Low depth algorithms for quantum amplitude estimation

Tudor Giurgica-Tiron[2,3], Iordanis Kerenidis[1,5], Farrokh Labib[2,4], Anupam Prakash[1], and William Zeng[2]

[1]QC Ware Corp., Palo Alto, USA and Paris, France.

[2]Goldman, Sachs & Co., New York, USA.

[3]Stanford University, Palo Alto, USA.

[4]CWI Amsterdam, Netherlands.

[5]CNRS, Université Paris, France.

We design and analyze two new low depth algorithms for amplitude estimation (AE) achieving an optimal tradeoff between the quantum speedup and circuit depth. For $\beta \in (0, 1]$, our algorithms require $N = \tilde{O}(\frac{1}{\epsilon^{1+\beta}})$ oracle calls and require the oracle to be called sequentially $D = O(\frac{1}{\epsilon^{1-\beta}})$ times to perform amplitude estimation within additive error ϵ. These algorithms interpolate between the classical algorithm ($\beta = 1$) and the standard quantum algorithm ($\beta = 0$) and achieve a tradeoff $ND = O(1/\epsilon^2)$. These algorithms bring quantum speedups for Monte Carlo methods closer to realization, as they can provide speedups with shallower circuits.

The first algorithm (Power law AE) uses power law schedules in the framework introduced by Suzuki et al [24]. The algorithm works for $\beta \in (0, 1]$ and has provable correctness guarantees when the log-likelihood function satisfies regularity conditions required for the Bernstein Von-Mises theorem. The second algorithm (QoPrime AE) uses the Chinese remainder theorem for combining lower depth estimates to achieve higher accuracy. The algorithm works for discrete $\beta = q/k$ where $k \geq 2$ is the number of distinct coprime moduli used by the algorithm and $1 \leq q \leq k - 1$, and has a fully rigorous correctness proof. We analyze both algorithms in the presence of depolarizing noise and provide numerical comparisons with the state of the art amplitude estimation algorithms.

Tudor Giurgica-Tiron: tgt@stanford.edu
Iordanis Kerenidis: iordanis.kerenidis@qcware.com
Farrokh Labib: farrokhlabib@gmail.com
Anupam Prakash: anupam.prakash@qcware.com
William Zeng: william.zeng@gs.com

1 Introduction

Amplitude estimation [6] is a fundamental quantum algorithm that allows a quantum computer to estimate the amplitude $\langle 0| U |0\rangle$ for a quantum circuit U to additive error ϵ with $O(1/\epsilon)$ calls to U. The algorithm offers a quadratic advantage over classical sampling and has many applications including speedups for Monte Carlo methods [22] and approximate counting [7].

To be more precise, we consider an amplitude estimation setting where the algorithm is given access to a quantum circuit U such that $U|0^t\rangle = \cos(\theta)|x, 0\rangle + \sin(\theta)|x', 1\rangle$ where $|x\rangle, |x'\rangle$ are arbitrary states on $(t-1)$ qubits. The algorithm's goal is to estimate the amplitude θ within an additive ϵ. The closely related approximate counting problem corresponds to the special case where $U|0^t\rangle = \frac{1}{2^{(t-1)/2}}(\sum_{i\in S}|i\rangle|0\rangle + \sum_{i\notin S}|i\rangle|1\rangle)$ is a uniform superposition over bit strings of length $(t-1)$ and the binary label on the second register is the indicator function for $S \subseteq \{0,1\}^{t-1}$. Amplitude estimation in this setting provides an estimate for $|S|$. Approximate counting in turn generalizes Grover's search [12] and the problem of finding the number of marked elements in a list.

We briefly discuss two applications of amplitude estimation, to quantum Monte Carlo methods and inner product estimation that are particularly relevant for applications of quantum computing to finance and machine learning.

Quantum Monte Carlo methods are an important application of amplitude estimation to the problem of estimating the mean of a real valued random variable by sampling. Let $f(x, S)$ be a real valued function where x is the input and S is the random seed and let σ be the variance of $f(x, S)$. Classically estimating the mean $E_S f(x, S)$ to additive error ϵ is known to require $N = O(\sigma^2/\epsilon^2)$ samples. Montanaro showed that there is a quantum algorithm for estimating the mean that required $N = \widetilde{O}\left(\frac{\sigma}{\epsilon}\right)$ samples, thus quadratically improving the dependence on σ and $1/\epsilon$ [22]. This quantum Monte-Carlo algorithm builds on prior works for estimating the mean of real valued functions using quantum computers [2, 12, 5] and has been further generalized to settings when more information about the distribution such as upper and lower bounds on the mean are known [21, 13].

Amplitude estimation also has applications which are not reducible to approximate counting, where the goal is to estimate $\langle 0| U |0\rangle$ for a unitary U without additional structure. Some examples of this kind include applications of amplitude estimation to quantum linear algebra and machine learning. For example, quantum procedures for estimating the inner product $\langle x|y\rangle$ between vectors $x, y \in \mathbb{R}^n$ have been found to be useful for quantum classification and clustering algorithms [18, 27]. Amplitude estimation is the source of the quantum speedup for inner product estimation, with the quantum algorithms requiring $O(1/\epsilon)$ samples as opposed to the $O(1/\epsilon^2)$ samples required classically for estimating the inner product to error ϵ. Amplitude estimation variants are also important for reducing the condition number dependence in the quantum linear system solvers from $O(\kappa^2)$ to $O(\kappa)$ [14, 3].

In recent times, there has been a lot of interest in reducing the resource requirements for amplitude estimation algorithms, moving towards the goal of finding an AE algorithm compatible with noisy intermediate scale quantum (NISQ) devices [23]. The major limitation for adapting amplitude estimation to nearer term devices is that the circuit U needs to be run sequentially $O(1/\epsilon)$ times resulting in a high circuit depth for the algorithm.

The classical amplitude estimation algorithm [6] invokes the controlled quantum circuit U at least $O(1/\epsilon)$ times in series followed by a quantum Fourier transform for estimating amplitudes to error ϵ using the phase estimation algorithm [19]. This makes applications of amplitude estimation like Monte Carlo methods or approximate counting prohibitive for near term hardware as, even in cases where the oracle itself does not have significant depth, the high number of repetitions in series of the oracle makes the overall depth of the circuits prohibitive for the high noise rates of current NISQ devices. A significant amount of recent work has tried to make amplitude estimation nearer term. The known results on amplitude estimation along with their resource requirements in terms of qubits, depth, and total number of oracle calls (number of times the circuit U is run) are summarized in Table 1.

Algorithm	**Qubits**	**Depth**	**Number of calls**
Amplitude estimation [6]	$n + \log(1/\epsilon)$	$d \cdot \frac{1}{\epsilon} + \log\log(1/\epsilon)$	$\frac{1}{\epsilon}$
QFT free amplitude estimation [24, 1]	n	$d \cdot \frac{1}{\epsilon}$	$\frac{1}{\epsilon}$
IQAE [11]	n	$d \cdot \frac{1}{\epsilon}$	$\frac{1}{\epsilon}\log(1/\epsilon)$
Power-law AE [This paper]	n	$d \cdot \left(\frac{1}{\epsilon}\right)^{1-\beta}$	$\left(\frac{1}{\epsilon}\right)^{1+\beta}$
QoPrime AE [This paper]	n	$d \cdot \left(\frac{1}{\epsilon}\right)^{1-q/k}$	$\left(\frac{1}{\epsilon}\right)^{1+q/k}$

Table 1: Asymptotic tradeoffs of amplitude estimation algorithms. Parameters: n is the number of qubits and d is the circuit depth for a single application of U, ϵ is the additive error, $\beta \in (0, 1]$, $k \geq 2$, $q \in [k-1]$.

A number of recent works [24, 1, 11] have given QFT free amplitude estimation algorithms, that is algorithms that do not require a Quantum Fourier transform (QFT) at the end of the computation. The QFT circuit is applied to a register with $O(\log(1/\epsilon))$ qubits and adds an asymptotic factor of $O(\log\log(1/\epsilon))$ to the overall circuit depth of the algorithm. Eliminating the QFT does not significantly lower the depth of the AE algorithm, but it is an important step towards making amplitude estimation nearer term as it removes the need to apply controlled versions of the oracle. Controlled oracles incur considerable overhead as they require adding multiple controls to every gate in the oracle circuit. The iterative quantum amplitude estimation (IQAE) algorithm in Table 1 has the same asymptotic performance as [24, 1], however it can be viewed as the state of the art AE algorithm in practice, as it has a provable analysis with the lowest constant overheads among the known QFT free amplitude estimation methods.

In this work, we take a different view and we ask the following question: can quantum amplitude estimation algorithms that use quantum circuits with depth asymptotically less than $O(1/\epsilon)$ provide any speed up with respect to classical algorithms? We respond to this question by focusing on algorithms that interpolate between classical and quantum amplitude estimation. At a high level, classical amplitude estimation requires $O(1/\epsilon^2)$ applications of the oracle, but these applications can be performed in parallel. The standard amplitude estimation algorithm [6] on the other hand requires $O(1/\epsilon)$ serial applications of the oracle, but one only needs to perform this computation once. In this paper, we present two different algorithms that interpolate between the classical and quantum settings with an optimal tradeoff $ND = O(1/\epsilon^2)$, where N is the total number of oracle calls and D is the maximum number of sequential oracle calls.

That said, it is known that there are limitations to depth reduction for amplitude estimation and that the circuit depth for AE cannot be reduced generically while maintaining the entire speedup. Zalka first established a tradeoff between the depth and the number of executions for Grover's search [28]. Recent work of Burchard [8], building upon [17] extends the tradeoffs in Zalka's work to the setting of approximate counting and shows that for approximate counting, depth-D parallel runs of the algorithm can at most achieve a D-fold speedup over the classical algorithm. Burchard's work [8] also suggests an algorithm matching his lower bound in settings where the approximate counting domain can be partitioned into equally sized pseudorandom subdomains. However, this method is restricted to approximate counting as it assumes a discrete domain that can be partitioned into pseudorandom parts and it incurs a number of overheads that are significant for near term devices, we refer to [4] for a more detailed discussion. The amplitude estimation algorithms that we propose in this paper match the Burchard-Zalka lower bounds exactly, and

are applicable to the more general setting of amplitude estimation where no additional structure required.

Tradeoffs between depth and the number of oracle calls for quantum algorithms have also been considered in some recent works. Variational quantum eigensolver (VQE) algorithms interpolating between classical sampling and phase estimation with circuit depth $O(1/\epsilon^{\alpha})$ and $O(1/\epsilon^{2(1-\alpha)})$ oracle calls were proposed in [26]. Fisher information calculations similar to the ones used here have been to analyze engineered quantum likelihood functions [20] and to to design experiments to best estimate the period for time invariant single qubit Hamiltonian systems [10].

The Power law Amplitude Estimation algorithm

Our first algorithm utilizes the framework proposed by Suzuki et al. [24] for QFT free amplitude estimation. A higher-level description of this algorithm, under the name of the Kerenidis-Prakash approach, appears in the recent survey paper [4]. In the Suzuki et al. framework the oracle is invoked with varying depths, and then measurements in the standard basis are performed, followed by classical maximum likelihood post-processing to estimate the amplitude. Algorithms in this framework are specified as schedules (m_k, N_k) where the oracle is applied m_k times in series for N_k iterations and, at the end, the results are post-processed classically using maximum likelihood estimation.

The main idea behind these algorithms is the following. The classical amplitude estimation procedure uses $O(1/\epsilon^2)$ calls to the circuit U and measurements in the standard basis, which is equivalent to sampling from a Bernoulli random variable with success probability $\alpha = \cos^2(\theta)$. The quantum amplitude estimation algorithms on the other hand, use quantum circuits that sequentially perform oracle calls at all depths up to $O(1/\epsilon)$. If the quantum circuits have depth k then a quantum algorithm samples from a Bernoulli random variable with success probability $\cos^2((2k+1)\theta)$. Suzuki et al [24] observed that samples from a Bernoulli random variable are more informative for estimating θ if the success probability is $\cos^2((2k+1)\theta)$, and this can be made precise using the notion of Fisher information $I_f(\alpha)$ for a schedule.

The QFT free amplitude estimation algorithm [24] uses an exponential schedule with depths $m_k = 2^k$ all the way up to the maximum depth of $O(1/\epsilon)$ and chooses $N_k = N_{shot}$ to be a constant. The quantum Fourier transform step at the end of the algorithm is eliminated, however the asymptotic efficiency of max-likelihood post-processing is not established rigorously [1] . Linear schedules with $m_k = k$ were also considered in [24]. Subsequently, Aaronson and Rall [1] provided a provable QFT free amplitude estimation algorithm that does not require max-likelihood estimation. The state of the art QFT free algorithm is the IQAE [11], which improves upon the large constant factors required for the analysis in [1] and has the best performance among all known QFT free variants of amplitude estimation.

