Quantum Algorithms for Interpolation and Sampling

by

Arsalan Motamedi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics (Quantum Information)

Waterloo, Ontario, Canada

Author's Declaration

This thesis consists of material all of which I co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis is based on an arXiv post, which is co-authored by Dr. Ronagh: https://arxiv.org/abs/2210.08104. In this work, AM performed the technical analysis and obtained the results. PR conceived the project centered around Gibbs sampling, and provided an innovative idea that laid the foundation for this research.

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. Despite suffering from an exponentially long mixing time, this algorithm allows for exponentially improved precision in sampling, and polynomial quantum speedups in mean estimation in the general case, and particularly under geometric conditions we identify for the critical points of the energy function.

Acknowledgements

I thank my supervisors Dr. Pooya Ronagh and Prof. Raymond Laflamme for their academic guidance, research advise, and suggestions. I also thank my committee members Prof. David Gosset and Prof. Roger Melko for the interesting questions and future works they suggested.

I would also like to thank my partner, Mahsa, for all her love and affection. I also thank my friends Shlok, Amit, Sanchit, Amolak, Pulkit, Alev, Einar, Yash, Matt, Kishor, and Devashish, for the fun and laughter we had together. I should also express my gratitude to my parents, Mehrdad and Arezu, and family in Canada, Marjan, Amir, Elena, and Elara, for all their unconditional support.

Dedication

To my brother, Saman.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CDF cumulative distribution function 2, 37, 38

EBM Energy-based Models 1, 6, 75

FPE Fokker–Planck Equation 1, 4, 52, 57, 58, 69

iid indpendent and identically distributed 38, 46, 75

MALA Metropolis adjusted Langevin algorithm 46, 48, 49

ODE Ordinary Differential Equation 49

PDE Partial Differential Equation 1, 40, 52, 57

SDE Stochastic Differential Equation 1, 43–45, 58

Chapter 1

Introduction

In recent advances in machine learning (ML), a reincarnation of EBMs has provided state-of-the-art performance in generative modeling [GWJ+19, SK21, SSDK+20, HJA20]. Unlike the traditional EBMs such as Boltzmann machines and Hopfield neural networks [AHS85, Hop82, HOT06, Hin07] these models require Gibbs sampling from continuous real-valued functions parameterized by large deep neural networks. However, the training of these models is extremely difficult and numerically unstable despite using state-of-the-art ML accelerators such as graphical and tensor processing and streaming units.

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

$$p_\theta(x) = \exp(-E_\theta(x))/Z_\theta \tag{1.1}$$

of an associated d-dimensional *energy potential*, $E_\theta : \mathbb{R}^d \to \mathbb{R}$. Here θ denotes a vector of model parameters $\theta \in \mathbb{R}^m$ and the normalizing constant $Z_\theta = \int e^{-E_\theta(x)}\, dx$ is the partition function of p_θ. This is done using Monte Carlo integration of (the overdamped) Langevin dynamics [DM19, NHH+20]–a Stochastic Differential Equation (SDE) governing diffusion processes, which is extremely costly and numerically unstable [NHZW19]. Despite the limitations imposed by this computational bottleneck, EBMs have been shown to provide improved representations of classical data. For example, [GWJ+19, HMZ20] 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 work, we propose a quantum algorithm for Gibbs sampling from a periodic continuous potential. We use finite difference techniques for solving differential equations on quantum computers [BCOW17, CL20, CLO21, LKK+21] to solve the 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 [Eva23]. 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 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.

1.1 Notation

In this thesis, we say a function $f : \mathbb{R}^d \to \mathbb{R}$ is periodic with a period of length ℓ if it satisfies the condition $f(x + \ell e_i) = f(x)$ for all $i \in [d]$, where $(e_i)_{i=1}^d$ represents the standard basis for $\mathbb{R}^d$. Moreover, we commonly refer to such a function as an ℓ-periodic function. The Schatten norm, denoted as $\|v\|_p$, characterizes the ℓ_p norm of a vector v. Specifically, when $p = 2$, the subscript is omitted, resulting in the notation $\|v\| \equiv \|v\|_2$.. The operator norm of an operator A is denoted by

$$\|A\| = \sup_v \frac{\|Av\|}{\|v\|}.$$

For a square-integrable function, say $f \in L^2(\mathbb{R}^d)$, we denote its norm by

$$\|f\| = \sqrt{\int_{x\in\mathbb{R}^d} |f(x)|^2 \mathrm{d}x}.$$

For an ℓ-periodic function, say f, we define

$$\|f\| = \sqrt{\int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} |f(x)|^2 \mathrm{d}x}.$$

For $\ell > 0$, and $N \in \mathbb{Z}_{\geq 0}$, we let $[-N..N] := \{-N, -N+1, \cdots, N\}$, and

$$V_N^{(\ell)} := \left\{ \left(\frac{\ell n_i}{2N+1} \right)_{i=1}^d : (n_i)_{i=1}^d \in [-N..N]^d \right\} \subset \left(-\frac{\ell}{2}, \frac{\ell}{2} \right)^d.$$

We often omit superscript ℓ as it is clear from the context. Furthermore, we denote the Hilbert space $\mathbb{C}^{V_N}$ (i.e., the space of functions from V_N to $\mathbb{C}$) by $\mathcal{V}_N$. Given a tuple of indices $m = (m_1, \cdots, m_d) \in \{0, \ldots, 2N\}^{\times d}$, we denote the associated computational basis state in the Hilbert space $\mathcal{V}_N \cong \left(\mathbb{C}^{2N+1}\right)^{\otimes d}$ by $|m\rangle$. For an ℓ-periodic function u, we denote by $\overrightarrow{u_N} \in \mathcal{V}_N$, the discretization of it on the lattice $V_N^{(\ell)}$; i.e., $u_N[n] = u(\frac{\ell n}{2N+1})$ for all $n \in [-N..N]^d$. Moreover, we denote the unit vector parallel to $\overrightarrow{u_N}$ by $|u_N\rangle$. Thus

$$|u_N\rangle \propto \sum_{m\in\{0,\ldots,2N\}^{\times d}} u(x_{m-N}) \, |m_1\rangle \otimes |m_2\rangle \otimes \cdots \otimes |m_d\rangle \tag{1.2}$$

where $x_n = \frac{\ell n}{2N+1}$ (note that our indices n and m only differ by a shift).

