Gibbs Sampling of Continuous Potentials on a Quantum Computer

Arsalan Motamedi [1 2] **Pooya Ronagh** [1 2 3 4]

Abstract

Gibbs sampling from continuous real-valued functions is a challenging problem of interest in machine learning. Here we leverage quantum Fourier transforms to build a quantum algorithm for this task when the function is periodic. We use the quantum algorithms for solving linear ordinary differential equations to solve the Fokker–Planck equation and prepare a quantum state encoding the Gibbs distribution. We show that the efficiency of interpolation and differentiation of these functions on a quantum computer depends on the rate of decay of the Fourier coefficients of the Fourier transform of the function. We view this property as a concentration of measure in the Fourier domain, and also provide functional analytic conditions for it. Our algorithm makes zeroeth order queries to a quantum oracle of the function and achieves polynomial quantum speedups in mean estimation in the Gibbs measure for generic non-convex periodic functions. At high temperatures the algorithm also allows for exponentially improved precision in sampling from Morse functions.

1. Introduction

In recent advances in machine learning (ML), a reincarnation of energy-based models (EBM) has provided state-of-the-art performance in generative modeling (Grathwohl et al., 2019; Song & Kingma, 2021; Song et al., 2020; Ho et al., 2020). Unlike the traditional EBMs such as Boltzmann machines and Hopfield neural networks (Ackley et al., 1985; Hopfield, 1982; Hinton et al., 2006; Hinton, 2007) these models require Gibbs sampling from continuous real-valued functions parameterized by large deep neural networks. However, both the training of and inference from these models is extremely computationally expensive and numerically unstable despite using state-of-the-art ML accelerators such as graphical and tensor processing and streaming units (Nijkamp et al., 2019).

The computational challenge in training EBMs is sampling from the canonical distribution represented by the model which is the Gibbs distribution

$$p_\theta(x) = \exp(-E_\theta(x))/Z_\theta \tag{1}$$

of an associated d-dimensional *energy potential*, $E_\theta : \mathbb{R}^d \to \mathbb{R}$. Here $\theta \in \mathbb{R}^m$ denotes a vector of model parameters and the normalizing constant $Z_\theta = \int e^{-E_\theta(x)}\, dx$ is the partition function of p_θ. Gibbs sampling is typically done using Monte Carlo integration of (the overdamped) Langevin dynamics (Du & Mordatch, 2019; Nijkamp et al., 2020), a stochastic differential equation (SDE) governing diffusion processes. Despite the limitations imposed by this computational bottleneck, EBMs have been shown to provide improved representations of classical data. For example, Grathwohl et al. (2019) and Hill et al. (2020) overcome the instabilities of the training process on particular datasets to provide numerical evidence that EBMs can achieve more calibrated and adversarially robust representations compared to conventional classifiers. We refer the reader to Appendix C for more details on the usage of Gibbs sampling in the training of and inference from EBMs.

In this paper, we investigate using quantum computation for Gibbs sampling from continuous energy potentials. We use finite difference techniques for solving differential equations on quantum computers (Berry et al., 2017; Childs & Liu, 2020; Childs et al., 2021; Liu et al., 2021; Krovi, 2023) to solve the Fokker–Planck equation (FPE), a second-order partial differential equation (PDE) admitting the Gibbs distribution as its steady state solution. The FPE and Langevin dynamics are associated with each other through Itô integration (Evans, 2023). Interestingly, directly solving for the steady state of the FPE requires solving linear systems with exponentially poor condition numbers. We therefore also have to integrate the FPE for a long enough time to asymptotically converge to the Gibbs state; as such, we do

[1]Institute for Quantum Computing, University of Waterloo, Waterloo, ON, Canada [2]Department of Physics & Astronomy, University of Waterloo, Waterloo, ON, Canada [3]Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada [4]Irréversible, Vancouver, BC, Canada. Correspondence to: Arsalan Motamedi <arsalan.motamedi@uwaterloo.ca>, Pooya Ronagh <pooya.ronagh@uwaterloo.ca>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

not achieve a shortcut to the problem of long mixing time in equilibrium dynamics. We do, however, achieve a high precision approximation to the Gibbs state in total variation distance.

Technical setting and contributions. In the Euclidean space, the Gibbs measure is only well-defined for an unbounded potential. This setting is not suitable for finite difference methods, therefore we are interested in probability density functions with compact support. This imposes boundary conditions on the FPE. In this paper we consider periodic boundary conditions since they allow us to leverage quantum Fourier transforms. In other words, we focus on potentials defined on high dimensional tori. We assume that the energy function takes values in a real interval of diameter Δ which we refer to as the diameter of the potential hereon.[1] For a periodic function $f : \mathbb{R}^d \to \mathbb{R}$ the Fourier transform coefficients correspond to points on the lattice $\mathbb{Z}^d$. We show that these coefficients decay sub-exponentially away from the origin assuming a bound on the growth of the derivatives of f (Definition A.1). The latter is a functional analytic condition similar to analyticity, however our definition is milder and therefore we call it *semi-analyticity*. We also show that if the Fourier coefficients on $\mathbb{Z}^d$ are viewed as the density of a probability measure defined on this lattice, then semi-analyticity is equivalent to the concentration of this measure (Theorem A.13).

Many periodic functions are semi-analytic. For example any function with finitely many non-zero Fourier coefficients is semi-analytic (Example A.1). Semi-analyticity is determined using two parameters which we denote as C and a in this paper. The first parameter represent the scale of the function (i.e., scales linearly with scalar multiplication), but the second parameter bears information about the geometry of the function and can be viewed as an inverse radius of convergence of its Taylor expansion (see Appendix A). We show some of the basic properties of analytic and semi-analytic functions; i.e., how C and a change under arithmetic operations and compositions (Proposition A.23). Consequently, we show that parameterized families such as deep neural networks with analytic activation functions are analytic. However, activation functions such as the sigmoid function creates sharp ramps in the energy landscape which shrinks the radius of convergence (Corollary A.26). Beside the algorithmic contributions in this paper, the above insights can shed light on suitable alternatives to deep neural networks as parameterized models for energy-based learning. We first show that real analytic periodic functions admit efficient quantum algorithms for interpolation and differentiation provided quantum states encoding very coarse discretizations of such functions (Theorem 3.2). We also provide a lower bound result in Theorem 3.3 showing the optimality of our discretization.

We refer the reader to Examples A.2 and A.3 for further preliminary examples of semi-analytic functions. More systematically, in Appendix A.3, we show several constructions of semi-analytic functions from other ones (e.g., using compositions, or basic operations between functions). We then continue with transcendental functions such as the exponentiation and sigmoid functions (as they relates to machine learning models in particular). As a corollary of these properties, we compute the analytic parameters of neural networks constructed via sigmoid activation functions in Corollary A.26. Additionally, as concrete numerical demonstrations, Figs. 2 and 3 showcase the error of our interpolation approach applied on the functions considered in Examples A.2 and A.3.

Given a parametrization θ of an energy potential[2] $E_\theta : \mathbb{R}^d \to \mathbb{R}$, one can construct a quantum oracle

$$|\theta\rangle\,|x\rangle\,|y\rangle \mapsto |\theta\rangle\,|x\rangle\,|y \oplus E(x)\rangle \tag{2}$$

realizing such a function with only a polylogarithmic overhead compared to the gate complexity of the corresponding classical boolean circuit (Nielsen & Chuang, 2002). The quantum circuit of this oracle is depicted in Fig. 1(a). In what follows, we ignore the model parameter register when the context is clear.

We note that if E has a positive minimum value $E_* \geq 0$, a naïve classical rejection sampling routine, consisting of drawing uniform random samples $x \in \mathbb{R}^d$ and accepting the sample with probability $\exp(-E(x))$ will require $O(e^{\Delta+E_*})$ iterations to generate an accepted sample from the Gibbs distribution. Similarly, the mixing time of Langevin dynamics is $O(\kappa_E)$ where κ_E is the *Poincaré constant* of the potential E and for generic non-convex functions this constant is in $O(e^{\Delta})$. We prove the counterparts to these two facts for periodic functions in Proposition B.3 and Proposition B.5. The latter bound can be interpreted as a Eyring-Kramers law for periodic functions (Berglund, 2011).

We use our Fourier interpolation techniques to achieve high precision Gibbs sampling as summarized in Table 1. Our algorithm uses $\widetilde{O}\left(\kappa_{E/2} e^{\Delta/2} d^7 \operatorname{polylog}(1/\varepsilon)\right)$ queries to the oracle O_E for Gibbs sampling with an approximation error ε in total variation distance (TV) (Theorem D.2). This provides an opportunity for quantum speedup for families of functions with small Poincaré constants. However, for a generic non-convex periodic function E the relevant Poincaré constant (i.e., that of $E/2$, due to Born rule) is in $O(e^{\Delta/2})$. Therefore, our complexity factor $\kappa_{E/2} e^{\Delta/2} = e^{\Delta}$

[1] For simplicity, we use a constant thermodynamic $\beta = 1$ throughout (otherwise $\beta\Delta$ can be thought of as a single parameter).

[2] For example, the weights and biases of a deep neural network, or any other parametrized ansatz.

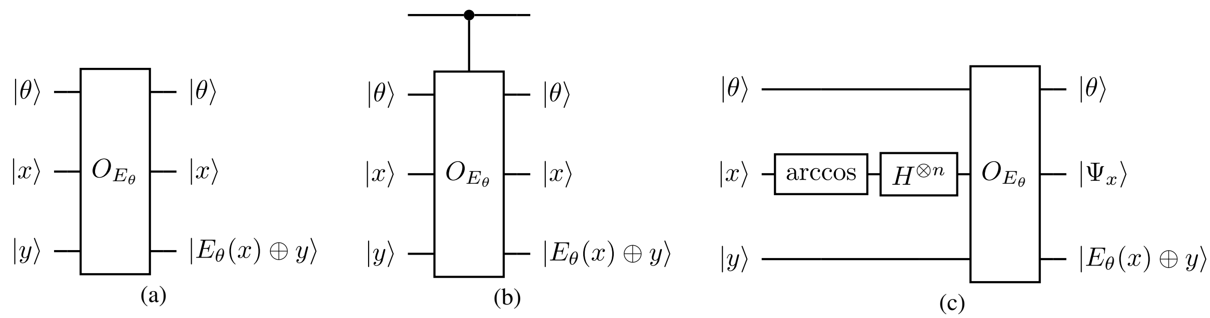

Figure 1. Schematics of the circuits of quantum oracles. All registers receive float-point representations of real numbers. (a) The oracle for an energy function E_θ. The first register receives the model parameters $\theta \in \mathbb{R}^m$, the second register receives a data sample $x \in \mathbb{R}^d$. And, the last register is used to evaluate the energy function. (b) A controlled variant of the same oracle, controlled on a single qubit represented by the top wire. (c) The oracle of figure (a) receiving an augmented sample in superposition as the state $|\Psi_x\rangle = \sum_{b \in \{0,1\}^d} |(-1)^{b_i} \arccos x_i\rangle$.

in the first row of Table 1 saturates the complexity bound of classical rejection sampling for periodic functions in absence of any additional structure. However for families of functions with a Poincaré constant better than the Eyring-Kramers bound we may achieve a quantum advantage. For example, for Morse functions with unique global minima we achieve the query complexities in the third row of the table (Corollary 4.2). The classical counterpart for this result is (Li & Erdogdu, 2020, Theorem 2.4) which results in the complexity bounds of the fourth row. In comparison, our algorithm achieves an exponential advantage in precision ε while only consuming zeroeth order queries to the function.

The prior results, Metropolized random walk (MRW), the underdamped Langevin dynamics (ULD) and the Metropolis-adjusted Langevin algorithm (MALA), together with their accelerations via quantum random walk in (Childs et al., 2022) are also included in Table 1. We note that all these algorithms require convexity assumptions on the potential. While convex potentials are of theoretical interest for analyzing and comparing different sampling algorithm, they do not pertain to practical ML scenarios as real-world distributions are not uni-modal.

The oracle of Fig. 1(a) is sufficient for sampling from the Gibbs distribution of E_θ which pertains to inference from EBMs. However, in practice we often collect samples in order to estimate the mean of another quantity $f : \mathbb{R}^d \to \mathbb{R}$ with respect to the Gibbs measure. For example, in training EBMs, the expectations of the gradients of the energy function is desired (see Appendix C). In this case, a *white-box* access to the energy oracle will allow us to use amplitude estimation techniques to achieve quantum advantage. By this we mean the ability to construct a controlled variant (as in Fig. 1(b)) of the unitary in Fig. 1(a) and its inverse. Given white-box access to both the energy and random variable oracles (i.e., an oracle for evaluating f), our algorithm achieves a quadratic advantage in mean estimation for generic periodic functions and a quartic advantage in the case of Morse functions with unique global minima (Corollary 4.3) as reported in the last column of Table 1.

Practical applicability. Our algorithm outperforms other methods in the high temperature and high precision regime. This is the sweet spot for generative modeling using energy-based techniques since in ML the Gibbs sampling temperature is often set to $\beta = 1$ or tuned as a finite hyperparameter, and the parameterized models representing the energy potential are regularized to attain bounded values (e.g., using L_2 regularization). We also note that our quantum advantage in precision is at the cost of high order polynomial dependence on dimension d. However, for applications processing classical data in computational basis states, at least a linear dependence on d for state preparation and measurement is unavoidable. Moreover, in ML applications Langevin dynamics is typically run on a latent space of much lower dimensions than the input data (Pang et al., 2020; Vahdat et al., 2021; Rombach et al., 2022). Generated latent samples are then scaled up to high resolution results through decoding techniques. Therefore, our exponential quantum advantage in precision can be of critical significance in ML use-cases.

Many real-world datasets are naturally represented using periodic degrees of freedom (e.g., rotation angles), particularly including problems within physics. For example, many-body systems such as the XY Hamiltonian (Araki & Matsui, 1985), and quantum field theories such as the ϕ^4 field theory (Jordan et al., 2012) are of this kind. Nevertheless, this is not typically the case for conventional ML training datasets (e.g., for image and audio datasets represented by various channels of colour intensities and acoustic amplitudes, respectively) (Ziyin et al., 2020). In such cases,

Table 1. Summary of the query complexities of some of the classical and quantum algorithms for sampling from a d-dimensional Gibbs distribution, highlighting the cases of significant quantum advantage in boldface. ε denotes the error in the designated norm (TV for total variation distance, and W_2 for 2-Wasserstein distance), and κ_f denotes the Poincaré constant of a function f. Our result (first row) corresponds to non-convex periodic functions and the relevant Poincaré constant is that of the function $E/2$ due to Born rule(see Section 2). Therefore, for families of functions that the Poincaré constant is better than the Eyring-Kramers bound we may achieve a quantum advantage in sampling. For example, for Morse functions with unique global minima we achieve the query complexities in the third row (Corollary 4.2). In comparison to the classical counterpart (fourth row), our algorithm achieves an exponential advantage in precision ε while only consuming zeroeth order queries to the function. For estimating means of random variables of the Gibbs state, we achieve a quadratic advantage for generic periodic functions (last column of the first and second rows) and a quartic advantage in the case of Morse functions with unique global minima. Here Δ_f is the diameter of the range of values a function f attains. The prior results (rows five to nine) all require convexity assumptions on the potential. For μ-strongly convex functions the Poincaré constant κ_E is $1/\mu$, famously known as the Bakry-Émery criterion (Bakry et al., 2014).

Method	Potential type	Query order	Sampling complexity	Norm	Mean estimation complexity
This paper	non-convex periodic	zeroeth	$\widetilde{O}\left(\kappa_{E/2}e^{\Delta/2}d^7\right)$	TV	$\boldsymbol{\widetilde{O}\left(\kappa_{E/2}e^{\Delta/2}d^7\Delta_f\varepsilon^{-1}\right)}$
Rejection sampling	non-convex	zeroeth	$O\left(e^{\Delta}\right)$	TV	$O\left(e^{\Delta}\Delta_f^2\varepsilon^{-2}\right)$
This paper	Morse and periodic	zeroeth	$\boldsymbol{\widetilde{O}\left(\lambda^{-2}e^{\Delta/2}d^7\right)}$	TV	$\boldsymbol{\widetilde{O}\left(\lambda^{-2}e^{\Delta/2}d^7\Delta_f\varepsilon^{-1}\right)}$
Cl. RLA (Li & Erdogdu, 2020)	Morse and periodic	first	$\widetilde{O}\left(\lambda^{-4}L^4d^3\varepsilon^{-2}\right)$	TV	$\widetilde{O}\left(\lambda^{-4}L^4d^3\Delta_f^2\varepsilon^{-4}\right)$
Cl. MRW (Dwivedi et al., 2018)	convex	first	$\widetilde{O}\left(L^2d^3\varepsilon^{-2}\right)$	TV	$\widetilde{O}\left(L^2d^3\Delta_f^2\varepsilon^{-4}\right)$
Q. ULD (Childs et al., 2022)	strongly convex	zeroeth	$\widetilde{O}\left(\mu^{-2}L^2d^{1/2}\varepsilon^{-1}\right)$	W_2	–
Cl. ULD (Cheng et al., 2017)	strongly convex	first	$\widetilde{O}\left(\mu^{-2}L^2d^{1/2}\varepsilon^{-1}\right)$	W_2	–
Q. MALA (Childs et al., 2022)	strongly convex	first	$\widetilde{O}\left(\mu^{-1/2}L^{1/2}d\right)$	TV	$\widetilde{O}\left(\mu^{-1/2}L^{1/2}d\Delta_f\varepsilon^{-1}\right)$
Cl. MALA (Lee et al., 2020)	strongly convex	first	$\widetilde{O}\left(\mu^{-1}Ld\right)$	TV	$\widetilde{O}\left(\mu^{-1}Ld\Delta_f^2\varepsilon^{-2}\right)$

we may assume that data is always normalized within a hypercube $[-1,1]^d$. One solution is to extend our algorithm to functions $f:[-1,1]^d\to\mathbb{R}$. We can do so by constructing the composite function $g(x) := f(\cos x)$ and applying our algorithm to g. We show in Appendix A.5 that this procedure is not efficient unless data is concentrated in the boundaries of the hypercube or possesses a dimension independent variance. Since our algorithm is more suitable for periodic functions, the alternative approach is to augment the dataset by inverting a trigonometric function of its components as shown in Fig. 1(c). Here each dimension of data is lifted to two branches of the $\arccos$ function. Therefore each sample $|x\rangle$ in the computational basis corresponds to 2^d inverse images, and an equal superposition $|\Psi_x\rangle$ of these exponentially many inverse images is efficiently constructed using Hadamard gates. By employing this approach, the resulting energy function learns the correct function across all branches of $\arccos$.

We refer the reader to Appendix C for a concise introduction to training and inference from EBMs. We also provide a detailed construction of the oracle used for the simulation of the Fokker–Planck equation using the energy oracles as building blocks (Fig. 4).

Related works. To the best of our knowledge, this work and the independent paper Childs et al. (2022) are the first efforts to analyze quantum algorithms for the problem of Gibbs sampling from a continuous real-valued function. Childs et al. (2022) achieve a quadratic speedup in expediting a Monte Carlo simulation of Langevin dynamics using quantum random walks but their result is restricted to strongly convex potentials. Similarly classical algorithms that achieve high precision Gibbs sampling also make assumptions about convexity or satisfaction of isoperimetric inequalities (Roberts & Tweedie, 1996; Dwivedi et al., 2018; Chewi et al., 2021). However, ML applications demand highly non-convex potentials that can capture complex modes of data in a multimodal landscape. Fortunately, our algorithm provides high precision Gibbs sampling of such potentials as long as they satisfy periodic boundary conditions. In addition, the quantum speedup observed by Childs et al. (2022) assumes accurate access to the gradients of the potential whereas we only perform zeroeth order queries to the potential.

As another interesting point of comparison, Ge et al. (2018) focus on Gibbs sampling from finite mixtures of log-concave distributions and incur very high-order polynomial costs in dimension d and inverse precision $1/\varepsilon$. Moreover, Fan et al. (2023) consider smooth potentials that satisfy Poincaré inequalities and achieve polylogarithmic dependence on precision, similar to us. However, we have an exponential improvement with respect to the Lipschitz constant, and more importantly, Fan et al. (2023) have an additional logarithmic dependence on the χ^2-divergence of a warm-start initial distribution and the desired target distribution. As mentioned in their work, one has merely an $\exp(d)$ bound on this divergence, which results in higher-order polynomial overhead with respect to d. We also mention the results of Li & Zhang (2022) and Ozgul et al. (2023) who tackle quantum optimization and sampling from non-convex functions with additional global geometric assumptions.

Finally we note that prior quantum algorithms (Terhal & DiVincenzo, 2000; Poulin & Wocjan, 2009; Chowdhury & Somma, 2016; van Apeldoorn et al., 2017) apply amplitude amplification to achieve a Grover speedup in preparing the Gibbs state of *discrete* spin systems. However, a naïve application of these techniques to (say, a discretization of) the continuous domain, will at best result in a query complexity that scales with $\sqrt{\exp(d)}$.

2. The algorithm

Our goal is to obtain the steady state solutions to the Fokker–Planck equation (FPE)

$$\partial_t \rho(x,t) = \nabla \cdot \left(e^{-E(x)}\, \nabla \left(e^{E(x)}\, \rho(x,t)\right)\right) \tag{3}$$

corresponding to a *toroidal diffusion process*. That is, a diffusion process obtained by projecting Langevin dynamics

$$\mathrm{d}Y_t = -\nabla E(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t \tag{4}$$

on a high dimensional torus $\mathbb{T} = \mathbb{R}^d/\mathbb{Z}^d$. By this notation we mean the topological quotient of $\mathbb{R}^d$ under the action of $\mathbb{Z}^d$ via $g : x \mapsto g + x$ for all $g \in \mathbb{Z}^d$. We refer the reader to Appendix B for more details on such stochastic processes. Here $(W_t)_{t\geq 0}$ is a Wiener process and the drift term $-\nabla E(Y_t)$ is along the gradient of a periodic smooth function $E : \mathbb{R}^d \to \mathbb{R}$, which is called the energy function, energy potential, or the potential, for short. We assume that the fundamental domain of this quotient is of length ℓ and more specifically $[-\ell/2, \ell/2]^d \subset \mathbb{R}^d$. The unique steady state solution ρ_s to (3) corresponds to the Gibbs state $\rho_s(x) \propto e^{-E(x)}$ of the potential. We intend to find this distribution by solving (3) using a uniform distribution $\rho_0(x) \propto 1$ as an initial condition and accessing the long time $T \gg 0$ asymptotes of the solution. We refer the reader to Algorithm 1 for a pseudo-code of our algorithm.

By solving the FPE we mean preparing a quantum state that encodes the solution of the PDE (3) on a discrete lattice. When discretizing an ℓ-periodic function, we first consider a lattice obtained from taking an odd number, $2N+1$, of equidistant points along each axis; $x_n = \frac{\ell n}{2N+1}$ for all $n \in [-N..N]^d$. We denote this discrete lattice by V_N and the Hilbert space $\mathbb{C}^{V_N}$ (i.e., the space of functions from V_N to $\mathbb{C}$) by $\mathcal{V}_N$. Our discretization scheme transforms the generator

$$\mathcal{L}(-) = \nabla \cdot \left(e^{-E}\, \nabla \left(e^{E} -\right)\right) \tag{5}$$

of the FPE to a linear operator $\mathbb{L} : \mathcal{V}_N \to \mathcal{V}_N$. An explicit construction of $\mathbb{L}$ is elaborated in Appendix B.1. We then solve the linear ordinary differential equation

$$\frac{d}{dt}\overrightarrow{u(t)} = \mathbb{L}\overrightarrow{u(t)} \tag{6}$$

using the results of Berry et al. (2017) and Krovi (2023) to find a high precision approximation of the solution, $u(T)$ (line 3 of Algorithm 1).

In solving this linear system we use the Fourier pseudo-spectral method to achieve high precision finite difference approximations of the derivatives of u using merely a coarse lattice V_N. To this end we require a tameness condition on the growth of the higher derivatives of u. We call this condition *semi-analyticity* for its close resemblance to the notion of analyticity in real functional analysis. In Section 3 we discuss the connections between semi-analyticity and the concentration of measure for a random variable we define from the Fourier transform of u. We also provide examples of semi-analytic functions. In particular, we show that all periodic real analytic functions are semi-analytic as well.

We note that sampling from the discretization of the Gibbs distribution

$$|\rho_s\rangle \propto \sum_x e^{-E(x)}\, |x\rangle\,, \tag{7}$$

results in an ensemble at thermodynamic $\beta = 2$ instead of at $\beta = 1$. Here the normalization constant of this state is $1/\sqrt{\bar{Z}_{\beta=2}}$, where the $\bar{Z}$ notation represents the partition function of the discretized probability measure. To overcome this problem, throughout we set the energy function of our interest to be $\frac{1}{2}E$. In the notation $|x\rangle := |x_1\rangle \otimes \cdots \otimes |x_d\rangle \in \mathcal{V}_N$ for addressing the points on the lattice, each $|x_i\rangle = |n_i + N\rangle$ is the one-hot encoding of the index $n_i + N$ where $n_i \in \{-N, -N+1, \ldots, N\}$.

Moreover, the discretization of the Gibbs state will result in sampling from each point of the lattice according to the discrete probability distribution

$$p(x) = \frac{1}{\bar{Z}_{\beta=2}} e^{-2E(x)} \simeq \frac{\ell^d}{(2N+1)^d Z_{\beta=2}} e^{-2E(x)}. \quad (8)$$

This is the case since $\bar{Z}_{\beta=2}\Delta x \simeq Z_{\beta=2}$. Therefore our proposed algorithm is to draw samples $x \in V_n$ via measurements in the computational basis states, and then, generate uniform samples from the box $\prod_{i=1}^{d}[x_i - \frac{\ell}{4n+2}, x_i + \frac{\ell}{4n+2}]$ (line 5 of Algorithm 1).

However, naïve usage of a small N combined with this uniform sampling strategy does not provide a good approximation to the Gibbs distribution. This is the second step in our algorithm wherein the semi-analyticity condition plays a critical role. We show that given a real periodic function $u : \mathbb{R}^d \to \mathbb{R}$, we can provide samples from a high-precision approximation of the distribution proportional to u^2 by querying very few points in the domain of the definition of u. Using a very coarse lattice V_N (i.e., with N being small) we achieve a sampling error of $O(e^{-N})$ by employing a technique from classical signal processing involving representation of u in the Fourier domain, although we use quantum Fourier transforms (QFT) for its implementation. We call this procedure *upsampling* of u (Oppenheim & Schafer, 1975) (line 3 of Algorithm 1).