Our power law amplitude estimation algorithm (Power law AE) uses power law schedules with constant $N_k = N_{shot}$ and $m_k = \lfloor k^{\frac{1-\beta}{2\beta}} \rfloor$, where k starts from 1 and increases until the maximum depth for the quantum circuit is $O(1/\epsilon^{1-\beta})$ for $\beta \in (0, 1]$, at a cost of more parallel runs with total number of oracle calls scaling as $O(1/\epsilon^{1+\beta})$. When β tends to 0, the schedule approaches the exponential schedule, while when β is equal to 1, the algorithm is the classical one. The analysis of the power law AE is based on the observation that maximum likelihood estimation in this setting is equivalent to sub-dividing the domain for the amplitude θ into $O(1/\epsilon)$ equal parts and performing Bayesian updates starting from a uniform prior. If the prior and the log-likelihood function are sufficiently regular, this allows us to use the Bernstein Von-Mises theorem [15], which can be viewed as a Bayesian central limit theorem that quantifies the rate of convergence to the Bayesian estimator to the normal distribution centered at the true value of θ with variance $1/N_{shot}I_f(\alpha)$, where $\alpha = \cos^2(\theta)$. The variant of the Bernstein Von-Mises theorem proved in [15] is particularly helpful for the analysis as it bounds the rate of convergence of the posterior distribution to the normal distribution in the ℓ_1 norm. The tradeoff $ND = O(1/\epsilon^2)$ follows from Fisher information calculations for the power law schedules.

Very recently, super-linear polynomial schedules in the presence of depolarizing noise have been considered by Tanaka et al. [25]. We also study the behaviour of our algorithm in the presence of depolarizing noise, and describe a way to make our algorithm robust to noise. In fact, we give a simple method for choosing the optimal parameter β given a desired accuracy and noise level.

One can see our algorithm as an optimal way of utilizing all the power of the available quantum circuit in terms of depth, meaning that instead of having to wait for quantum circuits to have good enough fidelity to apply sequentially a number of oracle calls of the order of $1/\epsilon$, which for Monte Carlo applications can grow between 10^3 to 10^6, our power-law AE algorithm makes it possible to use quantum circuits of any depth and provide a corresponding smaller speedup.

The theoretical analysis described above relies on strong regularity conditions on the log-likelihood and the prior required for the Bernstein Von-Mises theorem, which can be hard to verify rigorously, even though they seem to hold empirically for the log-likelihood function for amplitude estimation. It is therefore desirable, at least from a theoretical point of view, to have an AE algorithm achieving the same $ND = O(1/\epsilon^2)$ tradeoffs that does not rely on these conditions. If we attempt to find a schedule maximizing the Fisher information for a given number of oracle calls, the optimal solution is a schedule that makes oracle calls at the maximal possible depth. However, making oracle calls at a single depth are not sufficient for amplitude estimation due to the periodicity of the function $\cos^2((2k+1)\theta)$. This led us to consider AE algorithms that make queries at two (or more) depths and combine the results, leading to the QoPrime AE algorithm.

The QoPrime Amplitude Estimation algorithm

Our second amplitude estimation algorithm (QoPrime) uses a number theoretic approach to amplitude estimation that enables a fully rigorous correctness proof and has the same depth vs number of oracle calls tradeoff as the power-law AE for a discrete set of exponents.

The basic idea for the QoPrime algorithm is to choose k different co-prime moduli, each close to $O(1/\epsilon^{1/k})$ so their product is $N = O(1/\epsilon)$. Let the true value of the amplitude be $\pi M/2N$ where $M \in [0, N]$. The algorithm estimates $\lfloor M \rfloor \mod N_i$, where N_i is the product of q out of the k moduli using N/N_i sequential calls to the oracle followed by measurements in the standard basis. These low accuracy estimates are then combined using the Chinese remainder theorem to obtain $\lfloor M \rfloor \mod N$. We now sketch the main idea for the QoPrime algorithm for the simplest case of $k = 2$ moduli, this corresponds to an algorithm with $D = O(1/\epsilon^{1/2})$ and $N = O(1/\epsilon^{3/2})$. Let a and b be the two largest co-prime numbers upper bounded by the maximum depth $D = O(1/\sqrt{\epsilon})$. For simplicity, let us assume that the true value for θ is of the form $\frac{\pi M}{2(2a+1)(2b+1)}$ for some integer $M \in [(2a+1)(2b+1)]$. The QoPrime algorithm recovers $M \mod (2a+1)$ and $M \mod (2b+1)$ and then determines M using the Chinese remainder theorem.

Invoking the oracle a times in series followed by a measurement in the standard basis is equivalent to sampling from a Bernoulli random variable with the success probability $\cos^2((2a+1)\theta)$. As a function of θ, this probability is periodic with period $\frac{\pi}{2a+1}$ and is therefore a function of $\pm M \mod (2b+1)$. Subdividing the interval $[0, \pi/2]$ into $(2b+1)$ equal parts, it follows from the additive Chernoff bounds that $\pm M \mod (2b+1)$ can be recovered with $O((2b+1)^2)$ samples. The algorithm is able to estimate $M \mod (2b+1)$ with $O(2b+1)^2$ repetitions of a quantum circuit of depth $(2a+1)$ and, similarly, $M \mod (2a+1)$ with $O(2a+1)^2$ repetitions of a quantum circuit of depth $(2b+1)$. The Chinese remainder theorem is then used to combine these low precision estimates to obtain the integer $M \in [(2a+1)(2b+1)]$, thus boosting the precision for the estimation procedure. The total number of oracle calls made was $O(ab^2 + ba^2) = O(1/\epsilon^{1.5})$. The maximum depth for the oracle call was $O(1/\epsilon^{1/2})$. A recursive application of the algorithm at a lower precision is needed to determine the sign of the estimates, the details of which are described in section 3. The algorithm extends to the more general case where the true value for θ is $\frac{\pi M}{2N}$ where N is the product of $q \leq k$ coprime moduli and $M \in [0, N]$, in which case we also show how to pick these values q, k. Here, the maximum depth of the quantum circuit is $D = O(1/\epsilon^{1-q/k})$, and the total number of oracle calls are $N = O(1/\epsilon^{1+q/k})$.

Further, we study the accuracy of the algorithm in the presence of depolarizing noise and provide a number of graphs that show the behavior of the algorithm under noise. The analysis of the algorithm in a noisy setting shows that noise limits the depth of the oracle calls that can be made, but it also allows us to choose optimally the algorithm parameters to minimize the number of oracle calls for a given target approximation error and noise rate. The experiments show that the constant overhead for the QoPrime algorithm is reasonable ($C < 10$) for most settings of interest.

Last, we benchmark our two new low depth AE algorithms with the state of the art IQAE

algorithm [11] in noisy settings. Algorithms such as IQAE require access to a full circuit depth of $O(1/\epsilon)$, and this large depth is exponentially penalized by the depolarizing noise by requiring an exponentially large number of samples to achieve a precision below the noise level. In comparison, the power law and the QoPrime AE algorithms transition smoothly to a classical estimation scaling and do not suffer from an exponential growth in oracle calls. The Power law AE algorithm shows the best practical performance according to the simulations for different error rates and noise levels.

Overall, we present here two low depth algorithms for quantum amplitude estimation, thus potentially bringing a number of applications closer to the NISQ era. Of course, it is important to remember that even applying the oracle U once may already necessitate better quality quantum computers than the ones we have today so observing these quantum speedups in practice is still a long way ahead. Nevertheless, we believe the optimal tradeoff between the total number of oracle calls and the depth of the quantum circuit that is offered by our algorithms can be a powerful tool towards finding quantum applications for near and intermediate term devices.

The paper is organised as follows: In Section 2, we describe and analyze the power law amplitude estimation algorithm, while in Section 3, we describe and analyze the QoPrime amplitude estimation algorithm. In Section 4, we present empirical evidence of the performance of our algorithm and benchmarks with other state-of-the-art algorithms for amplitude estimation.

2 Amplitude estimation with power law schedules

2.1 Preliminaries

In this section, we introduce some preliminaries for the analysis of amplitude estimation with power law schedules. Let X be a random variable with density function $f(X, \alpha)$ that is determined by the single unknown parameter α. Let $l(X, \alpha) = \log f(X, \alpha)$ be the log-density function and let $l'(x, \alpha) = \frac{\partial l(x,\alpha)}{\partial \alpha}$. In this section, all expectations are with respect to the density function $f(X, \alpha)$ and $'$ denotes the partial derivative with respect to α.

The Fisher information $I_f(\alpha)$ is defined as the variance of the log-likelihood function, that is $I_f(\alpha) = Var[l'(X, \alpha)]$. It can also be equivalently defined as $I_f(\alpha) = -E_f[l''(X, \alpha)]$. More generally, for parameters $\alpha \in \mathbb{R}^n$, the Fisher information is defined as the covariance matrix of the second partial derivatives of $l(X, \alpha)$ with respect to α.

Let α^* be the true value for α and consider a Bayesian setting where a prior distribution on α is updated given i.i.d. samples $X_i, i \in [n]$ from a distribution $f(X, \alpha^*)$. The Bernstein-Von Mises theorem stated below quantifies the rate of convergence of the Bayesian estimator to the normal distribution with mean and variance $(\alpha^*, \frac{1}{nI_f(\alpha^*)})$ in the ℓ_1 norm for cases where the log-likelihood and the prior are sufficiently regular. The complete list of regularity conditions for the theorem is given in Appendix A.

Theorem 2.1. *[Bernstein Von-Mises Theorem [15]] Let $X_i, i \in [n]$ be independent samples from a distribution $f(X, \alpha^*)$ and let R_0 be the prior distribution on α. Let R_n be the posterior distribution after n samples and let Q_{n,α^*} be the Gaussian with mean and variance $(\alpha^*, \frac{1}{nI_f(\alpha^*)})$.*

If $f(X, \alpha^*)$ and R_0 satisfy the regularity conditions enumerated in the Appendix A, then there exists a constant $c > 0$ such that,

$$\Pr_{X_i \sim f, i \in [n]} \left[\|R_n - Q_{n,\alpha^*}\|_1 \geq c\sqrt{\frac{1}{n}} \right] = o\left(\frac{1}{\sqrt{n}}\right) \tag{1}$$

As we have defined before, the amplitude estimation algorithm has access to a quantum circuit U such that $U|0^t\rangle = \cos(\theta)|x, 0\rangle + \sin(\theta)|x', 0^\perp\rangle$ where $|x\rangle, |x'\rangle$ are arbitrary states on $(t-1)$ qubits. If the second register is measured in the standard basis, then the distribution of the measurement outcome $f(X, \alpha)$ is a Bernoulli random variable with success probability $\alpha = \cos^2(\theta) \in [0, 1]$. If a quantum circuit of k sequential calls to the circuit U is applied then the quantum state $\cos((2k+1)\theta)|x, 0\rangle + \sin((2k+1)\theta)|x', 0^\perp\rangle$ can be obtained. Measuring this state in the standard basis, the distribution of measurement outcome is again a Bernoulli random variable with success probability $\cos^2((2k+1)\theta)$.

A quantum AE algorithm therefore has access to samples from Bernoulli random variables with success probability $\cos^2((2k+1)\theta)$ where k is the number of sequential oracle calls in the quantum circuit, which corresponds to its depth. The higher depth samples are more informative for estimating θ. The next proposition quantifies the advantage for higher depth samples, showing that the Fisher information grows quadratically with the depth of the oracle calls.

Proposition 2.2. *Let $f(X,\alpha) = \beta^X(1-\beta)^{1-X}$ for parameter $\alpha = \cos^2(\theta)$ and $\beta = \cos^2((2m_k+1)\theta)$ for a positive integer m_k. The Fisher information is $I_f(\alpha) = \frac{(2m_k+1)^2}{\alpha(1-\alpha)}$.*

Proof. As $\alpha = \cos^2(\theta)$ we have $d\alpha = 2\cos(\theta)\sin(\theta)d\theta$. The log-likelihood function is $l(X,\alpha) = X\log\beta + (1-X)\log(1-\beta)$. Thus,

$$\begin{aligned} I_f(\alpha) &= -E_f\left[\frac{d}{d\alpha^2}(2X\log\cos((2m_k+1)\theta) + 2(1-X)\log\sin((2m_k+1)\theta))\right] \\ &= \frac{-1}{2\alpha(1-\alpha)}E_f\left[\frac{d}{d\theta^2}(X\log\cos((2m_k+1)\theta) + (1-X)\log\sin((2m_k+1)\theta))\right] \\ &= \frac{(2m_k+1)^2}{2\alpha(1-\alpha)}\left(\frac{E_f[X]}{\cos^2((2m_k+1)\theta)} + \frac{E_f[1-X]}{\sin^2((2m_k+1)\theta)}\right) \\ &= \frac{(2m_k+1)^2}{\alpha(1-\alpha)}. \end{aligned}$$

□

The Fisher information is defined as a variance and is therefore additive over independent samples that do not need to be identically distributed. The Fisher information of an amplitude estimation schedule (m_k, N_k) [24] is the sum of the Fisher informations for the individual samples.