We use the notation $\mathbb{S}^{d-1}$ to denote the unit ball in d-dimensional Euclidean domain i.e., $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$. Furthermore, we use $\mathbb{B}^d$ to denote the vectors of length less than or equal to 1; i.e., $\mathbb{B}^d := \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.

We deploy the usual notation for probability spaces $(\Omega, \Sigma, \mathbb{P})$, where Ω is the sample space, Σ is the σ-algebra, and $\mathbb{P}$ (or sometimes μ) is the probability measure. If $\Omega = \mathbb{R}^d$,

we implicitly use the Borrel sigma algebra; i.e., $\Sigma = \mathcal{B}(\mathbb{R}^d)$. Moreover, we use F to denote the cumulative distribution function (CDF), respectively. For a random variable X, we use $\mathbb{E}[X]$ and $\mathrm{Var}[X]$ to denote its expectation value and variance, respectively. For a bounded subset $K \subset \mathbb{R}^d$, we denote by $\mathrm{Unif}(K)$ the uniform distribution over K i.e.,

$$\mathrm{Unif}(K)(\mathrm{d}x) = \begin{cases} \frac{\mathrm{d}x}{\mathrm{Vol}(K)} & \text{if } x \in K \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathrm{Vol}(K)$ is the volume of K.

Also, 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:

$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}. \tag{1.3}$$

1.2 Summary of results

In this section, we present a concise overview of our main contributions and provide a comparison to the current state of the art.

1.2.1 Fourier interpolation

Definition 1.2.1. *Let the function $u : \mathbb{R}^d \to \mathbb{R}$ be ℓ-periodic along all axes, and moreover, let $X \sim \mathit{Unif}\left([-\ell/2, \ell/2)^d\right)$ 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*

$$\left(\frac{\ell}{2\pi}\right)^m \sqrt{\mathbb{E}\left(\sum_{\alpha:|\alpha|=m} |D^\alpha u(X)|^2\right)} \leq C\, a^m\, m!. \tag{1.4}$$

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

The Fourier transform of u

$$u(x) = \sum_{k \in \mathbb{Z}^d} \widehat{u}[k]\, e^{i\frac{2\pi\langle k,x\rangle}{\ell}} \tag{1.5}$$

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 2.2.2 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 2.2.1). Every periodic real-analytic function is semi-analytic (Proposition 2.2.1). We also show how semi-analyticity parameters change through basic operations like addition, multiplication,

and composition of functions (Proposition 2.2.7). In Corollary 2.2.5, 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. 1.1b 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, [RC22] has 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 1.2.1 (Main interpolation result)**.** *Given an L-Lipschitz (C, a)-semi-analytic periodic function u, an integer $N \geq 2ad$, and a quantum state $|\psi\rangle \in \mathcal{V}_N$, such that $\||\psi\rangle - |u_N\rangle\| \leq \delta$, 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 $\mathcal{U} = \sqrt{\mathbb{E}\,u^2(X)}$.

In Example 2.2.3 we show a family of functions that help us construct adversary witnesses that determine how coarse the discretization of our functions are allowed to be for upsampling to achieve arbitrarily small errors.

Theorem 1.2.2. *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 2.2.5). This is instrumental in constructing high precision approximations to the generator $\mathcal{L}$ of the FPE.

1.2.2 Gibbs sampling

Technical setting 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

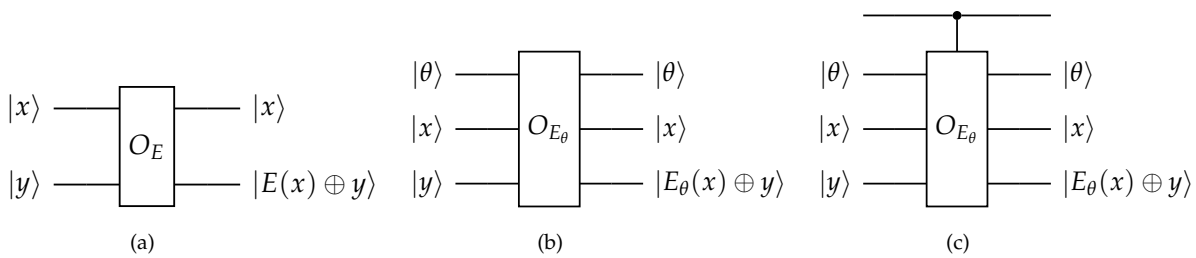

Figure 1.1: Schematics of the circuits of quantum oracles. All registers receive float-point representations of real numbers. (a) The oracle for an energy function E. The first register receives a data sample $x \in \mathbb{R}^d$. And, the second register is used to evaluate the energy function. (b) For a parameterized family of energy functions, an additional register may receive the model parameters $\theta \in \mathbb{R}^m$. (c) A controlled variant of the same oracle, controlled on a single qubit represented by the top wire.

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. Additionally, we use a constant thermodynamic β of 1 throughout (otherwise $\beta\Delta$ can be thought of as a single parameter). We show that the Fourier coefficients of f decay sub-exponentially away from the origin assuming a bound on the growth of the derivatives of f (Definition 2.2.1). 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 densities of a probability measure defined on this lattice, then semi-analyticity is equivalent to the concentration of this measure (Theorem 2.2.2).

Many periodic functions are semi-analytic. For example any function with finitely many non-zero Fourier coefficients is semi-analytic (Example 2.2.1). 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 **??**). 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 2.2.7). 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 2.2.5). 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.

Given a classical construction of the energy function (e.g., using the weights and biases of a deep neural network) $E : \mathbb{R}^d \to \mathbb{R}$, one can construct a quantum circuit

$$|x\rangle\,|y\rangle \mapsto |x\rangle\,|y \oplus E(x)\rangle \tag{1.6}$$

realizing such a function. A schematic of this circuit is provided in Fig. 1.1a. From the weights and biases $\theta \in \mathbb{R}^m$ of a classical deep neural network a similar oracle for the associated energy function

$$|\theta\rangle\,|x\rangle\,|y\rangle \mapsto |\theta\rangle\,|x\rangle\,|y \oplus E(x)\rangle \tag{1.7}$$

can be constructed if the circuit is augmented with an additional model parameter

Table 1.1: Summary of the query complexities of some of the classical and quantum algorithms for sampling from a d-dimensional Gibbs distribution. 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 3.2.1). 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 3.2.1). 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. The prior results (rows five to nine) all require convexity assumptions on the potential. Here Δ_f is the diameter of the range of values a function f attains.