To this end, in Section 3 we introduce *Fourier interpolation* and use this technique to upsample our quantum state in the Fourier domain (refer to Appendix A for further information). However, our interpolation technique is useful beyond the applications considered herein. For instance, quantum algorithms for solving partial differential equations (e.g., (Childs et al., 2021)) also prepare quantum states that encode the solutions of the equations on coarse discrete lattices. Our interpolation algorithm, applied as a post-processing quantum circuit, allows one to find approximate solutions on finer lattices and even on the continuous domain without discretization.

The quantum algorithm makes queries to oracles for the discrete generator $\mathbb{L}$ which themselves require access to $O(dN)$ oracles of the energy function at different points (see Eq. (62) in the appendix). In Appendix D.1 we show that assuming that the FPE generates a semi-analytic one-parameter family of probability measures $\{e^{\mathcal{L}t}\rho_0 : t \geq 0\}$, Algorithm 1 samples from a distribution ε-close to the Gibbs distribution (in total variation distance) by making $\widetilde{\mathcal{O}}(d^7 \frac{e^{\Delta/2}}{\ell^2} \kappa_{E/2} \,\mathrm{polylog}(1/\varepsilon))$ queries to the oracle (2) of the energy function. Note that dilating the domain of the definition of the energy function by a scalar α multiplies the Poincaré constant by α^2. This is why $\kappa_{E/2}$ is normalized by a factor of ℓ^2 in this complexity.

Extension to non-periodic domains. Let $f : [-1,1]^d \to \mathbb{R}$ be a (C,a)-analytic function with diameter Δ. Since $g(\theta) := f(\cos\theta)$ is 2π-periodic, applying Algorithm 1 to it allows us to sample from the distribution $\Theta \sim e^{-g}/Z_g$. The random variable $Y = \cos(\Theta)$, with the cosine function applied component-wise, follows the distribution $Y \sim e^{-f}/\widetilde{Z}_f \cdot \prod_{i=1}^{d} \frac{1}{\sqrt{1-x_i^2}}$. Note that

$$\begin{aligned}\widetilde{Z}_f &:= \int \frac{e^{-f(x)}}{\prod_i \sqrt{1-x_i^2}} \mathrm{d}x \\ &= Z_f \cdot \mathbb{E}_{X\sim e^{-f}/Z_f}\left[\prod_i (1-X_i^2)^{-1/2}\right].\end{aligned} \quad (9)$$

In Appendix A.5 we show that this procedure is equivalent to using Chebyshev interpolation instead of Fourier interpolation.

Using rejection sampling, we may accept samples from Y with probability $\frac{\prod_i \sqrt{1-x_i^2}}{\mathbb{E}[\prod_i (1-X_i^2)^{-1/2}]}$ to prepare samples from the original Gibbs distribution e^{-f}/Z_f on the hypercube $[-1,1]^d$. This adds a prefactor overhead of $\mathcal{O}(\mathbb{E}\left[\prod_i (1-X_i^2)^{-1/2}\right])$ for sampling (see Corollary A.31). Consequently, the rejection sampling algorithm is efficient only if the Gibbs measure is concentrated in the interior of the hypercube. For example, this includes quadratic potentials (i.e. Gaussian Gibbs measures) with bounded variance. However, for such families of functions $\Delta = \Omega(d)$ therefore the end-to-end algorithm is not efficient. For more general potentials, this process samples efficiently from the boundary of the hypercube as detailed in Proposition A.30.

3. Interpolation results

We now introduce the notion of semi-analyticity for smooth functions and prove several favourable properties of it using the Fourier spectral method. In what follows, for a string of length d of non-negative integers $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{Z}_+^d$ we define $\alpha! := \alpha_1! \cdots \alpha_d!$, and $|\alpha| := \alpha_1 + \cdots + \alpha_d$, and use the following notation for higher order derivatives:

$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}. \quad (10)$$

Definition 3.1. *Let the function $u : \mathbb{R}^d \to \mathbb{R}$ be ℓ-periodic along all axes, and moreover, let $X \sim Unif\big([-\ell/2, \ell/2)^d\big)$ be a uniform random variable. We say u is semi-analytic if there exists $C, a \in \mathbb{R}_+$, such that for any $m \in \mathbb{N}$ we have*

$$\left(\frac{\ell}{2\pi}\right)^m \sqrt{\mathbb{E}\left(\sum_{\alpha:|\alpha|=m} \left|D^\alpha u(X)\right|^2\right)} \leq C\, a^m\, m!. \quad (11)$$

Algorithm 1 Pseudocode of our Gibbs sampling algorithm.

input Energy function oracle O_E, lattice parameters $N, M \in \mathbb{N}$, solution time $T > 0$

1: Construct an oracle for the discretization $\mathbb{L}$ of the generator of the Fokker–Planck equation (see Fig. 4 in the appendix).
2: Deploy the algorithm of (Berry et al., 2017) to prepare a quantum state approximating $|u(T)\rangle$ pertaining to the solution of $\frac{d}{dt}\overrightarrow{u} = \mathbb{L}\overrightarrow{u}$, with $\overrightarrow{u}(0) = \mathbf{1}$, at time $t = T$.
3: Apply the upsampling isometry $F_M^{-1}\iota F_N$ involving quantum Fourier transformations on the prepared state (Theorem 3.2).
4: Measure the resulting state in the computational basis to obtain a lattice point $x \in [-\ell/2, \ell/2]^d$.
5: Draw a sample $\widetilde{x}$ uniformly at random from the box $\prod_{i=1}^d \left[x_i - \frac{\ell}{4M+2}, x_i + \frac{\ell}{4M+2}\right]$ around x.

output Sample point $\widetilde{x}$.

Furthermore, we refer to C and a as the semi-analyticity parameters.

The Fourier transform of u

$$u(x) = \sum_{k \in \mathbb{Z}^d} \widehat{u}[k]\, e^{i\frac{2\pi\langle k,x\rangle}{\ell}} \tag{12}$$

has coefficients $\widehat{u}[k] = \frac{1}{\ell^d}\int_{\mathbb{T}} u(x)e^{-i\frac{2\pi\langle k,x\rangle}{\ell}}\,dx$ assigned to the lattice points on $\mathbb{Z}^d$. The values $|\widehat{u}[k]|^2$ form a probability measure on this lattice. In Theorem A.13 we show that semi-analyticity is equivalent to the sub-exponential concentration of this measure.

We provide several examples of semi-analytic functions. Any function with finitely many non-zero Fourier coefficients is semi-analytic (Example A.1). Every periodic real-analytic function is semi-analytic (Proposition A.2). We also show how semi-analyticity parameters change through basic operations like addition, multiplication, and composition of functions (Proposition A.23). In Corollary A.26, we use these results to find the analyticity parameters of deep neural networks, as the de facto function approximators used in machine learning (which can act as the parameterized oracles shown in Fig. 1(a) for our quantum algorithm).

We now present our main result regarding upsampling of a quantum state represented on a discrete lattice to a target continuous distribution defined in the continuous ambient space of the lattice. Recently, Ramos-Calderer (2022) discussed the idea of upsampling in the context of efficient representation of classical data on a quantum computer although without a rigorous mathematical account. We, however, provide a rigorous analysis of the upsampling technique and its precision with respect to the target *continuous* distribution, rather than only to a discretization of it on a finer lattice. Given a tuple of indices $n = (n_1, \cdots, n_d) \in \{0, \ldots, 2N\}^{\times d}$, we denote the associated computational basis state in the Hilbert space $\mathcal{V}_N \cong \left(\mathbb{C}^{2N+1}\right)^{\otimes d}$ by $|n\rangle$. We further denote the discretization of u by $\overrightarrow{u} \in \mathcal{V}_N$, and the unit vector parallel to that by $|u_N\rangle \propto \sum_{m\in\{0,\ldots,2N\}^{\times d}} u(x_{m-N})\,|m\rangle$. We now state our main interpolation result.

Theorem 3.2 (Main interpolation result)**.** *Given an L-Lipschitz (C, a)-semi-analytic periodic function u, an integer $N \geq 2ad$, and a (previously prepared) quantum state $|\psi\rangle \in \mathcal{V}_N$, such that $\||\psi\rangle - |u_N\rangle\|_2 \leq \delta$, there exists a quantum algorithm with gate complexity $\mathcal{O}\left(\frac{dN}{a}\,\mathrm{polylog}\,(NdL\ell/C)\right)$ that returns samples from a distribution within at most ε total variation distance from the distribution proportional to u^2, where*

$$\varepsilon \leq \delta + \frac{16\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6\,\frac{N}{a}},$$

and $\mathcal{U} = \sqrt{\mathbb{E}\,u^2(X)}$, with $X \sim \mathrm{Unif}([-\frac{\ell}{2}, \frac{\ell}{2}]^d)$.

In Appendix A.5 we demonstrate that Chebyshev interpolation can also be performed efficiently by means of QFT and an analogous interpolation result holds as stated in Proposition A.29. Moreover, in Example A.3 we show a family of functions that provide adversary witnesses obstructing the usage of very coarse discretization for upsampling to arbitrarily small target errors.

Theorem 3.3. *Let u be a (C, a)-semi-analytic function. Consider any exact discretization $|u_N\rangle$ on the discrete lattice with $N \leq \theta a/16$, where $\theta \in (0, 1)$. There is no algorithm that can return samples close to the actual distribution (proportional to u^2) with a guaranteed error of less than $(1-\theta)^2\,\frac{1}{1024e}$.*

We conclude this section by noting that even the first and second order derivatives of a semi-analytic function u can be approximated with high precision (Proposition A.18). This is instrumental in constructing high precision approximations to the generator $\mathcal{L}$ of the FPE.

4. Algorithm complexity

We now state the computational complexity of our Gibbs sampler (Algorithm 1).

Theorem 4.1 (Main sampling result)**.** *Given an L-Lipschitz periodic potential E, suppose that the one-parameter family of all probability measures $\{e^{\mathcal{L}t}\rho_0 : t \geq 0\}$ consists of semi-analytic functions with parameters C and* *a. Algorithm 1 samples from a distribution ε-close to the Gibbs distribution (in total variation distance), by making $\widetilde{O}\left(a^4 d^7 \kappa_{E/2} e^{\Delta/2} \ell^{-2}\right)$ queries to the oracle of the energy function. The algorithm succeeds with bounded probability of failure and returns a flag indicating its success. In addition, the gate complexity of the algorithm is larger only by a factor of* $\mathrm{polylog}(Cade^{\Delta}(1+\ell L))/\varepsilon)$.

As a corollary, we show that the complexity of our algorithm improves under simplifying assumptions on the geometry of the saddle points of the energy function. Recall that a function E is called a Morse function if all its critical points are non-degenerate; i.e., if $\nabla^2 E(x)$ is non-singular whenever $\nabla E(x) = 0$. Mei et al. (2016) quantify this condition with additional parameters. We use a simplified definition compared to this reference, and call E to be a λ-strongly Morse function if the spectrum of $\nabla^2 E(x)$ is bounded below by $\lambda > 0$ in absolute value at every critical point; equivalently, if $\left\|\left(\nabla^2 E(x)\right)(v)\right\| \geq \lambda \|v\|$ for all critical points x and all vectors v.

The strong Morse condition allows every saddle point in the energy landscape to have *steep enough* escape directions. Therefore, intuitively, the dynamics opposite the gradient flow is not obstructed. Li & Erdogdu (2020) generalize this definition by allowing flat eigendirections in the saddle points as long as the exponentiation map along these directions leaves us inside the critical loci. We also note that for Morse function on compact domains, the strong Morse criteria is always satisfied for some parameter λ. Applying (Li & Erdogdu, 2020, Proposition 9.14) to weak Morse functions on the products of spheres results in a better Poincaré constant than the general bound in Proposition B.5. This is a generalization of the Bakry-Emery criterion (Bakry et al., 2014, Proposition 4.8.1) well beyond strong convexity. We have,

Corollary 4.2. *Let E be a λ-strongly Morse potential with a unique global minimum. Furthermore, assume that E, ∇E, and $\nabla^2 E$ are Lipschitz continuous with respective parameters L_1, L_2, and L_3. Letting $\mathcal{C}$ denote the set of critical points of the energy potential E, we also make an additional technical assumption as in (Li & Erdogdu, 2020, Proposition 9.14), namely that*

$$C_F = \min\left\{1, \frac{\lambda}{2}, \inf_{x:d(x,\mathcal{C})>\frac{\lambda}{L_3}} \frac{\|\nabla E(x)\|}{d(x,\mathcal{C})}\right\} \in (0,1]$$

satisfies $\max\left(\frac{4}{\lambda^2}, \frac{6L_2 d}{C_F^2}\right) \leq \frac{C_F^2}{12 L_2 L_3^2 d}$. *Then the query complexity of our algorithm reduces to* $\widetilde{O}\left(a^4 d^7 \lambda^{-2} e^{\Delta/2}\right)$.

This follows from the observation that E satisfies a Poincaré inequality with $\kappa_E = \mathcal{O}\left(\frac{1}{\lambda^2}\right)$ as per (Li & Erdogdu, 2020, Proposition 9.14). We note that this proposition is stated for $S^n \times \cdots \times S^n$ where $n \geq 2$. However, in presence of a unique global minimum the result remains valid for $n = 1$ as well; i.e., in the case of high-dimensional tori.[3]

We now investigate how the Gibbs sampler discussed earlier can be employed to calculate the expected values of random variables with bounded variance. Specifically, we consider a periodic function $f : [-\frac{\ell}{2}, \frac{\ell}{2}]^d \to \mathbb{R}$ that belongs to $L^2(\rho)$, and aim at estimating $\mathbb{E}[f(X)]$, where X is a random variable with distribution ρ. To this end we use the state-of-the-art mean estimation algorithm presented in Kothari & O'Donnell (2023).

Corollary 4.3 (Mean estimation)**.** *Let E be an energy function, satisfying the assumptions made in Theorem 4.1. Furthermore, let f be an L_f-Lipschitz ℓ-periodic function with diameter Δ_f. There is a quantum algorithm that returns an estimate $\widehat{\mu}$ to $\mathbb{E}[f(X)]$, with additive error at most $\varepsilon > 0$ and success probability at least $1-\delta$, making*

$$\widetilde{O}\left(a^4 d^7 e^{\Delta/2} \frac{\kappa_{E/2}}{\ell^2} \frac{\Delta_f}{\varepsilon} \log\left(\frac{1}{\delta}\right)\right) \tag{13}$$

queries to the controlled and standalone oracles of the energy function E and the function f.

Therefore a quantum computer can prepare a distribution ε-close in TV distance to the Gibbs distribution of Morse functions defined on tori using $\widetilde{O}\left(\lambda^{-2} e^{\Delta/2} d^7\right)$ queries to the energy oracle, while the Riemannian Langevin diffusion of Li & Erdogdu (2020) uses $\widetilde{O}\left(\lambda^{-4} L^4 d^3 \varepsilon^{-2}\right)$ classical queries to the energy function. For mean estimation on the same potentials, quantum computation requires $\widetilde{O}\left(\lambda^{-2} e^{\Delta/2} d^7 \Delta_f \varepsilon^{-1}\right)$ queries to the controlled and standalone energy oracles, while classically one requires $\widetilde{O}\left(\lambda^{-4} L^4 d^3 \Delta_f^2 \varepsilon^{-4}\right)$ queries to the energy function. This suggests an exponential quantum speedup in the sampling precision, and a quartic speedup in the precision of mean estimation.

For mean estimation we also obtain a quadratic speedup in the range Δ_f of the quantity f. However, since $\Delta = \mathcal{O}(L\ell\sqrt{d})$ we require to sample at temperatures in $\Omega(L\ell\sqrt{d})$ in order to avoid an exponentially poor performance in the dimension of the energy potential and its Lipschitz constant. Nevertheless, even at low temperatures this algorithm retains a quadratic advantage in comparison to classical rejection sampling.

5. Conclusion

In this paper, we propose a quantum algorithm for Gibbs sampling from a continuous potential defined on a d-dimensional torus. Our algorithm queries the quantum oracle of the energy potential $\widetilde{O}(d^7 \kappa_{E/2} e^{\Delta/2})$ times in the

[3] We thank Mufan (Bill) Li for confirming this fact.

most notable factors, with only polylogarithmic scaling with respect to the approximation error of the collected samples from the Gibbs distribution in total variation distance. Here Δ is the diameter of the range of the potential or alternatively the thermodynamic β if the potential was considered to be normalized in the range. We also provide examples of conditions under which at high enough temperatures our algorithm is suggestive of exponential quantum advantage at this task.

Our motivation for this research is to use quantum computation as a building block of learning schemes. For instance, the frontiers of research in energy-based learning can take advantage of improved Gibbs samplers from continuous potentials in order to both achieve a better representation of knowledge, and require significantly lower power consumption. Our algorithm achieves this end by solving the second-order PDE known as the Fokker–Planck equation (FPE). When incorporated into energy-based learning (Appendix C), the quantum algorithm does not use coherent queries to classical data, but rather use Hamiltonian simulation techniques to solve a PDE. Therefore, classical data does not need to be prepared in quantum random access memory (QRAM) as typically assumed in the literature on quantum algorithms.

This indicates that, more broadly, investigating steady states of PDEs other than the FPE can also be instrumental in designing classical and quantum machine learning algorithms. Our analysis made it apparent that except for the problem of long mixing time in equilibrium dynamics, the exponential hardness in Gibbs sampling at low temperatures exhibits itself when the eigendirections of the generator of the FPE are far from perpendicular. We believe that this technical constraint may be ramified for special families of potentials which ideally exhibit sufficient expressivity for learning tasks (or for other applications).

In order to obtain these results we take advantage of the efficiency of quantum Fourier transforms in manipulating functions in their Fourier representations. We show that this performance requires sub-exponential concentration of the Fourier components. We also show that this is equivalent to a condition milder than analyticity which we name semi-analyticity. We quantify analyticity and semi-analyticity of functions using parameters we introduce and track how these parameters change under arithmetic operations and compositions. However, many similar properties remain open to be investigated. We also generalize our upsampling results to non-periodic functions using Chebyshev polynomials.

Finally, we mention that our method makes queries directly to the oracle of the energy potential, and therefore is a zeroeth order method. This is unlike typical classical algorithms for Gibbs sampling, specially ones that use the stochastic integration of Langevin dynamics, the SDE associated to the FPE. It therefore remains open to investigate the opportunity for improving our results using quantum queries to the first order oracles of the potential.

Acknowledgments

AM and PR acknowledge the support of NSERC Discovery grant RGPIN-2022-03339. PR further acknowledges the support of Mike and Ophelia Lazaridis, Innovation, Science and Economic Development Canada (ISED), Irréversible Inc., and the Perimeter Institute for Theoretical Physics. Research at the Perimeter Institute is supported in part by the Government of Canada through ISED, and by the Province of Ontario through the Ministry of Colleges and Universities.

Impact statement

This paper presents work whose goal is to advance the fields of machine learning and qunatum computation. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Abramowitz, M. and Stegun, I. A. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1964.

Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. A learning algorithm for Boltzmann machines. *Cognitive science*, 9 (1):147–169, 1985.

Araki, H. and Matsui, T. Ground states of the XY-model. *Communications in mathematical physics*, 101(2):213–245, 1985.

Bakry, D., Gentil, I., Ledoux, M., et al. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

Berglund, N. Kramers' law: Validity, derivations and generalisations. *arXiv preprint arXiv:1106.5799*, 2011.

Berry, D. W., Childs, A. M., Ostrander, A., and Wang, G. Quantum algorithm for linear differential equations with exponentially improved dependence on precision. *Communications in Mathematical Physics*, 356(3):1057–1081, 2017.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Brassard, G., Hoyer, P., Mosca, M., and Tapp, A. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.

Cheng, X., Chatterji, N. S., Bartlett, P. L., and Jordan, M. I. Underdamped langevin MCMC: A non-asymptotic analysis. arxiv e-prints, page. *arXiv preprint arXiv:1707.03663*, 2017.

Chewi, S., Erdogdu, M. A., Li, M. B., Shen, R., and Zhang, M. Analysis of Langevin Monte Carlo from Poincaré to Log-Sobolev. *arXiv preprint arXiv:2112.12662*, 2021.

Childs, A. M. and Liu, J.-P. Quantum spectral methods for differential equations. *Communications in Mathematical Physics*, 375(2):1427–1457, 2020.

Childs, A. M., Liu, J.-P., and Ostrander, A. High-precision quantum algorithms for partial differential equations. *Quantum*, 5:574, 2021.

Childs, A. M., Li, T., Liu, J.-P., Wang, C., and Zhang, R. Quantum algorithms for sampling log-concave distributions and estimating normalizing constants. *arXiv preprint arXiv:2210.06539*, 2022.

Chowdhury, A. N. and Somma, R. D. Quantum algorithms for Gibbs sampling and hitting-time estimation. *arXiv preprint arXiv:1603.02940*, 2016.

Constantine, G. and Savits, T. A multivariate Faa di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, 1996.

Du, Y. and Mordatch, I. Implicit generation and generalization in energy-based models. *arXiv preprint arXiv:1903.08689*, 2019.

Dwivedi, R., Chen, Y., Wainwright, M. J., and Yu, B. Log-concave sampling: Metropolis-Hastings algorithms are fast! In *Conference on learning theory*, pp. 793–797. PMLR, 2018.

Evans, L. C. An introduction to stochastic differential equations. https://www.cmor-faculty.rice.edu/~cox/stoch/SDE.course.pdf, 2023. Accessed: 2023-05-10.

Fan, J., Yuan, B., and Chen, Y. Improved dimension dependence of a proximal algorithm for sampling. In *The Thirty Sixth Annual Conference on Learning Theory*, pp. 1473–1521. PMLR, 2023.

García-Portugués, E., Sørensen, M., Mardia, K. V., and Hamelryck, T. Langevin diffusions on the torus: estimation and applications. *Statistics and Computing*, 29(1): 1–22, 2019.

Ge, R., Lee, H., and Risteski, A. Simulated tempering langevin monte carlo ii: An improved proof using soft markov chain decomposition. *arXiv preprint arXiv:1812.00793*, 2018.

Gilyén, A., Su, Y., Low, G. H., and Wiebe, N. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 193–204, 2019.

Gradshteyn, I. and Ryzhik, I. Tables of integrals, series, and products, 2000.

Grathwohl, W., Wang, K.-C., Jacobsen, J.-H., Duvenaud, D., Norouzi, M., and Swersky, K. Your classifier is secretly an energy based model and you should treat it like one. *arXiv preprint arXiv:1912.03263*, 2019.

Grover, L. K. Fixed-point quantum search. *Physical Review Letters*, 95(15):150501, 2005.

Hamoudi, Y. and Magniez, F. Quantum Chebyshev's inequality and applications. *arXiv preprint arXiv:1807.06456*, 2018.

Hill, M., Mitchell, J., and Zhu, S.-C. Stochastic security: Adversarial defense using long-run dynamics of energy-based models. *arXiv preprint arXiv:2005.13525*, 2020.

Hinton, G. E. Learning multiple layers of representation. *Trends in cognitive sciences*, 11(10):428–434, 2007.

Hinton, G. E., Osindero, S., and Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural computation*, 18 (7):1527–1554, 2006.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

Hsu, E. P. *Stochastic analysis on manifolds*. Number 38. American Mathematical Soc., 2002.

Jordan, S. P., Lee, K. S., and Preskill, J. Quantum algorithms for quantum field theories. *Science*, 336(6085):1130–1133, 2012.

Komatsu, H. A characterization of real analytic functions. *Proceedings of the Japan Academy*, 36(3):90–93, 1960.

Kothari, R. and O'Donnell, R. Mean estimation when you have the source code; or, quantum Monte Carlo methods.

In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1186–1215. SIAM, 2023.

Krantz, S. G. and Parks, H. R. *A primer of real analytic functions*. Springer Science & Business Media, 2002.

Krovi, H. Improved quantum algorithms for linear and nonlinear differential equations. *Quantum*, 7:913, 2023.

Lee, Y. T., Shen, R., and Tian, K. Logsmooth gradient concentration and tighter runtimes for Metropolized Hamiltonian Monte Carlo. In *Conference on learning theory*, pp. 2565–2597. PMLR, 2020.

Li, M. B. and Erdogdu, M. A. Riemannian Langevin algorithm for solving semidefinite programs. *arXiv preprint arXiv:2010.11176*, 2020.

Li, T. and Wu, X. Quantum query complexity of entropy estimation. *IEEE Transactions on Information Theory*, 65(5):2899–2921, 2018.

Li, T. and Zhang, R. Quantum speedups of optimizing approximately convex functions with applications to logarithmic regret stochastic convex bandits. *Advances in Neural Information Processing Systems*, 35:3152–3164, 2022.

Liu, J.-P., Kolden, H. Ø., Krovi, H. K., Loureiro, N. F., Trivisa, K., and Childs, A. M. Efficient quantum algorithm for dissipative nonlinear differential equations. *Proceedings of the National Academy of Sciences*, 118 (35):e2026805118, 2021.

Markowich, P. A. and Villani, C. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp*, 19:1–29, 2000.

Mei, S., Bai, Y., and Montanari, A. The landscape of empirical risk for non-convex losses. *arXiv preprint arXiv:1607.06534*, 2016.

Montanaro, A. Quantum speedup of Monte Carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015.

Nielsen, M. A. and Chuang, I. Quantum computation and quantum information, 2002.

Nijkamp, E., Hill, M., Zhu, S.-C., and Wu, Y. N. Learning non-convergent non-persistent short-run MCMC toward energy-based model. *Advances in Neural Information Processing Systems*, 32, 2019.

Nijkamp, E., Hill, M., Han, T., Zhu, S.-C., and Wu, Y. N. On the anatomy of MCMC-based maximum likelihood learning of energy-based models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 5272–5280, 2020.

Oppenheim, A. V. and Schafer, R. W. Digital signal processing. *Research supported by the Massachusetts Institute of Technology, Bell Telephone Laboratories, and Guggenheim Foundation. Englewood Cliffs, N. J., Prentice-Hall, Inc., 1975. 598 p*, 1975.

Ozgul, G., Li, X., Mahdavi, M., and Wang, C. Stochastic quantum sampling for non-logconcave distributions and estimating partition functions. *arXiv preprint arXiv:2310.11445*, 2023.

Paley, R. E. and Zygmund, A. A note on analytic functions in the unit circle. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 28, pp. 266–272. Cambridge University Press, 1932.

Pang, B., Han, T., Nijkamp, E., Zhu, S.-C., and Wu, Y. N. Learning latent space energy-based prior model. *Advances in Neural Information Processing Systems*, 33: 21994–22008, 2020.

Pavliotis, G. A. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, volume 60. Springer, 2014.

Petrov, V. V. On lower bounds for tail probabilities. *Journal of statistical planning and inference*, 137(8):2703–2705, 2007.

Poulin, D. and Wocjan, P. Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer. *Physical review letters*, 103(22): 220502, 2009.

Ramos-Calderer, S. Efficient quantum interpolation of natural data. *arXiv preprint arXiv:2203.06196*, 2022.

