sqrt{\mathbb{E}[\|Y\|_2^2]}\log(d/\delta)}{\sqrt{2}n} \tag{1}$$

with probability at least $1 - \delta/2$. The theorem follows since $\|\widetilde{\mu} - \mu\|_\infty = \|\widetilde{\mu}_Y - \mu_Y\|_\infty$ and $\mathbb{E}[\|Y\|_2^2] = \mathbb{E}[\|X - \mu\|_2^2] + \|\mu - \eta\|_2^2 = \mathrm{Tr}(\Sigma) + \|\mu - \eta\|_2^2 \leq 2\,\mathrm{Tr}(\Sigma)$, where the last inequality holds with probability at least $1 - \delta/2$. The algorithms uses $\widetilde{O}(k2^{k/2}\log(k/\delta) + kn') = \widetilde{O}(n)$ quantum experiments and binary oracle queries to X.

We now turn to the proof of Equation (1). We make the assumption that all the subroutines used in step 4 are successful, which is the case with probability at least $(1-\delta/(5k))^{2k+2} \geq 1 - \delta/2$. The sequence $(a_j)_j$ of quantile estimates computed at step 4a satisfies $Q(2^{-j}) \leq a_j \leq Q(c2^{-j})$ for all $j \in \{0, \ldots, k\}$, where c is the constant mentioned in Theorem 2.9. On one hand, by Markov's inequality, $\Pr[\|Y\|_2 \geq a_j] = \Pr[\|Y\|_2^2 \geq a_j^2] \leq \mathbb{E}[\|Y\|_2^2]/a_j^2$. On the other hand, by definition of the quantile function, $\Pr[\|Y\|_2 \geq a_j] \geq \Pr[\|Y\|_2 \geq Q(c2^{-j})] \geq c2^{-j}$. Thus,

$$a_j \leq c^{-1/2}2^{j/2}\sqrt{\mathbb{E}[\|Y\|_2^2]}. \tag{2}$$

Since $\Pr[\|Y\|_2 > a_{j-1}] \leq \Pr[\|Y\|_2 > Q(2^{-(j-1)})] < 2^{-(j-1)}$, we also have that

$$\mathbb{E}[\|Y_j\|_2] < 2^{-(j-1)}. \tag{3}$$

Hence, $2^{-(j-1)}$ is a valid upper bound on the expectation of $\|Y_j\|_2$. Consequently, by Theorem 3.3, each estimate $\widetilde{\mu}_j$ satisfies $\|\widetilde{\mu}_j - \mathbb{E}[Y_j]\|_\infty \leq \frac{2^{-(j-1)/2}\log(5kd/\delta)}{n'}$. Moreover, the truncated random

variable $[\![Y]\!]_{a_k}^{+\infty}$ satisfies,

$$\|\mathbb{E}[[\![Y]\!]_{a_k}^{+\infty}]\|_\infty \leq \|\mathbb{E}[[\![Y]\!]_{a_k}^{+\infty}]\|_2 \leq \sqrt{\mathbb{E}[\|Y\|_2^2]\Pr[\|Y\|_2 > a_k]} \leq \frac{\sqrt{\mathbb{E}[\|Y\|_2^2]}}{2^{k/2}} \tag{4}$$

where the first step is by monotonicity of the norm and the second is by Cauchy-Schwartz inequality. Overall, the error is $\|\widetilde{\mu}_Y - \mu_Y\|_\infty \leq \sum_{j=0}^{k} a_j \frac{2^{-(j-1)/2}\log(5kd/\delta)}{n'} + \|\mathbb{E}[[\![Y]\!]_{a_k}^{+\infty}]\|_\infty \leq \frac{(k+1)\sqrt{2\mathbb{E}[\|Y\|_2^2]}\log(5kd/\delta)}{\sqrt{c}n'} + \frac{\sqrt{\mathbb{E}[\|Y\|_2^2]}}{2^{k/2}} \leq \frac{\sqrt{\mathbb{E}[\|Y\|_2^2]}\log(d/\delta)}{\sqrt{2}n}$. □

As a direct corollary of Proposition 2.8 and Theorem 3.4, we obtain the following result for estimating the mean in Euclidean norm. We prove in the next section that these bounds are optimal for all values of n and d, up to logarithmic factors.

Theorem 3.5 (Multivariate estimator in Euclidean norm)**.** *There exists a quantum estimator with the following properties. Let X be a d-dimensional random variable with mean μ and covariance matrix Σ. Given two reals $\delta \in (0,1)$ and $n \geq \log(d/\delta)$, the estimator outputs a mean estimate $\widetilde{\mu}$ such that*

$$\|\widetilde{\mu} - \mu\|_2 \leq \begin{cases} \sqrt{\frac{\mathrm{Tr}(\Sigma)}{n}} + \sqrt{\frac{\|\Sigma\|\log(1/\delta)}{n}}, & \text{if } n \leq d, \\ \frac{\sqrt{d\,\mathrm{Tr}(\Sigma)}\log(d/\delta)}{n}, & \text{if } n > d, \end{cases}$$

with probability at least $1-\delta$. It uses $\widetilde{O}(n)$ quantum experiments and binary oracle queries to X.

Proof. If $n \leq d$ we use any classical sub-Gaussian estimator (Proposition 2.8). If $n > d$ we use the quantum estimator of Theorem 3.4 and the norm inequality $\|\widetilde{\mu} - \mu\|_2 \leq \sqrt{d}\|\widetilde{\mu} - \mu\|_\infty$. □

3.3 Lower bounds

Our lower bounds are based on reductions from the following composition problem, where the goal is to approximate an N-bit string whose entries are determined by parities over M bits.

Problem 1 ($\mathsf{Search}^N \circ \mathsf{Parity}^M$)**.** Let $N, M \geq 1$ be two integers. Let $\mathcal{A}_{N,M}$ denote the set of all matrices $A \in \{0,1\}^{N\times M}$ such that $\lfloor N/2 \rfloor$ rows have Hamming weights $\lfloor M/2 \rfloor$, and the other rows have Hamming weights $\lfloor M/2 \rfloor + 1$. Define the vector $b^{(A)} \in \{0,1\}^N$ such that,

$$b_i^{(A)} = \begin{cases} 0, & \text{if the } i\text{-th row of } A \text{ has Hamming weight } \lfloor M/2 \rfloor, \\ 1, & \text{if the } i\text{-th row of } A \text{ has Hamming weight } \lfloor M/2 \rfloor + 1, \end{cases}$$

for each $i \in [N]$. Then, the $\mathsf{Search}^N \circ \mathsf{Parity}^M$ problem consists of finding a vector $\widetilde{b} \in \mathbb{R}^N$ that minimizes $\|\widetilde{b} - b^{(A)}\|_2$ given a quantum oracle $|i,j\rangle \mapsto (-1)^{A_{i,j}}|i,j\rangle$ to $A \in \mathcal{A}_{N,M}$.

We use the next lower bound for the $\mathsf{Search}^N \circ \mathsf{Parity}^M$ problem, which states that $\Omega(NM)$ queries are needed to approximate the vector $b^{(A)}$ with small error. The proof can be easily adapted from that of [Ape21, Lemma 11] or [CJ21, Lemma 5.7].

Lemma 3.6. *Let $\alpha > 1$ be a sufficiently large constant. Consider any quantum algorithm for the $\mathsf{Search}^N \circ \mathsf{Parity}^M$ problem that uses at most NM/α queries on all inputs. Then, there exists an input $A \in \mathcal{A}_{N,M}$ such that this algorithm returns a vector $\widetilde{b}$ satisfying $\|\widetilde{b} - b^{(A)}\|_2 \geq \Omega(\sqrt{N})$ with probability at least $2/3$.*

We now show that the quantum mean estimators developed in the previous sections are tight (up to logarithmic factors). For simplicity in the proof, we only consider the case of approximation in Euclidean norm. We first prove that the mean estimation problem admits no quantum advantage when the complexity parameter n is smaller than the dimension. The proof works by a reduction from the $\mathsf{Search}^{\alpha n} \circ \mathsf{Parity}^1$ problem (where α is the constant mentioned in Lemma 3.6).

Theorem 3.7 (Low-precision regime)**.** *Consider two integers n, d such that $n \leq d/\alpha$. Fix $\sigma > 0$ and let $\mathcal{P}_\sigma$ denote the set of all d-dimensional quantum random variables with covariance matrix Σ such that $\mathrm{Tr}(\Sigma) = \sigma^2$. Then, for any quantum estimator that uses at most n binary oracle queries, there exists $X \in \mathcal{P}_\sigma$ such that the estimator returns a mean estimate $\widetilde{\mu}$ of $\mu = \mathbb{E}[X]$ that satisfies*

$$\|\widetilde{\mu} - \mu\|_2 \geq \Omega\left(\sqrt{\frac{\mathrm{Tr}(\Sigma)}{n}}\right)$$

with probability at least $2/3$.

Proof. We assume for simplicity that α is even and d is a power of two (the other cases can be handled by simple padding arguments). Consider the partial Hadamard matrix $H \in \mathbb{R}^{\alpha n \times d}$ such that $H_{i,j} = \frac{1}{\sqrt{d}}(-1)^{\langle i,j \rangle}$, where $i, j \in \{0,1\}^{\log(d)}$ are written over $\log(d)$ bits. Note that $HH^\top = \mathbb{I}_{\alpha n}$ and the spectral norm of H is $\|H\| = 1$. Let $(\Omega, 2^\Omega, \mathbb{P})$ be the probability space such that $\Omega = [\alpha n]$ and $\mathbb{P}(i) = 1/(\alpha n)$ for all $i \in \Omega$. For any vector $b \in \{0,1\}^{\alpha n}$ with Hamming weight $\|b\|_1 = \alpha n/2$, define the random variable $X^{(b)} : \Omega \to \mathbb{R}^d$ such that,

$$X^{(b)}(i) = \alpha\sigma\sqrt{\frac{n}{(\alpha^2 n - \alpha)/2}} b_i H_i$$

where $H_i \in \mathbb{R}^d$ is the i-th row of H. The expectation of $X^{(b)}$ is $\mathbb{E}[X^{(b)}] = \frac{\sigma}{\sqrt{n(\alpha^2 n - \alpha)/2}} H^\top b$ and the trace of its covariance matrix is $\mathrm{Tr}(\Sigma) = \mathbb{E}[\|X^{(b)}\|_2^2] - \|\mathbb{E}[X^{(b)}]\|_2^2 = \frac{2\alpha\sigma^2}{\alpha^2 n - \alpha}\|b\|_1 - \frac{2\sigma^2}{n(\alpha^2 n - \alpha)}\|b\|_1 = \sigma^2$. Given any quantum estimator that uses n binary oracle queries to $X^{(b)}$ and outputs an estimate $\widetilde{\mu}$ of $\mu = \mathbb{E}[X^{(b)}]$, we can transform it into an algorithm for the $\mathsf{Search}^{\alpha n} \circ \mathsf{Parity}^1$ problem that uses n queries to b and returns the estimate $\widetilde{b} = \frac{\sqrt{n(\alpha^2 n - \alpha)/2}}{\sigma} H\widetilde{\mu}$ with error $\|\widetilde{b} - b\|_2 \leq \frac{\sqrt{n(\alpha^2 n - \alpha)/2}}{\sigma}\|H(\widetilde{\mu} - \mu)\|_2 \leq \frac{\alpha n}{\sigma}\|\widetilde{\mu} - \mu\|_2$. Thus, by Lemma 3.6, there exists an input b such that $\|\widetilde{\mu} - \mu\|_2 \geq \frac{\sigma}{\alpha n}\|\widetilde{b} - b\|_2 = \Omega\big(\frac{\sigma}{\sqrt{n}}\big)$. □

We now prove that the quantum estimator of Theorem 3.5 is optimal in the regime where n is larger than the dimension. The proof works by a reduction from the $\mathsf{Search}^d \circ \mathsf{Parity}^{\alpha n/d}$ problem.

Theorem 3.8 (High-precision regime)**.** *Consider two integers n, d such that $n > d/\alpha$. Fix $\sigma > 0$ and let $\mathcal{P}_\sigma$ denote the set of all d-dimensional quantum random variables with covariance matrix Σ such that $\mathrm{Tr}(\Sigma) = \sigma^2$. Then, for any quantum estimator that uses at most n binary oracle queries, there exists $X \in \mathcal{P}_\sigma$ such that the estimator returns a mean estimate $\widetilde{\mu}$ of $\mu = \mathbb{E}[X]$ that satisfies*

$$\|\widetilde{\mu} - \mu\|_2 \geq \Omega\left(\frac{\sqrt{d\,\mathrm{Tr}(\Sigma)}}{n}\right)$$

with probability at least $2/3$.

Proof. We assume for simplicity that αn is a multiple of d, and d is even (the other cases can be handled by simple padding arguments). Let $(\Omega, 2^\Omega, \mathbb{P})$ be the probability space such that $\Omega = [d] \times [\alpha n/d]$ and $\mathbb{P}(i,j) = 1/(\alpha n)$ for all $(i,j) \in \Omega$. For any input $A \in \mathcal{A}_{d,\alpha n/d}$ to the $\mathsf{Search}^d \circ \mathsf{Parity}^{\alpha n/d}$ problem, define the random variable $X^{(A)} : \Omega \to \mathbb{R}^d$ such that,

$$X^{(A)}(i,j) = \frac{\alpha\sigma n}{\sqrt{(\alpha n)^2 - 2d^2}}(-1)^{1+A_{i,j}} e_i$$

where $e_i \in \mathbb{R}^d$ is the i-th indicator vector. The expectation of $X^{(A)}$ is $\mathbb{E}[X^{(A)}] = \frac{2\sigma}{\sqrt{(\alpha n)^2 - d^2}} b^{(A)}$ and the trace of its covariance matrix is $\mathrm{Tr}(\Sigma) = \mathbb{E}[\|X^{(A)}\|_2^2] - \|\mathbb{E}[X^{(A)}]\|_2^2 = \frac{(\alpha\sigma n)^2}{(\alpha n)^2 - d^2} -$

$\frac{4\sigma^2}{(\alpha n)^2-d^2}\|b^{(A)}\|_1^2 = \sigma^2$ since $b^{(A)}$ has Hamming weight $d/2$. Given any quantum estimator that uses n binary oracle queries to $X^{(A)}$ and outputs an estimate $\widetilde{\mu}$ of $\mu = \mathbb{E}[X^{(A)}]$, we can transform it into an algorithm for the $\mathsf{Search}^d \circ \mathsf{Parity}^{\alpha n/d}$ problem that uses n queries to A and returns the estimate $\widetilde{b} = \frac{\sqrt{(\alpha n)^2-d^2}}{2\sigma}\widetilde{\mu}$. Thus, by Lemma 3.6, there exists an input A such that $\|\widetilde{\mu} - \mu\|_2 \geq \frac{2\sigma}{\sqrt{(\alpha n)^2-d^2}}\|\widetilde{b} - b^{(A)}\|_2 = \Omega\left(\frac{\sqrt{d}\sigma}{n}\right)$. □

4 Mean estimation with phase oracles

In some cases, we might not have access to the random variable X through a binary oracle $\mathcal{B}_X$, as considered in the previous part of this paper, but merely through a less powerful oracle. Several such input models arise naturally in the literature. For instance, in [GAW19], it is shown how *phase oracles* and *probability oracles* arise naturally in the context of variational quantum eigensolvers, QAOA, and quantum auto-encoders. In [Ape21], the author considers a quantum operation that prepares an unknown distribution, which we henceforth refer to as a *distribution oracle*. In [CJ21], we consider the multivariate mean estimation problem relative to all these input models.