2.2 The Power law Amplitude Estimation algorithm

The amplitude estimation algorithm using power law schedules is given as Algorithm 2.1. It is then analyzed to establish the tradeoff between the depth and the total number of oracle calls in a setting where the Bernstein Von Mises Theorem is applicable.

Algorithm 2.1 The Power law Amplitude Estimation

Require: Parameter $\beta \in (0,1]$, $N_{shot} \in \mathbb{Z}$ and desired accuracy ϵ for estimating θ.
Require: Access to a unitary U such that $U|0\rangle = \cos(\theta)|x,0\rangle + \sin(\theta)|x',1\rangle$.
Ensure: An estimate of θ within accuracy ϵ with high probability
1: Initialize the prior to be the uniform distribution on angles $\theta = \frac{\pi t\epsilon}{2}$.
2: **for** k=1 **to** $K = \max\left(\frac{1}{\epsilon^{2\beta}}, \log(1/\epsilon)\right)$ **do**
3: Initialize $N_{k_0} = 0$ and $N_{k_1} = 0$.
4: **for** i=1 **to** N_{shots} **do**
5: Apply $m_k = \lfloor k^{\frac{1-\beta}{2\beta}} \rfloor$ sequential oracle calls and measure last qubit of resulting quantum state in the standard basis.
6: If the outcome is 0, then $N_{k_0} = N_{k_0} + 1$, else $N_{k_1} = N_{k_1} + 1$.
7: **end for**
8: Perform Bayesian updates $p(\theta) \to p(\theta)\cos((2m_k+1)\theta)^{N_{k_0}}\sin((2m_k+1)\theta)^{N_{k_1}}$ for $\theta = \pi t\epsilon/2$ for integer valued $t \in [0, 1/\epsilon]$ and interpolate to obtain the posterior probability distribution.
9: **end for**
10: Output θ with the highest probability according to the posterior probability distribution.

The next theorem shows that Algorithm 2.1 achieves approximation error ϵ with parameters $N = O(\frac{1}{\epsilon^{1+\beta}})$ and $D = O(\frac{1}{\epsilon^{1-\beta}})$ where D is the maximum depth of the oracle calls and N is the total number of oracle calls made. The choice of $K = \max(\frac{1}{\epsilon^{2\beta}}, \log(1/\epsilon))$ ensures that our power law AE algorithm makes a sufficient number of queries for small β and approaches the exponential schedule of [24] as $\beta \to 0$.

Theorem 2.3. *The Power law Amplitude Estimation algorithm 2.1 outputs an ϵ accurate estimate with $N = O(\frac{1}{\epsilon^{1+\beta}})$ oracle calls and maximum depth $D = O(\frac{1}{\epsilon^{1-\beta}})$ with probability at least 0.9, that is the algorithm attains the tradeoff $ND = O(\frac{1}{\epsilon^2})$ in settings where the Bernstein-Von Mises theorem is applicable.*

Proof. The total number of oracle calls for Algorithm 2.1 is $N = \sum_{k\in[K]} N_{shot}(2m_k + 1)$ while the Fisher information for the power law schedule can be computed as $I_f(\alpha) = \frac{N_{shot}}{\alpha(1-\alpha)} \sum_{k\in[K]}(2m_k+1)^2$ using proposition 2.2. Approximating the sums in N and $I_f(\alpha)$ by the corresponding integrals we have $I_f(\alpha) = O(K^{1/\beta})$, $N = O(K^{(1+\beta)/2\beta})$ and maximum depth $D = O(K^{(1-\beta)/2\beta})$, so that $I_f(\alpha) = O(ND)$. Note that for $K = \max\left(\frac{1}{\epsilon^{2\beta}}, \log(1/\epsilon)\right)$, we have $N = O(\frac{1}{\epsilon^{1+\beta}})$ and $D = O(\frac{1}{\epsilon^{1-\beta}})$ and $I_f(\alpha) = O(\frac{1}{\epsilon^2})$.

It remains to show that with probability at least 0.9, the estimate output by the algorithm is within additive error ϵ of the true value. Applying the Bernstein Von-Mises Theorem, the ℓ_1 distance between the posterior distribution and the Gaussian with mean and variance $(\alpha^*, 1/\sqrt{N_{shot} I_f(\alpha^*)})$ is at most $c/\sqrt{N_{shot}}$ with probability at least $1 - c'/\sqrt{N_{shot}}$ for some constants $c, c' > 0$.

Choosing $N_{shot} > \left(\frac{\max(c,c')}{\delta}\right)^2$ for $\delta = 0.05$, the estimate is within $(\alpha^* \pm \frac{3\delta}{c\sqrt{I_f(\alpha^*)}})$ with probability at least $1 - \delta(1 + 1) - 0.0013 \geq 0.89$. The success probability can be boosted to $1 - \frac{1}{poly(\zeta)}$ for $\zeta > 0$ by running the algorithm $O(\log(1/\zeta))$ times and outputting the most frequent estimate.

□

The above proof analyzes Algorithm 2.1 in settings where the Bernstein Von Mises theorem is applicable. The complete list of regularity conditions required for the theorem are enumerated in Appendix A. In high level, for some neighborhood around the real value of θ they impose: the smoothness of the prior distribution; the smoothness of the density function $f(X, \theta)$, which will be satisfied if the norm of log-likelihood is bounded around θ; and the smoothness and differentiability of the Fisher information around θ, which will be true if the log-likelihood function has derivatives of order at least 3.

Figure 1 plots the log-likelihood function for the power law AE for a fixed exponent β and varying depths and a true value for θ chosen uniformly at random from $[0, \pi/2]$. The figure illustrates that the log-likelihood function is smooth in a neighborhood around the true value indicating that the regularity conditions for the Bernstein-Von Mises theorem are plausible in this setting. Adding noise may further regularize the log-likelihood functions and enforce the regularity conditions required for the Bernstein-Von Mises theorem.

The constants c, c' in the proof of Theorem can be determined explicitly if the regularity conditions are verified giving an explicit value for N_{shot}. In the absence of an explicit value, experimental results demonstrating convergence for a small value of N_{shot} can be taken as evidence that N_{shot} is a moderately small constant.

2.3 Power law amplitude estimation with noise

We perform here an analysis similar to the work in [25]. Noise provides a natural constraint on accessible circuit depths and noise models exponentially penalize larger depths, leading to an exponential decoherence of quantum states; therefore, exponentially more classical samples are needed to battle the noisy information. In this section, we explain this effect on our algorithm in the case of the depolarizing noise model.

Proposition 2.4 ([25])**.** *Assuming a depolarizing noise channel with a per-oracle-call rate of $\gamma \geq 0$, if measurements in the standard basis are performed on a quantum circuit of k sequential calls to the oracle U, the distribution of measurement outcomes is a Bernoulli random variable with success probability*

$$p = \frac{1}{2} - \frac{1}{2} e^{-\gamma(2k+1)} \cos(2(2k+1)\theta). \tag{2}$$

Let $\{m_k\}_{k=0,\ldots,K}$ be a schedule. The Fisher information with respect to the angle θ in the presence of depolarizing noise with parameter γ is given by

$$I_f(\theta) = 4N_{shot}\sum_{k=0}^{K}(2m_k+1)^2\frac{e^{-2\gamma(2m_k+1)}\sin^2(2(2m_k+1)\theta)}{1-e^{-2\gamma(2m_k+1)}\cos^2(2(2m_k+1)\theta)}, \tag{3}$$

see [25] for a proof. Recall that $\alpha = \sin^2(\theta)$ is the probability of success without depolarizing noise. For k such that γm_k becomes bigger than some large enough constant, the second term in the sum will be exponentially suppressed, the Fisher information will not increase significantly even if we keep increasing our depth. In practice, we do not want to use a Fisher information which is dependent on the parameter θ to be estimated; we can obtain a simple θ-independent upper bound to (3) by using the inequalities $\frac{1}{1-x} \leq 1+x$ and $e^x \geq 1+x$, which gives us:

$$I_f(\theta) \leq 4N_{\text{shot}}\sum_{k=0}^{K}(2m_k+1)^2 e^{-2\gamma m_k}\,. \tag{4}$$

Let us consider the power law schedule given by $m_k = \lfloor k^\eta \rfloor$ for $k = 0, 1, \ldots, K$ for $\eta = \frac{1-\beta}{2\beta}$ and $\beta \in (0, 1]$, as defined in Algorithm 2.1. We see that

$$\begin{aligned} I_f(\theta) &\leq \sum_{\gamma\lfloor k^\eta\rfloor \leq 1}(2\lfloor k^\eta\rfloor+1)^2 C + \text{ exponentially suppressed terms} \\ &\leq 10C\sum_{k=0}^{(1/\gamma)^{1/\eta}}(2k^\eta+1)^2 \leq C'/\gamma^{2+1/\eta} = C'/\gamma^{2/(1-\beta)}. \end{aligned}$$

Here C is a constant that can taken to be $C := \frac{4N_{shot}}{\sin^4(2\theta)}$ (if this diverges, add small random noise to θ). So by the Cramér-Rao bound, we have that

$$I_f(\theta) \leq \frac{C'}{\gamma^{2/(1-\beta)}} \Rightarrow \epsilon \geq c\gamma^{1/(1-\beta)}, \tag{5}$$

where $c := C'^{-1/2}$. This implies that given a noise level γ we cannot get an error rate ϵ smaller than $c\gamma^{1/(1-\beta)}$ by increasing the depth for a power law schedule with a *fixed* parameter β. Seeing this tradeoff from the other side, given a desired error rate ϵ and a noise level γ, we know how to pick the parameter β to match the lower bound in equation (5). Here we are assuming that the regularity conditions of Bernstein-von Mises also holds in the noisy setting, so that we are able to match the lower bound on ϵ.

If the noise level is smaller than the desired error rate, then we can pick the exponential schedule (for β equal to 0) since we can apply circuits of depth up to $O(1/\epsilon)$. For $\epsilon < \gamma$, we assume that we can match the lower bound in equation (5) so that β can be picked as follows.

Proposition 2.5. *Assume we are given as input the target error ϵ and noise level γ with $0 < \epsilon < \gamma < 1$ and that the regularity conditions of the Bernstein-von Mises hold in the noisy setting. The parameter β of the power law algorithm can be picked as*

$$\beta \geq 1 - \frac{\log\gamma}{\log\epsilon/c}$$

to achieve $\epsilon \leq c\gamma^{1/(1-\beta)}$ for the powerlaw schedule with parameter $\eta = \frac{1-\beta}{2\beta}$ (with high probability).

3 QoPrime: A number theoretic amplitude estimation algorithm

3.1 Preliminaries

We introduce the main technical tools needed for the QoPrime AE algorithm, these are the Chinese remainder theorem and the additive form of the Chernoff bounds.

Theorem 3.1. *[Chinese Remainder Theorem] Let $a_i \in \mathbb{N}, i \in [k]$ be pairwise coprime numbers such that for all $i \neq j$, $gcd(a_i, a_j) = 1$ and let $N = \prod_{i\in[k]} a_i$. Then for all $b_i \in [a_i], i \in [k]$ there is an efficient algorithm to find $M \in [N]$ such that $M \mod a_i = b_i$.*

Proof. First we provide the proof for $k = 2$. Applying the extended Euclidean algorithm we can find $u_1, u_2 \in \mathbb{Z}$ such that,

$$1 = gcd(a_1, a_2) = u_1 a_1 + u_2 a_2 \tag{6}$$

Then $M = (b_2 u_1 a_1 + b_1 u_2 a_2) \mod a_1 a_2$ satisfies $M = b_i \mod a_i$ for $i = 1, 2$.

For the proof of the general case, note that the $k - 1$ numbers $a_1 a_2, a_3, a_4, \cdots, a_k$ are coprime and by the above argument the constraints $M = b_i \mod a_i$ for $i = 1, 2$ is equivalent to $M = (b_2 u_1 a_1 + b_1 u_2 a_2) \mod a_1 a_2$. The procedure can therefore be repeated iteratively to find the desired M.

□

In this paper, we will be using relatively coprime moduli a_i, however the results can easily be adapted to a setting where the a_i are not pairwise coprime replacing $N = \prod_{i\in[k]} a_i$ by the least common multiple of the a_i. We define explicitly the bijection given by the Chinese remainder theorem as it is used later in the algorithm.

Definition 3.2. *Given pairwise co-prime moduli $N_i, i \in [k]$, let $N = \prod_{i\in[k]} N_i$. The function $CRT : \prod \mathbb{Z}_{N_i} \to Z_N$ on input $(M_1, M_2, \cdots, M_k)$ evaluates to the unique $M \in [N]$ given by Theorem 3.1 such that $M = M_i \mod N_i$.*

The second tool needed for the QoPrime algorithm is the additive form of the Chernoff bound and some supplementary calculations on the entropy of the binomial distribution.