Roberts, G. O. and Tweedie, R. L. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pp. 341–363, 1996.

Roman, S. The formula of Faa di Bruno. *The American Mathematical Monthly*, 87(10):805–809, 1980.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Shen, J., Tang, T., and Wang, L.-L. *Spectral methods: algorithms, analysis and applications*, volume 41. Springer Science & Business Media, 2011.

Song, Y. and Kingma, D. P. How to train your energy-based models. *arXiv preprint arXiv:2101.03288*, 2021.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Terhal, B. M. and DiVincenzo, D. P. Problem of equilibration and the computation of correlation functions on a quantum computer. *Physical Review A*, 61(2):022301, 2000.

Vahdat, A., Kreis, K., and Kautz, J. Score-based generative modeling in latent space. *Advances in neural information processing systems*, 34:11287–11302, 2021.

van Apeldoorn, J., Gilyén, A., Gribling, S., and de Wolf, R. Quantum SDP-solvers: Better upper and lower bounds. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pp. 403–414. IEEE, 2017.

van Handel, R. Probability in high dimension. https://web.math.princeton.edu/~rvan/APC550.pdf, 2016. Accessed: 2022-09-20.

Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Yoder, T. J., Low, G. H., and Chuang, I. L. Fixed-point quantum search with an optimal number of queries. *Physical review letters*, 113(21):210501, 2014.

Ziyin, L., Hartwig, T., and Ueda, M. Neural networks fail to learn periodic functions and how to fix it. *Advances in Neural Information Processing Systems*, 33:1583–1594, 2020.

A. Semi-analytic functions

In this section we analyze the effect of discretization on estimating distributions and derivatives of differentiable functions. Previous works (e.g., (Childs et al., 2021)) assume an upper bound for all derivatives of the function. This, however, is a restrictive assumption as it excludes simple functions such as $\cos(2x)$. We show that a milder condition such as analyticity (or an even a weaker condition we call semi-analyticity) is enough for such results to hold.

We now introduce the notion of semi-analyticity for smooth functions and prove several favourable properties of it using the Fourier spectral method. We borrow some of the ideas presented in Shen et al. (2011, Section 2.2), although ibid is only concerned with functions of a single variable and focused on interpolation errors. In what follows, for a string of length d of non-negative integers $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{Z}^d_{\geq 0}$ we define $\alpha! := \alpha_1! \cdots \alpha_d!$, and $|\alpha| := \alpha_1 + \cdots + \alpha_d$, and use the following notation for higher order derivatives:

$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}. \tag{14}$$

Definition A.1 (Definition 3.1 in the manuscript)**.** *Let $u : \mathbb{R}^d \to \mathbb{R}$ be ℓ-periodic along all axes, and moreover, let $X \sim Unif\big([-\ell/2, \ell/2)^d\big)$ be a uniform random variable. We say u is semi-analytic if there exists $C, a \in \mathbb{R}_+$, such that for any $m \in \mathbb{N}$ we have*

$$\left(\frac{\ell}{2\pi}\right)^m \sqrt{\mathbb{E}\left(\sum_{\alpha:|\alpha|=m} |D^\alpha u(X)|^2\right)} \leq C\, a^m\, m! \tag{15}$$

Furthermore, we refer to (C, a) as the semi-analyticity parameters.

Note that for a semi-analytic function (C, a) are scale invariant; i.e., replacing $u(\cdot)$ by $u(\frac{\cdot}{\alpha})$ and at the same time changing the fundamental domain to $[-\frac{\alpha\,\ell}{2}, \frac{\alpha\,\ell}{2}]^d$, for any $\alpha > 0$, would result in another (C, a)-semi-analytic function. One could absorb the coefficient $\left(\frac{\ell}{2\pi}\right)$ into a, however, we find our current formulation more convenient.

For simplicity, consider the case of having a univariate function f. We recall that the Taylor expansion around the point x_0 is $f(x) = f(x_0) + \sum_{m=1}^\infty \frac{f^{(m)}}{m!}(x - x_0)^m$. Hence, imposing the condition $\left|f^{(m)}(x_0)\right| \leq a^m\, m!$ on the growth of the derivatives guarantees convergence of this series for all $x \in (x - a^{-1}, x + a^{-1})$. Similarly, in the multi-variate case, imposing the condition $|D^\alpha f(x_0)| \leq \alpha!\, a^{|\alpha|}$ guarantees the convergence of the Taylor expansion in the box $\prod_{i=1}^d (x_{0,i} - a^{-1}, x_{0,i} + a^{-1})$, where $x_{0,i}$ denotes the i-th component of x_0. Although a rigorous connection between analyticity and semi-analyticity is provided below, we emphasize that we can understand the parameter a as an inverse convergence radius.

Recall that for a periodic function u over $[-\frac{\ell}{2}, \frac{\ell}{2}]^d$, the discretization on $2N + 1$ points along each axis results in a vector $\overrightarrow{u_N} \in \mathcal{V}_N$. We also use the notation $u_N[n] := u\left(\frac{\ell\, n}{2N+1}\right), \forall n \in [-N..N]^d$. We now define the Fourier transform of an ℓ-periodic function $u : \mathbb{R}^d \to \mathbb{R}$ via

$$\begin{aligned} u(x) &= \sum_{k \in \mathbb{Z}^d} \widehat{u}[k]\, e^{i\frac{2\pi\langle k, x\rangle}{\ell}}, \quad \text{where} \\ \widehat{u}[k] &= \frac{1}{\ell^d} \int_{\mathbb{T}} u(x) e^{-i\frac{2\pi\langle k, x\rangle}{\ell}}\, dx\,. \end{aligned} \tag{16}$$

Note that the Fourier transform of $D^\alpha u$ is $\left(\frac{2\pi}{\ell}\right)^{|\alpha|} (ik_1)^{\alpha_1} \cdots (ik_d)^{\alpha_d} \widehat{u}[k]$, and hence by Parseval's theorem we have $\left(\frac{\ell}{2\pi}\right)^{|\alpha|} \mathbb{E}\left[(D^\alpha u)^2\right] = \sum_{k\in\mathbb{Z}^d} k_1^{2\alpha_1} k_2^{2\alpha_2} \cdots k_d^{2\alpha_d} |\widehat{u}[k]|^2$. Moreover, since $\sum_{\alpha:|\alpha|=m} k_1^{2\alpha_1} \cdots k_d^{2\alpha_d} = \left(k_1^2 + \cdots + k_d^2\right)^m$, we conclude that Definition A.1 is equivalent to

$$|u|_m := \sqrt{\sum_{k\in\mathbb{Z}^d} \|k\|^{2m} \left|\widehat{u}[k]\right|^2} \leq C\, a^m\, m!. \tag{17}$$

It is straightforward to check that the quantity introduced above is a semi-norm.

We now provide examples of semi-analytic functions and show that this condition is quite mild. In particular, in Proposition A.2 we prove that every analytic function is semi-analytic (hence, the naming).

Example A.1. *Any function with finitely many non-zero Fourier coefficients is semi-analytic with* $C = \sqrt{\mathrm{Var}[u(X)]}$ *and* $a = k_0$*, where* $k_0 := \max\{\|k\| : \widehat{u}[k] \neq 0\}$.

Example A.2. *For any* $z > 0$*, the function* $u(x) = e^{z\cos(x)}$ *with domain* $[-\pi, \pi]^d$ *is* $\left(I_0(z)\, e^{z/2}, \max\{\frac{z}{2}, 1\}\right)$*-semi-analytic with* I_0 *being the modified Bessel function of the first kind. To see this, note that the Fourier coefficients of* u *are described by the modified Bessel function of the first kind (Abramowitz & Stegun, 1964, page 376):*

$$\widehat{u}[k] = \frac{1}{2\pi} \int_{x=-\pi}^{\pi} e^{-ikx}\, e^{z\cos(x)}\, dx = I_k(z). \tag{18}$$

Furthermore,

$$I_k(z) = \left(\frac{z}{2}\right)^k \sum_{\ell \geq 0} \frac{\left(\frac{z}{2}\right)^{2\ell}}{\ell!(k+\ell)!} \leq \left(\frac{z}{2}\right)^k \frac{1}{k!} \sum_{\ell \geq 0} \frac{\left(\frac{z}{2}\right)^{2\ell}}{(\ell!)^2} = \frac{1}{k!} \left(\frac{z}{2}\right)^k I_0(z). \tag{19}$$

Therefore, $|\widehat{u}[k]| \leq \left(\frac{z}{2}\right)^k \frac{I_0(z)}{k!}$*, which allows us to write*

$$\sum_{k\in\mathbb{Z}} |k|^{2m} \left|\widehat{u}[k]\right|^2 \leq (I_0(z))^2 \sum_{k\in\mathbb{Z}} \left(\frac{z}{2}\right)^{2k} \frac{k^{2m}}{(k!)^2}. \tag{20}$$

Hence, using the fact that the ℓ_1*-norm is larger in value than the* ℓ_2*-norm, we conclude that*

$$|u|_m \leq I_0(z) \sum_{k\in\mathbb{Z}} \left(\frac{z}{2}\right)^k \frac{k^m}{k!} \leq I_0(z)\, e^{z/2} \left(\max\{\frac{z}{2}, 1\}\right)^m m! \tag{21}$$

where the last inequality follows from Lemma E.1.

Proposition A.2. *Every periodic real-analytic function is semi-analytic.*

Proof. Let f be a real analytic and periodic function. It follows from Komatsu (1960, Lemma 1) that there exist $C, a > 0$ such that

$$\sup_{x\in\mathbb{T}} |D^\alpha f(x)| \leq C\, a^{|\alpha|}\, \alpha!. \tag{22}$$

From this we conclude that

$$\sqrt{\sum_{\alpha:|\alpha|=m} \mathbb{E}\left[(D^\alpha f)^2\right]} \leq C\, a^m \sqrt{\sum_{\alpha:|\alpha|=m} (\alpha!)^2} \leq C\, a^m \sum_{\alpha:|\alpha|=m} \alpha! \leq 3^{d-1} C\, a^m\, m! \tag{23}$$

where the last inequality follows from Lemma E.4. □

We let F_N denote the unitary representing the d-dimensional discrete Fourier transform and adopt the notation $\widetilde{u}_N := F_N \overrightarrow{u_N}$. We drop the subscript N when it is clear from the context. Moreover, consider $\Sigma \subseteq \Gamma$ as an embedding of a finite alphabet Σ in Γ, Γ being either a larger finite alphabet or $\mathbb{N}$. We also consider the natural embedding of spaces of functions $\iota : \ell^2(\Sigma) \hookrightarrow \ell^2(\Gamma)$ induced by the inclusion $\Sigma \subseteq \Gamma$ and the usual ℓ^2 norm $\|a\| = \sqrt{\sum_{n\in\Gamma} |a[n]|^2}$ and the induced metric $d(a, b) = \|a - b\|$.

We observe that when dealing with a periodic function having a Fourier spectrum with bounded support, such that $\widehat{u}[k] = 0$ for $k \notin [-k_0..k_0]^d$, Nyquist's well-known theorem guarantees exact recovery of the function (Oppenheim & Schafer, 1975). Specifically, given a discretization on a lattice V_N with $N \geq k_0$, there exists a classical algorithm to reconstruct the entire continuous function $u(x)$. In the following, we show that under the milder condition of semi-analyticity, one can still achieve approximate reconstructions. Even though the reconstructions will not be entirely accurate, we can limit the errors to a poly-logarithmic overhead by exploiting the fact that the values of $\widehat{u}[k]$ are exponentially small for sufficiently large k (as shown in Lemma A.3 below). Furthermore, it is worth noting that by measuring quantum states in the computational basis, we obtain a sample drawn from a distribution corresponding to the squared amplitudes. This feature enables us to develop a sampler in the continuum.

Lemma A.3. *For a (C, a)-semi-analytic periodic function u, with N an integer satisfying $N \geq 2a$, we have*

$$\sqrt{\sum_{k:\|k\|_2 \geq N} |\widehat{u}[k]|^2} \leq 2e^3\, C\, e^{-\frac{N}{a}\left(1-\frac{1}{2e}\right)}. \tag{24}$$

Proof. We have

$$\sum_{k:\|k\|_2 \geq N} |\widehat{u}[k]|^2 \leq N^{-2m} \sum_{k:\|k\|_2 \geq N} \|k\|^{2m}\, |\widehat{u}[k]|^2 \leq N^{-2m}\, C^2\, a^{2m}\, (m!)^2\,. \tag{25}$$

Now using $m! \leq \frac{m^{m+1}}{e^{m-1}}$ we obtain

$$N^{-m}\, a^m\, m! \leq \frac{e^2}{a} N^{-m} \left(\frac{a(m+1)}{e}\right)^{m+1}. \tag{26}$$

Setting $m+1 = \lfloor N/a \rfloor$ yields the bound

$$N^{-m}\, a^m\, m! \leq \frac{e^3}{a} N\, e^{-N/a}. \tag{27}$$

We can now use the inequality $x \leq \alpha e^{\frac{x}{\alpha\, e}}$ for all $x \in \mathbb{R}$ and all $\alpha > 0$ by setting $\alpha = 2$ and $x = N/a$ to complete the proof. $\square$

The reader may notice analogies between the result of Lemma A.3 and the sub-exponential decay bounds in the literature of concentration of measure. We discuss this connection in Appendix A.1.

Lemma A.4. *Let u be a (C, a)-semi-analytic periodic function with period $[-\frac{\ell}{2}, \frac{\ell}{2}]^d$, and let N be an integer such that $N \geq 2ad$. We have*

$$d\left(\frac{1}{(2N+1)^{d/2}}\, \widetilde{u}_N, \widehat{u}\right) \leq 2\sqrt{2}e^3\, Ce^{-\frac{3N}{5a}}. \tag{28}$$

Proof. We start by noting that

$$\begin{aligned}
\frac{1}{(2N+1)^{d/2}}\, \widetilde{u}_N[k] &= \frac{1}{(2N+1)^d} \sum_{n\in[-N..N]^d} u_N[n]\, e^{-i\frac{2\pi\langle k,n\rangle}{2N+1}} \\
&= \frac{1}{(2N+1)^d} \sum_{n\in[-N..N]^d} \sum_{k'\in\mathbb{Z}^d} \widehat{u}[k']\, e^{-i\frac{2\pi\langle k-k',n\rangle}{2N+1}} \\
&= \widehat{u}[k] + \sum_{p\in\mathbb{Z}^d\setminus\{0\}} \widehat{u}[k+(2N+1)p].
\end{aligned} \tag{29}$$

Therefore,

$$d\left(\frac{1}{(2N+1)^{d/2}}\, \widetilde{u}_N, \widehat{u}\right)^2 = \sum_{k\in[-N..N]^d} \left| \sum_{p\in\mathbb{Z}^d\setminus\{0\}} \widehat{u}[k+(2N+1)p] \right|^2 + \sum_{k\in\mathbb{Z}^d\setminus[-N..N]^d} |\widehat{u}[k]|^2 \tag{30}$$

where the second term was upper-bounded in Lemma A.3. As for the first term

$$\begin{aligned}\sum_{k\in[-N..N]^d}\left|\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\widehat{u}[k+(2N+1)p]\right|^2 &\leq \sum_{k\in[-N..N]^d}\left(\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+(2N+1)p\|^{-2m}\right.\\ &\qquad\left.\times\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+(2N+1)p\|^{2m}\left|\widehat{u}[k+(2N+1)p]\right|^2\right)\\ &\leq a^{2m}C^2\,(m!)^2\max_{k\in[-N..N]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+(2N+1)p\|^{-2m}.\end{aligned}\tag{31}$$

We note that

$$\begin{aligned}\max_{k\in[-N..N]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+(2N+1)p\|^{-2m} &\leq \max_{k\in[-N..N]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+2N\,p\|^{-2m}\\ &\leq N^{-2m}\max_{x\in[-1,1]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|x+2\,p\|^{-2m},\end{aligned}\tag{32}$$

where the first inequality follows from term-by-term comparison of the sums. Using Lemma E.5, if $m\geq d$, we get

$$d\left(\frac{1}{(2N+1)^{d/2}}\,\widetilde{u}_N,\widehat{u}\right)\leq 2\sqrt{2}a^m C\,N^{-m}\,2^{d/2}\,m!\leq 2\sqrt{2}a^m\,C\,N^{-m}\,2^{n/4a}\,m!\tag{33}$$

and setting $m=\lfloor N/a\rfloor-1$ (which guarantees $m\geq d$) yields

$$d\left(\frac{1}{(2N+1)^{d/2}}\,\widetilde{u}_N,\widehat{u}\right)\leq 2e^3\sqrt{2}\,Ce^{-\frac{n}{a}\left(1-\frac{1}{2e}-\frac{\ln(2)}{4}\right)}\tag{34}$$

which concludes the proof. $\square$

For our purposes, we will be applying Lemma A.4 to normalized vectors. Here we highlight the following distance bound as a corollary following immediately from Lemma A.4 and Lemma E.6.

Corollary A.5. *Let $N\geq 2ad$ and u be a (C,a)-semi-analytic periodic function. We have*

$$d\left(F_N\left|u_N\right\rangle,\frac{\widehat{u}}{\mathcal{U}}\right)\leq\frac{4\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6N/a}\tag{35}$$

where $\mathcal{U}=\left(\mathbb{E}\left[|u(X)|^2\right]\right)^{1/2}$, with $X\sim\mathrm{Unif}\left([-\ell/2,\ell/2]^d\right)$

Now we show that upsampling a semi-analytic function is useful in achieving minor aliasing effects.[4] Recall the sampling procedure introduced in Section 2. That is, we measure the output state of the algorithm, say $|\psi\rangle$, in the computational basis to obtain $x\in V_N$. We then sample uniformly at random from the box $\prod_{i=1}^d[x_i-\frac{\ell}{4N+2},x_i+\frac{\ell}{4N+2}]$. We call this procedure *continuous sampling* from $|\psi\rangle$.

Remark A.6. *The total variation distance of continuous sampling from two quantum states is upper bounded by the ℓ_2-norm of their difference. To see this, let $|\psi\rangle$ and $|\phi\rangle$ be two quantum states. We denote the probability density associated with the random variable obtained from continuous sampling from $|\psi\rangle$ by μ_ψ and we note that*

$$\mu_\psi(x)=\sum_{n\in[-N..N]^d}\mathbf{1}_{\{x\in\mathbb{B}_n\}}\left|\psi[n]\right|^2\left(\frac{2N+1}{\ell}\right)^d\tag{36}$$

where $\mathbf{1}_{\{x\in\mathbb{B}_n\}}$ is the identifier function; i.e., it is 1 if $x\in\mathbb{B}_n=\prod_{i=1}^d\left[\frac{n_i\ell}{2N+1}-\frac{\ell}{4N+2},\frac{n_i\ell}{2N+1}+\frac{\ell}{4N+2}\right]$ and 0 otherwise.

[4] In signal processing, aliasing effects refer to the errors caused by Fourier interpolation, specially when the tails of the Fourier transformation (that is, the very high and very low frequency components) have non-negligible amplitudes (Oppenheim & Schafer, 1975).

One can then write

$$\frac{1}{2}\int dx\ |\mu_\psi(x)-\mu_\phi(x)| = \frac{1}{2}\sum_{n\in[-N..N]^d}\left||\psi[n]|^2-|\phi[n]|^2\right| \le \||\psi\rangle-|\phi\rangle\| \tag{37}$$

where the inequality follows from Lemma E.7.

Proposition A.7. *Given a (C,a)-semi-analytic periodic function u, an integer $N\ge 2ad$, and a quantum state $|\psi_N\rangle\in\mathcal{V}_N$ satisfying $\||\psi_N\rangle-|u_N\rangle\|\le\delta$, there exists $M\in\mathbb{N}$ such that continuous sampling from $F_M^{-1}\iota F_N|\psi_N\rangle$ results in an ε-approximation to the continuous distribution proportional to u^2 in total variation distance, where*

$$\varepsilon\le\delta+\frac{8\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6\,\frac{N}{a}}. \tag{38}$$

Proof. Firstly, note that for any integer r, since sampling from $|u_r\rangle$ reaches the actual $|u|^2$ distribution as $r\to\infty$, there exists an integer M^* such that $|u_r\rangle$ for all $r\ge M^*$ gives δ-approximation of the continuous distribution. If $N\ge M^*$, then the statement is trivially satisfied after setting $M=N$. Otherwise, let $M=M^*$, and note that due to Corollary A.5 and by an application of the triangle inequality

$$\|\iota F_N|u_N\rangle-F_M|u_M\rangle\|\le\frac{8\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6\,\frac{N}{a}}. \tag{39}$$

Since isometries preserve the ℓ_2-norm, by another application of the triangle inequality we get

$$\left\|F_M^{-1}\iota F_N\ |\psi_N\rangle-|u_M\rangle\right\|\le\frac{8\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6\,\frac{N}{a}}. \tag{40}$$

Finally, using another triangle inequality for the total-variation distance and Remark A.6 we obtain the result. □

We now investigate the gate complexity of interpolating a semi-analytic function.

Theorem A.8 (Theorem 3.2 in the manuscript)**.** *Given an L-Lipschitz (C,a)-semi-analytic periodic function u, an integer $N\ge 2ad$, and a quantum state $|\psi\rangle\in\mathcal{V}_N$, such that $\||\psi\rangle-|u_N\rangle\|\le\delta$, there exists a quantum algorithm with gate complexity $\mathcal{O}\left(\frac{dN}{a}\operatorname{polylog}\left(NdL\ell/C\right)\right)$ that returns samples from a distribution within at most ε total variation distance from the distribution proportional to u^2, where*

$$\varepsilon\le\delta+\frac{16\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6\,\frac{N}{a}},$$

and $\mathcal{U}=\sqrt{\mathbb{E}\,u^2(X)}$, with $X\sim\mathrm{Unif}([-\frac{\ell}{2},\frac{\ell}{2}]^d)$.

Notice that by 'given a state $|\psi\rangle$' in the theorem above, we mean that we are given one copy of $|\psi\rangle$.

Proof. In this proof, for an integer $\alpha\in\mathbb{N}$ the notation $[\alpha]$ stands for $\{0,1,\dots,\alpha\}$. Using Lemma E.8, we set $M=\left\lceil\frac{1}{2\varepsilon'}\frac{L\ell d/2+10/3\sqrt{2}\,ae^4C}{\mathcal{U}}\right\rceil$ with $\varepsilon'=\frac{8\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6\,\frac{N}{a}}$. We now show that we can implement $F_M^{-1}\iota F_N$ with $\widetilde{\mathcal{O}}(d\log M)$ gates. Firstly, note that

$$F_N=\bigotimes_{i=1}^{d}f_N^{(i)} \tag{41}$$

where $f_N=\sum_{m,k\in[2N]}e^{-i\frac{2\pi(k-N)(m-N)}{2N+1}}|k\rangle\langle m|$, and the superscript i means that it acts non-trivially on the i-th register. We notice that $f_N=T_N\hat{f}_NT_N$, where $T_N=\sum_{k\in[2N]}e^{-i\frac{2\pi Nk}{2N+1}}|k\rangle\langle k|$, and $\hat{f}_N=\sum_{m,k\in[2N]}e^{-i\frac{2\pi km}{2N+1}}|k\rangle\langle m|$ is the usual quantum Fourier transform and thus can be implemented using $\mathcal{O}(\log N\log\log N)$ gates. Furthermore, it is straightforward to implement T_N in $\mathcal{O}(\log N)$. Overall, the gate complexity of F_N is $\widetilde{\mathcal{O}}(d\log N)$, since it can be implemented via d applications of f_N in parallel.

It remains to show that ι itself can also be implemented using $\mathcal{O}(d\log M)$ gates. Consider the isometry $\hat{\iota}:\mathcal{V}_N\to\mathcal{V}_M$ defined as $\hat{\iota}=\bigotimes\widehat{\iota'}^{(i)}$ via $\widehat{\iota'}:|n\rangle\mapsto|n\rangle$ (for $n\in[2N]$). We note that $\widehat{\iota'}$ can be performed by adding auxiliary qubits prepared in the $|0\rangle$ state. Also, from $\iota=\bigotimes_{i=1}^{d}\iota'^{(i)}$ with $\iota':|k+N\rangle\mapsto|k+M\rangle$ for $k\in[-N..N]$, we conclude that $\iota'=S\widehat{\iota'}$, where $S\in U(\mathcal{V}_{\mathcal{M}})$ is a shift operator for integers represented in the computational basis states, $S:|m\rangle\mapsto|m+M-N \bmod 2M+1\rangle$. Finally, note that $S=\hat{f}_N S'\hat{f}_N$, given $S'=\sum_{k\in[2N]}e^{-i\frac{2\pi(M-N)k}{2M+1}}|k\rangle\langle k|$, and the latter operator has gate complexity $\mathcal{O}(\log M)$. We therefore conclude that the complexity of implementing ι is $\mathcal{O}(d\log M)$. □

A.1. Concentration of measure

Sub-exponential distributions are studied in the context of high-dimensional probability theory. Intuitively, a random variable is considered sub-exponential if its probability distribution function has a tail that vanishes exponentially or faster (Vershynin, 2018). We make a connection between this concept and our notion of semi-analyticity, which will later allow us to better understand the latter class of functions. Let us recall the Bernstein random variables, which will appear to be useful later in this section.

Definition A.9. *X is a Bernstein random variable, if $X\geq 0$ almost surely, and for some $A,b>0$ its moments are upper bounded as*

$$\mathbb{E}\,X^m\leq A\,b^m\,m!, \tag{42}$$

for all positive integers m.

Following Boucheron et al. (2013) we prove concentration bounds on a Bernstein random variable.

Lemma A.10. *Let X be the Bernstein random variable defined in Definition A.9. It is the case that*

$$\mathbb{P}[X\geq t]\leq\begin{cases}\max(A,1)\,e^{-\frac{(t-b)^2}{8b^2}}, & \text{if } t\leq 3b,\\ e\,\max(A,1)\,e^{-\frac{t}{2b}}, & \text{if } t>3b.\end{cases} \tag{43}$$

Proof. Let us first upper bound the generating function corresponding to X. Let $0\leq\lambda<b^{-1}$, then

$$\mathbb{E}\,e^{\lambda X}=1+\sum_{m\in\mathbb{N}}\frac{\lambda^m\,\mathbb{E}\,X^m}{m!}\leq 1+A\sum_{m\in\mathbb{N}}(b\lambda)^m=1+A\frac{\lambda b}{1-\lambda b}. \tag{44}$$

Moreover, if $0\leq\lambda\leq\frac{1}{2b}$, we have $\frac{1}{1-\lambda b}\leq 1+2\lambda b$, which together with the identity $1+x\leq e^x$ yield

$$\mathbb{E}\,e^{\lambda X}\leq\max(A,1)\,\exp\{\lambda b+2\lambda^2b^2\},\quad\forall\lambda\in[0,\frac{1}{2b}]. \tag{45}$$

We may now upper bound the tail probability via Chernoff's bound

$$\mathbb{P}[X\geq t]\leq\inf_{0\leq\lambda}e^{-\lambda t}\,\mathbb{E}[e^{\lambda X}]\leq\max(A,1)\min_{0\leq\lambda\leq\frac{1}{2b}}\exp\{-\lambda(t-b-2\lambda b^2)\}. \tag{46}$$

For $t\leq 3b$, we make the choice $\lambda=\frac{t-b}{4b^2}$, and otherwise, we choose $\lambda=\frac{1}{2b}$ to conclude the result. □

Note that the tail of a Bernstein random variable shows a sub-exponential behavior eventually, as described by Eq. (43). Indeed, the set of Bernstein random variables coincides with the set of positive sub-exponential distributions as stated bellow.