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	$\widetilde{O}\left(\kappa_{E/2} e^{\Delta/2} d^7 \Delta_f \varepsilon^{-1}\right)$
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	$\widetilde{O}\left(\lambda^{-2} e^{\Delta/2} d^7\right)$	TV	$\widetilde{O}\left(\lambda^{-2} e^{\Delta/2} d^7 \Delta_f \varepsilon^{-1}\right)$
Cl. RLA [LE20]	Morse and periodic	first	$\widetilde{O}\left(\lambda^{-4} L^4 d^3 \varepsilon^{-2}\right)$	TV	$\widetilde{O}\left(\lambda^{-4} L^4 d^3 \Delta_f^2 \varepsilon^{-4}\right)$
Cl. MRW [DCWY18]	convex	first	$\widetilde{O}\left(L^2 d^3 \varepsilon^{-2}\right)$	TV	$\widetilde{O}\left(L^2 d^3 \Delta_f^2 \varepsilon^{-4}\right)$
Q. ULD [CLL+22]	strongly convex	zeroeth	$\widetilde{O}\left(\mu^{-2} L^2 d^{1/2} \varepsilon^{-1}\right)$	W_2	–
Cl. ULD [CCBJ17]	strongly convex	first	$\widetilde{O}\left(\mu^{-2} L^2 d^{1/2} \varepsilon^{-1}\right)$	W_2	–
Q. MALA [CLL+22]	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 [LST20]	strongly convex	first	$\widetilde{O}\left(\mu^{-1} L d\right)$	TV	$\widetilde{O}\left(\mu^{-1} L d \Delta_f^2 \varepsilon^{-2}\right)$

register as in Fig. 1.1b. It can also be shown that such a construction has only polylogarithmic overhead compared to the gate complexity of the corresponding classical boolean circuit [NC02].

The above setting is sufficient for sampling from the Gibbs distribution. 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. 1.1c) of the unitary in Fig. 1.1b and its inverse.

Contributions 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 1.2.1). We also provide a lower bound result in Theorem 1.2.2 showing the optimality of our discretization. We use these results to show that $\widetilde{O}\left(\kappa_{E/2} e^{\Delta/2} d^7\right)$ queries to the oracle O_E suffices for Gibbs

sampling with an approximation error ε in total variation distance (TV) (Theorem D.1.1). The exponentially poor dependence on Δ indicates the unsurprising difficulty of low-temperature Gibbs sampling.

Table 1.1 provides a comparison between our algorithm and previous results. 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. More rigorously, the mixing time of Langevin dynamics can be associated with the isoperimetric properties of the Gibbs measure [KHR22]. For example the mixing time is $\widetilde{O}(\kappa_E)$ where κ_E is the *Poincaré constant* of the potential E. We prove this fact for periodic functions in Proposition B.0.1 following the treatment of Euclidean domains in [MV00]. In Proposition B.0.2 we show that the Poincaré constant of periodic functions is bounded by $O(e^{\Delta})$. This bound can be interpreted as a Eyring-Kramers law for periodic functions [Ber11].

In our case the potentials are generic non-convex periodic functions and the relevant Poincaré constant is that of the function $E/2$ due to Born rule (see Section 3.2.1). Therefore, our complexity factor $\kappa_{E/2}e^{\Delta/2} = e^{\Delta}$ in the first row of Table 1.1 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 3.2.1). The classical counterpart for this result is [LE20, 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 [CLL+22] are also included in Table 1.1. We note that all these algorithms 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 [BGL+14].

Finally, for estimating means of random variables of the Gibbs state, given white-box access to both the energy and random variable oracles, we achieve a quadratic advantage for generic periodic functions and a quartic advantage in the case of Morse functions with unique global minima (Corollary 3.2.2) as reported in the last column of Table 1.1.

Related works To the best of our knowledge, this work and the independent paper [CLL+22] are the first efforts to analyze quantum algorithms for the problem of Gibbs sampling from a continuous real-valued function. [CLL+22] achieves a quadratic speedup in expediting a Monte Carlo simulation of Langevin dynamics using quantum random walks but it is restricted to strongly convex potentials. Similarly, classical algorithms that achieve high precision Gibbs sampling [RT96, DCWY18, CEL+21] also make assumptions about convexity or satisfaction of isoperimetric inequalities. See Section 3.1.1 for an overview of these algorithms. 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 in [CLL+22] assumes accurate access to the gradients of the potential whereas we only perform zeroeth order queries to the potential. Finally, we note that prior quantum algorithms such as [TD00, PW09, CS16, vAGGdW17] 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 $\sqrt{\exp(d)}$.

Chapter 2

Interpolation

Music is the space between the notes.

Claude Debussy

The problem of interpolation is a fundamental problem with applications spanning various fields, including statistics, signal processing, and cryptography. At its core, the problem involves determining a function given a set of function values at distinct points. However, the problem is inherently unsolvable without additional information about the underlying structure of the function. To address this challenge, one common assumption is that the function has a bounded degree. For example, when dealing with a degree d polynomial $f : \mathbb{R} \to \mathbb{R}$, the Lagrange interpolation technique can be employed. This technique allows us to approximate the function f using the formula:

$$f(x) = \sum_{i=1}^{d+1} f(x_i) \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}.$$

Here, the number of observation points (or queries to the function f) is $d + 1$. Indeed, it is impossible to determine f with d queries. One may contemplate a scenario in which the function is defined over a finite field $\mathbb{F}_q$. In this case, making d queries does not help to find the coefficients even with a bounded probability of error. This is the basis for a classical secret-sharing scheme proposed by [Sha79]. The scheme involves dividing a secret among $d + 1$ parties in such a way that any subset of d parties cannot recover the secret individually, but when all $d + 1$ parties collaborate, they can reconstruct the original secret. However, [CvDHS15] shows that $\lceil \frac{d+1}{2} \rceil$ queries suffice to determine f with bounded error. This query complexity is known to be tight [KK09, MP11].

In this chapter, we focus on the interpolation of functions with different structures. Specifically, we explore the interpolation of periodic functions and squared integrable functions. These types of functions find relevance in various domains, such as the field of signal processing [OS75]. This can be viewed as a generalization of [GR02], and hence, be used in a sampling algorithm. We indeed make use of this interpolation method in the next chapter.