Theorem 3.3 (Chernoff-Hoeffding Bound [16])**.** *Let X_i for $i \in [m]$ be i.i.d. random variables such that $X_i \in \{0, 1\}$ with expectation $E[X_i] = p$ and let $\epsilon > 0$ and $X = \frac{1}{m}\sum_{i\in[m]} X_i$. Then,*

1. $\Pr[X > p + \epsilon] \leq e^{-D(p+\epsilon||p)m}$.
2. $\Pr[X < p - \epsilon] \leq e^{-D(p-\epsilon||p)m}$.

where the relative entropy $D(x||y) = x \ln \frac{x}{y} + (1-x) \ln \frac{(1-x)}{(1-y)}$ where the relative entropy can be lower bounded using the inequality $D(x||y) \geq \frac{(x-y)^2}{2\max(x,y)}$ for all $x, y \in [0, 1]$.

The lower bound on the relative entropy is required to make the Chernoff bounds effective, we and derive a corollary that will be useful for the analysis of the QoPrime algorithm.

Corollary 3.4. *The following lower bound holds for the relative entropy for all $x, y \in [0, 1]$,*

$$D(x||y) \geq \frac{(x-y)^2}{2\max(1-x, 1-y)}$$

.

Proof. The relative entropy is symmetric under the substitution $(x, y) \to (1 - x, 1 - y)$, that is $D(x||y) = D((1-x)||(1-y))$. The result follows by applying the the inequality $D(x||y) \geq \frac{(x-y)^2}{2\max(x,y)}$ for $(x', y') = (1 - x, 1 - y)$. □

We also state the Multiplicative Chernoff bounds which will be used to compute some constants in the Qo-Prime algorithm.

Theorem 3.5 (Multiplicative Chernoff Bounds)**.** *Let X_i for $i \in [m]$ be independent random variables such that $X_i \in [0, 1]$ and let $X = \sum_{i\in[m]} X_i$. Then, $\Pr[|X - \mathbb{E}[X]| \geq \beta\mathbb{E}[X]] \leq e^{-\beta^2\mathbb{E}[X]/3}$ for $0 < \beta < 1$.*

3.2 The QoPrime AE algorithm

The QoPrime AE algorithm is presented as Algorithm 3.1. The implementation of the steps of the algorithm is described in greater detail below and the algorithm is analyzed to establish correctness and bound the running time.

An amplitude estimation algorithm has access to a quantum circuit U such that $U|0^t\rangle = \cos(\theta)|x,0\rangle + \sin(\theta)|x',1\rangle$ where $|x\rangle, |x'\rangle$ are arbitrary states on $(t-1)$ qubits. More precisely, let R_0 be the reflection in $|0^t\rangle$, that is $R_0|0^t\rangle = |0^t\rangle$ and $R_0|0^\perp\rangle = -|0^\perp\rangle$ and let S_0 be the reflection on $|0\rangle$ in the second register, that is $S_0|x,0\rangle = |x,0\rangle$ and $S_0|x,1\rangle = -|x,1\rangle$ for all $|x\rangle$. Like the standard AE algorithm, the QoPrime algorithm uses $(2k+1)$ sequential applications of the circuit for U to create the states,

$$|\phi_k\rangle := (UR_0U^{-1}S_0)^k U|0\rangle = \cos((2k+1)\theta)|x,0\rangle + \sin((2k+1)\theta)|x',1\rangle \tag{7}$$

An oracle call refers to a single application of the circuit U, the total number of oracle calls made is a measure of the running time for an AE algorithm. The maximum circuit depth for an amplitude estimation procedure is the number of sequential calls to U. The creation of a single copy of the state $|\phi_k\rangle$ in equation (7) requires $(2k+1)$ oracle calls.

The QoPrime algorithm is parametrized by integers (k,q) where $k \geq 2$ and $1 \leq q \leq (k-1)$, the parameter k is the number of moduli used for the reconstruction procedure while q determines the number of moduli that are grouped together. The algorithm starts by choosing nearby odd coprime integers $(n_1, n_2, \ldots, n_k)$, where each coprime is an integer approximately equal to $\left(\frac{\pi}{2\epsilon}\right)^{1/k}$. For the asymptotic purposes, all we need is that each coprime is therefore $n_i = \Theta(\epsilon^{-1/k})$. The k coprimes are partitioned into k/q groups, of size at most q. Let $\pi_i \subset [k]$ be the subset of coprimes in group i, and let $N_i = \prod_{j\in\pi_i} n_j$ for $i \in \lceil k/q \rceil$ be the product of the coprimes in each group. For each group i, we will sample at a depth of N/N_i.

Let $\theta = \frac{\pi M}{2N}$ be the true value of θ for $M = \lfloor M \rfloor + \{M\}$ where $M \in [0,N]$ and $N = \prod_{i\in[k]} n_i$. The QoPrime algorithm reconstructs estimates $\overline{M_i}$ that are within a $1/2$ confidence interval around $M \mod N_i$ with high probability. These estimates are constructed using samples from depth $N/N_i = O(1/\epsilon^{1-q/k})$ sequential calls to U to prepare the states in equation (7) followed by $O(N_i^2)$ measurements in the standard basis and reconstruction using the Chernoff bounds. For each group π_i of coprimes, measurements at depth N/N_i will give us information about the moduli $M \mod n_j$, for the coprimes in this particular group $n_j \in \pi_i$.

By repeating this for all groups π_i, we acquire information about all moduli $M \mod n_j$. These low-precision estimates are then combined using the Chinese remainder theorem to recover $M \mod N$, and therefore an ϵ-accurate estimation for $\theta = \frac{\pi M}{2N}$. As we shall see, there is a sign ambiguity associated with each group of coprimes, and a recursive call is needed to determine θ unambiguously. The QoPrime algorithm makes a total $\tilde{O}(1/\epsilon^{1+q/k})$ oracle calls for estimating θ within accuracy ϵ where $\tilde{O}$ hides factor that are logarithmic in k, q and the success probability for the algorithm. It trades off the maximum circuit depth needed for amplitude estimation against the total number of oracle calls.

Algorithm 3.1 The QoPrime Algorithm for Amplitude Estimation

Require: Accuracy ϵ for estimating θ and parameters (k, q) where $k \geq 2$ is the number of moduli and $1 \leq q \leq (k-1)$. Desired success probability p and value c such that $1 - 2ke^{-2c} > p$.
Require: Access to unitary U such that $U|0^T\rangle = \cos(\theta)|x, 0\rangle + \sin(\theta)|x', 1\rangle$.
Ensure: An estimate of θ within accuracy ϵ with probability at least p.
1: **if** $q/k > 1/3$ **then**
2: Using at most $\frac{24c}{\epsilon^{1+q/k}}$ samples from $U|0^T\rangle$, find θ' such that $|\theta' - \theta| \leq \frac{1}{2}\epsilon^{1-q/k}$ with probability at least $1 - e^{-2c}$.
3: **else**
4: Invoke the QoPrime algorithm recursively with accuracy $\frac{1}{2}\epsilon^{1-q/k}$ and parameters $(q', k') = (2q, k)$ such that $1 + q/k < 1 + q'/k' < \frac{(1+q/k)}{(1-q/k)}$ to obtain θ' such that $|\theta' - \theta| \leq \frac{1}{2}\epsilon^{1-q/k}$.
5: **end if**
6: Select k adjacent coprimes from the table in Section 3.3 starting at $\lfloor 1/\epsilon^{1/k} \rfloor$ with product closest to $\pi/(2\epsilon)$ in absolute value, and let $N = \prod_{i\in[k]} n_i$ be their product.
7: Partition $[k]$ into $K := \lceil k/q \rceil$ groups π_i of size at most q and let $N_i = \prod_{j\in\pi_i} n_j$.
8: Special case: If $|\theta' - \pi/4| \leq \frac{\min_i N_i}{4N}$, then invoke the QoPrime algorithm with the effective oracle U' using the exact amplitude amplification technique [6] as described in Lemma 3.9.
9: **for** i=1 **to** K **do**
10: Prepare $100cN_i^2$ copies of $|\phi_{(N-N_i)/2N_i}\rangle$ (see equation (7)) and measure in the standard basis.
11: Compute $\widehat{l_i} = \frac{2N_i}{\pi}\arccos(\sqrt{\widehat{p}})$ where $\widehat{p}$ is the observed probability of outcome 0.
12: **end for**
13: **for** all possible sign ambiguity resolutions $s \in \{-1, 1\}^K$ **do**
14: For all $j \in [K]$, compute $\overline{M_j} = s_j\widehat{l_j} \mod N_j$.
15: For all $j \in [K]$, compute $M_j = \lfloor \overline{M_j} + \beta_j \rfloor$ where $|\beta_j| \leq 1/4$ are such that the fractional parts $\{\overline{M_j} + \beta_j\} = \alpha$ for some $\alpha \in [0, 1]$.
16: Let $\overline{M} = CRT(M_1, M_2, \cdots, M_K)$, compute the sign-dependent estimate $\theta_s = \frac{\pi(\overline{M}+\alpha)}{2N}$.
17: **end for**
18: Output the choice of angle θ_s that minimizes $|\theta_s - \theta'|$ over all possible choices of $s \in \{-1, 1\}^K$.

The procedures used in the individual steps of the QoPrime Algorithm 3.1 are described next and correctness of the steps is established. Let $M = tN_i + l$ for some $t \in \mathbb{Z}$ and $0 \leq l \leq N_i$. Note that $(-1)^t M \mod N_i$ is the value being estimated in step 11 of the QoPrime algorithm making it necessary to determine the parity of t (cf. equation (8)). The approach taken to resolve this ambiguity is to compute all the estimates corresponding to the possible parities of t and comparing with an estimate of θ obtained recursively using the QoPrime algorithm with lower precision in steps 1-5 of the QoPrime algorithm. The first lemma in the analysis establishes the correctness of this procedure.

Lemma 3.6. *The recursive procedure of the QoPrime algorithm terminates in at most $O(\log k)$ iterations and outputs an additive error $\min_{i\in[k]}(N_i/2N)$ estimate for θ with probability at least $1 - 6e^{-2c}$.*

Proof. First we establish the correctness of the stopping condition in step 5. If $q/k \geq 1/3$ then $\frac{(1+q/k)}{2} \geq (1 - q/k)$ and thus by the multiplicative Chernoff bounds an additive error $\frac{\epsilon^{(1-q/k)}}{2}$ estimate for θ is obtained with $\frac{24c}{\epsilon^{(1+q/k)}}$ samples with probability at least $1 - e^{-2c}$ for some constant $c > 0$. It remains to show that the recursion terminates in at most $O(\log k)$ steps and that the total number of oracle calls used in the recursive step is upper bounded by $O(1/\epsilon^{1+q/k})$.

The recursive call to the QoPrime algorithm in step 4 uses $O(1/\epsilon'^{(1+q'/k')}) = O(1/\epsilon^{(1-q/k)(1+q'/k')})$ total oracle calls. As $(1 - q/k)(1 + q'/k') < (1 + q/k)$ the total number of oracle calls used by steps 4-8 is at most $\tilde{O}(1/\epsilon^{1+q/k})$. The extra oracle calls used for these steps do not change the asymptotic number of oracle calls used by the QoPrime algorithm. Further, as $1+q/k < 1+2q/k < (1+q/k)/(1-q/k)$ it is always feasible to choose $q' = 2q, k' = k$ and with this choice the stopping condition $q/k \geq 1/3$ in step 5 will hold after at most $O(\log k)$ iterations.

The success probability for these steps is the same as the success probability for the final

recursive call to the QoPrime algorithm in step 4, which uses at most $\lceil k/q \rceil \leq 3$ moduli by the stopping condition. Further, Theorem 3.10 shows that the the QoPrime algorithm succeeds with probability at least $1 - 2ke^{-2c}$, implying that the recursive procedure succeeds with probability at least $1 - 6e^{-2c}$. □

The analysis shows that the QoPrime correctly estimates θ apart from one exceptional case when the angle is close to $\pi/4$ that is dealt separately in step 8. This case will be discussed later in the analysis of the sign resolution procedure.

Step 10 of the QoPrime algorithm computes a quantum state which can be measured to sample from a Bernoulli random variable with success probability $p = \cos^2(\frac{(tN_i+l)\pi}{2N_i})$. Step 11 computes an estimate $\widehat{l} = \frac{2N_i}{\pi}\arccos(\sqrt{\widehat{p}})$ where $\widehat{p}$ is the observed probability of outcome 0. The analysis below shows that $|\widehat{l} - (-1)^t l \mod N_i| \leq 0.25$ with high probability.

In order to analyze these steps, we begin with the observation that if $p = \widehat{p}$ then $\widehat{l} = (-1)^t l \mod N_i$.

$$\frac{2N_i}{\pi}\arccos(\sqrt{p}) = \begin{cases} \frac{2N_i}{\pi}\arccos\left(\cos(\frac{l\pi}{2N_i})\right) & [\text{if } t = 0 \mod 2] \\ \frac{2N_i}{\pi}\arccos\left(\sin(\frac{l\pi}{2N_i})\right) & [\text{if } t = 1 \mod 2] \end{cases} = (-1)^t l \mod N_i. \tag{8}$$

The next Lemma quantifies the error made by the algorithm in approximating $(-1)^t l \mod N_i$ using the Chernoff bounds to bound the difference between $\widehat{p}$ and p.