There is one profound qualitative difference between all of these input models and the binary oracle setting considered in the previous section, which is that these input oracles in some sense preserve proximity. That is to say, if we have two random variables X and $\widetilde{X}$, whose values differ by at most ϵ in some norm, then the operator norm difference between their respective input oracles is bounded by $O(\text{poly}(\epsilon))$ too. This qualitatively differentiates this setting from the one considered in the previous sections, and we refer to input models satisfying this property as *analog* models.

For ease of exposition, in this part of the paper we only consider the case where our random variable X takes values bounded in the d-dimensional hypercube $[-1/4, 1/4]^d$, and can be accessed through a phase oracle $\mathcal{P}_X$, as defined in Definition 2.7. We refer to [CJ21] for a more elaborate exposition of the other input models mentioned at the start of this section. Just like in the previous section, we assume to have access to the probability space via the oracle $U_{\mathbb{P}}$ of Definition 2.5.

Since its values are contained in the hypercube $[-1/4, 1/4]^d$, the random variable X satisfies $\text{Var}[X_j] \leq 1/16$ for all $j \in [d]$, and hence $\text{Tr}[\Sigma] \leq d/16$. This suggests that we can use the results from the previous section naively, to obtain a multivariate mean estimator in this setting as well.

There are two naive ways of approaching this. First, we can simulate a call to the binary oracle $\mathcal{B}_X$, considered in the previous part of this paper, using d consecutive runs of phase estimation on this phase oracle. Thus, if one only cares about the performance of the multivariate mean estimator expressed in the number of calls to $U_{\mathbb{P}}$, then it is clear that the same results can be obtained, that is, with n calls to $U_{\mathbb{P}}$, one can obtain a multivariate mean estimator that finds an approximation $\widetilde{\mu}$ to μ which satisfies $\|\widetilde{\mu} - \mu\|_\infty = \widetilde{O}(\sqrt{d}/n)$.

Secondly, if one only cares about the number of calls to $\mathcal{P}_X$, then a little more elaborate construction also readily reduces to the binary oracle setting. Using Lemma 2.12, one can turn the phase oracle $\mathcal{P}_X$ into a probability oracle that given input $|\omega\rangle|j\rangle|0\rangle$ prepares the state $|\omega\rangle|j\rangle(\sqrt{1/2 + X(\omega)_j}|1\rangle + \sqrt{1/2 - X(\omega)_j}|0\rangle)$. This operation can be combined with $U_{\mathbb{P}}$ and an operation that prepares the uniform superposition over all $j \in [d]$, to obtain the operator U that acts as

$$U : |0\rangle|0\rangle|0\rangle \mapsto \sum_{\omega\in\Omega} \sqrt{\mathbb{P}(\omega)}|\omega\rangle \otimes \frac{1}{\sqrt{d}}\sum_{j=1}^{d}|j\rangle \otimes \left(\sqrt{\frac{1}{2} + X(\omega)_j}|1\rangle + \sqrt{\frac{1}{2} - X(\omega)_j}|0\rangle\right).$$

Now, let $\overline{\Omega} = \Omega \times [d] \times \{0, 1\}$, and let $\overline{\mathbb{P}}(\omega, j, b) = \mathbb{P}(\omega)/(2d) - (-1)^b\mathbb{P}(\omega)X(\omega)_j/d$ be a probability measure on $\overline{\Omega}$. Furthermore, let the random variable $Y : \overline{\Omega} \to \mathbb{R}^d$ be defined as

$Y(\omega, j, b) = (\mathbb{1}_{i=j\wedge b=1})_{i\in[d]}$. Then, implementing a binary oracle $\mathcal{B}_Y$ is trivial, and the operation U above implements the operation $U_{\mathbb{P}}$. Finally, observe that $\mathbb{E}[Y] = \vec{1}/(2d) + \mathbb{E}[X]/d$. Therefore, we can find an approximation $\widetilde{\mu}_Y$ to $\mu_Y = \mathbb{E}[Y]$ which satisfies $\|\widetilde{\mu}_Y - \mu_Y\|_\infty = \widetilde{O}(\sqrt{d}/n')$, with n' calls to U. Since U requires only polylogarithmically many calls to $\mathcal{P}_X$, and since $\mathbb{E}[Y]$ is shrunk by a factor of d compared to $\mu = \mathbb{E}[X]$, we can obtain an estimate $\widetilde{\mu}$ such that $\|\widetilde{\mu} - \mu\|_\infty = \widetilde{O}(d^{3/2}/n')$, with $\widetilde{O}(n')$ calls to $\mathcal{P}_X$. Note that this approach also uses $\widetilde{O}(n')$ calls to $U_{\mathbb{P}}$, so its performance in terms of number of quantum experiments is significantly worse compared to the previous approach.

The above two considerations lead to the natural question whether it is possible to combine both approaches, i.e., whether with n calls to $U_{\mathbb{P}}$ and n' calls to $\mathcal{P}_X$, it is possible to obtain an estimate $\widetilde{\mu}$ that satisfies $\|\widetilde{\mu} - \mu\|_\infty = \widetilde{O}(\max\{\sqrt{d}/n, d^{3/2}/n'\})$. In this section, we show that this is indeed possible, and that one can even shave off a factor of $\sqrt{d}$ in the second branch of the maximum. This result is displayed in Theorem 4.3.

Interestingly, the performance of this algorithm can only be shown to be optimal in the regime where $n \geq d$ and $n' \geq d$, which we refer to as the high-precision regime. In the low-precision regime, i.e., when either $n < d$ or $n' < d$, we spend some extra effort to characterize the optimal precision one can obtain. If $n' < d$, then the situation turns out to be simple, since one can only attain the trivial precision $\|\widetilde{\mu} - \mu\|_\infty = \widetilde{O}(1)$, which can even be achieved without making any queries at all. On the other hand, if $n < d$ and $n' \geq d$, then a small modification of the algorithm is enough to attain an optimal scaling of $\|\widetilde{\mu} - \mu\|_\infty = \widetilde{O}(1/\sqrt{n})$, up to polylogarithmic factors. The low-precision regime is included in the statement of Theorem 4.4.

As a final note, we remark that one can also use our techniques to obtain quantum mean estimators with performance guarantees in other ℓ_p-norms, i.e., where $p \in [1, \infty)$. All precision results follow directly from simple norm conversion, i.e., they are a multiplicative factor $\Theta(d^{1/p})$ worse compared to the ℓ_∞-case. We also show this to be optimal, up to polylogarithmic factors.

4.1 Near-optimal multivariate mean estimator in the high-precision regime

The main technical ingredient for the phase oracle setting is presented here, which is the construction of the multivariate mean estimator in ℓ_∞-norm, in the high-precision regime. Throughout, we let G be the same grid as in the previous section, i.e.,

$$G = \left\{ \frac{j}{m} - \frac{1}{2} + \frac{1}{2m} : j \in \{0, \dots, m-1\} \right\}^d \subset (-1/2, 1/2)^d.$$

where m is a number to be determined later. When we write $u \sim G$, we mean that u is taken uniformly over all elements of G, and again we let $\mathcal{H}_G$ be the Hilbert space spanned by mutually orthogonal computational basis states $|u\rangle$, for all $u \in G$.

We start by showing how one can use the phase oracle $\mathcal{P}_X$ to compute the inner product between any vector $u \in G$ and the outcome of the random variable $X(\omega)$, and prepare the result as a phase rotation. The techniques used here are similar to those exhibited in the proof of Proposition 3.2.

Lemma 4.1 (Directional phase oracle)**.** *Let $d \in \mathbb{N}$, $\epsilon \in (0,1)$, $m \geq 0$, and X a random variable bounded by $[-1/4, 1/4]^d$. Then there exists an operator $\mathcal{L}_{X,m,\epsilon} : |u\rangle|\omega\rangle|\vec{0}\rangle \mapsto |u\rangle|\omega\rangle|\varphi_{u,\omega}\rangle$ acting on $\mathcal{H}_G \otimes \mathcal{H}_\Omega \otimes \mathcal{H}_{\text{aux}}$, that can be implemented using $O((m + \log(1/\epsilon))\log(m/\epsilon))$ queries to $\mathcal{P}_X$, and such that*

$$\left\| |\varphi_{u,\omega}\rangle - e^{i\frac{m}{d}\langle u, X(\omega)\rangle}|\vec{0}\rangle \right\| \leq \epsilon.$$

Proof. For every $u \in G$, we let $u^{(+)}, u^{(-)} \in \mathbb{R}^d$ be defined as $u_j^{(+)} = \max\{u_j, 0\}$ and $u_j^{(-)} = -\min\{u_j, 0\}$, for all $j \in [d]$. Note $u = u^{(+)} - u^{(-)}$, and hence $\langle u, X(\omega)\rangle = \langle u^{(+)}, X(\omega)\rangle -$

$\langle u^{(-)}, X(\omega)\rangle$. Thus, it suffices to implement operations $\mathcal{L}_+ : |u\rangle|\omega\rangle|\vec{0}\rangle \mapsto |u\rangle|\omega\rangle|\varphi_{+,\omega,u}\rangle$ and $\mathcal{L}_- : |u\rangle|\omega\rangle|\vec{0}\rangle \mapsto |u\rangle|\omega\rangle|\varphi_{-,u,\omega}\rangle$ that satisfy

$$\left\||\varphi_{+,u,\omega}\rangle - e^{i\frac{m}{d}\langle u^{(+)},X(\omega)\rangle}|\vec{0}\rangle\right\| \leq \frac{\epsilon}{2}, \qquad \text{and} \qquad \left\||\varphi_{-,u,\omega}\rangle - e^{i\frac{m}{d}\langle u^{(-)},X(\omega)\rangle}|\vec{0}\rangle\right\| \leq \frac{\epsilon}{2},$$

because then $\mathcal{L}_{X,m,\epsilon} = \mathcal{L}_+\mathcal{L}_-^\dagger$. We now proceed to show how to implement $\mathcal{L}_+$, and omit the construction of $\mathcal{L}_-$ since it is completely analogous.

First, we note that by adding a global phase to every call of $\mathcal{P}_X$, and some local rotation on the control qubit at every controlled call of $\mathcal{P}_X$, we can just as well implement the operation

$$|\omega\rangle|j\rangle \mapsto e^{i\left(\frac{1}{2}+X(\omega)_j\right)}|\omega\rangle|j\rangle.$$

Next, we turn this operation into a probability oracle acting on $\mathcal{H}_\Omega \otimes \mathbb{C}^d \otimes (\mathbb{C}^2)^{\otimes(k+1)}$, using Lemma 2.12, with $\delta = 1/4$, and precision $\epsilon^2/(64m^2)$. This implements the operation $V_+ : |\omega\rangle|j\rangle|\vec{0}\rangle|0\rangle \mapsto |\omega\rangle|j\rangle(\sqrt{1-p_{\omega,j}}|\vec{0}\rangle|0\rangle + \sqrt{p_{\omega,j}}|\psi\rangle|1\rangle)$, for some state $|\psi\rangle$, using $O(\log(m/\epsilon))$ calls to $\mathcal{P}_X$, such that

$$\left|\sqrt{p_{\omega,j}} - \sqrt{\frac{1}{2} + X(\omega)_j}\right| \leq \frac{\epsilon^2}{64m^2}.$$

Next, we can implement the following operation without any queries,

$$|u\rangle|\omega\rangle|0\rangle|\vec{0}\rangle \mapsto |u\rangle|\omega\rangle\left(\sum_{j=1}^{d}\sqrt{\frac{u_j^{(+)}}{d}}|j\rangle + \sqrt{1-\frac{\|u^{(+)}\|_1}{d}}|0\rangle\right)|\vec{0}\rangle,$$

which is a valid operation since $\|u^{(+)}\|_1/d \leq 1/2 < 1$. By using next one call to V_+, we implement the operation $W_+ : |u\rangle|\omega\rangle|\vec{0}\rangle|0\rangle \mapsto |u\rangle|\omega\rangle\left(\sqrt{p_{u,\omega}}|\psi_{u,\omega}^1\rangle|1\rangle + \sqrt{1-p_{u,\omega}}|\psi_{u,\omega}^0\rangle|0\rangle\right)$, where $|\psi_{u,\omega}^0\rangle$ and $|\psi_{u,\omega}^1\rangle$ are unit vectors, with a single call to V_+, such that

$$\begin{aligned}
&\left|p_{u,\omega} - \frac{1}{d}\left\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega)\right\rangle\right| = \left|\sum_{j=1}^{d}\frac{u_j^{(+)}}{d}p_{\omega,j} - \sum_{j=1}^{d}\frac{u_j^{(+)}}{d}\left(\frac{1}{2}+X(\omega)_j\right)\right| \\
&\leq \sum_{j=1}^{d}\frac{u_j^{(+)}}{d}\left|p_{\omega,j} - \left(\frac{1}{2}+X(\omega)_j\right)\right| = \sum_{j=1}^{d}\frac{u_j^{(+)}}{d}\left|\sqrt{p_{\omega,j}} - \sqrt{\frac{1}{2}+X(\omega)_j}\right|\cdot\left|\sqrt{p_{\omega,j}} + \sqrt{\frac{1}{2}+X(\omega)_j}\right| \\
&\leq 2\frac{\|u^{(+)}\|_1}{d}\cdot\frac{\epsilon^2}{64m^2} \leq \frac{\epsilon^2}{64m^2}.
\end{aligned}$$