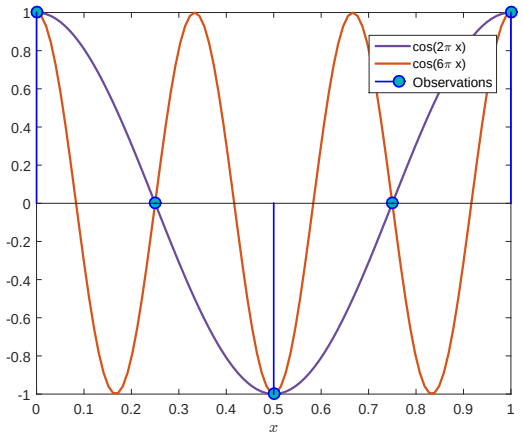

Figure 2.1: An example of interpolation for periodic functions. We consider two functions $u = \cos(2\pi x)$ and $v = \cos(6\pi x)$, and the sampling points $S = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ (note that 1 and 0 can be treated as the same point as the functions are 1-periodic). As the maximum k_0 for our functions u, v is 3, we need at least $2 \times 3 + 1 = 7$ points to interpolate each. These two functions are indistinguishable with our 4 observation (query) points $x \in S$.

2.1 Interpolation of periodic functions

In this section, we focus on the interpolation of periodic functions. As mentioned earlier, solving this interpolation problem requires the identification of underlying structural patterns within the function. Remarkably, the diameter of the support of the Fourier transform of a function plays a role analogous to the degree in polynomial interpolation problems. To elucidate this concept further, let us consider a periodic function $f : \mathbb{R} \to \mathbb{R}$ with a period of 1. Instead of employing Taylor expansions, we examine its Fourier series representation given by $f(x) = \sum_{-k_0 \le k \le k_0} \widehat{f}[k] e^{ikx}$. Notably, a classical algorithm exists that allows the recovery of the underlying function f by making $2k_0 + 1$ queries to the function f, and moreover, this query complexity is tight [OS75]. See Fig. 2.1 for an illustration.

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 is 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

$$\begin{aligned} u(x) &= \sum_{k \in \mathbb{Z}^d} \widehat{u}[k]\, e^{i\frac{2\pi\langle k,x\rangle}{\ell}}, \quad \text{where} \\ \widehat{u}[k] &= \frac{1}{\ell^d} \int_{\mathbb{T}} u(x) e^{-i\frac{2\pi\langle k,x\rangle}{\ell}}\, \mathrm{d}x\,. \end{aligned} \tag{2.1}$$

Also, let $\widetilde{u}_N$ be the discrete Fourier transform of u_N. Thus

$$\widetilde{u}_N[k] = \frac{1}{(2N+1)^{d/2}} \sum_{n \in [-N..N]^d} u_N[n] e^{-\frac{i2\pi\langle k,n\rangle}{2N+1}}. \tag{2.2}$$

Replacing u_N by the series obtained from $\widehat{u}$ (defined in (2.1)) yields

$$\widetilde{u}_N[k] = (2N+1)^{d/2} \left(\widehat{u}[k] + \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \widehat{u}[k + p(2N+1)] \right). \tag{2.3}$$

Note that (2.3) shows $\widetilde{u}_N$ is a wrapped version of $\widehat{u}$, and therefore, if $k_0 = \max\{\|k\|_\infty : \widehat{u}[k] \neq 0\}$, then, choosing $N \geq k_0$ ensures that $\widetilde{u}_N[k] = \widehat{u}[k]$ for all $k \in \mathrm{Supp}(\widehat{u})$. Hence, in case u has a Fourier transform with finitely many non-zero elements, an exact recovery is possible with $2k_0 + 1$ queries, as claimed earlier.

The task of our interest, however, differs from the conventional interpolation problem. We are looking for an answer to the following question.

Given a quantum register $|\psi\rangle \in \mathcal{V}_N$ such that $|\psi\rangle \approx |u_N\rangle$, how well can we generate samples from the continuous distribution determined by $|u|^2/\|u\|^2$?

Here is one motivation for this particular task. Many quantum algorithms have been developed to solve differential equations. In such algorithms, by 'solving' the problem one means preparing a state that is close to a discretization of the actual solution on a coarse lattice (e.g., see [CLO21]). But is it sufficient to have such values only on a coarse lattice? This is one motivation for this work. Notably, the algorithm of Section 3.2.1 also prepares a solution on a course lattice, which is then interpolated using the methodology and results of this section.

Let us now provide a concise explanation of how we perform such an interpolation. Our approach bears resemblance to the concept of 'upsampling' commonly employed in signal processing [OS75]. Initially, we construct an isometry denoted by W, which allows us to have an approximation of $|u_M\rangle$ with $M \gg N$. Concretely, we get $W^{\otimes d}|\psi\rangle \approx |u_M\rangle$. Subsequently, we measure the resulting state in the computational basis, resulting in a sample $x \in V_M$. From there, we proceed to uniformly sample a point within the box $\prod_{i=1}^d [x_i - \frac{\ell}{4N+2}, x_i + \frac{\ell}{4N+2}]$. This box encompasses the points that are closer to x than any other point of the discrete lattice V_M. We show that the isometry W exists (if N is large enough), and that we can implement it efficiently using 2-qubit gates.

Proposition 2.1.1. *Let u be an ℓ-periodic function, whose Fourier transform has a bounded support. Let $k_0 = \max\{\|k\|_\infty : \widehat{u}[k] \neq 0\}$, and $N \geq 2k_0 + 1$. There exists an isometry W that can be prepared by the application of $O(\log M \log\log M)$ many 2-qubit gates, such that*

$$W^{\otimes d}|u_N\rangle = |u_M\rangle \tag{2.4}$$

for any $M \geq N$.