Lemma 3.7. *Given integer N and $m = 100cN^2$ samples, steps 10-11 of the QoPrime algorithm finds an estimate such that $|\widehat{l} - (-1)^t l \mod N| \leq 0.25$ with probability at least $1 - 2e^{-2c}$.*

Proof. Define $F : [0,1] \to [0,N]$ as $F(p) = \frac{2N}{\pi}\arccos(\sqrt{p})$ so that $F(p) = (-1)^t l \mod N$ by equation (8) and $F(\widehat{p}) = \widehat{l}$. It is sufficient to show that $\Pr[|F(p) - F(\widehat{p})| \geq 0.25] \leq e^{-2c}$ for all $p \in [0,1]$.

The inverse function $G : [0,N] \to [0,1]$ such that $G(y) = \cos^2(\frac{\pi y}{2N})$ is monotonically decreasing. Let $F(p) = y$, then $F(p) - F(\widehat{p}) \geq 0.25$ is equivalent to $(y - 0.25) \geq F(\widehat{p})$, that is $G(y - 0.25) < \widehat{p}$. Applying the additive Chernoff bound,

$$\Pr[\widehat{p} > G(y - 0.25)] \leq e^{-mD(G(y-0.25)||G(y))} \tag{9}$$

The analysis splits into two cases, where the relative entropy is lower bounded using either the inequality in Theorem 3.3 or Corollary 3.4. Let $A = \frac{\pi y}{2N}$ and $B = \frac{\pi(y-0.25)}{2N}$, first consider the case $y > N/2$ or equivalently $A > \pi/4$,

$$\begin{aligned} mD(G(y-0.25)||G(y)) &\geq m\frac{(\cos^2(A) - \cos^2(B))^2}{2\cos^2(B)} = m\frac{(\cos(2A) - \cos(2B)^2}{8\cos^2(B)} \\ &= m\frac{\sin^2(A+B)\sin^2(A-B)}{2\cos^2(B)} \\ &= \frac{m\sin^2(A-B)}{2}(\sin(A) + \cos(A)\tan B))^2 \\ &\geq 50cN^2\sin^2(\frac{\pi}{8N}) \geq 50cN^2\frac{\pi^2}{128N^2} > 2c. \end{aligned} \tag{10}$$

Note that for the last two steps in the computation, the inequality $(\sin(A) + \cos(A)\tan B))^2 > 1$ for $A > \pi/4$ and $A - B = \frac{\pi}{8N}$ was used along with the lower bound $\sin(x) \geq x/2$ for $x \in [0, \pi/2]$. For the case $y < N/2$ we carry out a similar computation using Corollary 3.4 to lower bound the

relative entropy,

$$\begin{aligned} mD(G(y-0.25)||G(y)) &\geq m\frac{(\cos^2(A)-\cos^2(B))^2}{2\sin^2(A)} \\ &= m\frac{\sin^2(A+B)\sin^2(A-B)}{2\sin^2(A)} \\ &= \frac{m\sin^2(A-B)}{2}(\cos(B)+\sin(B)\cot A))^2 \\ &\geq 50cN^2\sin^2(\frac{\pi}{8N}) > 2c. \end{aligned} \tag{11}$$

For the last steps in the computation, the inequality $(\cos(B)+\sin(B)\cot(A))^2 > 1$ for $A < \pi/4$ and $A - B = \frac{\pi}{8N}$ was used along with the lower bound $\sin(x) > x/2$ for $x \in [0, \pi/2]$.

Similarly, $\widehat{F}(\widehat{p}) - F(p) \geq 0.25$ is equivalent to $G(y+0.25) > \widehat{p}$ and the probability can be bounded using the additive Chernoff bound,

$$\Pr[\widehat{p} < G(y+0.25)] \leq e^{-mD(G(y+0.25)||G(y))} \tag{12}$$

Further, $mD(G(y+0.25)||G(y)) \geq 2c$ for all $y \in [0, N]$ with probability at least $1 - e^{-2c}$, this follows from calculations similar to the ones in equations (10) and (11) with denominators $(\cos^2(B), \sin^2(A))$ replaced by $(\cos^2(A), \sin^2(B))$. From the union bound, it follows that $|\widehat{l} - (-1)^t l \mod N| \leq 0.25$ with probability at least $1 - 2e^{-2c}$.

□

The remaining steps of the algorithm resolves the sign ambiguity in the estimates obtained in steps 10-11 by comparing against the recursive estimate. The following lemma bounds smallest possible difference between possible reconstructions produced by the Chinese Remainder Theorem up to sign ambiguities.

Lemma 3.8. *For an integer $0 \leq M < N$, consider the Chinese remainder theorem bijection $M = CRT(m_1, \ldots, m_K)$, where $M \equiv m_i \mod N_i$. Let $N_1 < \cdots < N_K$ be the ordered K odd coprimes.*

The only possible integers $M' \neq M$ such that $|M - M'| < \min N_i$ such that M' can be obtained via CRT by changing some signs on the moduli, namely $M' = CRT(\pm m_1, \pm m_2, \ldots, \pm m_k)$ are contained in an interval around $N/2$,

$$M' \in \left[\frac{N-N_1}{2}+1,\ \frac{N+N_1}{2}-1\right] \cap \mathbb{Z}\,. \tag{13}$$

Proof. The CRT function is 'continuous' in the sense that if $CRT(m_1, \ldots, m_K) = M$ then $CRT(m_1 + a, \ldots, m_K + a) = M + a \mod N$ for any displacement a. We therefore examine the displacements of $-N_1 + 1 \leq a \leq N_1 - 1$ around M and seeing if they can be made to match with a different sign resolution. In other words, there should be such a choice of a such that $m_i + a \equiv \pm m_i \mod N_i$ for all the moduli N_i for $i \in [k]$. Since $|a| < N_1$ and N_1 can be assumed to be the smallest modulus, we see that this is not possible unless all the signs are flipped, i.e. $M' = \mathrm{CRT}(-m_1, \ldots, -m_K)$ and $m_i + a \equiv -m_i \mod N_i$, for all $i \in [K]$.

Second, since the N_i are all odd, there are two possibilities for possible solutions (M, M') (where M is assumed to be larger) in terms of the parity of a:

- If a is even, then $(M, M') = (N - a/2, a/2)$ are the unique solutions to the equations $2m_i + a = 0 \mod N_i$. In this case, $|M' - M| = N - a \geq N - N_1 > N_1$, contradicting the hypothesis in the theorem statement.
- If a is odd, then $(M, M') = ((N + a)/2, (N - a)/2)$ are the unique solutions to the equations $2m_i + a = 0 \mod N_i$. For the QoPrime algorithm, this case is resolved by the depth 0 measurements carried out in the recursive step.

□

It is this second set of $N_1 - 1$ special points (out of the N possibilities for M) which presents an issue to our algorithm from the point of view of resolving the sign ambiguity. These problematic angles lie in a small interval of $\theta \in \left[\frac{\pi(N-N_1)}{4N}, \frac{\pi(N+N_1)}{4N}\right]$ around the midpoint $\pi/4$. We use the technique of exact amplitude amplification which allows us to map an angle away from a known interval, should it happen to be there. Note that the size of the problematic interval is of the order of the estimation accuracy for the recursive estimate in step 1-5, so only a small fraction of points need to be handled.

Lemma 3.9. *[Exact amplitude amplification [6]] If $\theta \in \left[\frac{\pi(N-N_1)}{4N}, \frac{\pi(N+N_1)}{4N}\right]$, then appending an extra qubit in the state $(|0\rangle + |1\rangle)/\sqrt{2}$ and selecting on $|00\rangle$ results in an oracle $U'|0^t\rangle = \cos(\theta')|x, 00\rangle + \sin(\theta')|x', (00)^\perp\rangle$ where $|\pi/4 - \theta'| \geq 0.07\pi$ for $N_1 \leq 0.04N$.*

Proof. If the angle $\theta = \pi/4$, then $\cos(\theta') = \cos(\theta)/\sqrt{2} = 1/2$ and thus $\theta' = \pi/3$ and the difference $|\pi/4 - \theta'| \geq \pi/12 \geq 0.08\pi$. The forbidden interval is asymptotically smaller than $[\pi/4 - 0.01, \pi/4 + 0.01]$ for large enough N (that is $N_1 = O(N^{q/k}) \leq 0.04N$ for large enough N), the result follows. □

We next describe the procedure in steps 15-17 of the QoPrime algorithm for estimating the values M_i that will be used in the Chinese Remainder theorem. Define the confidence intervals $A_i = [\{\overline{M_i}\} - 0.25, \{\overline{M_i}\} + 0.25]$ corresponding to all the estimates $\{\overline{M_i}\}$ produced by the QoPrime algorithm. Let $I = \bigcap_i A_i$ be the intersection of the A_i. Applying the union bound and Lemma 3.7 it follows that I is non empty and the fractional part $\{M\} \in I$ with probability at least $1 - 2ke^{-c}$. Step 9 of the QoPrime algorithm is therefore able to find $\alpha \in I$. In Step 10, from α one finds $\beta_i \in [-0.25, 0.25]$ such that $\{\overline{M_i} + \beta_i\} = \alpha$ and then the value M_i is computed as $M_i = \lfloor \overline{M_i} + \beta_i \rfloor$. It remains to show that using the Chinese Remainder Theorem on these values produces an estimate with error ϵ with high probability.

Theorem 3.10. *The estimate output by the QoPrime algorithm is within additive error ϵ of the true value with probability at least $1 - 2ke^{-2c}$.*

Proof. Let $CRT(m_1, \ldots, m_k)$ be the function in Definition 3.2 denote the unique integer $m \mod N$ such that $m = m_i \mod N_i$. The Chinese remainder theorem shows that this function is invertible, specifically $CRT^{-1}(m) = (m \mod n_1, \ldots, m \mod n_k)$. The CRT function is continuous in the following sense $CRT^{-1}(m+a) = (m+a \mod n_1, \ldots, m+a \mod n_k)$, or equivalently $CRT(m_1 + a, m_2 + a, \ldots, m_k + a) = CRT(m_1, \ldots, m_k) + a$ for $a \in \mathbb{Z}$. For $M = \lfloor M \rfloor + \{M\}$, we have

$$M = CRT(\lfloor M \rfloor \mod N_1, \cdots, \lfloor M \rfloor \mod N_K) + \{M\} \tag{14}$$

The QoPrime algorithm instead outputs the reconstructed estimate,

$$\overline{M} = CRT(\lfloor \overline{M_1} + \beta_1 \rfloor, \cdots, \lfloor \overline{M_K} + \beta_i \rfloor) + \alpha \tag{15}$$

Let us establish next the correctness of the sign resolution procedure. The reconstruction θ_s for the correct sign pattern is within distance $N_1/2N$ of the estimate θ' produced by Lemma 3.6 with probability $1 - 6e^{-2c}$. Lemma 3.8 establishes that no other sign pattern can have comparable accuracy except for the exceptional case when the θ is close to $\pi/4$ which is handled by the exact amplitude amplification technique in Lemma 3.9. It follows that the sign resolution procedure in step 18 is correct works and it can be assumed for the analysis that the signs t are correct.

By Lemma 3.7 and the choice of β_j in step 15 of the Algorithm it follows that $|\overline{M_j} + \beta_j - (\lfloor M \rfloor \mod N_j + \{M\})| = \gamma < 0.5$ for all $j \in [k]$ where γ is independent of j with probability at least $1 - 2ke^{-2c}$. The accuracy of the estimates implies that $|\lfloor \overline{M_j} + \beta_j \rfloor - (\lfloor M \rfloor \mod N_j)| \leq 1$ for all $j \in [k]$. By continuity of the Chinese remainder theorem, the reconstruction error $|\overline{M} - M| \leq 1 + |\alpha - \{M\}| \leq 1.5$. The QoPrime algorithm therefore outputs an estimate $\frac{\pi\overline{M}}{2N}$ that differs from the true value $\frac{\pi M}{2N}$ by at most $\frac{\pi}{N} \leq \epsilon$.

□

3.3 Choosing the parameters

In this section, we further detail the choice of the parameters for the QoPrime algorithm. The small-ϵ asymptotics of the QoPrime algorithm rely on finding coprime moduli of similar magnitude and product of order $\Theta(\epsilon^{-1})$. To this effect, we formulate the following lemma:

Lemma 3.11. *[9] Given a fixed integer $k \geq 2$ and $N \in \mathbb{R}$, we can find k mutually coprime integers $n_1(N) < n_2(N) < \cdots < n_k(N)$ such that $\lim_{N\to\infty} \frac{n_1(N)\ldots n_k(N)}{N} = 1$ and $\lim_{N\to\infty} \frac{n_k(N)}{n_1(N)} = 1$.*

This lemma follows from the sub-linear scaling of the number of coprimes that can fit inside an interval of a given size, as studied in [9]. In practice, we pre-compute a table of k adjacent odd coprimes starting at each odd integer, stopping at large enough values of k and n_1 according to the target precision ϵ. It suffices to build the table for $k \leq 12$ and $n_1 \approx 10^5$ to achieve approximation error $\epsilon = 10^{-10}$, both with and without noise. Given the table, target precision ϵ and an integer $k \geq 2$, the implementation chooses k adjacent coprimes from the table starting at $\lfloor 1/\epsilon^{1/k} \rfloor$ with product closest to $\frac{\pi}{\epsilon}$ in absolute value. Figure 1 compares the table used in practice to the theoretical guarantees in Lemma 3.11.