Furthermore, for all $a, b > 0$, we have that $|\sqrt{a}+\sqrt{b}| \geq \max\{\sqrt{a},\sqrt{b}\} = \sqrt{\max\{a,b\}} \geq \sqrt{|a-b|}$, and hence

$$\left|\sqrt{a} - \sqrt{b}\right| = \frac{|a-b|}{\left|\sqrt{a}+\sqrt{b}\right|} \leq \frac{|a-b|}{\sqrt{|a-b|}} = \sqrt{|a-b|},$$

which implies that

$$\left|\sqrt{p_{u,\omega}} - \sqrt{\frac{1}{d}\left\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega)\right\rangle}\right| \leq \sqrt{\left|p_{u,\omega} - \frac{1}{d}\left\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega)\right\rangle\right|} \leq \frac{\epsilon}{8m}. \tag{5}$$

Next, we turn the operation W_+ back into a phase oracle using the amplitude-to-phase conversion algorithm of Lemma 2.11, with $t = m$ and accuracy $\epsilon/4$, whereby we implement the

unitary $\mathcal{P} : |u\rangle|\vec{0}\rangle \mapsto |u\rangle|\psi_{u,\omega}\rangle$ using $O(m + \log(1/\epsilon))$ applications of W_+, which by the triangle inequality satisfies

$$\begin{aligned}
\left\| |\psi_{u,\omega}\rangle - e^{i\frac{m}{d}\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega)\rangle}|\vec{0}\rangle \right\| &\leq \| |\psi_{u,\omega}\rangle - e^{imp_{u,\omega}}|\vec{0}\rangle \| + \left| e^{imp_{u,\omega}} - e^{i\frac{m}{d}\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega)\rangle} \right| \\
&\leq \frac{\epsilon}{4} + \left| mp_{u,\omega} - \frac{m}{d}\left\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega) \right\rangle \right| \\
&= \frac{\epsilon}{4} + m\left| \sqrt{p_{u,\omega}} - \sqrt{\frac{1}{d}\left\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega) \right\rangle} \right| \cdot \left| \sqrt{p_{u,\omega}} + \sqrt{\frac{1}{d}\left\langle u^{(+)}, \frac{\vec{1}}{2} + X(\omega) \right\rangle} \right| \\
&\leq \frac{\epsilon}{4} + 2m \cdot \frac{\epsilon}{8m} = \frac{\epsilon}{2}.
\end{aligned}$$

Finally, just like at the start of this proof, we can get rid of the extra global phase $m\langle u^{(+)}, \vec{1}/2\rangle$ by adding a phase gate to the control qubit whenever we call $\mathcal{P}$ in a controlled manner. The resulting operation implements $\mathcal{L}_+$.

It remains to check the number of calls to $\mathcal{P}_X$ we made throughout this proof. The number of calls to W_+ is $O(m+\log(1/\epsilon))$, each of which makes 1 call to V_+, which again performs $O(\log(m/\epsilon))$ calls to $\mathcal{P}_X$. Thus, the total number of calls to $\mathcal{P}_X$ amounts to $O((m + \log(1/\epsilon))\log(m/\epsilon))$. This completes the proof. □

One important subtlety that is a possible source for confusion is that in Equation (5), the thing that $p_{u,\omega}$ approximates is $\langle u^{(+)}, \vec{1}/2 + X(\omega)\rangle/d$, and not $\langle u, \vec{1}/2 + X(\omega)\rangle/d$. From Lemma 3.1, we know that the typical value of the latter would be $\|\vec{1}/2 + X(\omega)\|_2/d \leq 1/\sqrt{d}$, and hence if we were approximating this, we could amplify away this subnormalization of $1/\sqrt{d}$ before converting everything back into a phase oracle. We cannot use this trick, however, since the typical value of $\langle u^{(+)}, \vec{1}/2 + X(\omega)\rangle$ can be much bigger than $1/\sqrt{d}$. In fact, our optimality results later on in this section show that there is indeed no way to circumvent this.

Next, we show how the directional phase oracle we constructed in Lemma 4.1 can be used to construct a directional means oracle, in a similar spirit as in Proposition 3.2. This is the objective of the following Lemma.

Lemma 4.2 (Directional mean oracle constructed from phase oracle queries)**.** *Let $d \in \mathbb{N}$, $\epsilon, \eta \in (0,1)$, $m \geq \epsilon/(6\sqrt{d})$, and X a random variable bounded by $[-1/4, 1/4]^d$. There exists a unitary operator $\widetilde{\mathcal{P}}_{X,m,\eta,\epsilon} : |u\rangle|0\rangle \mapsto |u\rangle|\varphi_u\rangle$ acting on $\mathcal{H}_G \otimes \mathcal{H}_{\text{aux}}$ that can be implemented using $\widetilde{O}(\sqrt{d}m\log^2(1/(\epsilon\eta)))$ quantum experiments and $\widetilde{O}(dm\log^4(1/(\epsilon\eta)))$ queries to $\mathcal{P}_X$, and such that*

$$\left\| |\varphi_u\rangle - e^{im\langle u, \mathbb{E}[X]\rangle}|\vec{0}\rangle \right\| \leq \epsilon,$$

for a fraction at least $1 - \eta/2$ of all $u \in G$.

Proof. Let $K_1, K_2 > 0$ be constants to be fixed later. By setting

$$m' = \sqrt{\frac{d}{\log\left(\frac{144dm^2\left(\frac{1}{2}+\sqrt{d}\right)}{\epsilon^2\eta}\right)}}, \qquad \text{and} \qquad \epsilon' = \frac{1}{4K_1\log\left(\frac{m\sqrt{d}}{\epsilon\eta}\right)\cdot K_2\left(\frac{m\sqrt{d}}{\eta} + \log\left(\frac{1}{\epsilon}\right)\right)}$$

in Lemma 4.1, we can implement a directional phase oracle, i.e., an operation that acts as $|u\rangle|\omega\rangle|\vec{0}\rangle \mapsto |u\rangle|\omega\rangle|\chi_{u,\omega}\rangle$, such that

$$\left\| |\chi_{u,\omega}\rangle - e^{i\frac{m'}{d}\langle u, X(\omega)\rangle}|\vec{0}\rangle \right\| \leq \epsilon',$$

with $O((m' + \log(1/\epsilon'))\cdot\log(m'/\epsilon'))$ calls to $\mathcal{P}_X$.

Without incurring any extra overhead or error, we can also implement the operation $|u\rangle|\omega\rangle|\vec{0}\rangle \mapsto |u\rangle|\omega\rangle|\psi_{u,\omega}\rangle$, such that

$$\left\| |\psi_{u,\omega}\rangle - e^{i\left(\frac{1}{2}+\frac{m'}{d}\langle u, X(\omega)\rangle\right)}|\vec{0}\rangle \right\| \leq \epsilon',$$

since we can always apply some Z-rotation with angle $1/2$ to the control qubit if we want to implement this mapping in a controlled fashion.

Next, using Lemma 2.12 with $\delta = 1/4$ and precision $(m')^2\epsilon^2/(144d^2m^2)$, we can turn the above operation into a probability oracle, acting as $V_+ : |u\rangle|\omega\rangle|\vec{0}\rangle|0\rangle \mapsto |u\rangle|\omega\rangle|\varphi_{u,\omega}\rangle$, with $C_1 = O(\log(dm/(m'\epsilon)))$ calls to the directional phase oracle, and we let K_1 be the constant suppressed by the big-O-notation. It follows that

$$\left\| |\varphi_{u,\omega}\rangle - \sqrt{1-p_{u,\omega}}|\vec{0}\rangle|0\rangle - \sqrt{p_{u,\omega}}|\psi\rangle|1\rangle \right\| \leq C_1\epsilon',$$

and

$$\left| \sqrt{p_{u,\omega}} - \sqrt{\frac{1}{2} + \frac{m'}{d}\langle u, X(\omega)\rangle} \right| \leq \frac{(m')^2\epsilon^2}{144d^2m^2}, \qquad \text{if } 4m'\,|\langle u, X(\omega)\rangle| \leq d. \tag{6}$$

Let $B_{u,\omega} \in \{0,1\}$ be 1 whenever $4m'\,|\langle u, X(\omega)\rangle| > d$. Then for all $\omega \in \Omega$, we have using Lemma 3.1,

$$\Pr_{u\sim G}[B_{u,\omega}] = \Pr_{u\sim G}\left[4m'\,|\langle u, X(\omega)\rangle| > d\right] \leq \Pr_{u\sim G}\left[\frac{m'}{\sqrt{d}}\,|\langle u, X(\omega)\rangle| > \|X(\omega)\|_2\right] \leq 2e^{-\frac{2d}{(m')^2}},$$

and using $xe^{-x} \leq e^{-x/2}$, for all $x \geq 0$, we find

$$\begin{aligned}\Pr_{u\sim G}[B_{u,\omega}] &\leq 2\cdot\frac{(m')^2}{2d}\cdot\frac{2d}{(m')^2}e^{-\frac{2d}{(m')^2}} \leq \frac{(m')^2}{d}e^{-\frac{d}{(m')^2}} \\ &\leq \frac{(m')^2}{d}e^{-\log\left(\frac{144dm^2\left(\frac{1}{2}+\sqrt{d}\right)}{\epsilon^2\eta}\right)} = \frac{(m')^2\epsilon^2\eta}{144d^2m^2\left(\frac{1}{2}+\sqrt{d}\right)}.\end{aligned}$$

Thus, by averaging over all ω's, we find that

$$\frac{1}{|G|}\sum_{u\in G}\sum_{\omega\in\Omega}\mathbb{P}(\omega)B_{u,\omega} \leq \frac{(m')^2\epsilon^2\eta}{144d^2m^2\left(\frac{1}{2}+\sqrt{d}\right)},$$

and hence by the pigeonhole principle, we have that for at least a $(1-\eta/2)$-fraction of $u \in G$,

$$\sum_{\omega\in\Omega}\mathbb{P}(\omega)B_{u,\omega} \leq \frac{(m')^2\epsilon^2}{72d^2m^2\left(\frac{1}{2}+\sqrt{d}\right)},$$

i.e., for a $(1-\eta/2)$-fraction of $u \in G$, the probability of sampling an ω such that the approximation from Equation (6) holds is lower bounded by the right-hand side of the above equation.

Next, we prepend V_+ with the mapping $|u\rangle|0\rangle|\vec{0}\rangle|0\rangle \mapsto |u\rangle\sum_{\omega\in\Omega}\sqrt{\mathbb{P}(\omega)}|\omega\rangle|\vec{0}\rangle|0\rangle$, which can be implemented with one quantum experiment. The combined operation performs the mapping $W_+ : |u\rangle|\vec{0}\rangle|0\rangle \mapsto |u\rangle|\psi_u\rangle$, with one call to V_+, where

$$\left\| |\psi_u\rangle - \sqrt{1-p_u}|\psi_u^0\rangle|0\rangle + \sqrt{p_u}|\psi_u^1\rangle|1\rangle \right\| \leq C_1\epsilon',$$

for some states $|\psi_u^0\rangle$ and $|\psi_u^1\rangle$, and such that for at least a $(1-\eta/2)$-fraction of $u \in G$,

$$\begin{aligned}
\left|p_u - \mathbb{E}\left[\frac{1}{2} + \frac{m'}{d}\langle u, X\rangle\right]\right| &= \left|\sum_{\omega\in\Omega}\mathbb{P}(\omega)p_{u,\omega} - \sum_{\omega\in\Omega}\mathbb{P}(\omega)\left(\frac{1}{2} + \frac{m'}{d}\langle u, X(\omega)\rangle\right)\right| \\
&\leq \sum_{\omega\in\Omega}\mathbb{P}(\omega)\left|p_{u,\omega} - \left(\frac{1}{2} + \frac{m'}{d}\langle u, X(\omega)\rangle\right)\right| \\
&\leq \sum_{\omega\in\Omega}\mathbb{P}(\omega)B_{u,\omega}\cdot\left[\frac{1}{2} + m'\right] + \sum_{\omega\in\Omega}\mathbb{P}(\omega)(1-B_{u,\omega})\left|\sqrt{p_{u,\omega}} - \sqrt{\frac{1}{2} + \frac{m'}{d}\langle u, X(\omega)\rangle}\right|\cdot 2 \\
&\leq \frac{(m')^2\epsilon^2\left(\frac{1}{2}+m'\right)}{72d^2m^2\left(\frac{1}{2}+\sqrt{d}\right)} + 2\cdot\frac{(m')^2\epsilon^2}{144d^2m^2} \leq \frac{(m')^2\epsilon^2}{36d^2m^2},
\end{aligned}$$

where in the last line we used that $m' \leq \sqrt{d}$. We conclude that for at least a $(1-\eta/2)$-fraction of $u \in G$,

$$\left|\sqrt{p_u} - \sqrt{\frac{1}{2} + \frac{m'}{d}\langle u, \mathbb{E}[X]\rangle}\right| \leq \sqrt{\left|p_u - \mathbb{E}\left[\frac{1}{2} + \frac{m'}{d}\langle u, X\rangle\right]\right|} \leq \frac{m'\epsilon}{6dm},$$

following a same argument as in the proof of the previous lemma.