Proof. W applies a quantum Fourier transform, then adds ancillary qubits, and then performs an inverse quantum Fourier transform on the larger space. In particular, it maps $(u_N[n])_{n=-N}^N$ to $(u_M[m])_{m=-M}^M$ via $u_M[m] = \sum_{n=-N}^N W_{mn} u_N[n]$. Here the elements of W are

$$W_{mn} = \frac{1}{\sqrt{(2N+1)(2M+1)}} \sum_{k=-N}^{N} e^{i2\pi k\left(\frac{m}{2M+1} - \frac{n}{2N+1}\right)} \tag{2.5}$$

$$= \frac{(-1)^n \sin\left(\pi m \frac{N+\frac{1}{2}}{M+\frac{1}{2}}\right)}{\sqrt{(2N+1)(2M+1)}\sin\left(\pi\left(\frac{m}{2M+1} - \frac{n}{2N+1}\right)\right)}. \tag{2.6}$$

The complexity of the implementation of W is dominated by the complexity of the quantum Fourier transform over M, which is $O(\log M \log\log M)$. □

Corollary 2.1.1. *Let u be an ℓ-periodic function, whose Fourier transform has a bounded support. Let $k_0 = \max\{\|k\|_\infty : \widehat{u}[k] \neq 0\}$, and $N \geq 2k_0 + 1$. Then, there exists an algorithm generating samples from a distribution ε-close to $|u|^2/\|u\|^2$, with the gate complexity $O(d \log \frac{L\ell d}{2\|u\|\varepsilon} \log\log \frac{L\ell d}{2\|u\|\varepsilon})$.*

Proof. Note that $\|u_N\| = (2N+1)^{d/2}\|u\|$ for all $N \geq 2k_0 + 1$. This together with the analysis in Lemma A.0.8 shows that $M = \lceil \frac{L\ell d}{2\|u\|\varepsilon} \rceil$ suffices to achieve an error less than ε. Hence, the result follows from Proposition 2.1.1. □

It is important to highlight that the findings in signal processing directly imply that if the support of $\widehat{u}$ is bounded and $N \geq k_0$, then there exists an isometry W such that $W^{\otimes d}|u_N\rangle = |u_M\rangle$. However, in order to generate samples from the exact distribution determined by the density $|u|^2/\|u\|^2$, we must take the limit $M \to \infty$. This is infeasible in practice due to the requirement of infinite memory and time resources. Nevertheless, as we demonstrate later in Section 2.2, it is possible to generate samples from a distribution that is ε-close in terms of total variation distance to the desired density using a circuit composed of $O(d\,\mathrm{polylog}(\frac{1}{\varepsilon})$ number of 2-qubit gates. Consequently, a pertinent question arises: under what conditions can we achieve interpolation with a similar complexity?

We demonstrate that it is possible to attain the same level of precision by relaxing certain conditions. Specifically, the condition of *bounded Fourier spectrum* can be replaced by *semi-analytity* (a weaker condition than analyticity), and the condition of $N \geq k_0$ can be replaced by $N \geq 2ad$, where a can be interpreted as the (mean) inverse convergence radius of the function over one period. We introduce semi-analyticity and prove the complexity results in Section 2.2. Moreover, we show impossibility results in Section 2.2.1. Specifically, we show that given $|u_N\rangle$ with $N \leq ca$, for a constant $c > 0$, there is no algorithm that outputs samples from the desired density within the total variation distance of $\frac{1}{5000}$. In Section 2.2.3 we show how analyticity parameters of our interest change under multiplication, addition, and composition. In Section 2.3 we extend our results to non-periodic functions that are in $L^2(\mathbb{R}^d)$.

2.2 Semi-Analyticity (for periodic functions)

> The beauty of mathematics only shows itself to more patient followers.
>
> Maryam Mirzakhani

In this section, we analyze the effect of discretization on estimating distributions and derivatives of differentiable functions. Previous works (e.g., [CLO21]) 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 [STW11, 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:

$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}. \tag{2.7}$$

Definition 2.2.1. *Let $u : \mathbb{R}^d \to \mathbb{R}$ be ℓ-periodic along all axes, and moreover, let $X \sim Unif\left([-\ell/2, \ell/2)^d\right)$ 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*

$$\left(\frac{\ell}{2\pi}\right)^m \sqrt{\mathbb{E}\left(\sum_{\alpha:|\alpha|=m} |D^\alpha u(X)|^2\right)} \leq C\, a^m\, m! \tag{2.8}$$

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 multivariate 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

$$\begin{aligned} u(x) &= \sum_{k\in\mathbb{Z}^d} \widehat{u}[k]\, e^{i\frac{2\pi\langle k,x\rangle}{\ell}}, \quad \text{where} \\ \widehat{u}[k] &= \frac{1}{\ell^d}\int_{\mathbb{T}} u(x) e^{-i\frac{2\pi\langle k,x\rangle}{\ell}}\, dx\,. \end{aligned} \tag{2.9}$$

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 2.2.1 is equivalent to

$$|u|_m := \sqrt{\sum_{k\in\mathbb{Z}^d} \|k\|^{2m}\,|\widehat{u}[k]|^2} \le C\, a^m\, m!. \tag{2.10}$$

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 2.2.1 we prove that every analytic function is semi-analytic (hence, the naming).

Example 2.2.1. *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 2.2.2. *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 [AS64, page 376]:*

$$\widehat{u}[k] = \frac{1}{2\pi}\int_{x=-\pi}^{\pi} e^{-ikx}\, e^{z\cos(x)}\, dx = I_k(z). \tag{2.11}$$

Furthermore,

$$I_k(z) = \left(\frac{z}{2}\right)^k \sum_{\ell\ge 0} \frac{\left(\frac{z}{2}\right)^{2\ell}}{\ell!(k+\ell)!} \le \left(\frac{z}{2}\right)^k \frac{1}{k!}\sum_{\ell\ge 0}\frac{\left(\frac{z}{2}\right)^{2\ell}}{(\ell!)^2} = \frac{1}{k!}\left(\frac{z}{2}\right)^k I_0(z). \tag{2.12}$$

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

$$\sum_{k\in\mathbb{Z}} |k|^{2m}\,|\widehat{u}[k]|^2 \le (I_0(z))^2 \sum_{k\in\mathbb{Z}} \left(\frac{z}{2}\right)^{2k} \frac{k^{2m}}{(k!)^2}. \tag{2.13}$$

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

$$|u|_m \leq I_0(z) \sum_{k\in\mathbb{Z}} \left(\frac{z}{2}\right)^k \frac{k^m}{k!} \leq I_0(z)\, e^{z/2} \left(\max\{\frac{z}{2}, 1\}\right)^m m! \tag{2.14}$$

where the last inequality follows from Lemma A.0.1.