Proposition A.11. *The set of Bernstein random variables is the set of almost surely positive random variables that are sub-exponential.*

Proof. This follows from the characterization of sub-exponential random variables in (Vershynin, 2018, Proposition 2.7.1), according to which, the positive random variable X is sub-exponential if and only if $\mathbb{E}\, X^m \le Q^m\, m^m$ for some $Q \ge 0$.[5] Now, let X be a sub-exponential distribution. Using $\frac{m^m}{e^{m-1}} \le m!$, we have

$$\mathbb{E}\, X^m \le Q^m\, m^m \le e^{-1}(Qe)^m\, m!, \tag{47}$$

which concludes that X has a Bernstein property.

Conversely, assume X has a Bernstein property. From $m! \le m^m$, one concludes that

$$\mathbb{E}\, X^m \le Ab^m m! \le (\max(A,1)\, b)^m\, m^m, \tag{48}$$

which provides that X is a sub-exponential random variable. □

Now we show the connection between the notions of concentration of measure and the semi-analyticity condition in Definition A.1. The next definition allows us to make this connection clear.

Definition A.12. *Consider the Fourier transform of an ℓ-periodic function $u : [-\frac{\ell}{2}, \frac{\ell}{2}]^d \to \mathbb{R}$, denoted by $(\hat{u}[k])_k$. Note that $(|\hat{u}[k]|^2/\mathcal{U}^2)_k$ defines a probability distribution on the sample space $\mathbb{Z}^d$. We call the random variable K_u corresponding to this distribution the Fourier random variable of u.*

With this definition at hand, we make the following connection between semi-analyticity and the Bernstein random variables.

Theorem A.13. *A periodic function u is semi-analytic if and only if $\|K_u\|$ has the Bernstein property. In particular*

- *if u is (C,a)-semi-analytic, then $\|K_u\|$ has a Bernstein property with parameters $(C\mathcal{U}^{-1}, a)$; and*
- *if $\|K_u\|$ has a Bernstein property with parameters (A,b), then u is $(\sqrt{2Ae}, 4b)$-semi-analytic.*

Proof. u is semi-analytic $\Rightarrow$ $\|K_u\|$ has a Bernstein property: From (19), we have

$$\sqrt{\mathbb{E}\,\|K_u\|^{2m}} \le \mathcal{U}^{-1}\, C\, a^m\, m!. \tag{49}$$

Putting this together with the Jensen inequality $\mathbb{E}\,\|K_u\|^m \le \sqrt{\mathbb{E}\,\|K_u\|^{2m}}$, proves that $\|K_u\|$ is a Bernstein random variable with parameters $(C\mathcal{U}^{-1}, a)$.

$\|K_u\|$ has a Bernstein property $\Rightarrow$ u is semi-analytic: By definition we have $\mathbb{E}\,\|K_u\|^{2m} \le A\, b^{2m}\,(2m)!$. Note that

$$(2m)! \le \frac{(2m)^{2m+1}}{e^{2m-1}} \le 2e\, 4^{2m}\, \frac{m^{2m-2}}{e^{2m-2}} \le 2e\, 4^{2m}(m!)^2 \tag{50}$$

where the second inequality uses $4^m \ge m^3$. This implies $\sqrt{\mathbb{E}\,\|K_u\|^{2m}} \le \sqrt{2Ae}\,(4b)^m\, m!$. □

Making this connection allows us to obtain a Fourier concentration result similar to Lemma A.3.

Corollary A.14. *Let u be (C,a)-semi-analytic. It is the case that*

$$\sum_{k:\|k\|\ge t} |\hat{u}[k]|^2 \le \begin{cases} \max(C,\mathcal{U})\, e^{-\frac{(t-a)^2}{8a^2}}, & \text{if } t \le 3a, \\ e\, \max(C,\mathcal{U})\, e^{-\frac{t}{2a}}, & \text{if } t > 3a. \end{cases} \tag{51}$$

Proof. This follows directly from the first implication in Theorem A.13 and Lemma A.10. □

[5] Vershynin (2018) take the exponent m to be any real number larger than 1, but one can readily observe that $m \in \mathbb{N}$ is also a sufficient condition in their proofs.

Note that the result of Lemma A.3 can also be proven using the Markov inequality $\mathbb{P}\left[\|K_u\| > t\right] = \mathbb{P}\left[\|K_u\|^m > t^m\right] \leq \frac{\mathbb{E}\,\|K_u\|^m}{t^m}$ by a suitable choice of m. This correspondence further lets us find functions that saturate the semi-analyticity condition, as in the following example.

Example A.3. *Let $z > 1$. The 2π-periodic function $u(x) = \frac{z-1}{1-2\sqrt{z}\cos(x)+z}$ satisfies the following inequality.*

$$\frac{1}{(1+z^{-1})^{1/2}} \frac{m!}{(1-z^{-1})^m} \leq \sqrt{\mathbb{E}\,\|K_u\|^{2m}} \leq \sqrt{\frac{2e}{1+z^{-1}}} \max\left(8, \frac{8}{z-1}\right)^m m! \tag{52}$$

Therefore u is both upper bounded and lower bounded by growth rates in the definition of semi-analyticity, although for different choices of parameters. We refer the reader to Fig. 2 for visual demonstrations.

To obtain (52), note that the Fourier transform of $u(x)$ is $\widehat{u}[k] = z^{-\frac{|k|}{2}}$ since

$$1 + 2\sum_{k=1}^{\infty} \cos(kx) z^{-\frac{k}{2}} = \frac{z-1}{1+z-2\sqrt{z}\cos x}. \tag{53}$$

This implies $\mathcal{U}^2 = \frac{1+z^{-1}}{1-z^{-1}}$, and moreover the moments of $\|K_u\|$ can be lower bounded as follows

$$\frac{1+z^{-1}}{1-z^{-1}}\,\mathbb{E}\,\|K_u\|^m = \sum_{k=0}^{\infty} k^m z^{-k} \tag{54}$$

$$\geq \frac{\partial^m}{\partial (z^{-1})^m}\left(\sum_{k\geq 0} z^{-k}\right) = \frac{\partial^m}{\partial (z^{-1})^m}\left(\frac{1}{1-z^{-1}}\right) = \frac{m!}{(1-z^{-1})^{m+1}}. \tag{55}$$

We, therefore, have

$$\frac{1}{1+z^{-1}} \frac{m!}{(1-z^{-1})^m} \leq \mathbb{E}\,\|K_u\|^m \leq \frac{1}{1+z^{-1}} \max\{2, \frac{2}{z-1}\}^m m! \tag{56}$$

where the upper bound follows from Lemma E.2.

We now use another result from probability theory, namely the Paley–Zygmund lower bound (Paley & Zygmund, 1932; Petrov, 2007), to prove that taking $\Omega(a)$ points is necessary to produce samples from a distribution arbitrarily close to the one generated by the underlying distribution.

Lemma A.15 (Paley–Zygmund)**.** *Let X be a non-negative random variable (that is, $X \geq 0$ almost surely). For any $\theta \in (0,1)$, it is the case that*

$$\mathbb{P}\left[X > \theta\,\mathbb{E}[X]\right] > (1-\theta)^2 \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}. \tag{57}$$

Proof. Note that $X = X\mathbf{1}_{X \leq \theta\,\mathbb{E}[X]} + X\mathbf{1}_{X > \theta\,\mathbb{E}[X]}$, from which it is straightforward to conclude

$$\mathbb{E}[X] \leq \theta\,\mathbb{E}[X] + \sqrt{\mathbb{E}[X^2]\mathbb{P}\left[X > \theta\,\mathbb{E}[X]\right]}, \tag{58}$$

where the second term is due to the Cauchy–Schwartz inequality. □

Theorem A.16 (Theorem 3.3 in the manuscript)**.** *Let u be a (C,a)-semi-analytic function. Consider any exact discretization $|u_N\rangle$ on the discrete lattice with $N \leq \theta a/16$, where $\theta \in (0,1)$. There is no algorithm that can return samples close to the actual distribution (proportional to u^2) with a guaranteed error of less than $(1-\theta)^2\,\frac{1}{1024e}$.*

Proof. Note that $f(x) = \frac{C}{\sqrt{e}} \frac{z-1}{1-2\sqrt{z}\cos(2\pi x/\ell)+z}$ is (C,a)-semi-analytic for $z = 1 + \frac{8}{a}$. From the example above, we have

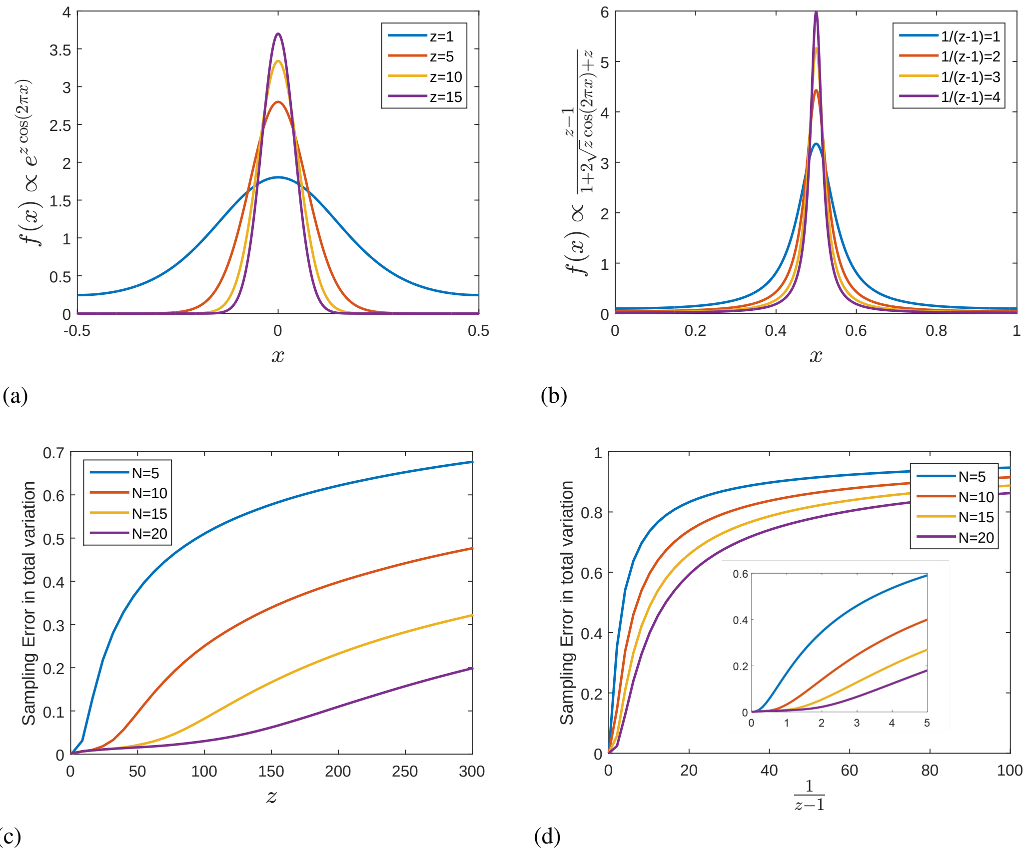

Figure 2. (a) and (b) show two families of functions considered respectively in Example A.2 and Example A.3. The functions are normalized so that $\int_{x\in[0,1]} \mathrm{d}x\ (f(x))^2 = 1$. That is, f^2 represents a distribution over one period. Note how in (a) the smoothness of the functions is controlled by the parameter z and in (b) it is controlled by $(z-1)^{-1}$. (c) and (d) show the Fourier interpolation accuracy on the two respective families of functions considered in (a) and (b). We demonstrate the interpolation error given the state $|f_N\rangle$ for different N. Note that in both cases having N larger than our upper bounds on a results in a sampling error less than 0.1. The sampling error is shown with respect to the smoothness parameters z and $(z-1)^{-1}$, obtained by the application of the upsampling algorithm using $M = 200$. Recall from Example A.2 and Example A.3 that we may think of $\max(1, \frac{z}{2})$ and $\max(8, \frac{8}{z-1})$ as upper bounds on the (average) inverse convergence radius of the respective functions in panels (a) and (b).

$\mathbb{E}\,\|K_f\| \geq \frac{a}{16}$, therefore

$$\mathbb{P}\left[\|K_u\| > \theta\frac{a}{16}\right] > \mathbb{P}\left[\|K_f\| > \theta\,\mathbb{E}\,\|K_f\|\right] > (1-\theta)^2\frac{(a/16)^2}{2ea^2} = (1-\theta)^2\frac{1}{512e}. \tag{59}$$

Hence, $\|K_f\|$ is large with a considerable probability. Let $g : [-\ell/2, \ell/2] \to \mathbb{R}$ be a function with the Fourier transform

$$\widehat{g}[k] = \alpha\begin{cases}\sum_{p\in\mathbb{Z}^d}\widehat{f}[k+p(2N+1)], & \text{if } k\in[-N..N]^d,\\ 0, & \text{otherwise.}\end{cases} \tag{60}$$

Here α is a normalization constant chosen such that $\mathbb{E}_{X\sim\text{Unif}}[g(X)^2] = C^2$. One can readily verify that $g\left(\frac{n\ell}{2N+1}\right) = \alpha f\left(\frac{n\ell}{2N+1}\right)$. Therefore, $|f_N\rangle = |g_N\rangle$. Moreover, note that g is also (C,a)-semi-analytic due to Example A.1 and that the total variation distance between the distributions whose densities are proportional to $|f|^2$ and $|g|^2$ is at least $\mathbb{P}\left[\|K_f\| > N\right]$, which is itself lower bounded by $(1-\theta)^2\frac{1}{512e}$ due to $N < \theta\frac{a}{16}$ and (59). Let $|\psi_N\rangle \in \mathcal{V}_N$ denote the discretization of f and g (so $|\psi_N\rangle = |f_N\rangle = |g_N\rangle$). Given the promises and the state $|\psi_N\rangle$, any algorithm will sample from a distribution, say $\mathcal{P}$, which is at least $\frac{1-\theta^2}{1024e}$ away from at least one of P_f and P_g. Hence, the algorithm fails as stated upon processing either g or f as the underlying functions. □

A.2. The Fourier differentiation method

Here we describe the Fourier pseudo-spectral method used in our work and prove several useful properties of it. Let $u : \mathbb{R}^d \to \mathbb{R}$ be ℓ-periodic in all dimensions. We define the Fourier derivatives on the discretized lattice as follows:

$$\widetilde{\partial}_j u_N[n] := F_N^{-1}\left(\frac{i2\pi k_j}{\ell}\left(F_N \overrightarrow{u_N}\right)[k]\right). \tag{61}$$

A straightforward calculation yields the following convolution relation

$$\widetilde{\partial}_j u_N[n] = \sum_{m\in[-N..N]} u_N[n_1,\cdots,n_{j-1},m,n_{j+1},\cdots,n_d]\, a[n_j - m] \tag{62}$$

where

$$a[m] = \begin{cases} 0, & \text{if } m = 0,\\ \frac{\pi\,(-1)^{m+1}}{\ell\,\sin\left(\frac{\pi m}{2N+1}\right)}, & \text{otherwise.}\end{cases} \tag{63}$$

Higher order derivatives can then be defined as consecutive applications of the first order operators:

$$\begin{aligned}\widetilde{\partial}_j^r u_N[n] :=& F_N^{-1}\left[\left(\frac{i2\pi k_j}{\ell}\right)^r \left(F_N \overrightarrow{u_N}\right)[k]\right]\\ =& \sum_{m\in[-N..N]} u_N[n_1,\cdots,n_{j-1},m,n_{j+1},\cdots,n_d]\, a^{(r)}[n_j - m],\end{aligned} \tag{64}$$

where $a^{(r)} = a * a * \cdots * a$ is the r-fold convolution. This means that taking the r-th derivatives in the j-th dimension is identical to r consecutive applications of the first derivative in direction j. However, if the number of discretization points is even the Fourier derivatives may be define differently (as in Shen et al. (2011)) in which case this composability property may not hold. Note that $\left\|\widetilde{\partial}_j u\right\|_2 \le 2\pi\, N/\ell\, \|u\|_2$ since each derivative is an operator with eigenvectors being the Fourier basis with eigenvalues $\frac{i2\pi k_j}{\ell}$. In what follows, we discuss some properties of this differentiation operation. Most notably, we show that it respects the Leibniz product rule and that the maximum derivative is at most $\mathcal{O}(N\log N)$ bigger than the largest value the function attains. Note that analogously when using $Df := \frac{f(x+h)-f(x)}{h}$ for finite difference approximation of conventional derivatives, using $h = \frac{1}{2N+1}$, the approximation is at most $\mathcal{O}(N)$ larger than the maximum value of $f(x)$.

Proposition A.17. *Let u and v be two ℓ-periodic functions in all dimensions. The Fourier derivatives $\widetilde{\partial}_j$ have the following properties:*

(a) The product rule: $\widetilde{\partial}_j(u\cdot v) = (\widetilde{\partial}_j u)\cdot v + u\cdot(\widetilde{\partial}_j v)$.

(b) $\left\|\widetilde{\partial}_j u\right\|_\infty \le \frac{2\pi}{\ell}\|u\|_\infty (2N+1)\left[\frac{1}{\pi}\ln\left(\frac{4N+2}{\pi}\right)+\frac{1}{2}\right]$. Also, if $N>3$, one obtains a simpler (but worse) upper bound $\left\|\widetilde{\partial}_j u\right\|_\infty \le \frac{48}{\ell}\|u\|_\infty N\ln N$.

(c) $\sum_{n\in[-N..N]^d}\left(\widetilde{\partial}_j u\right)_{[n]} = 0$.

(d) $\widetilde{\partial}_j^{\,2}$ is a symmetric operator.

Proof. (a) It suffices to show that the Fourier transforms of the two sides coincide.

$$\begin{aligned}(2N+1)^{d/2}\Big\{F_N\left(\widetilde{\partial}_j(u\cdot v)\right)\Big\}[k] &\overset{(1)}{=} \sum_{q\in[-N..N]^d}\frac{i2\pi k_j}{L}\,\widehat{u}[q]\,\widehat{v}[k-q]\\ &= \sum_{q\in[-N..N]^d}\frac{i2\pi(q_j+k_j-q_j)}{L}\,\widehat{u}[q]\,\widehat{v}[k-q]\\ &= \sum_{q\in[-N..N]^d}\left(\frac{i2\pi q_j}{L}\widehat{u}[q]\right)\widehat{v}[k-q]+\left(\frac{i2\pi(k_j-q_j)}{L}\widehat{v}[k-q]\right)\widehat{u}[q]\\ &\overset{(2)}{=} (2N+1)^{d/2}\left[F_N\left((\widetilde{\partial}_j u)\cdot v\right)+F_N\left((\widetilde{\partial}_j v)\cdot u\right)\right]\end{aligned} \tag{65}$$

Here (1) and (2) follow from the fact that the Fourier transform of the pointwise multiplication of two functions is the convolution of their Fourier transforms (up to the normalization factor $(2N+1)^{d/2}$).

(b) To show this, we make use of equation (62):

$$\begin{aligned}\left|(\widetilde{\partial}_j u)[n]\right| &= \left|\sum_{m\in[-N..N]} u[n_1,n_2,\cdots,m,\cdots,n_d]\,a[n_j-m]\right|\\ &\overset{(1)}{\leq} \frac{2\pi}{\ell}\,\|u\|_\infty\sum_{m=1}^{N}\frac{1}{\sin\left(\frac{\pi\,m}{2N+1}\right)}\\ &\leq \frac{2\pi}{\ell}\,\|u\|_\infty\left(\int_{x=1}^{N}\frac{dx}{\sin\left(\frac{\pi\,x}{2N+1}\right)}+\frac{1}{\sin\left(\pi/(2N+1)\right)}\right)\\ &\overset{(2)}{\leq} \frac{2\pi}{\ell}\,\|u\|_\infty\left(-\frac{(2N+1)}{\pi}\ln\left(\tan\left(\frac{\pi}{4N+2}\right)\right)+\frac{2N+1}{2}\right)\\ &\overset{(3)}{\leq} \frac{2\pi}{\ell}\,\|u\|_\infty(2N+1)\left[\frac{1}{\pi}\ln\left(\frac{4N+2}{\pi}\right)+\frac{1}{2}\right]\end{aligned} \tag{66}$$

where (1) follows from Hölder's inequality, and (2) follows from noting that $\sin(\pi/(2N+1)) \geq \frac{2}{2N+1}$. Finally, (3) follows from the fact that $\tan(x) \geq x$ for $0 \leq x < \pi/2$. The claim follows since these inequalities hold for any $n \in [-N..N]^d$. For $N > 3$, in order to simplify the right hand side of (3) we use the fact that $1 + x \leq 2x$ if $x \geq 1$. This implies that

$$\left|(\widetilde{\partial}_j u)[n]\right| \leq \frac{48}{\ell}\|u\|_\infty N\ln N. \tag{67}$$

(c) Note that for any vector v defined on the discrete lattice, one has

$$\sum_{n\in[-N..N]^d} v[n] = \left(\sqrt{2N+1}\right)^d\,\hat{v}[0] \tag{68}$$

where $\hat{v}$ represents the Fourier transform of v. Noting that $\left(F_N\widetilde{\partial}_j v\right)_{[0]} = \left(\frac{2\pi k_j}{\ell}\,\hat{u}[k]\right)_{[0]} = 0$ completes the proof.

(d) (62) and (63) show that $\widetilde{\partial}_j$ is anti-symmetric. And since composition of an anti-symmetric operator with itself is symmetric the result follows. ☐

So far we talked about the interpolation results for semi-analytic functions. We may now show that the Fourier differentiation technique is able to estimate the first and second order differentiation with high accuracy. Note that this is non-trivial as the Fourier differentiation operator is not bounded. The proof of this result borrows ideas from (Shen et al., 2011). Fig. 3

depicts an example of the Fourier interpolation and this derivative estimation method.

Proposition A.18. *Let u be (C,a)-semi-analytic and periodic, and let $N \geq 4ad$. It is the case that*

$$\sqrt{\sum_{j=1}^{d} \left\| \overrightarrow{\partial_j u_N} - \overrightarrow{\widetilde{\partial}_j u_N} \right\|^2} \leq \frac{40\sqrt{2}\pi e^3\, a}{\ell}\, C\, (2N+1)^{d/2}\, e^{-\frac{N}{2a}},\ \textit{and} \tag{69}$$

$$\left\| \overrightarrow{\nabla^2 u_N} - \overrightarrow{\widetilde{\nabla}^2 u_N} \right\| \leq \frac{200\sqrt{2}\pi^2 e^3\, a^2}{\ell}\, C^2\, (2N+1)^{d/2}\, e^{-0.4\frac{N}{a}}. \tag{70}$$

Proof. We first prove (69). We have

$$\partial_j u[n] = \sum_{k \in \mathbb{Z}^d} \frac{i2\pi k_j}{\ell} \widehat{u}[k]\, e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \tag{71}$$

$$= \sum_{k \in [-N..N]^d} e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \sum_{p \in \mathbb{Z}^d} \frac{i2\pi(k_j + p_j(2N+1))}{\ell} \widehat{u}[k+(2N+1)p]. \tag{72}$$

And similarly,

$$\widetilde{\partial}_j u[n] = \frac{1}{(2N+1)^{d/2}} \sum_{k \in [-N..N]^d} \frac{i2\pi k_j}{\ell} \widetilde{u}[k]\, e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \tag{73}$$

$$= \sum_{k \in [-N..N]^d} e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \sum_{p \in \mathbb{Z}^d} \frac{i2\pi k_j}{\ell} \widehat{u}[k+(2N+1)p] \tag{74}$$

where (74) follows from (29). Hence

$$\partial_j u[n] - \widetilde{\partial}_j u[n] = \sum_{k \in [-N..N]^d} e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \frac{i2\pi p_j(2N+1)}{\ell} \widehat{u}[k+(2N+1)p] \tag{75}$$

which using Parseval's theorem gives

$$\begin{aligned}\sum_{j=1}^{d} \left\| \overrightarrow{\partial_j u_N} - \overrightarrow{\widetilde{\partial}_j u_N} \right\|^2 &= (2N+1)^d \sum_{j=1}^{d} \sum_{k \in [-N..N]^d} \left| \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \frac{i2\pi(2N+1)p_j}{\ell} \widehat{u}[k+(2N+1)p] \right|^2 \\ &\leq \frac{4\pi^2}{\ell^2}(2N+1)^d \sum_{j=1}^{d} \sum_{k \in [-N..N]^d} \left\{ \left(\sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|k+(2N+1)p\|^{-2m} \right) \right. \\ &\quad \left. \times \left(\sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|k+(2N+1)p\|^{2m} |(2N+1)p_j|^2\, |\widehat{u}[k+(2N+1)p]|^2 \right) \right\}.\end{aligned} \tag{76}$$

Then, using inequality (32) together with Lemma E.5, along with the fact that for each $j \in [d]$ we have $|(2N+1)p_j| \leq 2|(2N+1)p_j + k_j|$, we get

$$\sum_{j=1}^{d} \left\| \overrightarrow{\partial_j u_N} - \overrightarrow{\widetilde{\partial}_j u_N} \right\|^2 \leq \frac{32\pi^2}{\ell^2} C^2\, (2N+1)^d\, 2^d\, N^{-2m}\, [(m+1)!]^2 a^{2m+2} \tag{77}$$

$$\leq \frac{32\pi^2}{\ell^2} C^2\, (2N+1)^d\, 2^{N/2a}\, N^{-2m}\, [(m+1)!]^2\, a^{2m+2} \tag{78}$$