Then, we convert the resulting operation W_+ back into a phase oracle, using Lemma 2.11 with precision $\epsilon/4$, and $t = dm/m'$. Then, the resulting operation performs $|u\rangle|0\rangle \mapsto |u\rangle|\chi_u\rangle$ with $C_2 = O(dm/m' + \log(1/\epsilon))$ calls to W_+, and we let K_2 be the constant suppressed by the big-O-notation. Then we find, by the triangle inequality, that for at least a $(1-\eta/2)$-fraction of $u \in G$,

$$\begin{aligned}
&\left\|\,|\chi_u\rangle - e^{i\frac{dm}{m'}\left(\frac{1}{2}+\frac{m'}{d}\langle u,\mathbb{E}[X]\rangle\right)}|\vec{0}\rangle\right\| \\
&\quad\leq C_1C_2\epsilon' + \left\||\chi_u\rangle - e^{i\frac{dm}{m'}p_u}|\vec{0}\rangle\right\| + \left|e^{i\frac{dm}{m'}p_u} - e^{i\frac{dm}{m'}\left(\frac{1}{2}+\frac{m'}{d}\langle u,\mathbb{E}[X]\rangle\right)}\right| \\
&\quad\leq \frac{\epsilon}{4} + \frac{\epsilon}{4} + \left|\frac{dm}{m'}p_u - \frac{dm}{m'}\left(\frac{1}{2} + \frac{m'}{d}\langle u, \mathbb{E}[X]\rangle\right)\right| \\
&\quad= \frac{\epsilon}{2} + \frac{dm}{m'}\left|\sqrt{p_u} - \sqrt{\frac{m'}{d}\langle u, \mathbb{E}[X]\rangle}\right|\cdot\left|\sqrt{p_u} + \sqrt{\frac{m'}{d}\langle u, \mathbb{E}[X]\rangle}\right| \\
&\quad\leq \frac{\epsilon}{2} + \frac{dm}{m'}\cdot\frac{m'\epsilon}{6dm}\cdot\left(2 + \frac{m'\epsilon}{6dm}\right) \leq \frac{\epsilon}{2} + \frac{\epsilon}{6}\cdot 3 = \epsilon,
\end{aligned}$$

where we used that $m' \leq \sqrt{d}$ and $m \geq \epsilon/(6\sqrt{d})$ in the last inequality.

Finally, we can remove the unnecessary global phase $dm/(2m')$ by applying some Z-rotation on any control qubit when we call the above operation in a controlled manner, which does not incur any additional overhead in terms of the number of queries or error. Thus, we have shown how to implement $\widetilde{P}_{X,m,\eta,\epsilon}$.

It remains to check how many quantum experiments and queries to $\mathcal{P}_X$ we have performed throughout its construction. Multiplying the complexities appearing earlier in this proof together results in

$$O\left(\left(\frac{dm}{m'} + \log\left(\frac{1}{\epsilon}\right)\right)\cdot\log\left(\frac{dm}{m'\epsilon}\right)\right) = \widetilde{O}\left(\sqrt{d}m\log^2\left(\frac{1}{\epsilon\eta}\right)\right)$$

quantum experiments, and

$$O\left(\left(\frac{dm}{m'} + \log\left(\frac{1}{\epsilon}\right)\right)\cdot\log\left(\frac{dm}{m'\epsilon}\right)\cdot\left(m' + \log\left(\frac{1}{\epsilon'}\right)\right)\cdot\log\left(\frac{m'}{\epsilon'}\right)\right),$$

queries to $\mathcal{P}_X$, which after substitution of m' and ϵ' can be upper bounded by

$$\widetilde{O}\left(dm\log^4\left(\frac{1}{\epsilon\eta}\right)\right).$$

This completes the proof. □

Now, we are ready to put everything together, and provide a full construction of a multivariate quantum mean estimator using phase oracles. The core idea is to use the bounded multivariate estimator from Algorithm 1, with slightly tweaked constants to accommodate for the slight differences in the guarantees we have on the precision of the directional means oracle. The full algorithm is presented in Algorithm 3.

1. Set $k = \left\lfloor \min\left\{n, \frac{n'}{\sqrt{d}}\right\}\right\rfloor$, $\eta = \frac{1}{288}$ and $m = 2^{\left\lceil \log\left(\frac{8\pi k}{\sqrt{d}\log(d/\delta)}\right)\right\rceil}$.
2. For $\ell = 1, \ldots, \lceil 18\log(d/\delta)\rceil$:
   (a) Compute the uniform superposition $|G\rangle := \frac{1}{m^{d/2}}\sum_{u\in G}|u\rangle$ over G.
   (b) Compute the state $|\psi\rangle := \widetilde{\mathcal{P}}_{X,m,\eta,\epsilon}|G\rangle|\vec{0}\rangle \in \mathcal{H}_G \otimes \mathcal{H}_{\text{aux}}$, where $\widetilde{\mathcal{P}}_{X,m,\eta,\epsilon}$ is the directional means oracle constructed in Lemma 4.2 with $\epsilon = 1/(12\sqrt{2})$.
   (c) Compute the state $|\phi\rangle := (\mathrm{QFT}_G^{-1} \otimes \mathbb{I}_{\text{aux}})|\psi\rangle$ where the unitary $\mathrm{QFT}_G : |u\rangle \mapsto \frac{1}{m^{d/2}}\sum_{v\in G} e^{2i\pi m\langle u,v\rangle}|v\rangle$ is the quantum Fourier transform over G.
   (d) Measure the $\mathcal{H}_G$ register of $|\phi\rangle$ in the computational basis and let $\widetilde{v}^{(\ell)} \in G$ denote the obtained result. Set $\widetilde{\mu}^{(\ell)} = 2\pi\widetilde{v}^{(\ell)}$.
3. Output the coordinate-wise median, $\widetilde{\mu} = \mathrm{median}(\widetilde{\mu}^{(1)}, \ldots, \widetilde{\mu}^{(\lceil 18\log(d/\delta)\rceil)})$.

Algorithm 3: Multivariate mean estimator with phase oracles, $\mathsf{QPhase}_d(X, n, n', \delta)$.

Theorem 4.3 (High-precision multivariate mean estimator with phase oracles)**.** *Let $d \in \mathbb{N}$, $\delta \in (0,1)$, $n \geq \log(d/\delta)$, $n' \geq \sqrt{d}\log(d/\delta)$, and X a random variable bounded by $[-1/4, 1/4]^d$, with $\mu = \mathbb{E}[X]$. Then the multivariate mean estimator with phase oracles, $\mathsf{QPhase}_d(X, n, \delta)$ (Algorithm 3), finds an approximation to the mean, $\widetilde{\mu}$, that with probability at least $1-\delta$ satisfies*

$$\|\widetilde{\mu} - \mu\|_\infty \leq \max\left\{\frac{\sqrt{d}}{n}, \frac{d}{n'}\right\} \cdot \log\left(\frac{d}{\delta}\right),$$

with $\widetilde{O}(n)$ calls to $U_{\mathbb{P}}$ and $\widetilde{O}(n')$ calls to $\mathcal{P}_X$.

Proof. We follow the general proof strategy from Theorem 3.3, and let

$$|\psi'\rangle = \frac{1}{\sqrt{m^d}}\sum_{u\in G} e^{im\langle u, \mathbb{E}[X]\rangle}|u\rangle|\vec{0}\rangle.$$

From the performance guarantee on $\widetilde{\mathcal{P}}_{X,m,\eta,\epsilon}$ that we proved in Lemma 4.2, we find that $\||\psi\rangle - |\psi'\rangle\|^2 \leq \sum_{u\in G}\||\varphi_u\rangle - e^{im\langle u,\mathbb{E}[X]\rangle}|\vec{0}\rangle\|^2/m \leq \eta + \epsilon^2 \leq 1/144$.

We now analyze the remainder of the algorithm as if the state $|\psi'\rangle$ was prepared, instead of $|\psi\rangle$. Let $\widetilde{v} \in G$ be the outcome of the measurement performed in step 2d of the algorithm. Since $\|\mathbb{E}[X]\|_\infty \leq 1/4 < 2\pi/3$, by the standard analysis of the phase estimation algorithm, as can for instance be found in Equation (5.34) in [NC11], for every $j \in [d]$ we have $|\widetilde{v}_j - \mathbb{E}[X]/(2\pi)| \leq 4/m$

with probability at least 5/6. If we now factor in that we start with $|\psi\rangle$ rather than $|\psi'\rangle$, the probability goes down from 5/6 to 2/3. Finally, from the Chernoff bound, it follows that $\|\widetilde{\mu} - \mathbb{E}[X]\|_\infty \leq 8\pi/m \leq \sqrt{d}\log(d/\delta)/k$ with probability at least $1-\delta$, from which the claim follows.

It remains to analyze the query complexity claims. We make $O(\log(d/\delta))$ calls to the directional means oracle from Lemma 4.2, from which we find that the number of quantum experiments is $\widetilde{O}(\sqrt{d}m\log^2(1/(\epsilon\eta))\log(d/\delta)) = \widetilde{O}(k) = \widetilde{O}(n)$, and similarly the number of calls to $\mathcal{P}_X$ is $\widetilde{O}(dm\log^4(1/(\epsilon\eta))\log(d/\delta)) = \widetilde{O}(k\sqrt{d}) = \widetilde{O}(n')$, completing the proof. □

Later on in this section, we find corresponding lower bounds on the precision that scale as $\Omega(\max\{\sqrt{d}/n, d/n'\})$, implying that the performance guarantee we obtain here is optimal up to polylogarithmic factors. However, this lower bound only holds in the regime where both $n \geq d$ and $n' \geq d$, and surprisingly it turns out that one can do better in the case where either $n < d$ or $n' < d$. We show this in the next section.

4.2 Near-optimal multivariate mean estimator in the low-precision regime

It turns out that the performance of Algorithm 3 is only optimal in the regime where $n \geq d$ and $n' \geq d$. In this section, we take a look at the regime where we have very few calls to the input oracles to spend, more specifically where $n < d$ or $n' < d$. We refer to this regime as the low-precision regime.

If $n' < d$, then the performance bound on Algorithm 3 becomes at least $d/n' > 1$. However, we know a priori that $\mathbb{E}[X]$ is contained in $[-1/4, 1/4]^d$, so if we just output the all-zeros vector, we will do better than what Theorem 4.3 suggests. Moreover, we will show in the next section that this is actually optimal, i.e., if one has less than d queries to $\mathcal{P}_X$ to spend, one might as well just output the all-zeros vector, since there is nothing one can do that will provably result in a significantly better estimate.

This leaves the regime where $n < d$ and $n' \geq d$, and it turns out that in this regime there is indeed a non-trivial approach that beats the complexity obtained by Algorithm 3. The modification is very simple – one just samples from the probability space n times, and then runs Algorithm 3 with the empirical distribution. The algorithm is presented in Algorithm 4, and the performance guarantees are presented in Theorem 4.4.

1. Set $k' = \lfloor 2n/\log(d/\delta) \rfloor$.
2. For $\ell = 1, \ldots, \lceil 32\log(d/\delta) \rceil$:
   (a) Obtain samples $\omega^{(1)}, \ldots, \omega^{(k')}$ from the probability space, and let $\overline{\mathbb{P}}$ be the empirical distribution based on the observed samples.
   (b) Run steps 2a to 2d of Algorithm 3, with $k = 2n'/\sqrt{d}$, all other parameters chosen identically, and the quantum experiment oracle $U_{\overline{\mathbb{P}}}$ constructed from the observed samples. Denote the outcome by $\widetilde{\mu}^{(\ell)}$.
3. Output the coordinate-wise median, $\widetilde{\mu} = \text{median}(\widetilde{\mu}^{(1)}, \ldots, \widetilde{\mu}^{(\lceil 18\log(d/\delta)\rceil)})$.

Algorithm 4: Low-precision multivariate mean estimator, $\mathsf{QLowPrecPhase}_d(X, n, n', \delta)$.

Theorem 4.4 (Low-precision analog mean estimator). *Let $d \in \mathbb{N}$, $\delta \in (0,1)$, $n \geq \log(d/\delta)$, $n' \geq \sqrt{d}\log(d/\delta)$, and X a random variable with values contained in $[-1/4, 1/4]^d$, with $\mu = \mathbb{E}[X]$. Then, $\mathsf{QLowPrecPhase}_d(X, n, n', \delta)$ (Algorithm 4) finds an approximation to the mean, $\widetilde{\mu}$, that*

with probability at least $1-\delta$ *satisfies*

$$\|\widetilde{\mu}-\mu\|_\infty \leq \max\left\{\frac{1}{\sqrt{n}}, \frac{d}{n'}\right\}\cdot \log\left(\frac{d}{\delta}\right),$$

with $\widetilde{O}(n)$ *calls to* $U_{\mathbb{P}}$*, and* $\widetilde{O}(n')$ *calls to* $\mathcal{P}_X$*.*

Proof. Since we only call $U_{\mathbb{P}}$ in step 2a, it is clear we perform a total of $O(k'\log(d/\delta)) = O(n)$ quantum experiments. Similarly, the number of calls to the phase oracle $\mathcal{P}_X$ is twice that in a run of Algorithm 3 with $n > n'/\sqrt{d}$, from which we readily deduce that it is indeed $\widetilde{O}(n')$.

It remains to check the precision guarantee. To that end, let

$$\overline{\mathbb{P}}_\omega = \frac{|\{j\in[k'] : \omega^{(j)} = \omega\}|}{k'},$$

and let $\overline{\mu} = \mathbb{E}_{\overline{\mathbb{P}}}[X]$, i.e., the mean of X under this empirical probability distribution. Note that $\overline{\mu}$ itself is also a random variable, since it depends on the ω's observed in the first stage of the algorithm. From the analysis in Theorem 4.3, we find that for all $j\in[d]$ and $\ell \in [\lceil 32\log(d/\delta)\rceil]$, $|\widetilde{\mu}_j - \overline{\mu}_j| \leq d\log(d/\delta)/(2n')$ with probability at least $2/3$. Thus, it remains to show that with high probability $\overline{\mu}$ approximates $\mu = \mathbb{E}[X]$ well, that is, it remains to show that for all $j\in[d]$, $|\overline{\mu}_j - \mu_j| \leq \log(d/\delta)/(2\sqrt{n})$ with high probability.