Proposition 2.2.1. *Every periodic real-analytic function is semi-analytic.*

Proof. Let f be a real analytic and periodic function. It follows from [Kom60, Lemma 1] that there exist $C, a > 0$ such that

$$\sup_{x\in\mathbb{T}} |D^\alpha f(x)| \leq C\, a^{|\alpha|}\, \alpha!. \tag{2.15}$$

From this we conclude that

$$\sqrt{\sum_{\alpha:|\alpha|=m} \mathbb{E}\left[(D^\alpha f)^2\right]} \leq C\, a^m \sqrt{\sum_{\alpha:|\alpha|=m} (\alpha!)^2} \leq C\, a^m \sum_{\alpha:|\alpha|=m} \alpha! \leq 3^{d-1} C\, a^m\, m! \tag{2.16}$$

where the last inequality follows from Lemma A.0.4. □

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 [OS75]. 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 2.2.1 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 2.2.1. *For a (C,a)-semi-analytic periodic function u, with N an integer satisfying $N \geq 2a$, we have*

$$\sqrt{\sum_{k:\|k\|_2\geq N} |\widehat{u}[k]|^2} \leq 2e^3\, C\, e^{-\frac{N}{a}\left(1-\frac{1}{2e}\right)}. \tag{2.17}$$

Proof. We have

$$\sum_{k:\|k\|_2\geq N} |\widehat{u}[k]|^2 \leq N^{-2m} \sum_{k:\|k\|_2\geq N} \|k\|^{2m} |\widehat{u}[k]|^2 \leq N^{-2m}\, C^2\, a^{2m}\, (m!)^2\,. \tag{2.18}$$

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

$$N^{-m}\, a^m\, m! \leq \frac{e^2}{a} N^{-m} \left(\frac{a(m+1)}{e}\right)^{m+1}. \tag{2.19}$$

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

$$N^{-m}\, a^m\, m! \leq \frac{e^3}{a} N\, e^{-N/a}. \tag{2.20}$$

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. □

The reader may notice analogies between the result of Lemma 2.2.1 and the sub-exponential decay bounds in the literature of concentration of measure. We discuss this connection in Section 2.2.1.

Lemma 2.2.2. *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*

$$d\left(\frac{1}{(2N+1)^{d/2}}\, \widetilde{u}_N, \widehat{u}\right) \leq 2\sqrt{2} e^3\, C e^{-\frac{3N}{5a}}. \tag{2.21}$$

Proof. We start by noting that

$$\begin{aligned} \frac{1}{(2N+1)^{d/2}}\, \widetilde{u}_N[k] &= \frac{1}{(2N+1)^d} \sum_{n \in [-N..N]^d} u_N[n]\, e^{-i\frac{2\pi\langle k,n\rangle}{2N+1}} \\ &= \frac{1}{(2N+1)^d} \sum_{n \in [-N..N]^d} \sum_{k' \in \mathbb{Z}^d} \widehat{u}[k']\, e^{-i\frac{2\pi\langle k-k',n\rangle}{2N+1}} \\ &= \widehat{u}[k] + \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \widehat{u}[k + (2N+1)p]. \end{aligned} \tag{2.22}$$

Therefore,

$$d\left(\frac{1}{(2N+1)^{d/2}}\, \widetilde{u}_N, \widehat{u}\right)^2 = \sum_{k \in [-N..N]^d} \left| \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \widehat{u}[k + (2N+1)p] \right|^2 + \sum_{k \in \mathbb{Z}^d \setminus [-N..N]^d} |\widehat{u}[k]|^2 \tag{2.23}$$

where the second term was upper-bounded in Lemma 2.2.1. As for the first term

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

We note that

$$\max_{k\in[-N..N]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+(2N+1)p\|^{-2m} \leq \max_{k\in[-N..N]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+2N\,p\|^{-2m} \leq N^{-2m}\max_{x\in[-1,1]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|x+2\,p\|^{-2m}. \tag{2.24}$$

Using Lemma A.0.5, if $m \geq d$, we get

$$d\left(\frac{1}{(2N+1)^{d/2}}\,\widetilde{u}_N,\widehat{u}\right) \leq 2\sqrt{2}a^m C\,N^{-m}\,2^{d/2}\,m! \leq 2\sqrt{2}a^m\,C\,N^{-m}\,2^{n/4a}\,m! \tag{2.25}$$

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

$$d\left(\frac{1}{(2N+1)^{d/2}}\,\widetilde{u}_N,\widehat{u}\right) \leq 2e^3\sqrt{2}\,Ce^{-\frac{n}{a}\left(1-\frac{1}{2e}-\frac{\ln(2)}{4}\right)} \tag{2.26}$$

which concludes the proof. □

For our purposes, we will be applying Lemma 2.2.2 to normalized vectors. Here we highlight the following distance bound as a corollary following immediately from Lemma 2.2.2 and Lemma A.0.6.

Corollary 2.2.1. *Let $N \geq 2ad$ and u be a (C,a)-semi-analytic periodic function. We have*

$$d\left(F_N\,|u_N\rangle\,,\frac{\widehat{u}}{\mathcal{U}}\right) \leq \frac{4\sqrt{2}e^3\,C}{\mathcal{U}}e^{-0.6N/a} \tag{2.27}$$

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

Now we show that upsampling a semi-analytic function is useful in achieving minor aliasing effects.[1] Recall the sampling procedure introduced in Section 3.2.1. 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 $|\psi\rangle$.

[1] 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 [OS75].

Remark 2.2.1. *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*

$$\mu_\psi(x) = \sum_{n\in[-N..N]^d} \mathbf{1}_{\{x\in\mathbb{B}_n\}} |\psi[n]|^2 \left(\frac{2N+1}{\ell}\right)^d \tag{2.28}$$

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

$$\frac{1}{2}\int dx\ \left|\mu_\psi(x) - \mu_\phi(x)\right| = \frac{1}{2}\sum_{n\in[-N..N]^d} \left||\psi[n]|^2 - |\phi[n]|^2\right| \leq \||\psi\rangle - |\phi\rangle\| \tag{2.29}$$

where the inequality follows from Lemma A.0.7.