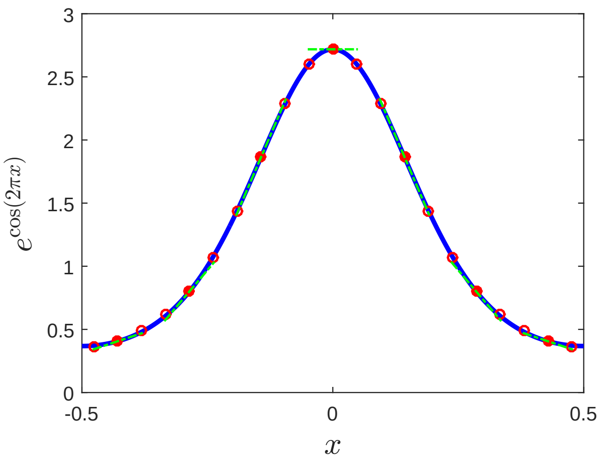

Figure 3. Applying the Fourier interpolation of Theorem A.8 on the input function $u(x) = e^{\cos 2\pi x}$ of Example A.2. The plot shows the interpolation results with $N = 3$ and $M = 10$. Filled circles correspond to the initial samples, and the hollow circles represent the interpolation output. The solid blue line represents the graph of the underlying function u. And the dashed green lines show the Fourier derivative estimations.

for $N \geq 2ad$. Hence, by choosing $m = \lfloor N/a \rfloor - 2$ one achieves the upper bound

$$\sqrt{\sum_{j=1}^{d} \left\| \overrightarrow{\partial_j u_N} - \overrightarrow{\widetilde{\partial}_j u_N} \right\|^2} \leq \frac{8\sqrt{2}\pi e^3}{\ell} C\,(2N+1)^{d/2}\, N\, e^{-0.6\frac{N}{a}}. \tag{79}$$

Again, using the inequality $x \leq \alpha\, e^{\frac{x}{e\alpha}}$ for all x and all positive α, and setting $\alpha = 5$ completes the proof. Now we prove (70). As in above, we start by writing the Fourier transform of the Laplacians:

$$\widetilde{\nabla^2} u_N[n] = \sum_{k \in [-N..N]^d} \frac{-4\pi^2}{\ell^2} \|k\|^2 \sum_{p \in \mathbb{Z}^d} \widehat{u}[k + (2N+1)p] \tag{80}$$

$$\nabla^2 u_N[n] = \sum_{k \in [-N..N]^d} \frac{-4\pi^2}{\ell^2} \sum_{p \in \mathbb{Z}^d} \|k + (2N+1)p\|^2\, \widehat{u}[k]. \tag{81}$$

Therefore,

$$\begin{aligned} \left\| \overrightarrow{\widetilde{\nabla^2} u_N} - \overrightarrow{\nabla^2 u_N} \right\|^2 &= \frac{16\pi^4}{\ell^4}(2N+1)^d \sum_{k \in [-N..N]^d} \left| \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \left(\|k + (2N+1)p\|^2 - \|k\|^2 \right) \widehat{u}_N[k] \right|^2 \\ &\overset{(a)}{\leq} \frac{16\pi^4}{\ell^4}(2N+1)^d \sum_{k \in [-N..N]^d} \left\{ \left(\sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|k + p(2N+1)\|^{-2m} \right) \right. \\ &\qquad \left. \times \left(\sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|k + (2N+1)p\|^4 \|k + (2N+1)p\|^{2m} \left|\widehat{u}_N[k]\right|^2 \right) \right\} \end{aligned} \tag{82}$$

where (a) uses the Cauchy-Schwartz inequality along with the fact that $\|k\| \leq \|k + p(2N+1)\|$ for all $k \in [-N..N]^d$. Again, we use Lemma E.5 and set $m = \lfloor N/a \rfloor - 3$ (which is guaranteed to be a natural number since $N \geq 4ad$) to conclude the proof. □

A.3. Construction of semi-analytic functions

One question that arises in our study of semi-analyticity is the behavior of the semi-analyticity parameters C and a under composition rules. For example, we may be interested in the semi-analyticity parameters of the function approximators represented by deep neural networks. Recall that the definition of semi-analyticity involves taking high order derivatives. Therefore, we make multiple uses of the Faà di Bruno formula (Roman, 1980; Krantz & Parks, 2002), according to which we have

$$\frac{d^m}{dx^m} f\left(g(x)\right) = \sum_{\substack{i_1, i_2, \cdots, i_m \in \{0, \cdots, m\} \\ i_1 + 2i_2 + \cdots + m i_m = m}} \frac{m!}{i_1! \cdots i_m!} f^{(i_1 + \cdots + i_m)}(g(x)) \left(\frac{dg/dx}{1!}\right)^{i_1} \cdots \left(\frac{d^m g/dx^m}{m!}\right)^{i_m} \tag{83}$$

for any pair of smooth functions $f, g : \mathbb{R} \to \mathbb{R}$. Working in d dimensions, we need to apply the multivariate Faà di Bruno formula, which is provided below.

Proposition A.19. *Let $g : \mathbb{R}^d \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be smooth functions and let $\alpha \in \mathbb{Z}^d_{\geq 0}$. We have*

$$D^\alpha (f \circ g)(x) = \alpha! \sum_{\lambda=1}^{|\alpha|} f^{(\lambda)}(g(x)) \sum_{s=1}^{|\alpha|} \sum_{p_s(\lambda,\alpha)} \prod_{j=1}^{s} \frac{1}{k_j!} \left(\frac{D^{\ell_j} g}{\ell_j!}\right)^{k_j}, \tag{84}$$

where

$$p_s(\lambda, \alpha) := \left\{ (k_1, \cdots, k_s, \ell_1, \cdots, \ell_s) : k_i > 0,\ 0 \prec \ell_1 \prec \cdots \prec \ell_s,\ \sum_{j=1}^{s} k_j \ell_j = \alpha,\ \sum_{j=1}^{s} k_j = \lambda \right\}.$$

Here, for $\mu, \nu \in \mathbb{Z}^d_{\geq 0}$, we say $\mu \prec \nu$ if the 1-norms compare as (i) $|\mu| < |\nu|$, or (ii) if $|\mu| = |\nu|$ then use lexicographic ordering.

Proof. This proposition is obtained by setting $m = 1$ in Theorem 2.1 of Constantine & Savits (1996). □

Reference (Krantz & Parks, 2002, Lemma 1.4.1) proves that the coefficients in (83) follow

$$\sum_{\substack{i_1, i_2, \cdots, i_m \in \{0, \cdots, m\} \\ i_1 + 2i_2 + \cdots + m i_m = m}} \frac{(i_1 + \cdots + i_m)!}{i_1! \cdots i_m!} R^{i_1 + \cdots + i_m} = \frac{R}{R+1} (R+1)^m, \tag{85}$$

for any $R > 0$. We use similar ideas to extend this result to the multivariate case.

Lemma A.20. *Let α be a d-dimensional vector of non-negative integers. We have*

$$\sum_{\lambda=1}^{|\alpha|} \lambda!\, R^\lambda \sum_{s=1}^{|\alpha|} \sum_{p_s(\lambda,\alpha)} \prod_{j=1}^{s} \frac{1}{k_j!} = \frac{R}{R+1} (R+1)^{|\alpha|}. \tag{86}$$

Proof. The proof is a generalization of (Krantz & Parks, 2002, Lemma 1.4.1). Let $g(x) = \frac{1}{1 - \sum_{i=1}^d x_i}$ and $f(x) = \frac{1}{1 - R(x-1)}$. In what follows, we consider $f(g(x))$, and its Taylor expansion, and subsequently, will apply (84) to get the desired relation.

To begin with, we observe the following. Let $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{Z}^d_{\geq 0}$. One can readily verify that

$$D^\alpha (x_1 + \cdots + x_d)^n \big|_{x=0} = \alpha!\, \mathbf{1}_{\{|\alpha| = n\}} \tag{87}$$

where $\mathbf{1}$ is the identifier function (i.e., it is 1 if the condition inside the brackets is satisfied, and 0 otherwise). Therefore, as

$g(x) = \sum_{n=0}^{\infty} \left(\sum_{i=1}^{d} x_i \right)^n$ in a neighbourhood of $x = 0$, we conclude

$$(D^\alpha g)(0) = \alpha!. \tag{88}$$

Additionally, it is transparent that $(f \circ g)(x) = \frac{1 - \sum_{i=1}^{d} x_i}{1 - (R+1) \sum_{i=1}^{d} x_i}$, which provides the following expansion on a neighbourhood of $x = 0$.

$$(f \circ g)(x) = 1 + \frac{R}{R+1} \sum_{n=1}^{\infty} (R+1)^n \left(\sum_{i=1}^{d} x_i \right)^n \tag{89}$$

Combining (87) and (89) provides $D^\alpha (f \circ g)(0) = \frac{R}{R+1} (R+1)^{|\alpha|} \alpha!$. Furthermore, it is straightforward to find $f^{(\lambda)}(g(0)) = \lambda!\, R^\lambda$, from which the lemma follows by substitutions into (84). ☐

We are now ready to study the composition of analytic functions of many variables. To make our claims easier to state and comprehend, let us first introduce some notation.

Definition A.21. *Let f be analytic on an open set U. We say $f \in A_U(C, a)$, if*

$$\sup_{x \in U} D^\alpha f(x) \leq C\, a^{|\alpha|}\, \alpha!, \tag{90}$$

and we say that $g \in S_U(C, a)$ if

$$\sqrt{\mathop{\mathbb{E}}_{X \sim \mathrm{Unif}(U)} \left(\sum_{\alpha : |\alpha| = m} |D^\alpha g(X)|^2 \right)} \leq C\, a^m\, m!. \tag{91}$$

Note that if f is periodic and analytic in its fundamental domain, then it is analytic on the entire domain $\mathbb{R}^d$. We use the notation $B_\infty(M)$ for the open ℓ_∞-ball $(-M, M)^d \subset \mathbb{R}^d$, and the notation A_M as a shorthand for $A_{B_\infty(M)}$.

Remark A.22. *Using the same argument as in Proposition A.2, for any open set U, we have $A_U(C, a) \subseteq S_U(3^{d-1} C, a)$.*

Proposition A.23. *Let $f_1 \in A_U(C_1, a_1)$ and $f_2 \in A_V(C_2, a_2)$ for two open domains U and V. The following statements hold.*

(a) $f_1 + f_2 \in A_{U \cap V}(C_1 + C_2, \max\{a_1, a_2\})$ if $U, V \subseteq \mathbb{R}^d$ are in the same Euclidean domains. Also, the same property holds for the semi-analytic families.

(b) $f_1 \cdot f_2 \in A_{U \cap V}(C_1 C_2, a_1 + a_2)$.

(c) Let V contain the image of f_1, that is $f_2 \in A_{f_1(U)}(C_2, a_2)$. Then $f_2 \circ f_1 \in A_U \left(\frac{C_1 a_2 C_2}{1 + C_1 a_2}, a_1 (1 + C_1 a_2) \right)$.

Proof. (a) Note that for any $\alpha \in \mathbb{Z}_+^d$ the quantity $\sup_{x \in U \cap V} |D^\alpha f(x)|$ defines a semi-norm. Using the triangle inequality of this semi-norm, we can obtain the result. For the semi-analyticity part, let $|u|_m := \sqrt{\mathbb{E} \left(\sum_{\alpha : |\alpha| = m} D^\alpha u(x) \right)^2}$ and note that $|\cdot|_m$ is also a semi-norm, and in particular, $|f_1 + f_2|_m \leq |f_1|_m + |f_2|_m$.

(b) We note that

$$D^\alpha (f_1 \cdot f_2)(x) = \sum_{\beta \in \prod_{i=1}^{d} \{0, \cdots, \alpha_i\}} \binom{\alpha}{\beta} D^{\alpha - \beta} f_1(x) \cdot D^\beta f_2(x), \tag{92}$$

where we have used the convention $\binom{\alpha}{\beta} := \prod_{i=1}^{d} \binom{\alpha_i}{\beta_i}$. Using the upper bounds on the derivatives of f_1 and f_2, we get

$$\begin{aligned}\sup_{x\in U\cap V} \left|D^\alpha(f_1\cdot f_2)(x)\right| &\le C_1C_2a_1\prod_{i=1}^{d}\sum_{\beta_i=0}^{\alpha_i} a_1^{\alpha_i-\beta_i}a_2^{\beta_i}\\ &\le C_1C_2\prod_{i=1}^{d}(a_1+a_2)^{\alpha_i} = C_1C_2(a_1+a_2)^{|\alpha|}.\end{aligned} \tag{93}$$

(c) Using Proposition A.19, we have

$$\sup_{x\in U} D^\alpha(f_2\circ f_1) = \sup_{x\in U} \alpha! \sum_{\lambda=1}^{|\alpha|} f_2^{(\lambda)} \sum_{s=1}^{\alpha}\sum_{p_s(\lambda,\alpha)}\prod_{j=1}^{s}\frac{1}{k_j!}\left(\frac{D^{\ell_j}f_1}{\ell j!}\right)^{k_j} \tag{94}$$

$$\le \alpha! C_2 a_1^{|\alpha|}\sum_{\lambda=1}^{\alpha}\lambda! a_2^\lambda C_1^\lambda \sum_{s=1}^{|\alpha|}\sum_{p_s(\lambda,\alpha)}\prod_{j=1}^{s}\frac{1}{k_j!} \tag{95}$$

$$\stackrel{(1)}{=} \alpha!\frac{C_2C_1a_2}{1+a_2C_1}\left[a_1(1+a_2C_1)\right]^{|\alpha|}, \tag{96}$$

where (1) follows from Lemma A.20. □

Corollary A.24. *Let $f\in A_U(C,a)$. Then, $e^f\in A_U(\frac{C}{1+C}e^\Delta,(1+C)a)$, where $\Delta=\sup_{x\in U} f(x)$ (compare this with Example A.2).*

Proof. This follows from Proposition A.23(c) and the fact that $g(x)=e^x$ is in $A_\Delta(e^\Delta,1)$. □

Corollary A.25. *Let $f \in A_M(C,a)$. Then, $\sigma(f(x)) := \frac{1}{1+e^{-f(x)}} \in A_M\left(1, a(1+C(1+e^\Delta))\right)$, where $\Delta = \sup_{x\in(-M,M)^d}|f(x)|$.*

Proof. This follows Proposition A.23(c) and noting that the function $g(x)=\frac{1}{1+x}$ is in $A_{\mathbb{R}_+}(1,1)$ because $g^{(m)}(x) = (-1)^m\frac{m!}{(1+x)^m}$. □

In the final corollary of this section we find the analyticity parameters of deep neural networks which are the de facto function approximators in deep learning.

Corollary A.26. *Let $f:\mathbb{R}^d\to\mathbb{R}$ be the function represented by a deep neural network consisting of D fully-connected layers with sigmoid activation functions. We denote the i-th layer weights matrix with $W^{(i)}$ and the bias vector with $b^{(i)}$. Then $f\in A_1(\widetilde{C},\widetilde{a})$ is with $\widetilde{C}=1$ and*

$$\widetilde{a}\le 2^D\exp\left(\sum_{k=1}^{D}2\left\|W^{(k)}\right\|_\infty+\left\|b^{(k)}\right\|_\infty\right) \tag{97}$$

Here the norm $\|\cdot\|_\infty$ is the maximum absolute row sum i.e., $\|X\|_\infty=\max_i\sum_j|X_{ij}|$ for any matrix X. Also, note that A_1 in the statement above could be generalized to A_M for arbitrary $M>0$, by rescaling the weights and biases of the first layer.

Proof. Let us denote the input and output of the i-th neuron of the k-th layer be denoted by $f_i^{(k)}$ and $g_i^{(k)}=\sigma(f^{(k)})_i$, respectively. We prove the result by induction on D. As for the base case, note that the input functions to the neurons of the first layer are $f_i^{(1)} := \langle w_i^{(1)},x\rangle+b_i^{(1)}$, where $w_i^{(1)} := (W_{ij}^{(1)})_j$ is the j-th row of the weight matrix of the first layer. Hence, by Proposition A.23(a) we get $f_i^{(1)}\in A(\left\|w_i^{(1)}\right\|_1,1)$. Also, note that $\sup_{x\in[-1,1]^d}\left|f_i^{(1)}(x)\right|\le\left\|w_i^{(1)}\right\|_1+\left|b_i^{(1)}\right|$. Applying Corollary A.25 yields $g_i^{(1)}\in A_1(1,a_i)$ with $a_i=2(1+\|w_i\|_1)e^{\left\|w_i^{(1)}\right\|_1+\left|b_i^{(1)}\right|}\le 2e^{2\left\|w_i^{(1)}\right\|_1+\left|b_i^{(1)}\right|}$. Taking a maximum over i proves the base case.

Now, assuming the bounds are valid for the neural network consisting of only the first k layers, we prove the bound for the first $k+1$ layers. By assumption $g_j^{(k)} \in A_1(1, \widetilde{a}_k)$, where

$$\widetilde{a}_k \leq 2^k \exp\left(\sum_{\ell=1}^{k} 2\left\|W^{(\ell)}\right\|_\infty + \left\|b^{(\ell)}\right\|_\infty\right),$$

and that $f_i^{(k+1)} = \langle w_i^{(k+1)}, g\rangle + b_i^{(k+1)}$. This, together with Proposition A.23(a) implies $f_i^{(k+1)} \in A_1\left(\left\|w_i^{(k+1)}\right\|_\infty, \widetilde{a}_k\right)$. As $g_i^{(k+1)} = \frac{1}{1+e^{-f_i^{(k+1)}}}$, we may use Corollary A.25 once more to complete our induction. □

A.4. Extension to infinite lattices

In this subsection, we extend the definition of semi-analyticity and the interpolation results to non-periodic functions. Consider a function $u : \mathbb{R}^d \to \mathbb{R}$, such that u and all its derivatives are in $L^2(\mathbb{R}^d)$. We are given the following quantum state, which encodes the values of u at certain points

$$|u_H\rangle \propto \sum_{j\in\mathbb{Z}^d} u(jH) |j_1\rangle \otimes |j_2\rangle \otimes \cdots \otimes |j_d\rangle,$$

where $H > 0$ is the discretization parameter, and $j = (j_1, j_2, \cdots, j_d)$ is a vector of integers. Note that the Fourier transform in this case is a function $\widehat{u} \in L^2(\mathbb{R}^d)$ that satisfies the following equations.

$$u(x) = \frac{1}{(2\pi)^{d/2}} \int_{\omega\in\mathbb{R}^d} e^{i\langle\omega,x\rangle} \widehat{u}(\omega)\, \mathrm{d}\omega \tag{98}$$

$$\widehat{u}(\omega) = \frac{1}{(2\pi)^{d/2}} \int_{x\in\mathbb{R}^d} e^{-i\langle\omega,x\rangle} u(x)\, \mathrm{d}x \tag{99}$$

Moreover, associated to $\overrightarrow{u} \in \ell^2(\mathbb{Z}^d)$ is a Fourier transform $\widetilde{u} : \mathbb{R}^d \to \mathbb{R}$ that satisfies

$$\widetilde{u}(\omega) := \left(\frac{H}{\sqrt{2\pi}}\right)^d \sum_{j\in\mathbb{Z}^d} e^{-i\langle\omega,jH\rangle} u(jH). \tag{100}$$

The coefficients of (98), (99), and (100) are chosen so that

$$\int_{x\in\mathbb{R}^d} |u(x)|^2\, \mathrm{d}x = \int_{\omega\in\mathbb{R}^d} |\widehat{u}(\omega)|^2\, \mathrm{d}\omega, \tag{101}$$

$$H^d \sum_{j\in\mathbb{Z}^d} |u(jH)|^2 = \int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widetilde{u}(\omega)|^2\, \mathrm{d}\omega. \tag{102}$$

Note the similarities between (99) and (100), and that $\widetilde{u} \to \widehat{u}$ as $H \to 0$ pointwise. Indeed, it is transparent that $\widetilde{u}(\omega)$ is periodic with period $\frac{2\pi}{H}$ along each axis, and that

$$\widetilde{u}(\omega) = \sum_{k\in\mathbb{Z}^d} \widehat{u}\left(\omega + \frac{2\pi k}{H}\right). \tag{103}$$

Note that $\widetilde{u}$ depends only on the values of u at the lattice points $H\mathbb{Z}^d$. Moreover, if $\hat{u}$ has a bounded support circumscribed within a fundamental domain of $\widetilde{u}$ then (103) implies that $\widetilde{u}(\omega) = \widehat{u}(\omega)$ for all $\omega \in [-\frac{\pi}{H}, \frac{\pi}{H}]^d$, and therefore one can exactly recover the function u (i.e., $u(x)$ can be found within arbitrarily small error at any $x \in \mathbb{R}^d$). This is indeed a restatement of the Nyquist theorem.

In what follows, we focus on the case where the support of $\widehat{u}$ is possibly the entire domain $\mathbb{R}^d$, but an interpolation with exponentially small error is still feasible. The arguments closely follow those of the previous subsections regarding periodic functions, and hence, we keep our proofs brief. From hereon, we use $\|\cdot\|$ to refer to the 2-norm for functions in $L^2(\mathbb{R}^d)$.

Definition A.27. *A function $u \in L^2(\mathbb{R}^d)$ is said to be semi-analytic if*

$$\left\| \sum_{\alpha:|\alpha|=m} D^\alpha u \right\| \le C\, a^m\, m! \tag{104}$$

for some $C, a \ge 0$. As before, we refer to C and a as the semi-analyticity parameters.

We note that Definition A.27 is equivalent to

$$\sqrt{\int_{\omega\in\mathbb{R}^d} \|\omega\|^{2m} |\widehat{u}(\omega)|^2 \,\mathrm{d}\omega} \le C\, a^m\, m!. \tag{105}$$

While establishing a connection between this notion of semi-analyticity and analyticity is challenging, we provide several examples of such functions.

Example A.4. *Any function which has a Fourier transform with bounded support is $(\|u\|, \omega_0)$-semi-analytic, where $\omega_0 = \sup\{\|\omega\|_2 : \widehat{u}(\omega) \neq 0\}$.*

Example A.5. *The Gaussian function $u(x) := \sqrt{\frac{2a}{\pi}} e^{-\frac{a}{2}x^2}$ with inverse variance a is $(1, a)$-semi-analytic. This is due to the fact that $|\widehat{u}(\omega)|^2$ corresponds to the probability distribution function of $\mathcal{N}(0, a)$, and hence*

$$\sqrt{\int_{\omega\in\mathbb{R}} |\omega|^{2m} |\widehat{u}(\omega)|^2 \,\mathrm{d}\omega} = \sqrt{\mathop{\mathbb{E}}_{X\sim\mathcal{N}(0,a)}[X^{2m}]} = \left(\frac{a}{2}\right)^m \sqrt{(2m-1)!!} \le a^m\, m!. \tag{106}$$

Example A.6. *The function $u(x) = \frac{2\lambda}{\lambda^2+x^2}$ is $(\sqrt{e/\lambda}, 2/\lambda)$-semi-analytic. That is due to the fact that $\widehat{u}(\omega) = e^{-\lambda|\omega|}$. Hence,*

$$\sqrt{\int_{\omega\in\mathbb{R}} |\omega|^{2m} |\widehat{u}(\omega)|^2 \,\mathrm{d}\omega} = \sqrt{\int_{\omega} \omega^{2m} e^{-2\lambda|\omega|} \,\mathrm{d}\omega} = \frac{1}{(2\lambda)^{m+\frac{1}{2}}} \sqrt{(2m)!} \le \sqrt{\frac{e}{\lambda}} \left(\frac{2}{\lambda}\right)^m m!.$$

We may now show the interpolation result of this section.

Theorem A.28. *Let $0 < h < H$. Further, let $V : \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ be an isometry defined by its action on the computational basis*

$$V : |j\rangle \mapsto \sum_{k\in\mathbb{Z}} V_{jk} |k\rangle \tag{107}$$

where $V_{jk} := \left(\frac{\sqrt{Hh}}{\pi}\right) \frac{(-1)^k \sin\left(\frac{\pi h k}{H}\right)}{hk - Hj}$. Then, for a (C, a)-semi-analytic function $u \in L^2(\mathbb{R}^d)$, it is the case that

$$\left\| V^{\otimes d} |u_H\rangle - |u_h\rangle \right\| \le \frac{8\sqrt{2}e^3\, C}{\|u\|} e^{-0.6\,\frac{\pi}{aH}} \tag{108}$$

if $\frac{\pi}{H} \ge 2ad$.

Proof. To better understand the proof, it is helpful to provide some intuition beforehand. Although the technical details closely resemble the previous results, we discuss the reasoning behind it. Note that our aim is to approximate the function values at the lattice points $h\mathbb{Z}^d$, given the values on $H\mathbb{Z}^d$, for $h < H$. To do so, we note that

$$\begin{aligned} u(hj) &= \frac{1}{(2\pi)^{d/2}} \int_{\omega\in\mathbb{R}^d} e^{i\langle jh,\omega\rangle}\widehat{u}(\omega)\,\mathrm{d}\omega \\ &\approx \frac{1}{(2\pi)^{d/2}} \int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} e^{i\langle jh,\omega\rangle}\widetilde{u}(\omega)\,\mathrm{d}\omega \\ &= \left(\frac{H}{2\pi}\right)^d \int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} e^{i\langle jh,\omega\rangle} \sum_{k\in\mathbb{Z}^d} e^{-i\langle Hk,\omega\rangle} u(kH)\,\mathrm{d}\omega \\ &= \sum_{k\in\mathbb{Z}^d} u(kH) \prod_{a=1}^{d} W_{j_a k_a}, \end{aligned} \tag{109}$$

where $W_{jk} := \left(\frac{H}{2\pi}\right)\int e^{i(jh-Hk)\omega} = \left(\frac{H}{\pi}\right)\frac{(-1)^k \sin\left(\frac{jh\pi}{H}\right)}{jh-kH}$. A straightforward computation reveals that $\sum_{j\in\mathbb{Z}} W_{jk}\overline{W}_{j\ell} = \sqrt{\frac{H}{h}}\delta_{k\ell}$, and hence, we conclude that V is an isometry.

It remains to prove that the approximation error in (109) is small. With a similar argument as the one used in Lemma A.3, we can show that

$$\sqrt{\int_{\omega\notin[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widehat{u}(\omega)|\,\mathrm{d}\omega} \leq 2e^3 C\, e^{-\frac{0.6\pi}{Ha}}, \tag{110}$$

and also, following the proof of Lemma A.4, we obtain

$$\sqrt{\int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widehat{u}(\omega)-\widetilde{u}(\omega)|^2\,\mathrm{d}\omega + \int_{\omega\notin[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widehat{u}(\omega)|^2\,\mathrm{d}\omega} \leq 2\sqrt{2}e^3\, Ce^{-\frac{3\pi}{5aH}}. \tag{111}$$

The rest of the proof follows directly from the arguments made in Corollary A.14, and Proposition A.7. □

We also note that the concentration results of Appendix A.1 are readily extendable to the non-periodic cases studied here. In particular, one can show that a function u is semi-analytic if and only if $|\widehat{u}|^2/\|u\|$ defines a sub-exponential distribution. We leave it open to investigate whether the foundations provided in the subsection can be used as the building blocks of quantum algorithms using registers of quantum modes with infinitely many levels (such as bosonic quantum computers made from quantum harmonic oscillators).