To that end, observe that for all $j\in[d]$,

$$\left|\mathbb{E}[\overline{\mu}_j] - \mu_j\right| = \left|\sum_{\omega\in\Omega}\overline{\mathbb{P}}_\omega X_j(\omega) - \mathbb{P}(\omega)X_j(\omega)\right| \leq \frac{1}{4}\left|\sum_{\omega\in\Omega}\overline{\mathbb{P}}_\omega - \mathbb{P}(\omega)\right|,$$

and also, using that $\mathbb{E}[\overline{\mathbb{P}}_\omega] = \mathbb{P}(\omega)$,

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{\omega\in\Omega}\overline{\mathbb{P}}_\omega - \mathbb{P}(\omega)\right)^2\right] &= \mathrm{Var}\left[\sum_{\omega\in\Omega}\overline{\mathbb{P}}_\omega - \mathbb{P}(\omega)\right] = \sum_{\omega\in\Omega}\mathrm{Var}\left[\overline{\mathbb{P}}_\omega\right] \\ &= \sum_{\omega\in\Omega}\frac{\mathbb{P}(\omega)(1-\mathbb{P}(\omega))}{k'} \leq \sum_{\omega\in\Omega}\frac{\mathbb{P}(\omega)}{k'} = \frac{1}{k'}.\end{aligned}$$

Therefore, by Markov's inequality, for all $j\in[d]$,

$$\mathbb{P}\left[\left|\overline{\mu}_j - \mu_j\right| > \frac{\log(d/\delta)}{2\sqrt{n}}\right] \leq \mathbb{P}\left[\left|\sum_{\omega\in\Omega}\overline{\mathbb{P}}_\omega - \mathbb{P}(\omega)\right| > \frac{4}{\sqrt{k'}}\right] = \mathbb{P}\left[\left|\sum_{\omega\in\Omega}\overline{\mathbb{P}}_\omega - \mathbb{P}(\omega)\right|^2 > \frac{16}{k'}\right] \leq \frac{1}{16}.$$

Thus, by the triangle inequality, for every $j\in[d]$ and $\ell\in[\lceil 32\log(d/\delta)\rceil]$, with probability at least $2/3\cdot 15/16 = 5/8$ we have that $|\widetilde{\mu}_j^{(\ell)} - \mu_j| \leq |\widetilde{\mu}_j^{(\ell)} - \overline{\mu}_j| + |\overline{\mu}_j - \mu_j| \leq (d/(2n') + 1/(2\sqrt{n}))\cdot\log(d/\delta) \leq \max\{1/\sqrt{n}, d/n'\}\cdot\log(d/\delta)$. Finally, it follows from the Chernoff bound that after $\lceil 32\log(d/\delta)\rceil$ iterations, we obtain that $\|\widetilde{\mu}-\mu\|_\infty \leq \max\{1/\sqrt{n}, d/n'\}\cdot\log(d/\delta)$, completing the proof. □

We have now described all algorithms. For convenience, we aggregate all algorithmic results in one self-contained statement.

Theorem 4.5. *Let* $d, n, n' \in \mathbb{N}$*,* $\delta\in(0,1)$*, and* X *a random variable with values contained in* $[-1/4, 1/4]^d$*, with* $\mu = \mathbb{E}[X]$*. Then, we can find an approximation to the mean,* $\widetilde{\mu}$*, that with probability at least* $1-\delta$ *satisfies*

$$\|\widetilde{\mu}-\mu\|_\infty \leq \begin{cases} 1, & \text{if } n' < d \text{ or } n < \log(d/\delta), \\ \max\left\{\frac{1}{\sqrt{n}}, \frac{d}{n'}\right\}\cdot\log\left(\frac{d}{\delta}\right), & \text{if } n'\geq d \text{ and } \log(d/\delta)\leq n < d, \\ \max\left\{\frac{\sqrt{d}}{n}, \frac{d}{n'}\right\}\cdot\log\left(\frac{d}{\delta}\right), & \text{if } n'\geq d \text{ and } n \geq d. \end{cases}$$

with $\widetilde{O}(n)$ calls to $U_{\mathbb{P}}$, and $\widetilde{O}(n')$ calls to $\mathcal{P}_X$. Furthermore, for all $p \in [1,\infty)$, we obtain the same performance guarantees on $\|\widetilde{\mu} - \mu\|_p$, but multiplied with $d^{1/p}$.

Proof. All statements are already present in Theorem 4.3 and Theorem 4.4, except for the case where $p \in [1,\infty)$, which follows from the norm inequality $\|\mathbf{x}\|_p \leq d^{1/p}\|\mathbf{x}\|_\infty$, for all $\mathbf{x} \in \mathbb{R}^d$. □

There exist at least three ways in which this result can be generalized. For instance, one can ask the question how many queries are required if instead of assuming that the random variable is bounded by $[-1/4, 1/4]^d$, it is instead bounded by some ℓ_q-ball of radius $1/4$, with $q \in [1,\infty)$. One can also wonder how many queries are required when one wants to obtain an approximation of $O\mathbb{E}[X]$, where O is some d-dimensional rotation matrix, i.e., a matrix that satisfies $O^T O = I$. Finally, if one has access to X through some other oracle than a phase oracle, then one can also wonder how many queries to such an oracle are required to solve the multivariate mean estimation problem. These questions are all addressed in [CJ21], albeit only in the high-precision regime. It is an interesting direction for further research to tightly characterize the query complexities of these problems in the low-precision setting as well.

4.3 Lower bounds

We now turn our attention to proving lower bounds on the number of queries required to the input oracles. We first focus our attention on the high-precision regime, and will show later on that the lower bounds in the low-precision regimes follow via some reduction from those that we prove in the high-precision regime.

As is customary with lower bounding, we would like to embed a problem whose hardness has already been shown before in the setting we consider here, in order to conclude that this problem must be at least as hard to solve. We start by considering the problem of recovering a constant fraction of the bits in a bit string when we are given access to it by means of a fractional phase oracle.

Lemma 4.6. *Let $\epsilon \in (0,\pi]$, $d \in \mathbb{N}$, and suppose that we have access to a bit string $\mathbf{b} \in \{0,1\}^d$ through controlled calls to a fractional phase oracle $\mathcal{F}_\epsilon : |j\rangle \mapsto e^{i\epsilon b_j}|j\rangle$. Then, in order to find a bit string $\widetilde{\mathbf{b}} \in \{0,1\}^d$ such that $\|\widetilde{\mathbf{b}} - \mathbf{b}\|_1 \leq d/4$ with probability at least $2/3$, we must make at least $\Omega(d/\epsilon)$ calls to $\mathcal{F}_\epsilon$.*

Proof. First, we argue that it is sufficient to consider the case where $\epsilon = \pi$. Indeed, in general, the query complexity of any problem is increased by a multiplicative factor of $\Theta(1/\epsilon)$, when one changes the input model from a regular phase oracle $\mathcal{F}_\pi$ to a fractional phase oracle $\mathcal{F}_\epsilon$. In Appendix B of [LMR+11], this is proven for problems that can be phrased as a binary function. However, since the problem we consider here does not have a unique correct output on every given input, we must combine their technique with the general adversary bound for relations, as derived by [Bel15], to arrive at the desired result. More details can be found in [CJ21].

Thus, it remains to focus on the case where $\epsilon = \pi$. Suppose that we have an algorithm $\mathcal{A}$ that finds a bit string $\widetilde{\mathbf{b}} \in \{0,1\}^d$ such that with probability at least $2/3$, we have $\|\widetilde{\mathbf{b}} - \mathbf{b}\|_1 \leq d/4$, i.e., $\widetilde{\mathbf{b}}$ and $\mathbf{b}$ differ in at most $d/4$ bits. Then, we can let $\mathcal{B}$ be the quantum algorithm that first runs $\mathcal{A}$ to obtain such a bit string $\widetilde{\mathbf{b}}$, and then selects uniformly at random a bit string $\overline{\mathbf{b}} \in \{0,1\}^d$ that satisfies $\|\overline{\mathbf{b}} - \widetilde{\mathbf{b}}\|_1 \leq d/4$. We have $M = \sum_{t=0}^{\lfloor d/4 \rfloor} \binom{d}{t}$ possible choices, which implies that the probability of this algorithm outputting $\mathbf{b}$ exactly is lower bounded by $2/3 \cdot 1/M$. By the information theoretic lower bound, i.e., Equation (4) in [FGGS99], the number of queries to $\mathcal{F}_\pi$, performed by $\mathcal{B}$ and hence also by $\mathcal{A}$, denoted by Q, satisfies

$$2^d \leq \frac{3}{2} \cdot \sum_{t=0}^{\lfloor \frac{d}{4} \rfloor} \binom{d}{t} \sum_{t=0}^{Q} \binom{d}{t} \leq \frac{3}{2} \cdot 2^{d\left(H\left(\frac{1}{4}\right) + H\left(\frac{Q}{d}\right)\right)},$$

where in the final inequality, we used a well-known approximation sums of binomial coefficients, as proven for instance in Lemma 16.19 in [FG06], and $H(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function. Taking logarithms on both sides yields that $H(Q/d) \geq 1 - H(1/4) + o(1)$, which implies that $Q = \Omega(d)$, completing the proof. □

The hardness of this problem can be used as a black box to show the high-precision lower bound on the precision we can attain, expressed in the number of calls to $\mathcal{P}_X$, as is shown in the theorem below.

Theorem 4.7. *Let $d \in \mathbb{N}$, $n' \geq d$, and suppose that we have a quantum algorithm that finds an approximation $\widetilde{\mu}$ of the mean μ of any random variable with values contained in $[-1/4, 1/4]^d$, using n' queries to $\mathcal{P}_X$. Then, there exist instances in which case the error between $\widetilde{\mu}$ and μ satisfies*

$$\|\widetilde{\mu} - \mu\|_1 = \Omega\left(\frac{d^2}{n'}\right),$$

with probability at least $2/3$.

Proof. Let $\epsilon = d/n'$, and $\mathbf{b} \in \{0,1\}^d$ a bit string, that we can access through controlled calls to a fractional phase oracle $\mathcal{F}_\epsilon : |j\rangle \mapsto e^{i\epsilon b_j}|j\rangle$. Then, we know from Lemma 4.6 that it takes $\Omega(d/\epsilon) = \Omega(n')$ calls to $\mathcal{F}_\epsilon$ to find a bit string $\widetilde{\mathbf{b}} \in \{0,1\}^d$ such that $\|\widetilde{\mathbf{b}} - \mathbf{b}\|_1 \leq d/4$.

Now, let $\Omega = \{0\}$, with $\mathbb{P}(0) = 1$, which implies that $U_{\mathbb{P}} = I$. Let the random variable $X : \Omega \to \mathbb{R}^d$ be defined as $X(0)_j = \epsilon b_j$. This implies that $\mathcal{P}_X : |0\rangle|j\rangle \mapsto e^{i\epsilon b_j}|0\rangle|j\rangle$, and hence $\mathcal{P}_X$ can be implemented with one call to $\mathcal{F}_\epsilon$. Furthermore, we have $\mu = \mathbb{E}[X] = \epsilon\mathbf{b}$, and hence if we can find an approximation $\widetilde{\mu}$ to the mean satisfying $\|\widetilde{\mu} - \mu\| \leq d^2/(8n')$, then we have

$$\min_{\widetilde{\mathbf{b}} \in \{0,1\}^d} \left\| \frac{\widetilde{\mu}}{\epsilon} - \widetilde{\mathbf{b}} \right\|_1 = \frac{1}{\epsilon} \min_{\widetilde{\mathbf{b}} \in \{0,1\}^d} \|\widetilde{\mu} - \epsilon\widetilde{\mathbf{b}}\|_1 \leq \frac{1}{\epsilon}\|\widetilde{\mu} - \mu\|_1,$$

and hence, if we take $\widetilde{\mathbf{b}}$ to be the bit string attaining the minimum in the left-hand side, we find by the triangle inequality

$$\|\widetilde{\mathbf{b}} - \mathbf{b}\|_1 \leq \left\| \frac{\widetilde{\mu}}{\epsilon} - \widetilde{\mathbf{b}} \right\|_1 + \left\| \frac{\widetilde{\mu}}{\epsilon} - \mathbf{b} \right\|_1 \leq \frac{1}{\epsilon}\|\widetilde{\mu} - \mu\|_1 + \frac{1}{\epsilon}\|\widetilde{\mu} - \mu\|_1 = \frac{2}{\epsilon}\|\widetilde{\mu} - \mu\|_1 \leq \frac{d}{4}.$$

But we know that this takes at least $\Omega(n')$ calls to $\mathcal{F}_\epsilon$, and hence obtaining an estimate $\widetilde{\mu}$ that satisfies $\|\widetilde{\mu} - \mu\|_1 \leq d^2/(8n')$ requires at least $\Omega(n')$ queries to $\mathcal{P}_X$ as well. Thus, if we only have n' queries to spend, there must exist instances such that $\|\widetilde{\mu} - \mu\|_1 = \Omega(d^2/n')$ with high probability, completing the proof. □

In order to give a similar lower bound in terms of the number of quantum experiments, we need to subtly change the problem to finding a vector $\widetilde{\mathbf{b}} \in \{0,1\}^d$ such that $\|H(\widetilde{\mathbf{b}} - \mathbf{b})\|_1 \leq d/4$, where H is a d-dimensional normalized Hadamard matrix, i.e., the entries of H are all $\pm 1/\sqrt{d}$, and $H^T H = I$. This problem is quite different in nature compared to the problem considered in Lemma 4.6, since the number of bits in which $\mathbf{b}$ and $\widetilde{\mathbf{b}}$ differ seems to no longer tells us anything useful about whether this condition is met. For instance, if $\mathbf{b} = \mathbf{0}$ and $\widetilde{\mathbf{b}} = \mathbf{1}$, i.e., they differ in *all* bits, then $H(\widetilde{\mathbf{b}} - \mathbf{b}) = \sqrt{d}\mathbf{e}_1$, and so the condition is met as long as d is large enough. Suprisingly, however, we are able to show that this problem is equally hard as the non-rotated problem up to constants, i.e., it still takes $\Omega(d/\epsilon)$ calls to $\mathcal{F}_\epsilon$ to find a $\widetilde{\mathbf{b}}$ that satisfies this rotated condition. The details can be found in the lemma below.