We can summarize the asymptotics of the algorithm in the following table:

$$\begin{array}{lll} \text{parameters:} & \epsilon, k, q & \\ \text{coprimes:} & n_1, \ldots, n_k & = \Theta(1/\epsilon^{1/k}) \\ \text{sampling depth:} & \max_i \frac{N}{N_i} & = \Theta(1/\epsilon^{1-q/k}) \\ \text{oracle calls:} & O\left(\sum_{i=1}^{\lceil k/q \rceil} N_i^2 \times \frac{N}{N_i}\right) & = O\left(\left\lceil \frac{k}{q} \right\rceil 1/\epsilon^{1+q/k}\right) \end{array} \tag{16}$$

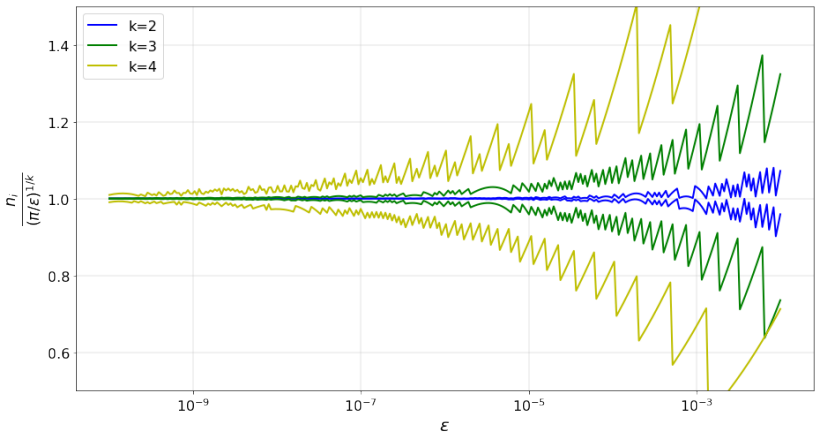

Figure 1: Convergence of the coprime finding routine. The algorithm finds k adjacent coprimes n_1, n_2, ..., n_k such that their product $N = n_1 \ldots n_k$ is close to π/ϵ. Both the smallest coprime n_1 (approaching 1 from below) and the largest coprime n_k (approaching 1 from above) are shown to converge to $(\pi/\epsilon)^{1/k}$ matching the convergence in Lemma 3.11 for several values of k.

Note that a bound onthe number of total classical sampling sessions used in the algorithm can be obtained easily using our geometric schedule described in Section 3.2. Since for every recursive call the ratio q/k is divided by two, we have that for the nth recursive step there will be $\lceil k/2^n q \rceil$ sampling sessions. Therefore the total number of sampling sessions can be bounded by the simple geometric sum $\lceil k/q \rceil + \lceil k/2q \rceil + \lceil k/2^2 q \rceil + \cdots \leq 2\lceil k/q \rceil$. While simple, this fact is important because it tells us that the complexity of the algorithm can be described only by the first level of the recursion, up to a small order-one factor. Therefore, in choosing the algorithm's parameters such as k and q, we will only optimize over the behavior of the first level of the recursion. This analysis means that, for a given target precision ϵ and overall failure probability δ, the total number of oracle calls can be bounded by:

$$\mathcal{N}(k, q, \epsilon, \delta) \leq C \times \left\lceil \frac{k}{q} \right\rceil \times \frac{1}{\epsilon^{1+q/k}} \times \log\left(\frac{4}{\delta}\left\lceil \frac{k}{q} \right\rceil\right) \tag{17}$$

The constant C which tightens this bound does not depend on the parameters of the algorithm, in fact experimental results provide evidence that C is a small constant and that in practice, we can expect $C < 10$. (see Figure 9).

In practice, given target precision ϵ and accepted failure probability δ, we can find optimal values of k and q such that the oracle calls in (17) is minimized. In this noiseless scenario, it can be shown that the optimal parameter q is always 1 (i.e. it is best to access the largest allowed depth). Minimizing the oracle call number in (17) by choosing k leads to (ignoring the subleading contribution from the failure probability δ):

$$k^*(\epsilon) \approx \log \frac{1}{\epsilon} \tag{18}$$

Plugging this back into (17) leads to an asymptotic dependency of oracle calls of $1/\epsilon$, up to logarithmic factors, as expected in the quantum regime:

$$\min_{(q,k)} \mathcal{N} = O\left(\epsilon^{-1} \log \epsilon^{-1} \log\left(\frac{4 \log \epsilon^{-1}}{\delta}\right)\right) \tag{19}$$

3.4 QoPrime algorithm with noise

In this section, we study the performance of the QoPrime AE algorithm under the same depolarizing noise model as we studied for the power law AE algorithm. Inverting the noisy probabilistic model (2), the classical inference problem in the algorithm, when sampling at depth N/N_i, becomes:

$$\overline{M_j} \equiv \pm \frac{2N_i}{\pi} \arccos \sqrt{\frac{1}{2} + e^{\gamma N/N_i}\left(\hat{p} - \frac{1}{2}\right)} \mod N_i \tag{20}$$

As before, we can use this relation to translate a confidence interval on the coin toss probability $\hat{p}$, computed by classical postprocessing of measurement samples, to a confidence interval on $\overline{M_j}$. The difference in the noisy case is the exponential stretch factor of $e^{\gamma N/N_i}$ enhancing the angle confidence interval. Since a classical confidence interval shrinks as the square root of the number of samples, the required number of samples will pick up a factor of $e^{2\gamma N/N_i}$ under this noise model in order to guarantee the noiseless confidence intervals. Similar to the noiseless algorithm (16), we can therefore summarize the asymptotics of the noisy algorithm as follows:

$$\begin{array}{lll} \text{parameters:} & \epsilon, k, q, \gamma & \\ \text{coprimes:} & n_1, \ldots, n_k & = \Theta(1/\epsilon^{1/k}) \\ \text{sampling depth:} & \max_i \frac{N}{N_i} & = \Theta(1/\epsilon^{1-q/k}) \\ \text{oracle calls:} & \Theta\left(\sum_{i=1}^{\lceil k/q \rceil} N_i^2 e^{2\gamma N/N_i} \frac{N}{N_i}\right) & = \Theta\left(\left\lceil \frac{k}{q} \right\rceil \frac{1}{\epsilon^{1+q/k}} e^{2\gamma (\frac{\pi}{2\epsilon})^{1-q/k}}\right) \end{array} \tag{21}$$

Therefore, for a given target precision ϵ, depolarizing noise level γ, and overall failure probability δ, the total number of oracle calls scales as:

$$\mathcal{N}(k, q, \gamma, \epsilon, \delta) \leq C \times \left\lceil \frac{k}{q} \right\rceil \times \frac{1}{\epsilon^{1+q/k}} \times \exp\left(2\gamma \left(\frac{\pi}{2\epsilon}\right)^{1-q/k}\right) \times \log\left(\frac{4}{\delta}\left\lceil \frac{k}{q} \right\rceil\right) \tag{22}$$

where C is the same constant overhead as in equation (17). Given target precision ϵ, noise level γ, and accepted failure probability δ, we can find optimal values of k and q such that the number of oracle calls in (22) is minimized. See also Figure 2 for the behaviour of the number of oracle calls for different values of k and q.

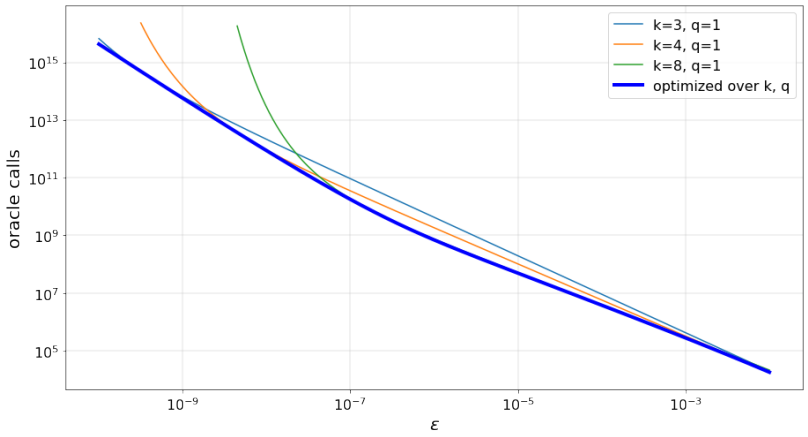

Figure 2: The behavior of the oracle call dependency in (22) for several values of parameters k and q, and for fixed noise level $\gamma = 10^{-5}$ and probability of failure $\delta = 10^{-5}$. When taking the minimum over the family of curves parametrized by all valid k and q (here assumed continuous for simplicity), we obtain the envelope of the optimal QoPrime algorithm (thick blue line). This emergent behavior smoothly interpolates between a classical $1/\epsilon^2$ scaling in the noise-dominated region $\epsilon \ll \gamma$ and a quantum scaling $1/\epsilon$ in the coherent region $\epsilon \gg \gamma$.

Optimizing over k and q in this manner is the step which ensures that the effective scaling of oracle calls as a function of ϵ is always between the classical scaling of $1/\epsilon^2$ and the quantum scaling $1/\epsilon$ (see Figure 3). Specifically, this can be formulated as a bound on the instantaneous exponent:

$$-2 \leq \lim_{\epsilon \to 0} \frac{d \inf_{k,q} \log \mathcal{N}(\epsilon, \gamma, \delta, k, q)}{d \log \epsilon} \leq -1 \tag{23}$$

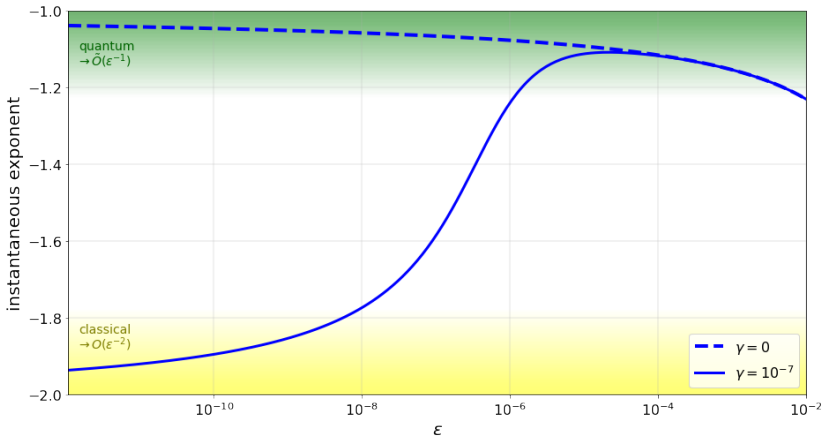

Figure 3: The transition from coherent to noise-dominated as measured by the instantaneous exponent $\frac{d \log \mathcal{N}}{d \log \epsilon}$, where $\mathcal{N}$ is the number oracle calls optimized over parameters k and q.

It can be shown that in the noise-dominated limit, the optimal parameter q tends to its upper bound $q = k - 1$ (corresponding to the shallowest accessible circuits). Using this observation, we can analytically study the optimization over k using the dependency in (22) and obtain that the optimal k parameter will have the form (in a continuous approximation):

$$k^*(\epsilon, \gamma) \approx \frac{\log \frac{\pi}{2\epsilon}}{\log \frac{\log \frac{1}{\epsilon}}{2\gamma \log \frac{\pi}{2\epsilon}}} \tag{24}$$

This allows us to study the asymptotic dependency of oracle calls on the target precision ϵ analyt-

ically; extracting the low-ϵ limit yields:

$$\lim_{\gamma \gg \epsilon} \mathcal{N}(\epsilon, \gamma, \delta) \leq 4Ce\,\gamma\,\epsilon^{-2} \log\left(\frac{4}{\delta}\right) , \tag{25}$$

where C is the constant prefactor introduced in (22). This classical-limit curve can be used to compare the asymptotic runtime of our algorithm to classical Monte Carlo techniques.

Outside of the two noise limits, specifically noise-dominated (25) and noiseless (19), the optimal parameters k and q depend non-trivially on the problem, and they can be found numerically. Example (k, q) optimal trajectories obtained by optimizing (22) are shown in Figure 4.