A.5. Chebyshev interpolation

In polynomial interpolation, the question of interest is the following. Given values of a function f at $x_1 < x_2 < \cdots < x_N$, find a polynomial that passes through all point $(x_i, f(x_i))$ and approximates f everywhere else. It is by Weierstrass approximation theorem that we know such an interpolation is feasible with arbitrary precision. Moreover, Legendre's interpolating approach provides the generic solution

$$p(x) = \sum_i f(x_i) \prod_{j\neq i} \frac{x-x_j}{x_i-x_j}. \tag{112}$$

The error of the above interpolation at a point x, for an N times differentiable function f is given by

$$e(x) = \frac{f^{(N)}(x)}{N!} \prod_i (x-x_i). \tag{113}$$

It is easy to observe that $\max_{x\in[-1,1]} |e(x)| \sim f^{(N)}\frac{\ell^N}{N}$. However, since the N-th derivative of f can grow rapidly (e.g., as fast as $N!$), this interpolation scheme may fail to work on analytic functions. A well-known example is the Runge function. However, one can choose the points x_i more wisely to avoid such divergence in approximation, which is the topic of the rest of this section.

Chebyshev polynomials. Chebyshev polynomials are the functions

$$T_k(x) = \cos(k \arccos(x)), \tag{114}$$

defined on the domain $[-1, 1]$ and have the following recursive property

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x), \tag{115}$$

which readily shows that T_k is of degree k. Moreover, they satisfy the following orthogonality condition

$$\int_{-1}^{1} T_n(x)T_m(x)\frac{\mathrm{d}x}{\sqrt{1-x^2}} = \begin{cases} 0 & \text{if } n \neq m, \\ \pi & \text{if } n = m = 0, \\ \frac{\pi}{2} & \text{if } m = n \neq 0, \end{cases} \tag{116}$$

and form a complete basis for the function space $L^2(\mu)$, where $\mathrm{d}\mu = \frac{\mathrm{d}x}{\sqrt{1-x^2}}$. In the multi-variate case, this measure is of the form $\frac{\mathrm{d}x}{\prod_i \sqrt{1-x_i^2}}$. As a result, for any $f \in L^2(\mu)$, we have a decomposition

$$f(x) = \sum_{n \geq 0} c_n T_n(x), \tag{117}$$

where

$$c_n = \kappa_n \int_{-1}^{1} f(x)T_n(x)\frac{\mathrm{d}x}{\sqrt{1-x^2}}, \tag{118}$$

with $\kappa_0 = \pi$ and $\kappa_n = \pi/2$ for all $n > 0$. Note that with the change of variable $\cos\theta = x$, one may write the Chebyshev coefficients as

$$c_n = \frac{\kappa_n}{2}\int_{-\pi}^{\pi} f(\cos\theta)T_n(\cos\theta)\mathrm{d}\theta = \frac{\kappa_n}{2}\int_{-\pi}^{\pi} f(\cos\theta)\cos(n\theta)\mathrm{d}\theta \tag{119}$$

$$= \pi\int_{-\pi}^{\pi} f(\cos\theta)e^{in\theta}\mathrm{d}\theta. \tag{120}$$

Here the last equality is due to the fact that $g(\cdot) = f(\cos(\cdot))$ is an even function. Hence, with the notation of this appendix, we have $c_n = \pi\mathcal{F}\{g\}_{[n]}$.

Chebyshev grid. We consider the lattice defined by the N points $x_k = \cos\left(\frac{\pi}{2} + \frac{k\pi}{2n+1}\right)$ where $k \in \{1, \cdots, n\}$. Note that these points correspond to the roots of $T_{N+1}(x)$ and satisfy

$$\prod_{i=1}^{N}(x - x_i) = 2^{-N}T_{N+1}(x). \tag{121}$$

Hence, using this grid together with Lagrange interpolation, we obtain that the remainder formula reads as $e(x) \sim \frac{f^{(N)}}{2^N N!}T_{N+1}(x)$. Note that $|T_{N+1}(x)| \leq 1$ by definition from (114).

A similar interpolation to that of Theorem 3.2 can be done by means of QFT on the Chebyshev grid. We let $|v_N\rangle$ to be the N-point discretization of f on this grid. In other words, we have $|f_N\rangle := \sum_{k=1}^{N} f(x_k)\,|k\rangle$. The proposition below immediately follows from Theorem 3.2 and Proposition A.23.

Proposition A.29 (Chebyshev interpolation). *Let $f : [-1, 1]^d \to \mathbb{R}$ be a (C, a)-analytic function and $|\psi\rangle$ be ε-close to $|f_N\rangle$. Then, for any $M > N$ there is a quantum curcuit that consists of $\widetilde{\mathcal{O}}(d)$ elementary gates and prepares $|f_M\rangle$, with error at most $\varepsilon + \frac{16e^3}{\mathcal{U}}e^{-0.6N/(1+a)}$.*

Applying the Chebyshev interpolation to the Gibbs state e^{-f}/Z_f results in approximate samples from $\frac{e^{-f}}{\prod_{i=1}^d \sqrt{1-x_i^2}}$ on the hypercube $[-1,1]^d$. We now show that most of this mass is concentrated in the neighbourhood of the boundary of co-dimension $k = O(1)$ if $\Delta = o(d)$. Note that we cannot have a concentration with $\Delta = \Omega(d)$, as the measure defined by $f(x) = \beta\|x\|^2$ is concentrated at the origin if $\beta > 1$.

Proposition A.30. *The measure defined by*

$$\mathrm{d}\mu \propto \frac{e^{-f(x)}}{\prod_{i=1}^d \sqrt{1-x_i^2}}\,\mathrm{d}x \tag{122}$$

in high dimension is concentrated at around any boundary of co-dimension $k = O(1)$, if $\Delta = o(d)$. More precisely, let $\mathcal{C}_k^{(\varepsilon)} := \{x \in [-1,1]^d : |\{i : |1 - x_i| \le \varepsilon\}| \ge k\}$. We have

$$\int_{x\in\mathcal{C}_k^{(\varepsilon)}} \mathrm{d}\mu = \Omega(1), \tag{123}$$

for any $k = O(1)$, $\varepsilon = \omega(1/d^2)$, and $\Delta = o(d)$.

Proof. We observe that

$$\frac{\int_{x\notin\mathcal{C}_{k+1}^{(\varepsilon)}} \mathrm{d}\mu}{\int_{x\in\mathcal{C}_k^{(\varepsilon)}} \mathrm{d}\mu} \le e^{\Delta}\cdot\left(\frac{\pi - \sqrt{2\varepsilon}}{\pi}\right)^{d-k-1} = \exp\left(\Delta - (d-k-1)\log\left(1 - \frac{\sqrt{2\varepsilon}}{\pi}\right)\right). \tag{124}$$

It is also straightforward to show that

$$\int_{x\notin\mathcal{C}_1^{(\varepsilon)}} \mathrm{d}\mu \le \exp\left(\Delta - d\log\left(1 - \frac{\sqrt{2\varepsilon}}{\pi}\right)\right). \tag{125}$$

The result now follows from (125) and (124). □

Corollary A.31. *Let $E : [-1,1]^d \to \mathbb{R}$ be an L-Lipschitz potential function, and suppose that the one-parameter family of all measures $\{e^{\mathcal{L}t}\rho_0 : t \ge 0\}$ consists of analytic functions with parameters C and a. There is a quantum algorithm that samples from a distribution ε-close to the Gibbs measure (in total variation distance), by making*

$$\mathcal{O}\left(d^3\kappa'_{E/2}e^{\Delta/2}\,\mathbb{E}_{X\sim e^{-E}/Z}[e^{\|x\|^2}]\max\left\{(1+a)^4d^4, \log^4\left(\frac{\sqrt{d}e^{\frac{5\Delta}{4}}C(1+a)^3(1+\ell L)}{\varepsilon}\right)\right\}\right.$$
$$\left.\mathrm{polylog}\left(\frac{(1+a)de^{\Delta}\log(C(1+\ell L))}{\varepsilon}\right)\right). \tag{126}$$

queries to the energy oracle. Note that $\kappa'_{E/2}$ here is the Poincaré constant of the toric diffusion obtained by the drift $G(\theta) := E(\cos\theta)$. Moreover, one can estimate the mean of a random variable, say $f(X)$, where $X \sim e^{-E}/Z$, by making

$$\widetilde{\mathcal{O}}\left(d^7(1+a)^4e^{\Delta/2}\kappa'_{E/2}\,\mathbb{E}_{X\sim e^{-E}/Z}[e^{\|x\|^2}]\frac{\Delta_f}{\varepsilon}\right) \tag{127}$$

queries to the controlled and standalone oracles of the energy function E and the function f. The algorithm succeeds with high probability with a flag indicating its success.

B. Langevin diffusion on a torus

Let $\{Y_t\}_{t\ge0}$ be a continuous time stochastic process in $\mathbb{R}^d$ satisfying the overdamped Langevin dynamics at thermodynamic $\beta = 1$,

$$dY_t = -\nabla E(Y_t)\,dt + \sqrt{2}\,dW_t \tag{128}$$

where $(W_t)_{t\geq 0}$ is a Wiener process. This equation is well-studied in the literature. In particular, it is known that for confining energy potentials the process is time reversible, ergodic, with a unique stationary distribution proportional to e^{-E} (Pavliotis, 2014, Proposition 4.2). Note that the condition for being confining imposes the function to be non-periodic. However, we are interested in a counterpart to the same results on a torus. We refer the reader to García-Portugués et al. (2019) for notions of toroidal diffusions and wrapping of a diffusion process in the Euclidean domain on a torus (that is, the pushforward of the original process under the quotient map of the torus). More generally, Hsu (2002) discusses stochastic calculus on manifolds.

We start with a periodic energy function with period ℓ in all dimensions. It is shown (García-Portugués et al., 2019, Proposition 2) that the corresponding wrapped Langevin dynamics is Markovian, ergodic, time-reversible, and admitting a unique stationary distribution, if the second derivatives of e^{-E} are Hölder continuous. This condition is satisfied in our case since compactness of the torus implies that the third partial derivatives of e^{-E} attain their maxima. Therefore we can derive a Lipschitz property for the second derivatives, resulting in their Hölder continuity. For further clarity, we will denote the wrapped process by $\{X_t\}_{t\geq 0}$ and write the overdamped toroidal Langevin diffusion in the same form as (128) given by

$$dX_t = -\nabla E(X_t)\,dt + \sqrt{2}\,dW_t. \tag{129}$$

The Fokker–Planck equation associated to (128) viewed as an Itô stochastic differential equation

$$\partial_t \sigma(y,t) = \nabla\cdot\left(e^{-E}\,\nabla\left(e^{E}\,\sigma(y,t)\right)\right) \tag{130}$$

describes the evolution of the probability density function $\sigma(-,t)$ of Y_t (see (Pavliotis, 2014) for more details). The probability density $\rho_t = \rho(-,t)$ of the wrapped process X_t satisfies

$$\rho(x,t) = \sum_{k\in\mathbb{Z}^d} \sigma(y + k\,\ell) \tag{131}$$

and hence, one gets the following parabolic differential equation with periodic boundary conditions as the Fokker–Planck equation corresponding to the toroidal diffusion process X_t. We will call this the toroidal Fokker–Planck equation.

$$\partial_t \rho(x,t) = \nabla\cdot\left(e^{-E}\,\nabla\left(e^{E}\,\rho(x,t)\right)\right) \tag{132}$$

The corresponding generator $\mathcal{L}$ is also a well-defined operator

$$\mathcal{L}(-) = \nabla\cdot\left(e^{-E}\,\nabla\left(e^{E}\,-\right)\right). \tag{133}$$

Remark B.1. *It is worth mentioning that the Fokker–Planck generator is usually written in the form*

$$\mathcal{L}(-) = \nabla^2 E(-) + \nabla E\cdot\nabla(-) + \nabla^2(-) \tag{134}$$

however as we will see in Appendix B.1, the generator is better behaved under discretization when it is written in the form (133)*. Nevertheless, discretizing all the derivatives in the usual Fokker–Planck equation results in an operator of the same form*

$$\widetilde{\nabla^2} E(-) + \widetilde{\nabla} E\cdot\widetilde{\nabla}(-) + \widetilde{\nabla^2}(-) = \widetilde{\nabla}\cdot\left(e^{-E}\,\widetilde{\nabla}\left(e^{E}\,-\right)\right) \tag{135}$$

where the tilde on top represents Fourier differentiation operators. This is due to the fact that discrete Fourier differentiation obeys Leibniz's rules (see Appendix A.2 and Appendix B.1).

Remark B.2. *We can derive the uniqueness of the stationary state of the toroidal Fokker–Planck equation by considering the operator*

$$\mathcal{L}' := e^{E/2}\circ\mathcal{L}\circ e^{-E/2}. \tag{136}$$

Let π be a density function satisfying $\mathcal{L}\pi = 0$. It is straightforward to see that for any periodic density function ρ

$$\langle \rho, \mathcal{L}' \rho \rangle = - \int_{x \in [-L/2, L/2]^d} e^{-E} \left\| \nabla \left(e^{E/2} \rho \right) \right\|^2 \leq 0 \tag{137}$$

with equality happening if and only if $\rho(x) \propto e^{-E(x)/2}$. We apply this inequality to $e^{E/2}\pi$ and conclude that $\pi = \rho_s$.

The trend to equilibrium for this stochastic process is studied in the literature (Markowich & Villani, 2000; Berglund, 2011; Bakry et al., 2014; van Handel, 2016). A functional inequality known as the *Poincarè inequality* is equivalent to exponentially fast convergence of Langevin diffusion (Bakry et al., 2014, Theorem 4.2.5) with a rate known as the *Poincarè constant*. Here we show that an analogous inequality holds for potentials on a torus and we find a corresponding Poincarè constant. But first, we will argue that the aforementioned exponential decay property for diffusions in the Euclidean space translates to a counterpart on the torus for toroidal diffusions.

Proposition B.3. *Given the toroidal Fokker–Planck equation (132), let ρ_s be the corresponding steady distribution. Further, assume that for a constant $\lambda > 0$ any differentiable function $f \in L^2(\rho_s)$ satisfies*

$$\mathrm{Var}_{\rho_s}[f] \leq \lambda\, \mathbb{E}_{\rho_s}\left[\|\nabla f\|^2 \right]. \tag{138}$$

Then the following decay in the distance of ρ_t and ρ_s is satisfied:

$$\|\rho_t/\rho_s - 1\|_{L^2(\rho_s)} \leq e^{-t/\lambda} \|\rho_0/\rho_s - 1\|_{L^2(\rho_s)}. \tag{139}$$

Proof. Let us denote the ratio of the distributions by $h_t = h(-,t) = \rho_t/\rho_s$. Using the Fokker–Planck equation (130) one has $\partial_t h = \rho_s^{-1} \nabla \cdot (\rho_s \nabla h)$, which implies equality (a)

$$\frac{d}{dt} \int_{x \in [-\ell/2, \ell/2]^d} \rho_s (h-1)^2 = 2 \int_{x \in [-\ell/2, \ell/2]^d} \rho_s (h-1) \partial_t h \tag{140}$$

$$\stackrel{(a)}{=} -2 \int_{x \in [-\ell/2, \ell/2]^d} \rho_s \|\nabla h\|^2. \tag{141}$$

Note that $\mathbb{E}_{\rho_s}[h] = 1$ and hence, the left hand side of (140) is the time derivative of $\mathrm{Var}_{\rho_s}[h_t]$, while the right hand side of (141) is $-2\, \mathbb{E}_{\rho_s}\left[\|\nabla h\|^2 \right]$. We conclude that $\mathrm{Var}_{\rho_s}[h_t] \leq e^{-2\lambda t}\, \mathrm{Var}_{\rho_s}[h_0]$ and therefore the result follows. □

Remark B.4. *We note that by Jensen's inequality* [6]

$$\mathbb{E}_{\rho_s}[|\rho_t/\rho_s - 1|] \leq \sqrt{\mathbb{E}_{\rho_s}(\rho_t/\rho_s - 1)^2} \tag{142}$$

which yields

$$\int_{x \in [-\ell/2, \ell/2]^d} |\rho_t(s) - \rho_s(x)| \leq \sqrt{\int_{x \in [-\ell/2, \ell/2]} \rho_s(x) \left(\frac{\rho_t}{\rho_s} - 1 \right)^2}. \tag{143}$$

Note that the left hand side is twice the total-variation distance between the two distributions, hence

$$\mathrm{TV}(\rho_t, \rho_s) \leq \frac{1}{2} \sqrt{\mathrm{Var}_{\rho_s}[\rho_t/\rho_s]}. \tag{144}$$

We can now show that for all bounded energy functions on tori there exists a universal Poincarè constant exists.

Proposition B.5. *Let E be an ℓ-periodic energy potential with a bounded range Δ. Then for all ℓ-periodic $f \in L^2(\rho_s)$*

$$\mathrm{Var}_{\rho_s}[f(X)] \leq \frac{\ell^2 e^{\Delta}}{4\pi^2}\, \mathbb{E}_{\rho_s}\left[\|\nabla f\|_2^2 \right]. \tag{145}$$

[6] The quantity $\|p/q - 1\|_{L^2(p)}^2$ is referred to as the χ^2 divergence of distributions p and q.

Proof. We have

$$\begin{aligned}\mathrm{Var}_{\rho_s}\left[f(X)\right] &= \mathrm{Var}_{\rho_s}\left[f(X) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right] \\ &\leq \mathbb{E}_{\rho_s}\left[\left(f(X) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right)^2\right] \\ &\leq \frac{e^{-\min_x E(x)}}{Z} \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \left(f(x) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right)^2\end{aligned} \tag{146}$$

and also, due to Parseval's theorem

$$\int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \left(f(x) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right)^2 = \ell^d \sum_{k\in\mathbb{Z}^d\setminus\{0\}} \left|\widehat{f}[k]\right|^2, \tag{147}$$

where $\widehat{f}$ is the Fourier transform of f. Note that the $k=0$ term is excluded in (147) since we have subtracted the average of f on the left hand side. On the other hand, since the Fourier transform of ∇f is $\frac{2i\pi}{\ell^2} k\, \widehat{f}[k]$, one could again use Parseval's theorem to write

$$\int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \|\nabla f\|^2 = \ell^d \sum_{k\in\mathbb{Z}^d} \frac{4\pi^2}{\ell^2}\|k\|^2 \left|\widehat{f}[k]\right|^2 \geq \ell^d \sum_{k\in\mathbb{Z}^d\setminus\{0\}} \frac{4\pi^2}{\ell^2} \left|\widehat{f}[k]\right|^2. \tag{148}$$

Now, the inequality $\frac{e^{-\max_x E(x)}}{Z} \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \|\nabla f\|^2 \leq \mathbb{E}_{\rho_s}\left[\|\nabla f\|^2\right]$ together with (146), (147), and (148) prove the claim. □

Corollary B.6. *In Proposition B.3 we have* $\lambda = \frac{\ell^2 e^{\Delta}}{4\pi^2}$ *as the universal Poincaré constant.*

B.1. Discretization of the Fokker–Planck equation

We now introduce the discrete operator $\mathbb{L}$ obtained from the generator $\mathcal{L}$ of the diffusion process:

$$\begin{aligned}\mathbb{L} : \mathcal{V}_N &\to \mathcal{V}_N \\ \overrightarrow{f} &\mapsto \widetilde{\nabla}\cdot\left(e^{-E}\widetilde{\nabla}(e^{E}\overrightarrow{f})\right)\end{aligned} \tag{149}$$

where tilde on the top of derivative is used to represent Fourier derivative operators (see Appendix A.2). Note that by the product rule of the Fourier derivative operator (Proposition A.17a) we can rewrite $\mathbb{L}$ in terms of derivatives of e^{-E} and the function that $\mathbb{L}$ acts on as follows:

$$\mathbb{L}(-) = e^{E}\left(\left\|\widetilde{\nabla}e^{-E}\right\|^2 + \widetilde{\nabla^2}e^{-E}\right)(-) - e^{E}\,\widetilde{\nabla}e^{-E}\cdot\widetilde{\nabla}(-) + \widetilde{\nabla^2}(-). \tag{150}$$

In what follows, we denote the condition number of a matrix A, by κ_A. We also use the shorthand notation $[-N..N] := \{-N, -N+1, \cdots, N\}$.

Lemma B.7. *The discrete operator* (149) *has the following properties.*

(a) It is diagonalizable as $\mathbb{L} = V^{-1}\, D\, V$ *with* $\kappa_V \leq e^{\Delta/2}$, *where* D *is a negative semi-definite diagonal matrix (i.e.,* $D \leq 0$*).*

(b) The kernel of $\mathbb{L}$ *is one dimensional and is spanned by the discretized Gibbs distribution* $\overrightarrow{\rho_s}$.

(c) The operator norm of $\mathbb{L}$ *is bounded above via* $\|\mathbb{L}\| \leq \frac{dN^2}{\ell^2}\min\left\{4\pi^2 + 2606\Delta(\ln N)^2, 4\pi^2\, e^{\Delta}\right\}$ *for* $N > 3$.

Proof. Claims (a) and (b): Let $U = e^{-E/2}$ (i.e., U is a diagonal matrix with diagonal entries all equal to $e^{-E/2}$).

Considering the action of the operator $\mathbb{L}' = U^{-1}\,\mathbb{L}\,U$ on the vector $\overrightarrow{f}$ and by consecutive applications of Proposition A.17a we have

$$\begin{aligned}\mathbb{L}'\overrightarrow{f} &= e^{E/2}\widetilde{\nabla}\cdot\left(e^{-E}\widetilde{\nabla}(e^{E/2}\overrightarrow{f})\right)\\ &= e^{E/2}\widetilde{\nabla}\cdot\left(e^{-E}\widetilde{\nabla}e^{E/2}\right)\overrightarrow{f} + \left(e^{-E/2}\widetilde{\nabla}e^{E/2} + e^{E/2}\widetilde{\nabla}e^{-E/2}\right)\cdot\left(\widetilde{\nabla}\overrightarrow{f}\right) + \widetilde{\nabla}^2\overrightarrow{f}\\ &= e^{E/2}\widetilde{\nabla}\cdot\left(e^{-E}\widetilde{\nabla}e^{E/2}\right)\overrightarrow{f} + \widetilde{\nabla}^2\overrightarrow{f}.\end{aligned} \tag{151}$$

We now note that the first term above is symmetric since it is a diagonal operator, and so is the operator in the second term, i.e. $\widetilde{\nabla}^2$ (due to Proposition A.17d). Hence, $\mathbb{L}'$ is symmetric. Note that this concludes $\mathbb{L}$ being diagonalizable, and moreover, since $\kappa_U \le e^{\Delta/2}$, we also have $\kappa_V \le e^{\Delta/2}$. Next, we show that $\mathbb{L}' \preccurlyeq 0$.

$$\begin{aligned}\langle f, \mathbb{L}'\, f\rangle &= \sum_{n\in[-N..N]^d} e^{E[n]/2} f[n]\,\widetilde{\nabla}\cdot\left(e^{-E}\,\widetilde{\nabla}(e^{E/2} f)\right)_{[n]}\\ &\overset{(a)}{=} \sum_{n\in[-N..N]^d} \widetilde{\nabla}\cdot\left(e^{E/2}\, f\,\widetilde{\nabla}(e^{E/2} f)\right)_{[n]} - \sum_{n\in[-N..N]^d} e^{-E[n]}\left\|\widetilde{\nabla}e^{E/2} f\right\|^2_{[n]}\\ &\overset{(b)}{=} -\sum_{n\in[-N..N]^d} e^{-E[n]}\left(\left\|\widetilde{\nabla}e^{E/2} f\right\|^2\right)_{[n]} \le 0\end{aligned} \tag{152}$$

where (a) and (b) follow from Proposition A.17a and Proposition A.17c, respectively. Also, note that the expression is zero if and only if $f[n] \propto e^{-E[n]/2}$. Therefore, the only eigen-direction of $\mathbb{L}'$ corresponding to the eigenvalue 0 is that of $e^{-E/2}$. Using the similarity transform between $\mathbb{L}$ and $\mathbb{L}'$, this consequently implies that the kernel of $\mathbb{L}$ is the subspace spanned by the discretized Gibbs distribution.

Claim (c): We note that $\mathbb{L} = \left(\widetilde{\nabla}\cdot\right)\circ e^{-E}\circ\widetilde{\nabla}\circ e^{E}$, where $\left(\widetilde{\nabla}\cdot\right)$ is the discrete divergence operator. We can now upper bound the spectral norm of L by noting that $\left\|e^{-E}\right\|\left\|e^{E}\right\| \le e^{\Delta}$, and $\left\|\widetilde{\nabla}\right\| \le \sqrt{d}\,\frac{2\pi N}{\ell}$, and also $\left\|\left(\widetilde{\nabla}\cdot\right)\right\| \le \sqrt{d}\frac{2\pi N}{\ell}$. Furthermore, using (135), the triangle inequality, and Proposition A.17b, we conclude that $\|\mathbb{L}\| \le \frac{dN^2}{\ell^2}\left(4\pi^2 + 48^2(\ln N)^2\Delta + 96\pi\Delta\ln N\right)$, for $N > 3$. □

In the following lemma, we prove that evolution under $\mathbb{L}$, does not dramatically change the ℓ_2-norm of the state under evolution. We denote the vector of all ones by $\mathbf{1} \in \mathcal{V}_N$.

Lemma B.8. *Consider the differential equation $\frac{d}{dt}\overrightarrow{u} = \mathbb{L}\,\overrightarrow{u}$, with initial condition $\overrightarrow{u}(0) = \mathbf{1}$. We have*

$$\sup_{t\ge 0}\|\overrightarrow{u}(t)\| \le e^{\Delta/2}\,\|\mathbf{1}\|, \tag{153}$$

$$\inf_{t\ge 0}\|\overrightarrow{u}(t)\| = \|\mathbf{1}\| \tag{154}$$

Proof. We write the solution as $\overrightarrow{u}(t) = e^{\mathbb{L}t}\overrightarrow{u}(0)$. From Lemma B.7a we have $\left\|e^{\mathbb{L}t}\right\| = \left\|V^{-1}\,e^{Dt}\,V\right\| \le \|V\|\left\|V^{-1}\right\|\left\|e^{Dt}\right\|$. Therefore $\kappa_V = \|V\|\left\|V^{-1}\right\|$ and $\left\|e^{Dt}\right\| \le 1$ imply that $\left\|e^{\mathbb{L}t}\right\| \le \kappa_V \le e^{\Delta/2}$. This proves (153).