Lemma 4.8. *Let $\epsilon \in (0, \pi]$, $d \in \mathbb{N}$ a power of 2, and let H be a d-dimensional Hadamard matrix, i.e., all entries of H are $\pm 1/\sqrt{d}$, and $H^T H = I$. Suppose we have access to a bit string $\mathbf{b} \in \{0,1\}^d$ by means of a fractional phase oracle $\mathcal{F}_\epsilon : |j\rangle \mapsto e^{i\epsilon b_j}|j\rangle$. In order to find a bit string $\widetilde{\mathbf{b}} \in \{0,1\}^d$ such that $\|H(\widetilde{\mathbf{b}} - \mathbf{b})\|_1$, the number of calls to $\mathcal{F}_\epsilon$ satisfies $\Omega(d/\epsilon)$.*

Proof. Similarly as in the proof of Lemma 4.6, it suffices to consider the case where $\epsilon = \pi$. Suppose that we have a quantum algorithm $\mathcal{A}$ that finds a bit string $\widetilde{\mathbf{b}} \in \{0,1\}^d$ satisfying the condition posed in the statement of the lemma with probability at least $2/3$. Using heavy-duty tools from statistics, it is shown in Appendix B of [CJ21] that

$$\Pr_{\overline{\mathbf{b}} \sim \{0,1\}^d} \left[\|H(\widetilde{\mathbf{b}} - \overline{\mathbf{b}})\|_1 \leq \frac{d}{4} \right] \leq 2^{-Cd}, \qquad \text{with} \qquad C = \frac{\log(e)}{2} \left(\frac{1}{2\sqrt{2}} - \frac{1}{4} \right)^2 \in (0,1).$$

Thus, if know $\widetilde{\mathbf{b}} \in \{0,1\}^d$ and that $\|H(\widetilde{\mathbf{b}} - \mathbf{b})\|_1 \leq d/4$, then there are less than $2^{(1-C)d}$ possible choices left for $\mathbf{b}$. Thus, the algorithm $\mathcal{B}$, that first runs $\mathcal{A}$ to obtain a vector $\widetilde{\mathbf{b}}$ such that $\|H(\widetilde{\mathbf{b}} - \mathbf{b})\|_1 \leq d/4$, and subsequently takes any $\overline{\mathbf{b}}$ that satisfies $\|H(\widetilde{\mathbf{b}} - \overline{\mathbf{b}})\|_1 \leq d/4$ uniformly at random, will recover $\mathbf{b}$ exactly with probability lower bounded by $2/3 \cdot 2^{(C-1)d}$. Analogously as in the proof of Lemma 4.6, this implies that

$$2^d \leq \frac{3}{2} \cdot 2^{(1-C)d} \cdot \sum_{t=0}^{Q} \binom{d}{t} \leq \frac{3}{2} \cdot 2^{(1-C)d + dH\left(\frac{Q}{d}\right)},$$

and taking the logarithm on both sides implies that $H(Q/d) \geq C + o(1)$, which in turn implies that $Q = \Omega(d)$, completing the proof. □

We now show how the hardness of problem considered in the previous lemma can be used to lower bound the query complexity in the mean estimation problem.

Theorem 4.9. *Let $d \in \mathbb{N}$, $n \geq d$, and suppose that we have a quantum algorithm that finds an approximation $\widetilde{\mu}$ to the mean μ of any random variable with values contained in $[-1/4, 1/4]^d$, using n queries to $U_{\mathbb{P}}$. Then, there exist instances in which case the error between $\widetilde{\mu}$ and μ satisfies*

$$\|\widetilde{\mu} - \mu\|_1 = \Omega\left(\frac{d^{\frac{3}{2}}}{n} \right),$$

with probability at least $2/3$.

Proof. Let d' be the biggest power of 2 below or equal to d. Let $\epsilon = d'/n$, $\epsilon' = \arcsin(\epsilon)$, and suppose that we have access to some hidden bit string $\mathbf{b} \in \{0,1\}^{d'}$ by means of controlled calls to a fractional phase oracle $\mathcal{F}_{\epsilon'} : |j\rangle \mapsto e^{i\epsilon' b_j} |j\rangle$. We know from Lemma 4.8 that it takes $\Omega(d'/\epsilon') = \Omega(n)$ calls to find a bit string $\widetilde{\mathbf{b}}$ such that $\|H(\widetilde{\mathbf{b}} - \mathbf{b})\|_1 \leq d'/4$.

Now, let $\Omega = [d'] \times \{0,1\}$, and for every $\mathbf{b} \in \{0,1\}^{d'}$, let the probability measure $\mathbb{P}_{\mathbf{b}}$ on Ω be defined as

$$\mathbb{P}_{\mathbf{b}}(j,x) = \frac{1}{d'} \cos^2\left(\frac{\pi}{4} + (-1)^x \frac{\epsilon' b_j}{2} \right), \qquad \text{for all } j \in [d'], x \in \{0,1\}.$$

Observe that with one call to $\mathcal{F}_{\epsilon'}$, we can implement

$$\begin{aligned}
&\frac{1}{\sqrt{2d'}} \sum_{j=1}^{d'} (|j\rangle|0\rangle + i|j\rangle|1\rangle) \overset{\mathcal{F}_{\epsilon'}}{\mapsto} \frac{1}{\sqrt{2d'}} \sum_{j=1}^{d'} (|j\rangle|0\rangle + ie^{i\epsilon' b_j}|j\rangle|1\rangle) \\
&= \sum_{j=1}^{d'} \frac{e^{i\frac{\pi}{4} + i\frac{\epsilon' b_j}{2}}}{\sqrt{2d'}} \left(e^{-i\frac{\pi}{4} - i\frac{\epsilon' b_j}{2}} |j\rangle|0\rangle + e^{i\frac{\pi}{4} + i\frac{\epsilon' b_j}{2}} |j\rangle|1\rangle \right) \\
&\overset{I \otimes (SH)}{\mapsto} \sum_{j=1}^{d'} \frac{e^{i\frac{\pi}{4} + i\frac{\epsilon' b_j}{2}}}{\sqrt{d'}} \left(\cos\left(\frac{\pi}{4} + \frac{\epsilon' b_j}{2} \right) |j\rangle|0\rangle + \sin\left(\frac{\pi}{4} + \frac{\epsilon' b_j}{2} \right) |j\rangle|1\rangle \right) \\
&= \sum_{(j,x) \in \Omega} \frac{e^{i\frac{\pi}{4} + i\frac{\epsilon' b_j}{2}}}{\sqrt{d'}} \cos\left(\frac{\pi}{4} + (-1)^x \frac{\epsilon' b_j}{2} \right) |j\rangle|x\rangle = \sum_{(j,x) \in \Omega} \sqrt{\mathbb{P}_{\mathbf{b}}(j,x)} e^{i\frac{\pi}{4} + i\frac{\epsilon' b_j}{2}} |j\rangle|x\rangle,
\end{aligned}$$

and hence we can implement $U_{\mathbb{P}_{\mathbf{b}}}$ with one call to $\mathcal{F}_{\epsilon'}$.[1]

Next, let the random variable $X : \Omega \to \mathbb{R}^d$ be defined as

$$X(j,x) = \frac{x\sqrt{d'}}{4} H\mathbf{e}_j,$$

where H is a $(d' \times d')$-dimensional normalized Hadamard matrix, i.e., $H^T H = I$, whose first row and column have all positive signs. Such a Hadamard matrix exists because we know that d' is a power of 2. Furthermore,

$$\mu = \mathbb{E}[X] = \frac{1}{d'} \sum_{j=1}^{d'} \cos^2\left(\frac{\pi}{4} - \frac{\epsilon' b_j}{2}\right) \frac{\sqrt{d'}}{4} H\mathbf{e}_j = \frac{1}{8}\mathbf{e}_1 + \frac{\sin(\epsilon')}{8\sqrt{d'}} H\mathbf{b} = \frac{1}{8}\mathbf{e}_1 + \frac{\sqrt{d'}}{8n} H\mathbf{b},$$

Thus, if we find an approximation $\widetilde{\mu}$ to μ such that $\|H(\widetilde{\mu} - \mu)\|_1 \leq (d')^{3/2}/(16n)$, then

$$\min_{\widetilde{\mathbf{b}} \in \{0,1\}^d} \left\| \frac{8n}{\sqrt{d'}} \left(\widetilde{\mu} - \frac{1}{8}\mathbf{e}_1\right) - H\widetilde{\mathbf{b}} \right\|_1 = \frac{8n}{\sqrt{d'}} \min_{\widetilde{\mathbf{b}} \in \{0,1\}^d} \left\| \widetilde{\mu} - \frac{1}{8}\mathbf{e}_1 - \frac{\sqrt{d'}}{8n} H\widetilde{\mathbf{b}} \right\|_1 \leq \frac{8n}{\sqrt{d'}} \|H(\widetilde{\mu} - \mu)\|_1,$$

and hence if we let $\widetilde{\mathbf{b}}$ be the bit string for which the minimum in the above expression is attained, then we find that

$$\begin{aligned} \|H(\widetilde{\mathbf{b}} - \mathbf{b})\|_1 &\leq \left\| \frac{8n}{\sqrt{d'}} \left(\widetilde{\mu} - \frac{1}{8}\mathbf{e}_1\right) - H\widetilde{\mathbf{b}} \right\|_1 + \left\| \frac{8n}{\sqrt{d'}} \left(\widetilde{\mu} - \frac{1}{8}\mathbf{e}_1\right) - H\mathbf{b} \right\|_1 \\ &\leq \frac{8n}{\sqrt{d'}} \|H(\widetilde{\mu} - \mu)\|_1 + \frac{8n}{\sqrt{d'}} \|H(\widetilde{\mu} - \mu)\|_1 = \frac{4n}{\sqrt{d'}} \|H(\widetilde{\mu} - \mu)\|_1 \leq \frac{d'}{4}. \end{aligned}$$

We know that constructing such a bit string $\widetilde{\mathbf{b}}$ requires $\Omega(n)$ queries to $\mathcal{F}_{\epsilon'}$, and hence we find that in order to find an $(d')^{3/2}/(16n)$-precise ℓ_1-approximation of the mean of a random variable, we need to make at least $\Omega(n)$ calls to $U_{\mathbb{P}}$ as well. Thus, if we have only n queries to spend, there must be an instance for which the ℓ_1-approximation satisfies $\Omega(d^{3/2}/n)$. This completes the proof. □

We can now use the results obtained in Theorem 4.7 and Theorem 4.9 as black boxes to obtain results in different norms and regimes.

Theorem 4.10. *Let $d \in \mathbb{N}$, $n, n' \geq 1$, and suppose that we have a quantum algorithm that finds an approximation $\widetilde{\mu}$ to the mean μ of any random variable with values contained in $[-1/4, 1/4]^d$, using n queries to $U_{\mathbb{P}}$ and n' queries to $\mathcal{P}_X$. Then, there exist instances such that*

$$\|\widetilde{\mu} - \mu\|_1 = \begin{cases} \Omega\left(d\right), & \text{if } n' < d, \\ \Omega\left(\max\left\{\frac{d}{\sqrt{n}}, \frac{d^2}{n'}\right\}\right), & \text{if } n' \geq d \text{ and } n < d, \\ \Omega\left(\max\left\{\frac{d^{\frac{3}{2}}}{n}, \frac{d^2}{n'}\right\}\right), & \text{if } n' \geq d \text{ and } n \geq d, \end{cases}$$

with probability at least $2/3$. Moreover, the same expressions multiplied by $d^{1/p-1}$ can be obtained as lower bounds for $\|\widetilde{\mu} - \mu\|_p$, for all $p \in (1, \infty]$.

[1] Note that we don't have to worry about the extra global phase here – we can absorb it in the definition of the state $|\omega\rangle$, i.e., if $\omega = (j,x)$ we can define $|\omega\rangle = e^{i\pi/4 + i\epsilon' b_j/2}|j\rangle|x\rangle$, and then use all the machinery from the rest of this document.

Proof. If $n, n' \geq d$, then we know from Theorem 4.7 and Theorem 4.9 that

$$\|\widetilde{\mu} - \mu\|_1 = \Omega\left(\max\left\{\frac{d^{\frac{3}{2}}}{n}, \frac{d^2}{n'}\right\}\right).$$

Next, let $1 \leq n < d$, and let $k = \lfloor d/n \rfloor$. Let $X' : \Omega \to \mathbb{R}^n$ be a random variable, and define $X : \Omega \to \mathbb{R}^d$ by $X(\omega) = X'(\omega) \otimes \mathbb{1}_k$, padded with an appropriate number of zeros in the final entries. We find that

$$\frac{1}{k}\sum_{\ell=1}^{k} \mu_{k(j-1)+\ell} = \frac{1}{k} \cdot k \cdot \mu_j = \mu_j,$$

and hence, if we find an approximation of $\widetilde{\mu} \in \mathbb{R}^d$ to $\mu = \mathbb{E}[X]$, then we can define $\widetilde{\mu}' \in \mathbb{R}^n$ as $\widetilde{\mu}'_j = \sum_{\ell=1}^{k} \widetilde{\mu}_{k(j-1)+\ell}/k$, which implies

$$\begin{aligned}\|\widetilde{\mu}' - \mu'\|_1 &= \sum_{j=1}^{n}\left|\frac{1}{k}\sum_{\ell=1}^{k}\widetilde{\mu}_{k(j-1)+\ell} - \frac{1}{k}\sum_{\ell=1}^{k}\mu_{k(j-1)+\ell}\right| \leq \frac{1}{k}\sum_{j=1}^{n}\sum_{\ell=1}^{k}\left|\widetilde{\mu}_{k(j-1)+\ell} - \mu_{k(j-1)+\ell}\right| \\ &= \frac{1}{k}\|\widetilde{\mu} - \mu\|_1,\end{aligned}$$

and hence

$$\|\widetilde{\mu} - \mu\|_1 \geq k\|\widetilde{\mu}' - \mu'\|_1 = \Omega\left(\frac{d}{n} \cdot \frac{n^{\frac{3}{2}}}{n}\right) = \Omega\left(\frac{d}{\sqrt{n}}\right).$$

Finally, the result for different values for $p \in (1, \infty]$ follows directly from Hölder's inequality, since for all $\mathbf{x} \in \mathbb{R}^d$, we have $\|\mathbf{x}\|_p \geq d^{1/p-1}\|\mathbf{x}\|_1$. This completes the proof. □

We aggregate all our results in Figure 1.