Proposition 2.2.2. *Given a (C,a)-semi-analytic periodic function u, an integer $N \geq 2ad$, and a quantum state $|\psi_N\rangle \in \mathcal{V}_N$ satisfying $\||\psi_N\rangle - |u_N\rangle\| \leq \delta$, 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*

$$\varepsilon \leq 2\delta + \frac{8\sqrt{2}e^3\, C}{\mathcal{U}} e^{-0.6\frac{N}{a}}. \tag{2.30}$$

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 $r \geq M^*$ gives δ-approximation of the continuous distribution. If $N \geq M^*$, then the statement is trivially satisfied after setting $M = N$. Otherwise, let $M = M^*$, and note that due to Corollary 2.2.1 and by an application of the triangle inequality

$$\|\iota F_N |u_N\rangle - F_M |u_M\rangle\| \leq \frac{8\sqrt{2}e^3\, C}{\mathcal{U}} e^{-0.6\frac{N}{a}}. \tag{2.31}$$

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

$$\left\|F_M^{-1}\iota F_N\ |\psi_N\rangle - |u_M\rangle\right\| \leq \delta + \frac{8\sqrt{2}e^3\, C}{\mathcal{U}} e^{-0.6\frac{N}{a}}. \tag{2.32}$$

Finally, using another triangle inequality for the total-variation distance and Remark 2.2.1 we obtain the result. □

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

Theorem 2.2.1. *Given an L-Lipschitz (C,a)-semi-analytic periodic function u, an integer $N \geq 2ad$, and a quantum state $|\psi\rangle \in \mathcal{V}_N$, such that $\||\psi\rangle - |u_N\rangle\| \leq \delta$, 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 $\mathcal{U} = \sqrt{\mathbb{E}\, u^2(X)}$.

Proof. In this proof, for an integer $\alpha \in \mathbb{N}$ the notation $[\alpha]$ stands for $\{0, 1, \ldots, \alpha\}$. Using Lemma A.0.8, we set $M = \left\lceil \frac{1}{2\varepsilon'} \frac{L\ell d/2 + 10/3\sqrt{2}\, a e^4 C}{\mathcal{U}} \right\rceil$ 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

$$F_N = \bigotimes_{i=1}^{d} f_N^{(i)} \tag{2.33}$$

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 $f_N = T_N \hat{f}_N T_N$, 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 $\mathcal{O}(d \log M)$. □

2.2.1 Connections to 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 [Ver18]. 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 2.2.2. *X is a Bernstein random variable, if $X \geq 0$ almost surely, and for some $A, b > 0$ its moments are upper bounded as*

$$\mathbb{E}\, X^m \leq A\, b^m\, m!, \tag{2.34}$$

for all positive integers m.

Following [BLM13] we prove concentration bounds on a Bernstein random variable.

Lemma 2.2.3. *Let X be the Bernstein random variable defined in Definition 2.2.2. It is the case that*

$$\mathbb{P}[X \geq t] \leq \begin{cases} \max(A,1)\, e^{-\frac{(t-b)^2}{8b^2}}, & \text{if } t \leq 3b, \\ e\, \max(A,1)\, e^{-\frac{t}{2b}}, & \text{if } t > 3b. \end{cases} \tag{2.35}$$

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

$$\mathbb{E}\, e^{\lambda X} = 1 + \sum_{m \in \mathbb{N}} \frac{\lambda^m\, \mathbb{E}\, X^m}{m!} \leq 1 + A \sum_{m \in \mathbb{N}} (b\lambda)^m = 1 + A \frac{\lambda b}{1 - \lambda b}. \tag{2.36}$$

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

$$\mathbb{E}\, e^{\lambda X} \leq \max(A,1)\, \exp\Big\{\lambda b + 2\lambda^2 b^2\Big\}, \quad \forall \lambda \in [0, \frac{1}{2b}]. \tag{2.37}$$

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

$$\mathbb{P}[X \geq t] \leq \inf_{0 \leq \lambda} e^{-\lambda t}\, \mathbb{E}[e^{\lambda X}] \leq \max(A,1) \min_{0 \leq \lambda \leq \frac{1}{2b}} \exp\Big\{-\lambda(t - b - 2\lambda b^2)\Big\}. \tag{2.38}$$

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. (2.35). Indeed, the set of Bernstein random variables coincides with the set of positive sub-exponential distributions as stated bellow.

Proposition 2.2.3. *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 [Ver18, Proposition 2.7.1], according to which, the positive random variable X is sub-exponential if and only if $\mathbb{E}\, X^m \leq Q^m\, m^m$ for some $Q \geq 0$.[2] Now, let X be a sub-exponential distribution. Using $\frac{m^m}{e^{m-1}} \leq m!$, we have

$$\mathbb{E}\, X^m \leq Q^m\, m^m \leq e^{-1}(Qe)^m\, m!, \tag{2.39}$$

which concludes that X has a Bernstein property.

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

$$\mathbb{E}\, X^m \leq A b^m m! \leq (\max(A,1)\, b)^m\, m^m, \tag{2.40}$$

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

[2] In [Ver18] the exponent m is taken to be any real number larger than 1, but one could readily observe that $m \in \mathbb{N}$ is also a sufficient condition through the same proof provided in [Ver18].

Now we show the connection between the notions of concentration of measure and the semi-analyticity condition in Definition 2.2.1. The next definition allows us to make this connection clear.

Definition 2.2.3. *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 2.2.2. *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 (2.12), we have

$$\sqrt{\mathbb{E}\,\|K_u\|^{2m}} \leq \mathcal{U}^{-1}\, C\, a^m\, m!. \tag{2.41}$$

Putting this together with the Jensen inequality $\mathbb{E}\,\|K_u\|^m \leq \sqrt{\mathbb{E}\,\|K_u\|^{2m}}$, 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 $\mathbb{E}\,\|K_u\|^{2m} \leq A\, b^{2m}\, (2m)!$. Note that

$$(2m)! \leq \frac{(2m)^{2m+1}}{e^{2m-1}} \leq 2e\, 4^{2m}\, \frac{m^{2m-2}}{e^{2m-2}} \leq 2e\, 4^{2m} (m!)^2 \tag{2.42}$$

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

Making this connection allows us to obtain a Fourier concentration result similar to Lemma 2.2.1.

Corollary 2.2.2. *Let u be (C, a)-semi-analytic. It is the case that*

$$\sum_{k:\|k\|\geq t} |\hat{u}[k]|^2 \leq \begin{cases} \max(C, \mathcal{U})\, e^{-\frac{(t-a)^2}{8a^2}}, & \text{if } t \leq 3a, \\ e\, \max(C, \mathcal{U})\, e^{-\frac{t}{2a}}, & \text{if } t > 3a. \end{cases} \tag{2.43}$$

Proof. This follows directly from the first implication in Theorem 2.2.2 and Lemma 2.2.3. □

Note that the result of Lemma 2.2.1 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 2.2.3. *Let $z > 1$. The 2π-periodic function $u(x) = \frac{z-1}{1-2\sqrt{z}\cos(x)+z}$ satisfies the following inequality.*

$$\frac{1}{(1+z^{-1})^{1/2}}\frac{m!}{(1-z^{-1})^m} \leq \sqrt{\mathbb{E}\|K_u\|^{2m}} \leq \sqrt{\frac{2e}{1+z^{-1}}}\max\left(8, \frac{8}{z-1}\right)^m m! \tag{2.44}$$

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. 2.2 for visual demonstrations.