For (154) we use the fact that $\langle\mathbf{1}|\,\mathbb{L} = 0$ (which follows from Proposition A.17c), to conclude that $\langle\mathbf{1}, u(t)\rangle = \langle\mathbf{1}, u(0)\rangle$. Using the Cauchy-Schwartz inequality one has

$$\|u(t)\|\,\|\mathbf{1}\| \ge \langle\mathbf{1}, u(0)\rangle \tag{155}$$

and given $\langle\mathbf{1}, u(0)\rangle = \|\mathbf{1}\|^2$, we have $\|u(t)\| \ge \|\mathbf{1}\|$. The result follows by noting that $t = 0$ this inequality is an equality. □

B.2. Auxiliary lemmas

In this section we upper bound the error in solving the discretization of the Fokker–Planck equation. Here $\overrightarrow{u(t)}$ denotes the discretization of the actual solution to the Fokker–Planck equation(i.e., the differential equation in the continuous domain). We shorten our notation and denote this solution by $\overrightarrow{u}$. In contrast, we denote the solution to the discretized Fokker–Planck equation by $\overrightarrow{v}$, that is $\overrightarrow{v}$ satisfies the linear system $\frac{d\overrightarrow{v}}{dt} = \mathbb{L}\overrightarrow{v}$.

Lemma B.9. *Let $u(x,t)$ ($\forall t \geq 0$) be a solution to the Fokker–Planck equation (130). Then $\max_x e^E\, u(x,t)$ is a non-increasing function of time.*

Proof. Let $v(x,t) = e^E u(x,t)$. Using (130) v satisfies

$$\partial_t v = -\nabla E \cdot \nabla v + \nabla^2 v = \mathcal{L}^* v \tag{156}$$

which is the backward Kolmogorov equation with $\mathcal{L}^*$ the adjoint of the operator $\mathcal{L}$ (see for instance (Pavliotis, 2014) or (van Handel, 2016, Section 2.2)). From this we can generate two proofs to the lemma.

Let x^* be a local maximum of $v(\cdot,t)$. Since $\nabla v(x^*,t) = 0$, and $\nabla^2 v(x^*,t) \leq 0$, one concludes from (156) that the value of any local maximum of v can only decrease with time. Another argument relies on observing that the solution to the backward Kolmogorov equation is the expectation

$$v(x, t+s) = \mathbb{E}\left[v(X_{t+s}, t) \middle| X_t = x\right] \quad (\forall s, t \geq 0) \tag{157}$$

where $(X_t)_{t\geq 0}$ is a toroidal stochastic process. However, the expectation of a function is at most its maximum, therefore

$$v(x, t+s) \leq \max_{y\in\mathbb{T}} v(y,t) \quad (\forall x). \tag{158}$$

It now suffices to take the maximum of the left hand side of (158) to prove the claim. □

From hereon we assume E is a potential for which e^{-E} is (C,a)-semi-analytic. Note that for the shifted potential $-E+\delta$ the semi-analyticity parameter C may be replaced with Ce^{δ}. Therefore without loss of generality we assume E attains its minimum value at zero. Note also that in this case $\mathcal{U} \geq e^{-\Delta}$ as it pertains to Proposition A.7.

Lemma B.10. *Let $\overrightarrow{u}$ denote the discretization of the solution to the Fokker–Planck equation ((156)), with the initial condition $u(0) = \mathbf{1}$. Assuming e^{-E} and u are both (C,a)-semi-analytic, and given $N \geq \max(4ad, 4)$, we have*

$$\left\| \frac{d}{dt}\overrightarrow{u(t)} - \mathbb{L}\overrightarrow{u(t)} \right\| \leq 8 \times 10^5 \pi e^3\, \frac{\sqrt{d}\, e^{\Delta}\, C\,(1+\ell L/48)(a^3+a^2)}{\ell^2}\, (2N+1)^{d/2}\, e^{-0.4N} \tag{159}$$

for every point $t \geq 0$ in time, where L is the Lipschitz constant of E. [7]

Proof. We write

$$\mathbb{L}\overrightarrow{u} - \frac{d}{dt}\overrightarrow{u} = e^E\left(\left\|\widetilde{\nabla} e^{-E}\right\|^2 - \left\|\nabla e^{-E}\right\|^2\right)\overrightarrow{u} + e^E\left(\widetilde{\nabla^2} e^{-E} - \nabla^2 e^{-E}\right)\overrightarrow{u}$$
$$+ e^E\left(\nabla e^{-E}\cdot\nabla\overrightarrow{u} - \widetilde{\nabla} e^{-E}\cdot\widetilde{\nabla} u\right) + \left(\widetilde{\nabla^2}\overrightarrow{u} - \overrightarrow{\nabla^2 u}\right) \tag{160}$$

[7] More pedantically, we may let L be the maximum absolute value of the partial derivatives that E attains on the lattice $L = \max_{x\in\mathbb{T}}\max_{j\in[d]} |\partial_j E|$.

and bound every term on the right hand side as follows. For the first term

$$\begin{aligned}
&\sum_{n\in[-N..N]^d} \left| e^{E[n]}\, u[n] \left(\left\| \widetilde{\nabla} e^{-E} \right\|^2 - \left\| \nabla e^{-E} \right\|^2 \right)_{[n]} \right|^2 \\
&\quad \leq \left(\max_x e^{E(x)} u(x) \right)^2 \sum_{n\in[-N..N]^d} \left| \left\| \widetilde{\nabla} e^{-E} \right\|^2 - \left\| \nabla e^{-E} \right\|^2 \right|^2 \\
&\quad \overset{(a)}{\leq} e^{2\, \max_x E(x)} \left[\sum_{n\in[-N..N]^d} \left(\left\| \widetilde{\nabla} e^{-E} \right\| - \left\| \nabla e^{-E} \right\| \right)^2_{[n]} \left(\left\| \widetilde{\nabla} e^{-E} \right\| + \left\| \nabla e^{-E} \right\| \right)^2_{[n]} \right] \\
&\quad \overset{(b)}{\leq} d\, e^{2\Delta} \left(\frac{48}{\ell} N \ln N + L \right)^2 \sum_{n\in[-N..N]^d} \left\| \widetilde{\nabla} e^{-E} - \nabla e^{-E} \right\|^2_{[n]} \\
&\quad \overset{(c)}{\leq} A_1 \frac{dC^2 a^2\, e^{2\Delta}}{\ell^4} \left(N \ln N + \frac{\ell L}{48} \right)^2 (2N+1)^d\; e^{-N/a}
\end{aligned} \tag{161}$$

where we have used Lemma B.9 in (a), a triangle inequality of ℓ_2 norms together with Proposition A.17b in (b), and inequality (69) in (c). Here $A_1 = 3200 \times 48^2 \pi^2\, e^6 < 8 \times 10^6 \pi^2 e^6$ is a constant.

For the second term in (160), we use inequality (70) and Lemma B.9 to write

$$\sum_{n\in[-N.N]^d} \left| e^{E[n]}\, u[n] \left(\widetilde{\nabla^2} e^{-E} - \nabla^2 e^{-E} \right) \right|^2 \leq A_2 \frac{C^2 a^4\, e^{2\Delta}}{\ell^4}\; (2N+1)^d\; e^{-0.8N/a} \tag{162}$$

with $A_2 = 8 \times 10^4 \pi^4 e^6$ being a constant.

We rewrite the third term as

$$e^{E[n]} \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right)_{[n]} = e^{E[n]} \left[\left(\nabla e^{-E} - \widetilde{\nabla} e^{-E} \right) \cdot \widetilde{\nabla} u + \nabla e^{-E} \cdot \left(\nabla u - \widetilde{\nabla} u \right) \right] \tag{163}$$

which allows us to conclude that

$$\begin{aligned}
\left| e^{E[n]} \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right)_{[n]} \right| &\leq e^{E[n]} \left| \left(\nabla e^{-E} - \widetilde{\nabla} e^{-E} \right) \cdot \widetilde{\nabla} u \right| + \left| \nabla E[n] \cdot \left(\nabla u - \widetilde{\nabla} u \right) \right| \\
&\leq e^{E[n]} \left\| \nabla e^{-E} - \widetilde{\nabla} e^{-E} \right\| \left\| \widetilde{\nabla} u \right\| + \left\| \nabla E \right\| \left\| \nabla u - \widetilde{\nabla} u \right\|.
\end{aligned} \tag{164}$$

Hence (by the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$ and using Proposition A.17b) we have

$$\begin{aligned}
\left\| e^E \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right) \right\|^2 &\leq 2 \frac{d\, e^{2\, \max_x E(x)}}{\ell^4} (48 N \log N)^2 \left\| \nabla e^{-E} - \widetilde{\nabla} e^{-E} \right\|^2 \\
&\quad + 2dL^2 \left\| \nabla u - \widetilde{\nabla} u \right\|^2
\end{aligned} \tag{165}$$

which by applying (69) implies

$$\left\| e^E \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right) \right\|^2 \leq 2A_1 \frac{d\, e^{2\Delta}\, C^2 a^2}{\ell^4} \left(N \ln N + \frac{\ell L}{48} \right)^2 (2N+1)^d\; e^{-N/a}. \tag{166}$$

Finally the last term of (160) is taken care of directly using (70):

$$\left\| \widetilde{\nabla^2} u - \nabla^2 u \right\| \leq \sqrt{A_2} \frac{C\, a^2}{\ell^2}\; (2N+1)^{d/2}\, e^{-0.4N} \tag{167}$$

We now observe that $N \ln N + x \leq N \ln N (1+x)$ for all $x > 0$ and all $N \geq 3$. This together with $N \ln N \leq N^2 \leq$

$100a^2e^{0.1N/a}$, and combined with (161), (162), (166), (167), and (160) yield

$$\left\| \frac{d}{dt}\overrightarrow{u(t)} - \mathbb{L}\overrightarrow{u(t)} \right\| \leq A \frac{\sqrt{d}e^{\Delta}C\,(1+\frac{\ell L}{48})(a^3+a^2)}{\ell^2}\,(2N+1)^{d/2}\,e^{-0.4N} \tag{168}$$

where $A = 4 \times 100\sqrt{2A_1} < 1.6 \times 10^6 \pi e^3$. □

Proposition B.11. *Let $u(x,t)$ be the exact solution to the diffusion process and further let $\overrightarrow{v}$ be the solution to the discretized (in space) differential equation i.e., $\overrightarrow{v}$ satisfies $\frac{d\overrightarrow{v}}{dt} = \mathbb{L}\overrightarrow{v}$. Assume $\{u(\cdot,t) : t \in [0,T]\}$ consists of (C,a)-semi-analytic functions, and further assume $N \geq \max(4da, 4)$. Then,*

$$\|\overrightarrow{u}(T) - \overrightarrow{v}(T)\| \leq 1.6 \times 10^6 \pi e^3 \frac{T\,e^{3\Delta/2}\,C^2\,(a^3+a^2)\,(1+\frac{\ell L}{48})}{\ell^2}\,(2N+1)^{d/2}\,e^{-0.4\frac{N}{a}}. \tag{169}$$

Proof. For convenience we will denote the right hand side of (159) as $f[N]$ (which defers from the right hand side of (169) above by a factor of $Te^{\Delta/2}$).

We now write $\frac{d}{dt}\overrightarrow{u}(t) = \mathbb{L}\overrightarrow{u}(t) + \overrightarrow{e}(t)$ where $\|\overrightarrow{e}(t)\|$ is upper bounded in Lemma B.10. Note also that $\frac{d}{dt}\overrightarrow{v} = \mathbb{L}\overrightarrow{v}$ by definition. Hence, letting $\overrightarrow{\varepsilon} := \overrightarrow{u} - \overrightarrow{v}$ we get

$$\frac{d}{dt}\overrightarrow{\varepsilon}(t) = \mathbb{L}\overrightarrow{\varepsilon}(t) + \overrightarrow{e}(t). \tag{170}$$

By Lemma B.7a, $\mathbb{L} = V^{-1}\,D\,V$ where $D \leq 0$ and $\kappa_V \leq e^{\Delta/2}$. It is left to upper bound the norm of $\overrightarrow{\varepsilon}(T)$. We multiply both sides of (170) from left by V, and let $\overrightarrow{\mathcal{E}} := V\overrightarrow{\varepsilon}$ and $\overrightarrow{b} := V\overrightarrow{e}$ to get

$$\frac{d}{dt}\overrightarrow{\mathcal{E}}(t) = D\,\overrightarrow{\mathcal{E}}(t) + \overrightarrow{b}(t). \tag{171}$$

Now, we take the inner product with respect to $\overrightarrow{\mathcal{E}}$:

$$\langle \overrightarrow{\mathcal{E}}, \frac{d}{dt}\overrightarrow{\mathcal{E}} \rangle = \langle \overrightarrow{\mathcal{E}}, D\,\overrightarrow{\mathcal{E}} \rangle + \langle \overrightarrow{\mathcal{E}}(t), \overrightarrow{b}(t) \rangle. \tag{172}$$

Since D is negative semi-definite

$$\mathrm{Re}\left(\langle \overrightarrow{\mathcal{E}}, \frac{d}{dt}\overrightarrow{\mathcal{E}} \rangle\right) \leq \mathrm{Re}\left(\langle \overrightarrow{\mathcal{E}}(t), \overrightarrow{b}(t) \rangle\right) \leq \left\|\overrightarrow{\mathcal{E}}\right\| \left\|\overrightarrow{b}\right\|. \tag{173}$$

Therefore since $\frac{d}{dt}\left\|\overrightarrow{\mathcal{E}}\right\|^2 = 2\,\mathrm{Re}\left(\langle \overrightarrow{\mathcal{E}}, \frac{d}{dt}\overrightarrow{\mathcal{E}} \rangle\right)$, we conclude

$$\frac{d}{dt}\left\|\overrightarrow{\mathcal{E}}(t)\right\| \leq \left\|\overrightarrow{b}(t)\right\|. \tag{174}$$

Recalling $\overrightarrow{\mathcal{E}} = V\overrightarrow{\varepsilon}$ and $\overrightarrow{b} = V\overrightarrow{e}$ we get

$$\frac{d}{dt}\|V\,\overrightarrow{\varepsilon}\| \leq f[N]\|V\,|e\rangle\|. \tag{175}$$

where $|e(t)\rangle$ is $\overrightarrow{e}(t)$ normalized. We use $\|V\ |e\rangle\| \leq \sigma_{\max}(V)$ to conclude that

$$\|V\,\overrightarrow{\varepsilon}(T)\| \leq T f[N]\sigma_{\max}(V)\,. \tag{176}$$

Finally, using $\|V\overrightarrow{\varepsilon}\| \geq \|\overrightarrow{\varepsilon}\|\sigma_{\min}(V)$ and $\kappa_V \leq e^{\Delta/2}$ we complete the proof. □

C. Training energy-based models

In machine learning, energy-based models (EBM) are used to regress a Gibbs distribution

$$p_\theta(x) = \exp(-\beta E_\theta(x))/Z_{\beta,\theta} \tag{177}$$

known as the *model distribution* from an unknown distribution p_{data} represented by classical data. The EBM comprises a function approximator for an *energy potential* $E_\theta : \mathbb{R}^d \to \mathbb{R}$. Here $\theta \in \mathbb{R}^m$ denotes a vector of m model parameters (e.g., weights and biases of a deep neural network). Given a set of i.i.d training samples $D = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$, the goal of the learning procedure is to find a vector of model parameters $\theta^* \in \mathbb{R}^m$ that attain optimal regression of p_{data} via p_{θ^*} with respect to the Kullback-Leibler (KL) distance between the two distributions. It is easy to see that this is equivalent to maximizing the log-likelihood of the training data:

$$\mathrm{KL}(p_{\text{data}}(x)||p_\theta(x)) = -\mathbb{E}_{x\sim p_{\text{data}}}[\log p_\theta(x)] + \text{constant}. \tag{178}$$

However, we do not need access to the value of the likelihood directly but rather the gradient of the log-probability of the model. We have

$$\mathbb{E}_{p_{\text{data}}}[\nabla_\theta \log p_\theta(x)] = -\beta\mathbb{E}_{p_{\text{data}}}[\nabla_\theta E_\theta(x)] + \beta\mathbb{E}_{p_\theta}[\nabla_\theta E_\theta(x)]. \tag{179}$$

While the first term is easy to approximate using the data samples the second term is approximated through costly Gibbs sampling. Indeed, if we can efficiently draw samples from the model distribution p_θ, we have access to unbiased estimates of the log-likelihood gradient, which in turn can be used to train the EBM via stochastic gradient descent (Song & Kingma, 2021).

The energy function is a composition of linear functions (affine transformations) with nonlinear ones such as the sigmoid function [8] for which we can build a quantum oracle as in Fig. 1. Our algorithm works by queries to an oracle for the discretization of the generator of the Fokker–Planck equation L (see (149)). In order to construct this oracle, we use the expression (135) restated here as

$$\mathbb{L}(-) = \widetilde{\nabla^2}E(-) + \widetilde{\nabla}E \cdot \widetilde{\nabla}(-) + \widetilde{\nabla^2}(-) \tag{180}$$

constructed using $2d(2N+1)$ replicas of the energy oracle.

After iterative queries to the oracle of $\mathbb{L}$, Algorithm 1 returns a sample from the model distribution of the EBM. Repeated executions will then provide an approximation of the second term in (179). This in turn allows for updating the model parameters θ via stochastic gradient descent. And finally, repeated descent steps will result in an approximation for trained parameters θ^*.

Alternatively, we may use the controlled variant of Fig. 4 in order to perform mean estimations of the components of the gradient $\nabla_\theta E_\theta(x)$ as per Corollary 4.3 quantumly, as opposed to using samples provided by the quantum computer to perform classical estimation of the expectation $\mathbb{E}_{p_\theta}[\nabla_\theta E_\theta]$. The mean estimation algorithm queries this controlled oracle of $\mathbb{L}$ and additionally the controlled oracles of the m partial derivatives of E_θ. The construction of the latter oracles can be automated in the same fashion as automatic differentiation in ML.

D. Proofs of the results in Section 4

D.1. Gibbs sampling

Our algorithm benefits from the use of the high precision quantum linear differential equation solver developed by Berry et al. (2017). Krovi later proved in (2023) that this solver is more efficient than initially shown in the original work. The solver is designed to tackle the problem of interest, which is solving the following ODE at time $T > 0$:

$$\frac{d\overrightarrow{x}}{dt} = A\overrightarrow{x} + \overrightarrow{b},\ \overrightarrow{x}(0) = \overrightarrow{x}_{\text{init}}, \tag{181}$$

[8] Although the rectified linear unit (ReLU) is more typically used it does not fall under the semi-analyticity assumptions of our paper.

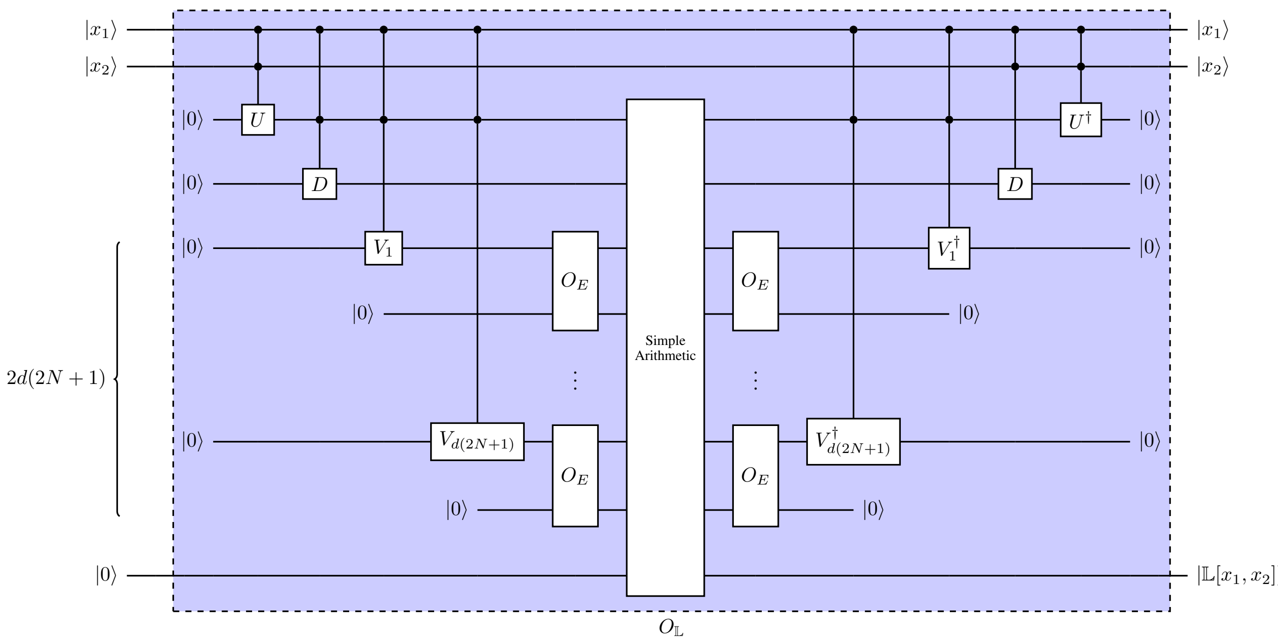

Figure 4. The circuit for the oracle of discrete generator $\mathbb{L}$ comprising $2d(2N+1)$ copies of the energy potential oracle, O_E. To query $\mathbb{L}[x_1, x_2] = \langle x_2| \mathbb{L} |x_1\rangle$, first the controlled-$U$ gate checks for the difference between x_1 and x_2: the third register is set to $|i\rangle$ if x_1 and x_2 defer only on their i-th entry. The state remains unchanged, if $x_1 = x_2$, and it is set to a null state $|\perp\rangle$ otherwise. Conditioned on this third register being at state $|i\rangle$, another register (the fourth register) computes the distance between x_1 and x_2 along the i-th axis on the lattice (the controlled-D gate). Again, conditioned on the state of the third register, we query the energy function at specific lattice points to compute either $\partial_i E(x_1)$ (if the third register is in $|i\rangle$), or $\nabla^2 E(x_1)$ (if the third register is in $|0\rangle$) using the sequence of controlled-V_j gates. The estimation of these derivatives exploits Fourier spectral method (see Appendix A.2) and is applied via a circuit performing simple arithmetic.

where $\overrightarrow{x}, \overrightarrow{b} \in \mathbb{C}^n$ and $A \in \mathbb{C}^{N\times N}$ is a matrix whose all eigenvalues have non-positive real parts. The aim of 'solving' the ODE is to prepare a quantum state that encodes the entries of $x(T)$. The main idea that is used in the construction of the algorithm is the truncation of the Taylor expansion of the exponential function, as the $\overrightarrow{x}(T)$ satisfies the following closed-form solution ((Krovi, 2023, Lemma 6)).

$$\overrightarrow{x}(T) = e^{At}\overrightarrow{x}(0) + \left[\int_0^T e^{As}\, \mathrm{d}s\right] \overrightarrow{b} \tag{182}$$

We directly apply their algorithm to solve Eq. (156). Hence, we restate their complexity result.

Theorem D.1 (Adoption of Theorem 7 of (Krovi, 2023))**.** *Suppose $A = VDV^{-1}$ is an $N \times N$ diagonalizable matrix, where $D = diag(\lambda_1, \cdots, \lambda_N)$ satisfies $\mathrm{Re}(\lambda_i) \leq 0$ for any $i \in [N]$. In addition, suppose A has at most s nonzero entries in any row and column, and we have an oracle O_A that computes these entries. Suppose $\overrightarrow{x}_{init}$ and $\overrightarrow{b}$ are N-dimensional vectors with known norms and that we have two controlled oracles, O_x and O_b that prepare states proportional to $\overrightarrow{x}_{init}$ and b, respectively. Let $\overrightarrow{x}$ evolve according to the differential equation*

$$\frac{d}{dt}\overrightarrow{x} = A\overrightarrow{x} + \overrightarrow{b} \tag{183}$$

with the initial condition $\overrightarrow{x}(0) = \overrightarrow{x}_{init}$. Let $T > 0$ and

$$g = \max_{t\in[0,T]} \|\overrightarrow{x}(t)\|/\|\overrightarrow{x}(T)\|. \tag{184}$$

Then there exists a quantum algorithm that produces a state ε-close to $\overrightarrow{x}(T)/\|\overrightarrow{x}(T)\|$ in ℓ^2-norm, succeeding with $\Omega(1)$ probability, with a flag indicating success, using

$$O\Big(gT\|A\|\kappa_V \operatorname{poly}\Big(s, \log d, \log\Big(1 + \frac{Te^2\|b\|}{\|x_T\|}\Big), \log\Big(\frac{1}{\varepsilon}\Big), \log(T\|A\|\kappa_V)\Big)\Big)\,, \tag{185}$$

queries overall, where $\kappa_V = \|V\|\|V^{-1}\|$ is the condition number of V. In addition, the gate complexity of the algorithm is larger than its query complexity by a factor of

$$O(\operatorname{polylog}(1 + \frac{Te^2\|b\|}{\|x_T\|}, 1/\varepsilon, T\|A\|))\,, \tag{186}$$

.

Proof. We can observe that the only difference in the complexity result between our theorem and the one presented in (Krovi, 2023, Theorem 7) is the substitution of $C(A)$ with κ_V. It is worth recalling from (Krovi, 2023, Definition 5) that $C(A)$ is defined as the maximum norm of e^{At} over the interval $[0, T]$. By imposing the condition $A = VDV^{-1}$ with $D \leq 0$, we readily obtain $C(A) \leq \kappa_V$. It is worth noting that in the case where A is not diagonalizable, the relation between $C(A)$ and κ_V is presented in (Krovi, 2023, Lemma 4). ☐

We now provide the proof of main result.

Theorem D.2 (Theorem 4.1 in the manuscript)**.** *Given an L-Lipschitz periodic potential E, suppose that the one-parameter family of all probability measures $\{e^{\mathcal{L}t}\rho_0 : t \geq 0\}$ consists of semi-analytic functions with parameters (C, a). Algorithm 1 samples from a distribution ε-close to the Gibbs distribution (in total variation distance), by making*

$$\mathcal{O}\left(d^3\, \frac{\kappa_{E/2}}{\ell^2}\, e^{\frac{\Delta}{2}} \max\left\{a^4 d^4, \log^4\left(\frac{\sqrt{d}e^{\frac{5\Delta}{4}} Ca^3(1 + \ell\, L)}{\varepsilon}\right)\right\} \operatorname{polylog}\left(\frac{ade^{\Delta} \log(C(1 + \ell L))}{\varepsilon}\right)\right)$$

queries to the oracle of the energy function. The algorithm succeeds with bounded probability of failure and returns a flag indicating its success. In addition, the gate complexity of the algorithm is larger only by a factor of $\operatorname{polylog}(Cade^{\Delta}(1 + \ell L))/\varepsilon)$.