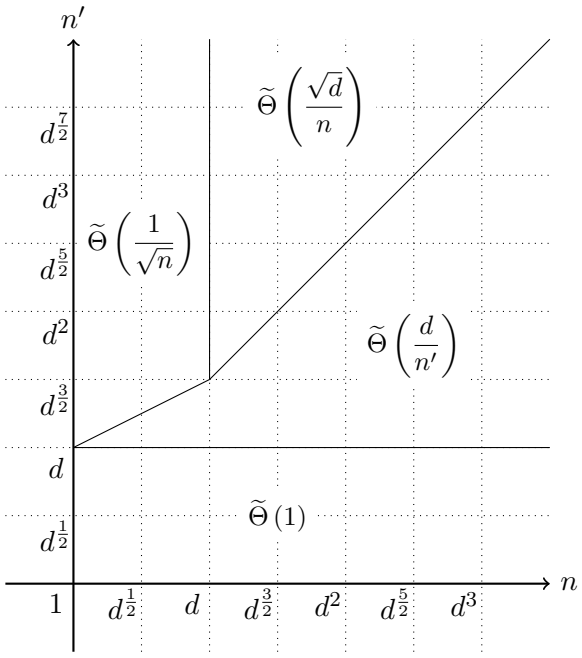

Figure 1: Overview of the different regimes of the mean estimation problem. The horizontal and vertical axes show n and n', i.e., the number of queries to $U_{\mathbb{P}}$ and $\mathcal{P}_X$, respectively. The complexities shown in the figure are the optimal error scaling of $\|\widetilde{\mu} - \mu\|_\infty$ that can be achieved with particular choices for n and n'. The tildes are hiding polylogarithmic factors in n, n', d and $1/\delta$. For the optimal error scalings of $\|\widetilde{\mu} - \mu\|_p$ for $p \in [1, \infty)$, one can simply multiply the expressions above by $d^{1/p}$.

5 Applications

In this section, we describe some applications of our results. We first explain how our formulation of the multivariate mean estimation problem covers the general task of estimating the expectation values of several mutually commuting observables with respect to a given quantum state. We then present several applications in the literature, and notably in quantum machine learning, where this problem arises.

5.1 Estimating expectation values of commuting observables

5.1.1 Classical versus quantum experiments

From a classical perspective, the mean estimation problem is commonly described by a random experiment (or Monte Carlo process) that draws a *classical* sample ω (e.g., a bit-string) from a certain probability space $(\Omega, 2^\Omega, \mathbb{P})$ and leads to the associated observation $X(\omega) \in \mathbb{R}^d$ (see Definition 2.4). It is then quite clear that our generalization to quantum experiments, defined by a unitary $U_\mathbb{P}$ that prepares a superposition over basis states $|\omega\rangle$, $\omega \in \Omega$, and a unitary $\mathcal{B}_X$ that evaluates X for the same basis states (see Definitions 2.5 and 2.6) can simulate such a random experiment. However, from a physical perspective, quantum experiments are also more general. The basis states $\{|\omega\rangle\}_{\omega\in\Omega}$ do not need to be computational basis states, and can be themselves superpositions of computational basis states or include arbitrary phases. Therefore, in our definition of quantum experiments, the unitaries $U_\mathbb{P}$ can indeed be arbitrary unitaries acting on a given Hilbert space $\mathcal{H}$, and what really matters here is the definition of the basis $\{|\omega\rangle\}_{\omega\in\Omega}$ of $\mathcal{H}$.

5.1.2 Problem definition

With this observation, we can now move to the problem of estimating expectation values of mutually commuting observables. Let U be a unitary transformation that prepares a given quantum state $|\psi\rangle = U|0\rangle$ in a given m-qubit Hilbert space $\mathcal{H}$, and let $O_1, \ldots, O_d$ be d *mutually commuting* observables (i.e., Hermitian operators) acting on $\mathcal{H}$. We want to compute estimates of the d expectation values $\langle O_i\rangle = \langle\psi|O_i|\psi\rangle$. Since the observables commute, they all share a common eigenbasis $\{|\phi_j\rangle\}_{1\leq j\leq 2^m}$, to which they assign eigenvalues $\lambda_i = (\lambda_{i,1}, \ldots, \lambda_{i,2^m}) \in \mathbb{R}^{2^m}$, respectively. Let us therefore look at the expression of $|\psi\rangle = U|0\rangle$ in this eigenbasis:

$$U : |0\rangle \mapsto \sum_{j=1}^{2^m} \sqrt{\mathbb{P}(\phi_j)} e^{i\varphi_j} |\phi_j\rangle \tag{7}$$

for some phases $\varphi_j \in [0, 2\pi]$ and real amplitudes $\sqrt{\mathbb{P}(\phi_j)}$ such that $\sum_{j=1}^{2^m} \mathbb{P}(\phi_j) = 1$. If we now take $U_\mathbb{P}$ to be the unitary U, $\{|\omega\rangle\}_{\omega\in\Omega}$ to be $\left\{e^{i\varphi_j}|\phi_j\rangle\right\}_{1\leq j\leq 2^m}$, and $X(\omega)$ to be $X(\phi_j) = (\lambda_{1,j}, \ldots, \lambda_{d,j})$ (i.e., the eigenvalues $\lambda_{i,j}$ assigned by each of the observables O_i to $|\phi_j\rangle$), the problem of estimating the d expectation values $\langle O_i\rangle = \langle\psi|O_i|\psi\rangle$ fits our formulation of the (quantum) mean estimation problem. Indeed, note that the phases $e^{i\varphi_j}$ do not contribute to the expectation values $\langle O_i\rangle = \langle\psi|O_i|\psi\rangle = \sum_{j=1}^{2^m} \mathbb{P}(\phi_j)\lambda_{i,j}$, and therefore absorbing them in the basis states $|\omega\rangle$ does not influence the mean values to be estimated (nor our algorithms).

5.1.3 Applicability assumptions

For the applicability of our algorithms to this problem, we assume that a description of the eigenbasis $\{|\phi_j\rangle\}_{1\leq j\leq 2^m}$, in terms of a unitary transformation $V : |j\rangle \mapsto |\phi_j\rangle$ from computational basis states $|j\rangle$ to eigenvectors $|\phi_j\rangle$, and the eigenvalues $\{\lambda_i\}_{1\leq i\leq d}$ of the observables $\{O_i\}_{1\leq i\leq d}$ are known. These are the same assumptions that one would have in quantum algorithms for the univariate version of this problem (i.e., with one observable) [KOS07; WCNA09] or in a

setting where one would *directly* estimate the expectation values $\langle O_i \rangle = \langle \psi | O_i | \psi \rangle = \langle \psi | V \Lambda_i V^\dagger | \psi \rangle$ for $\Lambda_i = \text{diag}(\lambda_{i,1}, \ldots, \lambda_{i,2^m})$ by applying $V^\dagger$ on $|\psi\rangle$, measuring computational basis states $|j\rangle$, and using several measurement outcomes $\{|j\rangle, (\lambda_{1,j}, \ldots, \lambda_{d,j})\}$ to simultaneously[2] compute these estimates. Note that, in practice, the same transformation $V^\dagger$ would be absorbed in the unitary $U_{\mathbb{P}}$ used in our algorithms, as to make the basis states $\{|\omega\rangle\}_{\omega \in \Omega}$ computational basis states, and ease the implementation of the unitaries $\mathcal{B}_X$ and $\mathcal{P}_X$ (using single-qubit rotations controlled by computational basis states).

5.2 Examples of applications

5.2.1 Training variational quantum circuits

A straightforward application appears in some variational quantum algorithms for machine learning [BLSF19]. In a multidimensional regression setting [MNKF18] or a reinforcement learning setting [JGM+21; SJD21], a variational quantum circuit defined by a parametrized and data-dependent unitary $U(x, \theta)$ and a set of observables $(O_1, \ldots, O_d)$ can be used as a hypothesis family $f_\theta(x) = (\langle O_1 \rangle_{x,\theta}, \ldots, \langle O_d \rangle_{x,\theta})$, for $\langle O_i \rangle_{x,\theta} = \langle 0^{\otimes n} | U^\dagger(x, \theta) O_i U(x, \theta) | 0^{\otimes n} \rangle$, to model target functions g with d-dimensional outputs. When the observables $O_1, \ldots, O_d$ all commute (e.g., commuting tensor products of Pauli operators or projectors on some basis states, for an arbitrary basis), the problem of estimating $f_\theta(x)$ fits the problem definition above.

5.2.2 Training Boltzmann machines

Another application considers the problem of estimating updates of a Boltzmann machine in a machine learning setting (e.g., a classification or generative modeling problem) [WKS16; WW19; KW17; JTN+21]. Take for instance a Boltzmann machine defined by a Hamiltonian:

$$H = \sum_{i<j} J_{i,j} \sigma_i^z \sigma_j^z + \sum_i b_i \sigma_i^z \tag{8}$$

where $J_{i,j}$ and b_i are real weights and biases and σ_i^z is a Pauli-Z operator acting on a qubit i out of n total qubits. The updates on the weights and biases of this Boltzmann machine take the form:

$$\Delta J_{i,j} = -\mathcal{L}(J, b) \langle \sigma_i^z \sigma_j^z \rangle, \quad \Delta b_i = -\mathcal{L}(J, b) \langle \sigma_i^z \rangle \tag{9}$$

where $\mathcal{L}(J, b)$ is a loss dependent on the Boltzmann machine performance at the machine learning task and the expectation values $\langle \sigma_i^z \rangle, \langle \sigma_i^z \sigma_z^j \rangle$ are with respect to the Gibbs state:

$$|\psi\rangle = \frac{1}{\text{Tr}_x[e^{-H}]} \sum_x \sqrt{e^{-H(x)}} |x\rangle \tag{10}$$

for computational basis states $|x\rangle$. Assume having access to a unitary U that prepares the Gibbs state of Equation (10), e.g., using one of the subroutines in [WKS16; WW19; KW17; JTN+21], then estimating the updates of the Boltzmann machine is an instance of the problem above for observables $\left\{ -\mathcal{L}(J, b) \sigma_i^z \sigma_j^z,\ -\mathcal{L}(J, b) \sigma_i^z \right\}_{i,j}$, i.e., weighted σ_i^z and $\sigma_i^z \sigma_j^z$ operators, which are all diagonal in the computational basis.

5.2.3 Training policies in reinforcement learning

In the context of reinforcement learning [SB18], an agent-environment interaction is described by a Monte Carlo process where, for a sequence of interactions, an agent acts probabilistically on its environment, the latter updates its state (probabilistically) depending on the actions of

[2] When the observables do not commute, one cannot "parallelize" measurements in such a manner, and would then be required to use different techniques like shadow tomography [Aar20; HKP20].

the agent and issues a real-valued reward r_t. The goal of the agent is to find a policy (i.e., a probability distribution $\pi(a_t|s_t)$ of actions a_t given states s_t) that maximizes its *expected* rewards $V(\pi) = \sum_{t=1}^{T} r_t$ for T interactions with the environment. To do this, policy-based reinforcement learning algorithms define a certain family of parametrized policies $\pi_\theta \in \Pi_\theta$ (e.g., deep neural networks) and explore this policy family using gradient ascent on the expected rewards $V(\pi_\theta)$. The so-called policy gradient theorem [SMSM99] gives a formulation of the gradient of the expected rewards $V(\pi_\theta)$ with respect to the parameters $\theta \in \mathbb{R}^d$ of the policy as:

$$\nabla_\theta V(\pi_\theta) = \mathop{\mathbb{E}}_{s_1,a_1,r_1,s_2,a_2,r_2,\ldots} \left[\sum_{t=1}^{T} \nabla_\theta \log(\pi_\theta(a_t|s_t)) \sum_{t'=t}^{T} r_{t'} \right]. \tag{11}$$

This gradient is therefore given by the expectation value of the d-dimensional random variable

$$X(s_1, a_1, r_1, s_2, \ldots) = \sum_{t=1}^{T} \nabla_\theta \log(\pi_\theta(a_t|s_t)) \sum_{t'=t}^{T} r_{t'}$$

(or equivalently, d observables that are all diagonal in the computational basis) with respect to all possible interactions with the environment, following a policy π_θ. In order for our mean estimators to be applicable here, our only assumption on the environment is that we have oracle access to its dynamics, notably its state-transitions

$$|s_t\rangle|a_t\rangle|0\rangle \mapsto \sum_{s_{t+1}} \sqrt{P(s_{t+1}|s_t, a_t)}|s_t\rangle|a_t\rangle|s_{t+1}\rangle$$

and its reward function

$$|s_t\rangle|0\rangle \mapsto |s_t\rangle|r_t\rangle.$$

As for the policy, we assume having the ability to implement π_θ coherently (i.e., similarly to $U_{\mathbb{P}}$), and to construct a (classical) circuit that computes the gradient $\nabla_\theta \log(\pi_\theta(a_t|s_t))$ given s_t, a_t, θ.

6 Discussion

In this work, we developed near-optimal quantum mean estimators in two different input models. In the binary oracle setting, we managed to obtain matching upper and lower bounds up to polylogarithmic factors, when we measure the performance of our estimator with respect to the Euclidean norm. We did not investigate the problem of deriving sharp bounds for other norms in this model. In the classical literature, sample-optimal estimators for general norms were given in [LM19b]. One case that could be interesting to further study is the ℓ_∞-norm, since it arises naturally in our quantum algorithm as well as in the applications we consider. By combining the one-dimensional result with a union bound, one can obtain a classical estimator that achieves a precision of $\sqrt{\max_{j\in[d]} \mathrm{Var}[X_j] \log(d/\delta)/n}$, whereas quantumly we obtained $\sqrt{\sum_{j\in[d]} \mathrm{Var}[X_j]} \log(d/\delta)/n$. It would be interesting to figure out whether some combination of these two approaches can be shown to be optimal in all regimes.

One further observation is that we do not assume to have any knowledge about Σ beforehand. Some preliminary considerations seem to indicate that in some ℓ_p-norms, especially where $p < 2$, it might be useful to know bounds on the individual diagonal entries of this covariance matrix. Whether these considerations are fundamental, or can be worked around, is also an interesting question to address in the future.

Acknowledgments

AC and SJ would like to thank Vedran Dunjko and Māris Ozols for pointing us in the direction of this problem, and for many insightful and motivating discussions. AC would also like to thank Ronald de Wolf and Joran van Apeldoorn for insightful discussions and helpful tips. Furthermore, AC would like to extend his gratitude to the anonymous legends that answered the Math Overflow post related to this research [MO21]. SJ acknowledges support from the Austrian Science Fund (FWF) through the projects DK-ALM:W1259-N27 and SFB BeyondC F7102. SJ also acknowledges the Austrian Academy of Sciences as a recipient of the DOC Fellowship.

References

[Aar20] S. Aaronson. "Shadow Tomography of Quantum States". In: *SIAM Journal on Computing* 49.5 (2020), STOC18-368–STOC18-394. DOI: 10.1137/18M120275X (cit. on p. 30).