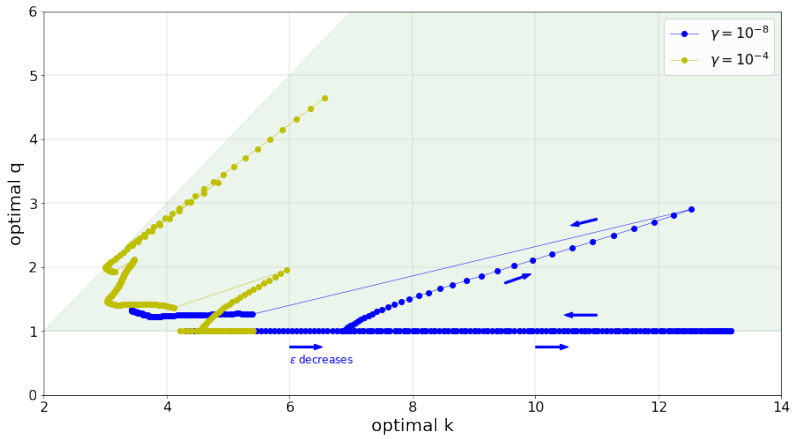

Figure 4: The trajectory of optimal k, q parameters chosen by minimizing the asymptotic dependency in (22), for two different noise levels. The optimization is over continuous k, q for simplicity. The green region marks the valid parameter region $k \geq 2$, $1 \leq q \leq k-1$. The arrow shows the direction of the optimal parameters as the target precision ϵ is being lowered from $\epsilon = 10^{-3}$ to $\epsilon = 10^{-10}$. While $\epsilon \ll \gamma$ (i.e. noiseless regime), we have that $q = 1$.

4 Empirical results

In this section, we present empirical results for the power law and the QoPrime AE algorithms and compare them with state of the art amplitude estimation algorithms [11]. The experimental results validate the theoretical analysis and provide further insight in the behaviour of these low depth algorithms in noisy regimes.

4.1 The power law AE

Figure 5 compares the theoretical and empirically observed scaling for the number of oracle calls N as a function of the error rate ϵ in the power law AE algorithm in the absence of noise, i.e. $\gamma = 0$. We numerically simulated the power law AE algorithm for randomly chosen $\theta \in [0, \pi/2]$ and with $m_k = \lfloor k^{\frac{1-\beta}{2\beta}} \rfloor$ for fixed values of parameter $\beta \in \{0.455, 0.714\}$, which make the exponent be $\{0.2, 0.6\}$ respectively. We also provide the extremal cases of $\beta \in \{0, 1\}$. The experimental results agree closely with the predictions of the theoretical analysis.

Figure 6 shows the scaling of the power law AE algorithm under several noise levels. Here, the parameter β is chosen adaptively, based on the error rate ϵ and noise level γ. For target errors below the noise level, we can use the exponential schedule to get the optimal quantum scaling. After this threshold, we use the power law schedules with exponents chosen as in Proposition 2.5. The result is that for these smaller target errors, the scaling is in between the optimal quantum and the classical scaling.

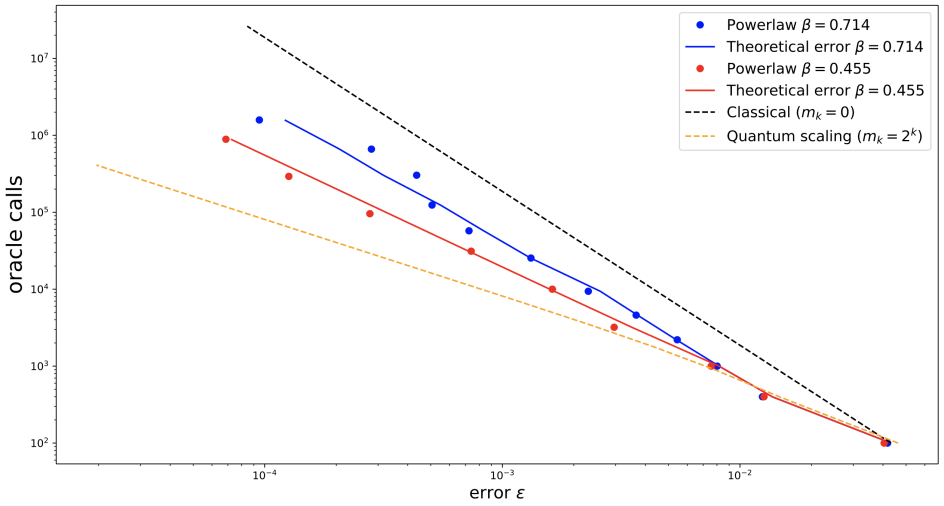

Figure 5: Performance of the power law AE algorithm in theory (solid) and practice (dots) using the schedule $m_k = \lfloor k^{\frac{1-\beta}{2\beta}} \rfloor$ for values $\beta = 0.455$ (red) and $\beta = 0.714$ (blue), for different error rates ϵ. The true value is θ^* is chosen at random. We took the number of shots $N_{shot} = 100$. Applying linear regression to these experimental data points, gives slopes -1.718 and -1.469, whereas the theoretical slopes are -1.714 and -1.455 for the blue and red points respectively.

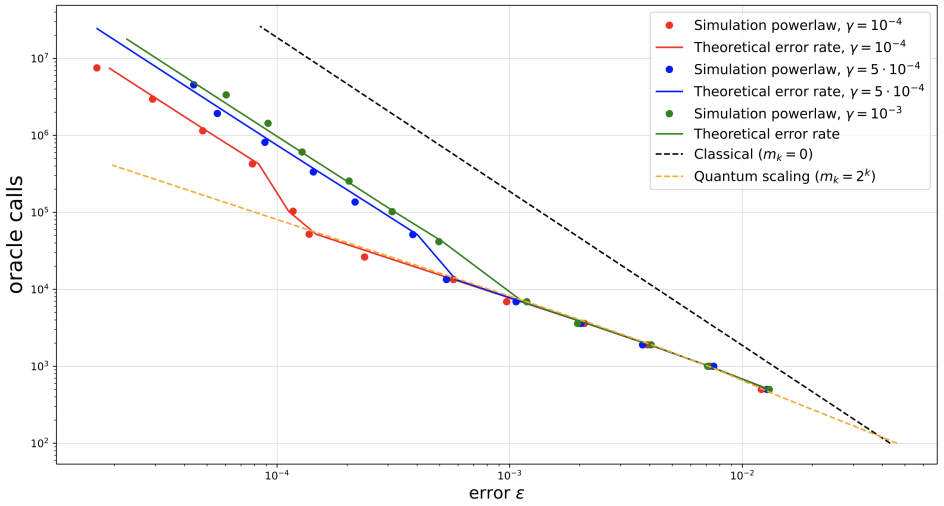

Figure 6: Performance of the power law AE algorithm in theory (solid) and practice (dots) using power law schedules where the parameter β is optimized for given ϵ and γ. Also the classical and quantum scalings are plotted for comparison. For small target errors, we obtain the optimal quantum scaling, while for smaller target errors we use power law exponents using Proposition 2.5. The result is that the scaling approaches the classical scaling as the target error goes to 0.

4.2 The QoPrime algorithm

The parameters k and q for the QoPrime algorithm are chosen by optimizing over the Chernoff upper-bounds obtained in Lemma 3.7 and described in (21). Figure 7 shows the theoretical upper bounds and the empirically observed number of oracle calls as a function of the accuracy ϵ for different noise rates. The algorithm in practice performs better than the theoretical bounds as it computes the confidence intervals using exact binomial distributions as opposed to the Chernoff bounds in the theoretical analysis.

Figure 8 plots the maximum oracle depth as a function of the target precision ϵ for the QoPrime algorithm in noiseless and noisy settings, as well as for the IQAE algorithm. Finally, Figure 9

provides empirical estimates for the constant factor C for the QoPrime algorithm in noisy settings. The observed value of C is a small constant and the simulations show that $C < 10$ over a wide range of ϵ and noise rates that cover most settings of interest.

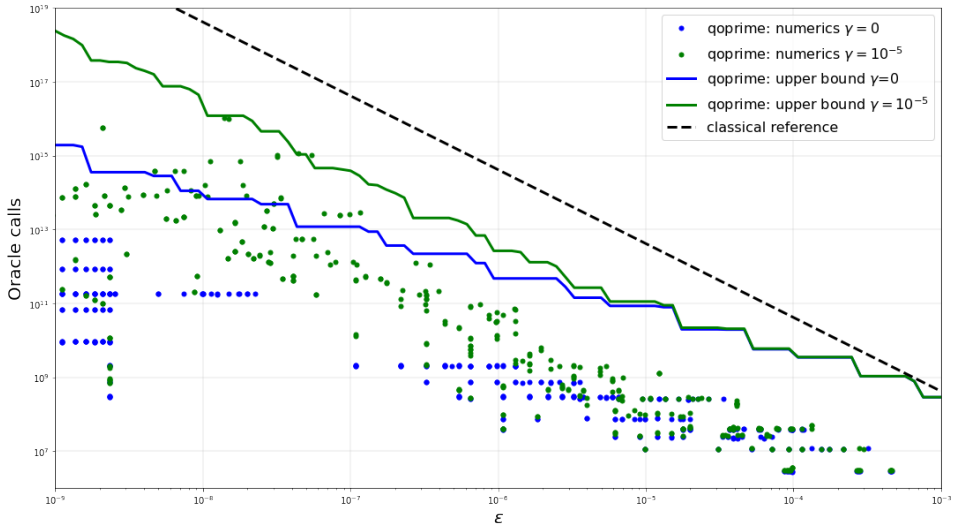

Figure 7: Performance of the QoPrime algorithm under two noise levels, and across various choices for the true angle θ. Shown are both theoretical upper bounds (solid) and exact simulated oracle calls (dots) for two scenarios: noiseless (blue), and depolarizing rate $\gamma = 10^{-5}$ (green). For each target precision, 20 values of the true angle spanning the $[0, \pi/2]$ interval have been selected to generate the samples; the horizontal axis represents realized approximation precision. The classical Monte Carlo curve (black) is obtained by assuming noiseless classical sampling from a constant oracle depth of 1. We see the curve follow a quantum ϵ^{-1} scaling for small errors $\epsilon \gg \gamma$, which transitions into a classical ϵ^{-2} dependency when the precision is much smaller than the noise level ($\epsilon \ll \gamma$).

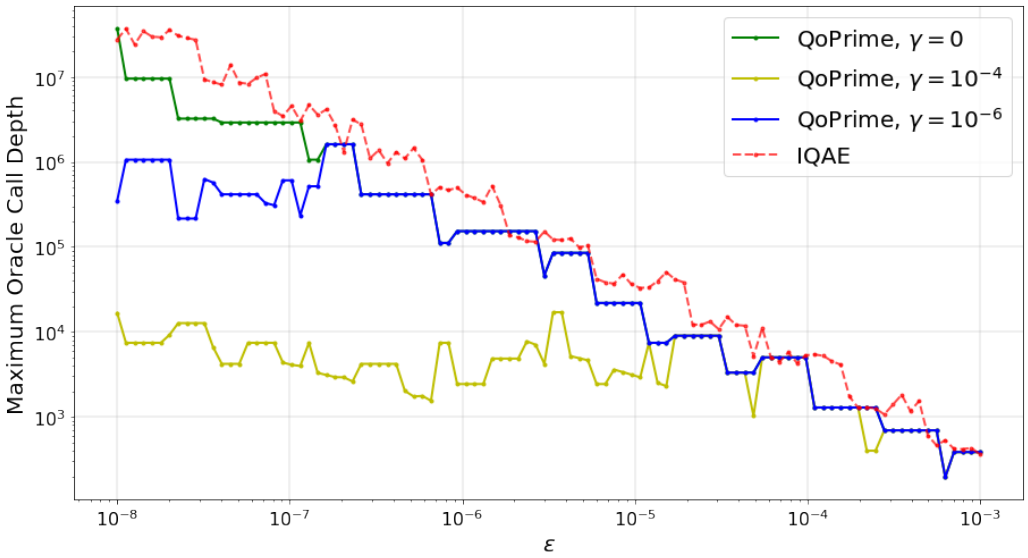

Figure 8: Maximum oracle depth as a function of the target precision ϵ. The noiseless QoPrime algorithm and the IQAE algorithm in [11] both have a similar scaling of depth as $O(\epsilon^{-1})$. However, introducing a depolarizing noise level γ provides a bound for the depth required by the QoPrime algorithm. Specifically, the QoPrime algorithm will not access depths higher than the noise scale γ, which would correspond to exponentially suppressed confidence intervals, and require exponentially more samples.

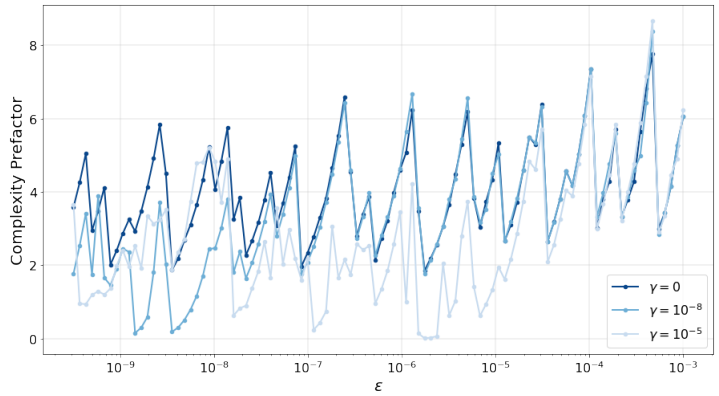

Figure 9: Empirical values of the constant prefactor C, as defined in (22) above, on the target precision ϵ for an arbitrary value of the true angle θ and failure probability $\delta = 10^{-5}$.