Proof. We first note that the energy function we consider in the Fokker–Planck equation is $E/2$. Here we provide the method to obtain a 6ε-approximate sampler. Using Proposition B.11 we have

$$\left\| \left|e^{\mathbb{L}T}\,\rho_0\right\rangle - \left|e^{\mathcal{L}T}\,\rho_0\right\rangle \right\| \leq 3.2\times 10^6 \pi e^3 \frac{\sqrt{d}Te^{3\Delta/4}\,C\,(1+\frac{\ell L}{48})(a^3+a^2)}{\ell^2} e^{-0.4N} \tag{187}$$

where we have also used Lemma E.6 and Lemma B.8. Hence, if we let $A = 3.2\times 10^6\pi e^3$, we may choose

$$N = \left\lceil \max\left\{ 0.4^{-1}\log\left(\frac{A\,\sqrt{d}T\,e^{3\Delta/4}\,C\,(a^2+a^3)(1+\ell\,L/48)}{\ell^2\,\varepsilon}\right), 4ad, 4 \right\}\right\rceil \tag{188}$$

to guarantee an at most ε distance between $\left|e^{\mathbb{L}T}\,\rho_0\right\rangle$ and $\left|\rho_T\right\rangle = \left|e^{\mathcal{L}T}\,\rho_0\right\rangle$. Now we apply Theorem D.1 to obtain an output state $|\mathcal{A}\rangle$, for which we have $\||\mathcal{A}\rangle - |\rho_T\rangle\| \leq 2\varepsilon$. Hence, by Theorem A.8, continuous sampling from the algorithm's output will provide samples from a 5ε-close distribution to the distribution proportional to ρ_T^2.

Now we need to set T. Lemma F.1, together with Corollary B.6 implies that choosing

$$T = \kappa_{E/2}\log\left(2e^{\Delta/2}/\varepsilon\right) \tag{189}$$

guarantees that the distribution proportional to ρ_T^2 is ε-close (in total variation distance) to the Gibbs distribution. Overall, our sampling procedure returns samples from a distribution 6ε-close to the Gibbs distribution.

The complexity of the algorithm according to Theorem D.1 is now obtained by noting firstly that $\kappa_V \leq e^{\Delta/4}$ from Lemma B.7a for $E/2$. Next, we note that the sparsity of $\mathbb{L}$ is $s = O(dN)$. Also $g = O(e^{\frac{\Delta}{4}})$ by Lemma B.8. The norm of $\mathbb{L}$ is bounded by Lemma B.7c as $\|\mathbb{L}\| = O(\Delta dN^2/\ell^2 \operatorname{polylog} N)$. Finally $\beta \leq 1$ for us also using Lemma B.8. This provides the complexity of every term in Theorem D.1 with respect to N. We also note that

$$N = \Theta^*\left(\max\left\{\log\left(\sqrt{d}\frac{e^{5\Delta/4}\,a^3C\,(1+\ell L)}{\varepsilon}\right), ad\right\}\right) \tag{190}$$

by our choice of T. Finally, a query to $\mathbb{L}$ requires $\mathcal{O}(dN)$ queries to $\mathcal{O}_E$ and this completes the proof. □

D.2. Mean estimation

In this section, we delve into how the Gibbs sampler discussed earlier can be employed to calculate the expected values of random variables with bounded variance. Specifically, we consider a periodic function $f : [-\frac{\ell}{2}, \frac{\ell}{2}]^d \to \mathbb{R}$ that belongs to $L^2(\rho)$, and we aim at estimating $\mathbb{E}_\rho f(X)$, where X is a random variable with distribution ρ. We utilize the state-of-the-art estimation algorithm presented in (Kothari & O'Donnell, 2023) to compute the expected value of our function.

The main problem of interest is that of the mean estimation of a classical random variable, whose classical probability amplitudes are encoded in a quantum state. Brassard et al. (2002) consider having access to a unitary U, that acts as $U\,|0\rangle = \sqrt{p}\,|0\rangle + \sqrt{1-p}\,\left|0^\perp\right\rangle$, where $\left|0^\perp\right\rangle$ is a vector orthogonal to $|0\rangle$. They prove that $\mathcal{O}(\frac{1}{\varepsilon})$ queries to controlled-U is sufficient to estimate p with precision ε, and with high probability. The proof is based on the fact that $|\psi\rangle := U\,|0\rangle$ can be viewed as

$$|\psi\rangle = \frac{1}{2}e^{i\theta}\left(|0\rangle - i|0^\perp\rangle\right) + \frac{1}{2}e^{-i\theta}\left(|0\rangle + i|0^\perp\rangle\right) \tag{191}$$

where $\sin\theta = \sqrt{p}$. We then note that $\left(|0\rangle \pm i|0^\perp\rangle\right)$ are eigenvectors of a rotation matrix with rotation angle ϕ in the $|0\rangle\,, |0^\perp\rangle$ plane, with eigenvalues $e^{\pm i\phi}$. As the Grover diffusion operator is itself a rotation with angle 2θ, the phase estimation algorithm (with the unitary being the Grover operator, and the input state being $|\psi\rangle$) will reveal θ, and consequently p. One can think of this algorithm as an estimation algorithm for a binary random variable. Note that classically one requires $\Omega(\frac{1}{\varepsilon^2})$ samples in order to achieve an ε-accurate estimation of p. This quadratic speedup with respect to the error parameter is sometimes referred to as the Heisenberg limit, and as we discuss later, is not restricted to the case of binary random variables.

Subsequently, Montanaro (2015) and Li & Wu (2018) extended the above algorithm and obtained mean estimation algorithms for more generic cases. Hamoudi & Magniez (2018) combines the latter algorithms and obtains a desired complexity of

$\mathcal{O}^*(\frac{1}{\varepsilon})$. The recent work of (Kothari & O'Donnell, 2023) is the state of the art and provides an algorithm that we directly apply for our mean estimation tasks.

Assume we have access to controlled-U, and its inverse, such that $U|0\rangle = \sum_{x\in\Omega}\sqrt{p_x}|x\rangle$. Further, one can assume having access to controlled versions of a unitary F and its inverse ($F^\dagger$) that acts as $F|x\rangle|0\rangle|0\rangle = |x\rangle|f(x)\rangle|0\rangle$, for some function f. Note that F is allowed to exploit auxiliary qubits for the evaluation of f. Having access to such quantum circuits is phrased as 'having the code' for the random variable $f(X)$ in (Kothari & O'Donnell, 2023). We restate the following theorem from their work.

Theorem D.3 (Theorem 1.3 of (Kothari & O'Donnell, 2023))**.** *There is a computationally efficient quantum algorithm with the following properties: Given 'the code' for a random variable $f(X)$, the algorithm makes $O(n\log\frac{1}{\delta})$ queries to the oracles for the controlled unitaries U, $U^\dagger$, F, and $F^\dagger$ to output an estimation $\widehat{\mu}$ such that*

$$\mathbb{P}\left[|\widehat{\mu} - \mathbb{E}[f(X)]| \geq \frac{\mathrm{Var}[f(X)]}{n}\right] \leq \delta. \tag{192}$$

The algorithm they propose is again based on Gorver's diffusion operators. However, they use unitaries with complex phases (as opposed to reflections). Let us now state our mean estimation result.

Corollary D.4 (Corollary 4.3 in the manuscript)**.** *Let E be an energy function, satisfying the assumptions made in Theorem D.2. Furthermore, let f be an L_f-Lipschitz ℓ-periodic function with diameter Δ_f. There is a quantum algorithm that returns an estimate $\widehat{\mu}$ to $\mathbb{E}[f(X)]$, with additive error at most $\varepsilon > 0$ and success probability at least $1-\delta$, making*

$$\mathcal{O}\left(d^7 a^4 e^{\Delta/2}\frac{\kappa_{E/2}}{\ell^2}\frac{\Delta_f}{\varepsilon}\log\left(\frac{1}{\delta}\right)\mathrm{polylog}\left(C, a, \frac{1}{\varepsilon}, \Delta_f, L_f\ell, L\ell\right)\right) \tag{193}$$

queries to the controlled and standalone oracles of the energy function E and the function f.

Proof. Consider the quantum circuit that implements line 2 of Algorithm 1, and call it U. We can manipulate this circuit to obtain a unitary $\widetilde{U}$ such that $\left\||u(T)\rangle - \widetilde{U}|u(0)\rangle\right\| \leq \varepsilon_1$, by making $\mathcal{O}\left(\log\frac{1}{\varepsilon_1}\right)$ calls to U, $U^\dagger$, and additional gates. This is achieved via fixed-point amplitude amplification algorithm[9] (Grover, 2005). Note that a total-variation distance of ε_1 between the two distributions, results in at most a $M\varepsilon_1$ distance between the expected values. Furthermore, we note that expectation with respect to the algorithm's output would be far from the actual value by at most

$$\frac{d/2(\ell L_f + \ell L\Delta_f + 10\sqrt{2}/3e^4\Delta_f aC)}{M} + 2\Delta_f\varepsilon_1, \tag{194}$$

which follows from the total variation bounds obtained above, and further, that of Lemma E.8. Hence, implementing line 3 of Algorithm 1 with $M = \frac{\mathrm{poly}(C,a,\Delta_f,L_f\ell,L\ell)}{\varepsilon}$, and $\varepsilon_1 = \frac{\varepsilon}{8\Delta_f}$, results in at most a distance $\varepsilon/2$ from the ideas expectation. Finally, setting $n = \frac{\Delta_f}{\varepsilon}$ and applying Theorem D.3 concludes the result. □

E. Lemmas used in Appendix A

Lemma E.1. *For any integer $m \geq 0$ and $z > 0$*

$$\sum_{k=0}^{\infty}\frac{z^k\,k^m}{k!} \leq e^z\,\max\{z^m, 1\}\,m!. \tag{195}$$

Proof. We note that

$$\sum_{k=0}^{\infty}\frac{z^k\,k^m}{k!} = \frac{\partial^m}{\partial x^m}\bigg|_{x=0}\sum_{k=0}^{\infty}\frac{z^k\,e^{kx}}{k!} = \frac{\partial^m}{\partial x^m}\bigg|_{x=0} e^{z\,e^x}. \tag{196}$$

[9] The $\frac{\pi}{3}$-amplitude amplification algorithm of (Grover, 2005) has a dependence on the success probability of the algorithm, which is later improved in the works of Yoder et al. (2014) and Gilyén et al. (2019). However, as the success probability of our circuit is $\Omega(1)$, we do not need to utilize the more complex algorithms.

Defining $f(x) := e^{z\, e^x}$ and $g(x) := z\, e^x$ we observe that $f'(x) = f(x)\, g(x)$ and further $g'(x) = g(x)$. Therefore, one can expand the s-th derivative as follows

$$\frac{\partial^s}{\partial x^s} f(x) = \sum_{r=1}^{s} C_s[r]\ (g(x))^r\ f(x) \tag{197}$$

Taking derivative of both sides yields the following recursive relations

$$C_{s+1}[r] = \begin{cases} C_s[1], & \text{if } r = 1, \\ r\, C_s[r] + C_s[r-1], & \text{if } 2 \le r \le s, \\ C_s[s], & \text{if } r = s+1. \end{cases} \tag{198}$$

Therefore

$$\sum_{r=1}^{s+1} C_{s+1}[r] = \sum_{r=1}^{s} C_s[r]\,(r+1) \le (s+1) \sum_{r=1}^{s} C_s[r] \tag{199}$$

which given $C_1[1] = 1$ implies that

$$\sum_{r=1}^{m} C_m[r] \le m!. \tag{200}$$

Lastly, we note that

$$\left.\frac{\partial^m}{\partial x^m}\right|_{x=0} e^{z\, e^x} \le e^z\ \max\{z^m, 1\} \sum_{r=1}^{m} C_m[r] \le e^z\, z^m\, m! \tag{201}$$

since $g(0) = z$, and $f(0) = e^z$, and the last inequality follows from (200). Combined with (196) the result follows. □

Lemma E.2. *For any $z > 1$ and $m \in \mathbb{N}$, the following holds*

$$\sum_{k=0}^{\infty} k^m\, z^{-k} \le \frac{1}{1 - z^{-1}} \max\left(2, \frac{2}{z-1}\right)^m\ m!. \tag{202}$$

Proof. First, we note that the function we are upper bounding is a special case of the Lerch transcendents (Gradshteyn & Ryzhik, 2000), for which we also provide a lower bound in Example A.3. In particular, $\Phi(z^{-1}, m, 0) = \sum_{k=0}^{\infty} k^m z^{-k}$. We shall now prove the result by setting $\alpha = \ln z$ and writing

$$\sum_{k \ge 0} k^m z^{-k} = \frac{\partial^m}{\partial \alpha^m} \sum_{k \ge 0} e^{\alpha k} = \frac{\partial^m}{\partial \alpha^m} \frac{1}{1 - e^{\alpha}}. \tag{203}$$

By a simple induction, we arrive at the following form

$$\frac{\partial^m}{\partial \alpha^m} \frac{1}{1 - e^{\alpha}} = \sum_{r=1}^{m} C_m[r] \frac{e^{r\alpha}}{(1 - e^{\alpha})^{r+1}}, \tag{204}$$

with the following recursive relation for the coefficients

$$C_{s+1}[r] = \begin{cases} C_s[1], & \text{if } r = 1, \\ r\,(C_s[r] + C_s[r-1]) & \text{if } 2 \le r \le s, \\ rC_s[s], & \text{if } r = s+1. \end{cases} \tag{205}$$

Hence, we have $\sum_{r=1}^{m} C_m[r] \leq 2m \sum_{r=1}^{m-1} C_{m-1}[r]$, and consequently $\sum_{r=1}^{m} C_m[r] \leq 2^m m!$, as $C_1[1] = 1$. Therefore,

$$\frac{\partial^m}{\partial \alpha^m} \frac{1}{1-e^{\alpha}} \leq \frac{1}{1-z^{-1}} \, \max\{1, \frac{1}{z-1}\}^m \, 2^m \, m!. \tag{206}$$

□

Lemma E.3. *For any $m \geq 1$*

$$\sum_{k=0}^{m} k!(m-k)! \leq 3 \, m!. \tag{207}$$

Proof. Note that the inequality could be checked by direct calculation for $m = 1, 2, 3$. We now consider $m \geq 4$ and note that

$$\sum_{k=0}^{m} k!(m-k)! = m! \sum_{k=0}^{m} \frac{1}{\binom{m}{k}} \tag{208}$$

However, we have $\min_{k=2,\cdots,m-2} \binom{m}{k} \geq m$ and hence

$$\sum_{k=0}^{m} \frac{1}{\binom{m}{k}} = 2 + \frac{2}{m} + \sum_{k=2}^{m-2} \frac{1}{\binom{m}{k}} \leq 2 + \frac{2}{m} + \frac{m-3}{m} \leq 3. \tag{209}$$

□

Lemma E.4. *For all $d \geq 1$*

$$\sum_{\substack{i_1, i_2, \cdots, i_d \in \{0, \cdots, m\} \\ i_1 + i_2 + \cdots + i_d = m}} i_1! \, i_2! \, \cdots \, i_d! \leq 3^{d-1} \, m!. \tag{210}$$

Proof. We prove this claim by induction. The base case ($d = 1$) is trivially true. Assuming the statement is correct for d, we have

$$\begin{aligned} \sum_{i_1 + i_2 + \cdots + i_d = m} i_1! \, i_2! \, \cdots \, i_{d+1}! &= \sum_{i_1=0}^{m} i_1! \left(\sum_{i_2 + \cdots + i_{d+1} = m - i_1} i_2! \cdots i_{d+1}! \right) \\ &\leq 3^{d-1} \sum_{i_1=0}^{m} i_1 \, (m - i_1)! \\ &\leq 3^d \, m! \end{aligned} \tag{211}$$

where the last step is due to Lemma E.3. □

Lemma E.5. *Let m be an integer greater than d. We have*

$$\max_{x \in [-1,1]^d} \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|x + 2\,p\|^{-2m} \leq 2^{d+1} \tag{212}$$

for a universal constant ξ.

Proof. Firstly, we note that

$$\max_{x \in [-1,1]^d} \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|x + 2\,p\|^{-2m} \leq \sum_{j=1}^{d} \binom{d}{j} \max_{x \in [-1,1]^j} \sum_{p \in (\mathbb{Z} \setminus \{0\})^j} \frac{1}{\|x + 2p\|^{2m}} \tag{213}$$

$$\leq \sum_{j=1}^{d} \binom{d}{j} \max_{x\in[-1,1]^j} \sum_{p\in(\mathbb{Z}\setminus\{0\})^j} \frac{1}{\|x+2p\|^{2j}}. \tag{214}$$

We now note that $\sum_{p\in(\mathbb{Z}\setminus\{0\})^j} \frac{1}{\|x+2p\|^{2m}}$ is maximized over $[-1,1]^j$ for x being one of the corner points. Hence, we can upper bound the summation by the following integral with respect to the volume form dV_j of the j-dimensional ball in the ℓ_2 norm:

$$\begin{aligned}\max_{x\in[-1,1]^d} \sum_{p\in\mathbb{Z}^d\setminus\{0\}} \|x+2\,p\|^{-2m} &\leq \sum_{j=1}^{d} \binom{d}{j} \left(\frac{1}{\sqrt{j}} + \frac{1}{2^j}\int_{r=1}^{\infty} \frac{dV_j}{r^{2j}}\right) \\ &\leq \sum_{j=1}^{d} \binom{d}{j} \left(\frac{1}{\sqrt{j}} + \frac{\pi^{j/2}}{2^j\Gamma(j/2+1)} \int_{r=1}^{\infty} r^{-1-j}\,dr\right) \\ &= \sum_{j=1}^{d} \binom{d}{j} \left(\frac{1}{\sqrt{j}} + \frac{\pi^{j/2}}{2^j\Gamma(j/2+1)} \frac{1}{j}\right) \\ &\leq 2^d \sup_{j\in\mathbb{N}} \left(\frac{1}{\sqrt{j}} + \frac{\pi^{j/2}}{2^j\Gamma(j/2+1)} \frac{1}{j}\right) = 2^{d+1}\end{aligned} \tag{215}$$

□

Lemma E.6 (Lemma 13 of (Berry et al., 2017))**.** *Let $\vec{a}$ and $\vec{b}$ be two vectors of the same vector space. It is the case that*

$$\left\|\frac{\vec{a}}{\|\vec{a}\|} - \frac{\vec{b}}{\|\vec{b}\|}\right\| \leq \frac{2\left\|\vec{a}-\vec{b}\right\|}{\max\left[\|\vec{a}\|, \|\vec{b}\|\right]}. \tag{216}$$

Proof. Without loss of generality, we assume $\vec{a}$ has a larger norm. We then write

$$\begin{aligned}\left\|\frac{\vec{a}}{\|\vec{a}\|} - \frac{\vec{b}}{\|\vec{b}\|}\right\| &= \left\|\frac{\vec{a}}{\|\vec{a}\|} - \frac{\vec{b}}{\|\vec{a}\|} + \frac{\vec{b}}{\|\vec{a}\|} - \frac{\vec{b}}{\|\vec{b}\|}\right\| \\ &\leq \left\|\frac{\vec{a}}{\|\vec{a}\|} - \frac{\vec{b}}{\|\vec{a}\|}\right\| + \left\|\frac{\vec{b}}{\|\vec{a}\|} - \frac{\vec{b}}{\|\vec{b}\|}\right\| \\ &\leq \frac{2\left\|\vec{a}-\vec{b}\right\|}{\|\vec{a}\|}\end{aligned} \tag{217}$$

where in the last inequality we have used the triangle inequality $\|\vec{a}\| - \left\|\vec{b}\right\| \leq \left\|\vec{a}-\vec{b}\right\|$. □

Lemma E.7. *Let $|\psi\rangle$ and $|\phi\rangle$ be two quantum states residing in a finite dimensional Hilbert space. Let us denote the output measurement probabilities in the computational basis by P_ψ and P_ϕ. Then we have*

$$\mathrm{TV}\left(P_\psi, P_\phi\right) \leq \||\psi\rangle - |\phi\rangle\|. \tag{218}$$

Proof. We have

$$\begin{aligned}
\mathrm{TV}\left(P_\psi, P_\phi\right) &= \frac{1}{2} \sum_{i \in I} \left| |\psi(i)|^2 - |\phi(i)|^2 \right| \\
&= \frac{1}{2} \sum_{i \in I} \left| \, |\psi(i)| - |\phi(i)| \right| \cdot \left| |\psi(i)| + |\phi(i)| \, \right| \\
&\leq \frac{1}{2} \sqrt{\sum_{i \in I} \left| \, |\psi(i)| - |\phi(i)| \right|^2} \sqrt{\sum_{i \in I} \left| \, |\psi(i)| + |\phi(i)| \right|^2} \\
&\overset{(a)}{\leq} \frac{1}{\sqrt{2}} \sqrt{\sum_{i \in I} \left| \psi(i) - \phi(i) \right|^2} \sqrt{\sum_{i \in I} |\psi(i)|^2 + |\phi(i)|^2} \\
&= \| |\psi\rangle - |\phi\rangle \|
\end{aligned} \tag{219}$$

where (a) follows from the basic inequalities $(a+b)^2 \leq 2a^2 + 2b^2$ and $\left| |a| - |b| \right| \leq |a - b|$. □

Lemma E.8. *Let $u : \mathbb{R}^d \to \mathbb{R}$ be ℓ-periodic along each dimension. Further let u be (C, a)-semi-analytic and L-Lipschitz. Also, let μ be the probability density proportional to u^2, and further, $\widehat{\mu}$ be the probability density associated with the continuous sampling from $|u_M\rangle$. Choosing M as*

$$M = \left\lceil \frac{1}{\delta} \frac{L\ell d/2 + 10/3\sqrt{2}\, a e^4 C}{\mathcal{U}} \right\rceil \tag{220}$$

where $\mathcal{U} = \sqrt{\mathbb{E}[u^2(X)]}$, we are guaranteed to have $\mathrm{TV}\left(\mu, \widehat{\mu}\right) \leq \delta$.

Proof. We have

$$\mathrm{TV}(\mu, \widehat{\mu}) = \frac{1}{2} \int dx \left| \frac{u^2}{\ell^d \mathcal{U}^2} - \sum_{n \in [-M..M]^d} \mathbf{1}_{\{x \in \mathbb{B}_n\}} \frac{u_M^2[n]}{\|\overrightarrow{u_M}\|^2} \left(\frac{2M+1}{\ell} \right)^d \right|. \tag{221}$$

Using the triangle inequality, we have

$$\begin{aligned}
\mathrm{TV}(\mu, \widehat{\mu}) &\leq \frac{1}{2} \int dx \left| \frac{u^2}{\ell^d \mathcal{U}^2} - \sum_{n \in [-M..M]^d} \mathbf{1}_{\{x \in \mathbb{B}_n\}} \frac{u_M^2[n]}{\mathcal{U}^2 \ell^d} \right| \\
&\quad + \frac{1}{2} \int \frac{dx}{\ell^d} \sum_{n \in [-M..M]^d} \mathbf{1}_{\{x \in \mathbb{B}_n\}} u_M^2[n] \cdot \left| \frac{1}{\mathcal{U}^2} - \frac{(2M+1)^d}{\|\overrightarrow{u_M}\|^2} \right| \\
&= \frac{1}{2\ell^d \mathcal{U}^2} \int dx \sum_{n \in [-M..M]^d} \mathbf{1}_{\{x \in \mathbb{B}_n\}} \left| u^2 - u_M^2[n] \right| + \frac{1}{2\mathcal{U}^2} \left| \frac{\|\overrightarrow{u_M}\|^2}{(2M+1)^d} - \mathcal{U}^2 \right| \\
&\overset{(a)}{\leq} \frac{L\ell d}{\mathcal{U}(2M+1)} + \frac{2\sqrt{2} e^3 C}{\mathcal{U}} e^{-3M/5a}
\end{aligned} \tag{222}$$

where in (a) we use Lemma A.4 with the choice of M in (220). Using the inequality $e^{-x/e} \leq 1/x$, we obtain

$$\mathrm{TV}(\mu, \widehat{\mu}) \leq \frac{L\ell d/2 + 10/3\sqrt{2}\, a e^4 C}{\mathcal{U} M} \tag{223}$$

which concludes the proof. □

F. Lemmas used in Appendix D.1

Lemma F.1. *Let $u(x)$ be an ℓ-periodic real-valued function satisfying*

$$\sqrt{\int_{\mathbb{T}} \rho_s \left(\frac{u}{\rho_s} - 1\right)^2} \leq \delta \sqrt{\int_{\mathbb{T}} \rho_s \left(\frac{1}{V\rho_s} - 1\right)^2} \tag{224}$$

for some $\delta > 0$, with $V = \ell^d$ (the volume of the torus $\mathbb{T}$). Then,

$$\frac{1}{2}\int_{\mathbb{T}} \left| \frac{u^2}{\int_{\mathbb{T}} u^2} - \frac{\rho_s^2}{\int_{\mathbb{T}} \rho_s^2} \right| \leq 2\,\delta\, e^{\Delta/2}. \tag{225}$$

Proof. Note that from the assumption

$$\sqrt{\int_{\mathbb{T}} (u - \rho_s)^2} \leq e^{\Delta/2}\,\delta \sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}. \tag{226}$$

By a similar argument as in Lemma E.7 we have

$$\frac{1}{2}\int_{\mathbb{T}} \left| \frac{u^2}{\int u^2} - \frac{\rho_s^2}{\int_{\mathbb{T}} \rho_s^2} \right| \leq \sqrt{\int_{\mathbb{T}} \left(\frac{u}{\sqrt{\int_{\mathbb{T}} u^2}} - \frac{\rho_s}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} \right)^2}. \tag{227}$$

Furthermore, using the triangle inequality (c.f., Lemma E.6)

$$\sqrt{\int_{\mathbb{T}} \left(\frac{u}{\sqrt{\int_{\mathbb{T}} u^2}} - \frac{\rho_s}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} \right)^2} \leq 2\,\frac{\sqrt{\int_{\mathbb{T}} (u-\rho_s)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}}. \tag{228}$$

Now, putting equations (226), (227), and (228) together, we have

$$\frac{1}{2}\int_{\mathbb{T}} \left| \frac{u^2}{\int u^2} - \frac{\rho_s^2}{\int_{\mathbb{T}} \rho_s^2} \right| \leq 2\delta e^{\Delta/2} \frac{\sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}}. \tag{229}$$

Now consider a random variable X drawn uniformly at random from $\mathbb{T}$, and define $Y := \rho_s(X)$. It is clear that

$$\frac{\sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} = \sqrt{\frac{\mathbb{E}\left[\left(Y - \frac{1}{V}\right)^2\right]}{\mathbb{E}\left[Y^2\right]}}. \tag{230}$$

Furthermore, note that $\mathbb{E}[Y] = \frac{1}{V}$, which implies

$$\frac{\sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} = \sqrt{\frac{\mathrm{Var}\left[Y\right]}{\mathbb{E}\left[Y^2\right]}} \leq 1. \tag{231}$$

Combining the latter equation with (229) completes the proof. ☐