[AFDJ03] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. "An Introduction to MCMC for Machine Learning". In: *Machine Learning* 50.1 (2003), pp. 5–43. DOI: 10.1023/a:1020281327116 (cit. on p. 1).

[AMS99] N. Alon, Y. Matias, and M. Szegedy. "The Space Complexity of Approximating the Frequency Moments". In: *Journal of Computer and System Sciences* 58.1 (1999), pp. 137–147. DOI: 10.1006/jcss.1997.1545 (cit. on p. 2).

[Ape21] J. van Apeldoorn. "Quantum Probability Oracles & Multidimensional Amplitude Estimation". In: *Proceedings of the 16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC)*. 2021, 9:1–9:11. DOI: 10.4230/LIPIcs.TQC.2021.9 (cit. on pp. 6, 13, 15).

[AW99] D. S. Abrams and C. P. Williams. *Fast Quantum Algorithms for Numerical Integrals and Stochastic Processes*. arXiv:quant-ph/9908083. 1999. DOI: 10.48550/arXiv.quant-ph/9908083 (cit. on pp. 2, 3, 5).

[BBHT98] M. Boyer, G. Brassard, P. Høyer, and A. Tapp. "Tight Bounds on Quantum Searching". In: *Fortschritte der Physik* 46.4-5 (1998), pp. 493–505. DOI: 10.1002/(SICI)1521-3978(199806)46:4/5<493::AID-PROP493>3.0.CO;2-P (cit. on p. 2).

[BDGT11] G. Brassard, F. Dupuis, S. Gambs, and A. Tapp. *An Optimal Quantum Algorithm to Approximate the Mean and its Application for Approximating the Median of a Set of Points over an Arbitrary Distance*. arXiv:1106.4267 [quant-ph]. 2011. DOI: 10.48550/arXiv.1106.4267 (cit. on pp. 2, 3, 5, 7).

[Bel15] A. Belovs. *Variations on Quantum Adversary*. arXiv:1504.06943 [quant-ph]. 2015. DOI: 10.48550/arXiv.1504.06943 (cit. on p. 24).

[BH10] K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics*. 5th ed. Graduate texts in physics. Springer, 2010. DOI: 10.1007/978-3-642-03163-2 (cit. on p. 1).

[BHH11] S. Bravyi, A. W. Harrow, and A. Hassidim. "Quantum Algorithms for Testing Properties of Distributions". In: *IEEE Transactions on Information Theory* 57.6 (2011), pp. 3971–3981. DOI: 10.1109/TIT.2011.2134250 (cit. on pp. 3, 7).

[BHMT02] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. "Quantum Amplitude Amplification and Estimation". In: *Contemporary Mathematics* 305 (2002), pp. 53–74. DOI: 10.1090/conm/305/05215 (cit. on pp. 2–5, 8, 9).

[BLSF19] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini. "Parameterized Quantum Circuits as Machine Learning Models". In: *Quantum Science and Technology* 4.4 (2019), p. 043001. DOI: 10.1088/2058-9565/ab4eb5 (cit. on p. 30).

[BV97] E. Bernstein and U. V. Vazirani. "Quantum Complexity Theory". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1411–1473. DOI: 10.1137/S0097539796300921 (cit. on pp. 3, 5, 10).

[CFMW10] S. Chakraborty, E. Fischer, A. Matsliah, and R. de Wolf. "New Results on Quantum Property Testing". In: *Proceedings of the 30th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*. 2010, pp. 145–156. DOI: 10.4230/LIPIcs.FSTTCS.2010.145 (cit. on p. 7).

[CJ21] A. Cornelissen and S. Jerbi. *Quantum Algorithms for Multivariate Monte Carlo Estimation*. arXiv:2107.03410 [quant-ph]. 2021. DOI: doi.org/10.48550/arXiv.2107.03410 (cit. on pp. 2, 13, 15, 24, 26).

[DH96] C. Dürr and P. Høyer. *A Quantum Algorithm for Finding the Minimum*. arXiv:quant-ph/9607014. 1996. DOI: 10.48550/arXiv.quant-ph/9607014 (cit. on p. 8).

[Fey82] R. P. Feynman. "Simulating Physics with Computers". In: *International Journal of Theoretical Physics* 21.6/7 (1982). DOI: 10.1007/BF02650179 (cit. on p. 1).

[FG06] J. Flum and M. Grohe. *Parameterized Complexity Theory*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2006. DOI: 10.1007/3-540-29953-X (cit. on p. 25).

[FGGS99] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. "Bound on the Number of Functions that Can Be Distinguished with k Quantum Queries". In: *Physical Review A* 60 (1999), pp. 4331–4333. DOI: 10.1103/PhysRevA.60.4331 (cit. on p. 24).

[GAW17] A. Gilyén, S. Arunachalam, and N. Wiebe. *Optimizing Quantum Optimization Algorithms via Faster Quantum Gradient Computation*. arXiv:1711.00465 [quant-ph]. 2017. DOI: 10.1137/1.9781611975482.87 (cit. on p. 8).

[GAW19] A. Gilyén, S. Arunachalam, and N. Wiebe. "Optimizing Quantum Optimization Algorithms via Faster Quantum Gradient Computation". In: *Proceedings of the 30th Symposium on Discrete Algorithms (SODA)*. 2019, pp. 1425–1444. DOI: 10.1137/1.9781611975482.87 (cit. on pp. 2, 4–8, 10, 11, 15).

[GL20] A. Gilyén and T. Li. "Distributional Property Testing in a Quantum World". In: *Proceedings of the 11th Innovations in Theoretical Computer Science Conference, (ITCS)*. 2020, 25:1–25:19. DOI: 10.4230/LIPIcs.ITCS.2020.25 (cit. on p. 8).

[Gla03] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer New York, 2003. DOI: 10.1007/978-0-387-21617-1 (cit. on p. 1).

[Gro98] L. K. Grover. "A Framework for Fast Quantum Mechanical Algorithms". In: *Proceedings of the 30th Symposium on Theory of Computing (STOC)*. 1998, pp. 53–62. DOI: 10.1145/276698.276712 (cit. on pp. 2, 5, 7).

[GSLW19] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe. "Quantum Singular Value Transformation and Beyond: Exponential Improvements for Quantum Matrix Arithmetics". In: *Proceedings of the 51st Symposium on Theory of Computing (STOC)*. 2019, pp. 193–204. DOI: 10.1145/3313276.3316366 (cit. on pp. 3, 5).

[Ham21] Y. Hamoudi. "Quantum Sub-Gaussian Mean Estimator". In: *Proceedings of the 29th European Symposium on Algorithms (ESA)*. 2021, 50:1–50:17. DOI: 10.4230/LIPIcs.ESA.2021.50 (cit. on pp. 2–6, 8).

[Hei02] S. Heinrich. "Quantum Summation with an Application to Integration". In: *Journal of Complexity* 18.1 (2002), pp. 1–50. DOI: 10.1006/jcom.2001.0629 (cit. on pp. 2–7).

[Hei04] S. Heinrich. "On the Power of Quantum Algorithms for Vector Valued Mean Computation". In: *Monte Carlo Methods and Applications* 10.3–4 (2004), pp. 297–310. DOI: 10.1515/mcma.2004.10.3-4.297 (cit. on pp. 1, 2, 6).

[HKP20] H.-Y. Huang, R. Kueng, and J. Preskill. "Predicting Many Properties of a Quantum System from Very Few Measurements". In: *Nature Physics* 16.10 (2020), pp. 1050–1057. DOI: 10.1038/s41567-020-0932-7 (cit. on p. 30).

[HM19] Y. Hamoudi and F. Magniez. "Quantum Chebyshev's Inequality and Applications". In: *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP)*. 2019, 69:1–69:16. DOI: 10.4230/LIPIcs.ICALP.2019.69 (cit. on pp. 2, 4–7).

[Hop20] S. B. Hopkins. "Mean Estimation with Sub-Gaussian Rates in Polynomial Time". In: *The Annals of Statistics* 48.2 (2020), pp. 1193–1213. DOI: 10.1214/19-AOS1843 (cit. on p. 2).

[JGM+21] S. Jerbi, C. Gyurik, S. C. Marshall, H. J. Briegel, and V. Dunjko. *Parametrized Quantum Policies for Reinforcement Learning*. arXiv:2103.05577 [quant-ph]. 2021. DOI: 10.48550/arXiv.2103.05577 (cit. on p. 30).

[Jor05] S. P. Jordan. "Fast Quantum Algorithm for Numerical Gradient Estimation". In: *Physical Review Letters* 95.5 (2005), p. 050501. DOI: 10.1103/PhysRevLett.95.050501 (cit. on pp. 3, 5, 6, 10).

[JTN+21] S. Jerbi, L. M. Trenkwalder, H. P. Nautrup, H. J. Briegel, and V. Dunjko. "Quantum Enhancements for Deep Reinforcement Learning in Large Spaces". In: *PRX Quantum* 2.1 (2021), p. 010328. DOI: 10.1103/prxquantum.2.010328 (cit. on p. 30).

[JVV86] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. "Random Generation of Combinatorial Structures from a Uniform Distribution". In: *Theoretical Computer Science* 43 (1986), pp. 169–188. DOI: 10.1016/0304-3975(86)90174-X (cit. on p. 2).

[Kit95] A. Kitaev. *Quantum Measurements and the Abelian Stabilizer Problem*. arXiv:quant-ph/9511026. 1995. DOI: 10.48550/arXiv.quant-ph/9511026 (cit. on pp. 1, 4).

[KOS07] E. Knill, G. Ortiz, and R. D. Somma. "Optimal Quantum Measurements of Expectation Values of Observables". In: *Physical Review A* 75.1 (2007), p. 012328. DOI: 10.1103/PhysRevA.75.012328 (cit. on p. 29).

[KW17] M. Kieferová and N. Wiebe. "Tomography and Generative Training with Quantum Boltzmann Machines". In: *Physical Review A* 96.6 (2017), p. 062327. DOI: 10.1103/physreva.96.062327 (cit. on p. 30).

[LM19a] G. Lugosi and S. Mendelson. "Mean Estimation and Regression Under Heavy-Tailed Distributions: A Survey". In: *Foundations of Computational Mathematics* 19.5 (2019), pp. 1145–1190. DOI: 10.1007/s10208-019-09427-x (cit. on pp. 1–3, 5, 8).

[LM19b] G. Lugosi and S. Mendelson. "Near-Optimal Mean Estimators with Respect to General Norms". In: *Probability Theory and Related Fields* 175.3-4 (2019), pp. 957–973. DOI: 10.1007/s00440-019-00906-4 (cit. on p. 31).

[LMR+11] T. Lee, R. Mittal, B. W. Reichardt, R. Špalek, and M. Szegedy. "Quantum Query Complexity of State Conversion". In: *Proceedings of the 52nd Symposium on Foundations of Computer Science (FOCS)*. 2011, pp. 344–353. DOI: 10.1109/FOCS.2011.75 (cit. on p. 24).

[Low17] G. H. Low. "Quantum Signal Processing by Single-Qubit Dynamics". PhD thesis. Massachusetts Institute of Technology, 2017 (cit. on p. 8).

[LW19] T. Li and X. Wu. "Quantum Query Complexity of Entropy Estimation". In: *IEEE Transactions on Information Theory* 65.5 (2019), pp. 2899–2921. DOI: 10.1109/TIT.2018.2883306 (cit. on p. 7).

[MNKF18] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. "Quantum Circuit Learning". In: *Physical Review A* 98.3 (2018), p. 032309. DOI: 10.1103/physreva.98.032309 (cit. on p. 30).

[MO21] *Probability of ℓ^1-norms of vertices of the rotated Hamming cube*. Available at https://mathoverflow.net/q/390129/92442. 2021 (cit. on p. 32).

[Mon15] A. Montanaro. "Quantum Speedup of Monte Carlo Methods". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471.2181 (2015), p. 20150301. DOI: 10.1098/rspa.2015.0301 (cit. on pp. 1–7, 9).

[NC11] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th ed. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511976667 (cit. on p. 21).

[NW99] A. Nayak and F. Wu. "The Quantum Query Complexity of Approximating the Median and Related Statistics". In: *Proceedings of the 31st Symposium on Theory of Computing (STOC)*. 1999, pp. 384–393. DOI: 10.1145/301250.301349 (cit. on pp. 7, 8).

[NY83] A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley & Sons, 1983 (cit. on p. 2).

[SB18] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. Second. The MIT Press, 2018 (cit. on p. 30).

[SJD21] A. Skolik, S. Jerbi, and V. Dunjko. *Quantum Agents in the Gym: A Variational Quantum Algorithm for Deep Q-Learning*. arXiv:2103.15084 [quant-ph]. 2021. DOI: 10.48550/arXiv.2103.15084 (cit. on p. 30).

[SMSM99] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour. "Policy Gradient Methods for Reinforcement Learning with Function Approximation". In: *Proceedings of the 12th International Conference on Neural Information Processing Systems (NIPS)*. 1999, pp. 1057–1063 (cit. on p. 31).

[Sze04] M. Szegedy. "Quantum Speed-Up of Markov Chain Based Algorithms". In: *Proceedings of the 45th Symposium on Foundations of Computer Science (FOCS)*. 2004, pp. 32–41. DOI: 10.1109/FOCS.2004.53 (cit. on p. 1).

[Ter99] B. M. Terhal. "Quantum Algorithms and Quantum Entanglement". PhD thesis. University of Amsterdam, 1999 (cit. on pp. 2, 4, 5, 7, 9).

[WCNA09] P. Wocjan, C.-F. Chiang, D. Nagaj, and A. Abeyesinghe. "Quantum Algorithm for Approximating Partition Functions". In: *Physical Review A* 80.2 (2009), p. 022340. DOI: 10.1103/PhysRevA.80.022340 (cit. on pp. 2, 4, 9, 29).

[WKS16] N. Wiebe, A. Kapoor, and K. M. Svore. "Quantum Deep Learning". In: *Quantum Information & Computation* 16.7&8 (2016), pp. 541–587. DOI: 10.26421/QIC16.7-8-1 (cit. on p. 30).

[WW19] N. Wiebe and L. Wossnig. *Generative Training of Quantum Boltzmann Machines with Hidden Units*. arXiv:1905.09902 [quant-ph]. 2019. DOI: 10.48550/arXiv.1905.09902 (cit. on p. 30).