4.3 Benchmarking

Last we compare the performance of the Power law and QoPrime AE algorithms against the state of the art amplitude estimation algorithm IQAE [11].

Figure 11 plots the performance of the Power Law, the QoPrime and the IQAE in noisy settings, where performance is measured by the number of oracle calls for target accuracy ϵ. The plot emphasizes the advantage of the Power law and the QoPrime over algorithms such as IQAE, which require access to a full circuit depth of $O(1/\epsilon)$. In this scenario, this large depth is exponentially penalized by the depolarizing noise by requiring an exponentially large number of classical samples to achieve a precision below the noise level. In comparison, the power law and the QoPrime AE algorithms transition smoothly to a classical estimation scaling and do not suffer from an exponential growth in oracle calls. The Power law AE algorithm has the best practical performance according to the simulations.

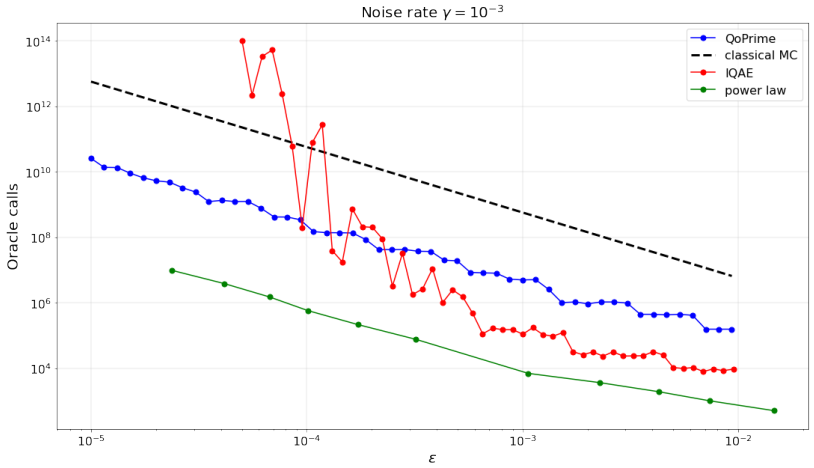

Figure 10: Comparison of the two algorithms introduced in this work (Power law and QoPrime) against the Iterative Quantum Amplitude Estimation algorithm (IQAE) introduced in [11]. A noise level of $\gamma = 10^{-3}$ is used for all three.

Acknowledgements: This work is a collaboration between Goldman Sachs and QCWare, it was carried out during TGT and FL's internships at Goldman Sachs. We acknowledge helpful discussions with Adam Bouland, Nikitas Stamatopolous, Rajiv Krishnakumar, and Paul Burchard. We thank an anonymous reviewer for helpful comments and for pointing out errors in a previous version of the QoPrime algorithm.

References

[1] S. Aaronson and P. Rall, "Quantum approximate counting, simplified," in *Symposium on Simplicity in Algorithms*. SIAM, 2020, pp. 24–32. [Online]. Available: https://dx.doi.org/10.1137/1.9781611976014.5

[2] D. S. Abrams and C. P. Williams, "Fast quantum algorithms for numerical integrals and stochastic processes," *arXiv:quant-ph/9908083*, 1999.

[3] A. Ambainis, "Variable time amplitude amplification and quantum algorithms for linear algebra problems," in *STACS'12 (29th Symposium on Theoretical Aspects of Computer Science)*, vol. 14. LIPIcs, 2012, pp. 636–647. [Online]. Available: https://dx.doi.org/10.4230/LIPIcs.STACS.2012.636

[4] A. Bouland, W. van Dam, H. Joorati, I. Kerenidis, and A. Prakash, "Prospects and challenges of quantum finance," *arXiv preprint arXiv:2011.06492*, 2020.

[5] 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*, 2011.

[6] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, "Quantum amplitude amplification and estimation," *Contemporary Mathematics*, vol. 305, pp. 53–74, 2002. [Online]. Available: https://dx.doi.org/10.1090/conm/305/05215

[7] G. Brassard, P. Høyer, and A. Tapp, "Quantum counting," in *International Colloquium on Automata, Languages, and Programming*. Springer, 1998, pp. 820–831. [Online]. Available: https://dx.doi.org/10.1007/BFb0055105

[8] P. Burchard, "Lower bounds for parallel quantum counting," *arXiv preprint arXiv:1910.04555*, 2019.

[9] P. Erdös and J. L. Selfridge, "Complete prime subsets of consecutive integers," *Proceedings of the Manitoba Conference on Numerical Mathematics, Winnipeg*, p. 13, 1971.

[10] C. Ferrie, C. E. Granade, and D. G. Cory, "How to best sample a periodic probability distribution, or on the accuracy of hamiltonian finding strategies," *Quantum Information Processing*, vol. 12, no. 1, pp. 611–623, 2013. [Online]. Available: https://dx.doi.org/10.1007/s11128-012-0407-6

[11] D. Grinko, J. Gacon, C. Zoufal, and S. Woerner, "Iterative quantum amplitude estimation," *arXiv preprint arXiv:1912.05559*, 2019. [Online]. Available: https://dx.doi.org/10.1038/s41534-021-00379-1

[12] L. K. Grover, "A framework for fast quantum mechanical algorithms," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, 1998, pp. 53–62. [Online]. Available: https://dx.doi.org/10.1145/276698.276712

[13] Y. Hamoudi and F. Magniez, "Quantum Chebyshev's inequality and applications," in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[14] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Physical review letters*, vol. 103, no. 15, p. 150502, 2009. [Online]. Available: https://dx.doi.org/10.1007/978-3-642-27848-8_771-1

[15] C. Hipp and R. Michel, "On the Bernstein-v. Mises approximation of posterior distributions," *The Annals of Statistics*, pp. 972–980, 1976.

[16] W. Hoeffding, "Probability inequalities for sums of bounded random variables," in *The collected works of Wassily Hoeffding*. Springer, 1994, pp. 409–426. [Online]. Available: https://doi.org/10.2307/2282952

[17] S. Jeffery, F. Magniez, and R. De Wolf, "Optimal parallel quantum query algorithms," *Algorithmica*, vol. 79, no. 2, pp. 509–529, 2017. [Online]. Available: https://dx.doi.org/10.1007/s00453-016-0206-z

[18] I. Kerenidis, J. Landman, A. Luongo, and A. Prakash, "q-means: A quantum algorithm for unsupervised machine learning," *Proceedings of Neural Information Processing Systems (NeurIPS)*, 2019.

[19] A. Y. Kitaev, "Quantum measurements and the abelian stabilizer problem," *arXiv preprint quant-ph/9511026*, 1995.

[20] D. E. Koh, G. Wang, P. D. Johnson, and Y. Cao, "A framework for engineering quantum likelihood functions for expectation estimation," *arXiv preprint arXiv:2006.09349*, 2020.

[21] T. Li and X. Wu, "Quantum query complexity of entropy estimation," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 2899–2921, 2018. [Online]. Available: https://dx.doi.org/10.1109/TIT.2018.2883306

[22] A. Montanaro, "Quantum speedup of Monte Carlo methods," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2181, p. 20150301, 2015. [Online]. Available: https://dx.doi.org/10.1098/rspa.2015.0301

[23] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, 2018. [Online]. Available: https://dx.doi.org/10.22331/q-2018-08-06-79

[24] Y. Suzuki, S. Uno, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto, "Amplitude estimation without phase estimation," *Quantum Information Processing*, vol. 19, no. 2, p. 75, 2020. [Online]. Available: https://dx.doi.org/10.1007/s11128-019-2565-2

[25] T. Tanaka, Y. Suzuki, S. Uno, R. Raymond, T. Onodera, and N. Yamamoto, "Amplitude estimation via maximum likelihood on noisy quantum computer," *arXiv preprint arXiv:2006.16223*, 2020. [Online]. Available: https://dx.doi.org/10.1007/s11128-021-03215-9

[26] D. Wang, O. Higgott, and S. Brierley, "Accelerated variational quantum eigensolver," *Physical review letters*, vol. 122, no. 14, p. 140504, 2019. [Online]. Available: https://doi.org/10.1103/PhysRevLett.122.140504

[27] N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning," *Quantum Information & Computation*, vol. 15, no. 3-4, pp. 316–356, 2015. [Online]. Available: https://dx.doi.org/10.26421/QIC15.3-4-7

[28] C. Zalka, "Grover's quantum searching algorithm is optimal," *Physical Review A*, vol. 60, no. 4, p. 2746, 1999. [Online]. Available: https://dx.doi.org/10.1103/PhysRevA.60.2746

A Appendix: Regularity conditions for Bernstein Von-Mises Theorem

We enumerate the regularity conditions for the Bernstein Von Mises theorem (Section 4, [HM75]) for power law schedules. Let us consider a schedule where k oracle calls in series are made N_k times followed by measurements in the standard basis and Bayesian updates. The random variable N_{k_0} and N_{k_1} represent the number of times outcomes 0 and 1 are observed out of the N_k measurements. The probability density function and the log likelihood are given by,

$$f(X,\theta) = \frac{1}{Z}\prod_k \cos^2((2k+1)\theta)^{N_{k_0}} \sin^2((2k+1)\theta)^{N_{k_1}} \tag{26}$$

$$l(X,\theta) = \sum_k 2N_{k_0}\log\cos((2k+1)\theta) + 2N_{k_1}\log\sin((2k+1)\theta) + \log Z \tag{27}$$

The regularity conditions state that there is a suitable domain Θ such that the following statements hold for all possible values of the measurement outcomes X and for some integer $s \geq 2$.

1. $f(X,\theta)$ is continuous on Θ.
2. $l(X,\theta)$ is continuous on $\overline{\Theta}$.
3. For every $\theta \in \Theta$ there is an open neighborhood U_θ such that for $\sigma, \tau \in U$ we have $E_\sigma(l(X,\tau)^s)$ is bounded. (more precisely, the supremum of $E_\sigma(l(X,\tau)^s)$ is finite).

4. For every $(\theta, \tau) \in (\Theta, \overline{\Theta})$ there exist neighborhoods U and V of θ, τ (these neighborhoods depend on θ and τ) such that the supremum $E_\sigma |\inf_{\delta \in V} l(X, \delta)|^s$ over all $\sigma \in U$ is finite.

5. $l(X, \theta)$ is twice differentiable on Θ.

6. For all $\theta \in \Theta$ there is a neighborhood U_θ such that for all $\tau \in U_\theta$,

$$0 < E_\tau[l''(X, \tau)^s] < \infty \tag{28}$$

   This is the s-th moment of the Fisher information.

7. There are neighborhoods U for all θ and a bounded function $k_\theta : X \to \mathbb{R}$ such that,

$$\|l''(X, \tau) - l''(X, \sigma)\| \leq \|\tau - \sigma\| k_\theta(X) \tag{29}$$

   for all $\tau, \sigma \in U$.

8. The prior probability λ is positive on Θ and 0 on $\mathbb{R} \setminus \Theta$.

9. For every $\theta \in \Theta$ there is a neighborhood U_θ such that for all $\sigma, \tau \in U_\theta$ and constant $c_\theta > 0$ such that,

$$|\log \lambda(\sigma) - \log \lambda(\tau)| \leq \|\sigma - \tau\| c_\theta \tag{30}$$

   This is stated as being equivalent to the continuity of λ' on Θ.

These regularity conditions can be sub-divided into three groups as follows:

1. Conditions 1-4 are about the smoothness of $f(X, \theta)$ and $l(X, \theta)$, they will be satisfied if the the norm of log-likelihood is bounded on Θ. We can choose Θ to be a subinterval around the true value for which the log-likelihood is bounded.

2. Conditions 5-7 are about the smoothness of the Fisher information on Θ. They assert that the Fisher information is bounded on Θ and is differentiable, this means that the log-likelihood function should have derivatives of order at least 3.

3. Conditions 8-9 are about the smoothness of the prior distribution, namely that the first derivative of the prior should be a continuous function. These are trivially true for the uniform distribution.

Figure 1 in Section 2 illustrates that the log-likelihood function is smooth over a neighborhood of the true value indicating that the regularity conditions for the Bernstein-Von Mises theorem are plausible in this setting, for a large neighborhood Θ to be around the true value. Algorithm 2.1 is stated with Θ as the entire $[0, \pi/2]$ interval as this choice seems to work in practice, one can also imagine a slightly modified algorithm where the first few sampling rounds are used to get a rough estimate for the true value lying in a large interval Θ and for subsequent rounds the prior is uniform on Θ, with convergence established using the Bernstein Von-Mises theorem. Adding noise may further regularize the log-likelihood functions and enforce the regularity conditions required for the Bernstein-Von Mises theorem.