To obtain (2.44), note that the Fourier transform of $u(x)$ is $\widehat{u}[k] = z^{-\frac{|k|}{2}}$ since

$$1 + 2\sum_{k=1}^{\infty}\cos(kx)z^{-\frac{k}{2}} = \frac{z-1}{1+z-2\sqrt{z}\cos x}. \tag{2.45}$$

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

$$\frac{1+z^{-1}}{1-z^{-1}}\mathbb{E}\|K_u\|^m = \sum_{k=0}^{\infty} k^m z^{-k} \tag{2.46}$$

$$\geq \frac{\partial^m}{\partial(z^{-1})^m}\left(\sum_{k\geq 0} z^{-k}\right) = \frac{\partial^m}{\partial(z^{-1})^m}\left(\frac{1}{1-z^{-1}}\right) = \frac{m!}{(1-z^{-1})^{m+1}}. \tag{2.47}$$

We, therefore, have

$$\frac{1}{1+z^{-1}}\frac{m!}{(1-z^{-1})^m} \leq \mathbb{E}\|K_u\|^m \leq \frac{1}{1+z^{-1}}\max\{2, \frac{2}{z-1}\}^m m! \tag{2.48}$$

where the upper bound follows from Lemma A.0.2.

We now use another result from probability theory, namely the Paley–Zygmund lower bound [PZ32, Pet07], 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 2.2.4 (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*

$$\mathbb{P}\left[X > \theta\,\mathbb{E}[X]\right] > (1-\theta)^2\frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}. \tag{2.49}$$

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

$$\mathbb{E}[X] \leq \theta\,\mathbb{E}[X] + \sqrt{\mathbb{E}[X^2]\mathbb{P}\left[X > \theta\,\mathbb{E}[X]\right]}, \tag{2.50}$$

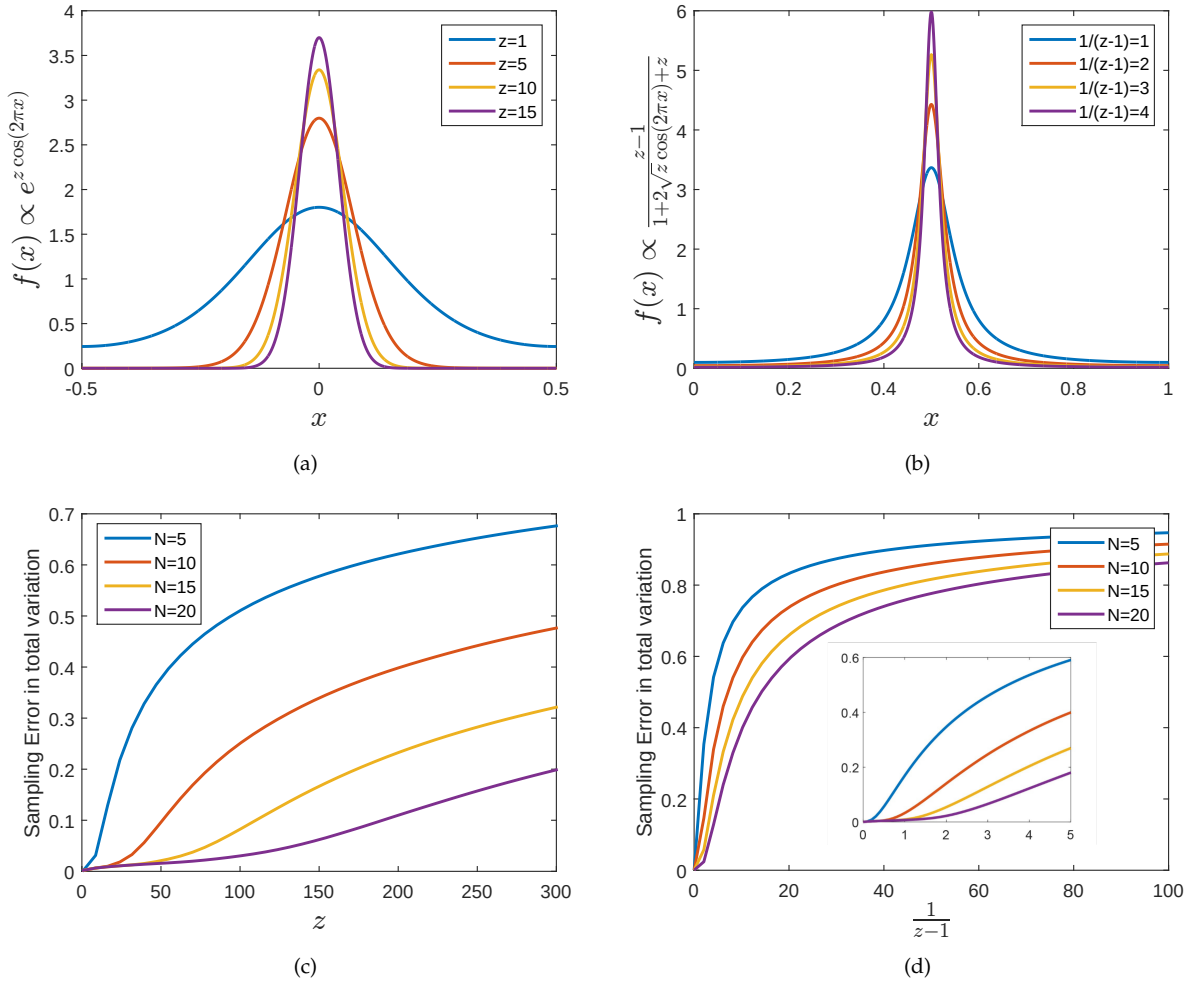

Figure 2.2: (a) and (b) show two families of functions considered respectively in Example 2.2.2 and Example 2.2.3. The functions are normalized so that $\int_{x\in[0,1]} \mathrm{d}x\ (f(x))^2 = 1$. That is, f^2 represents a distribution over one period. Note how in (a) the smoothness of the functions 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 2.2.2 and Example 2.2.3 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).

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

Theorem 2.2.3. *Let u be a (C,a)-semi-analytic function. Consider any exact discretization $|u_N\rangle$ on the discrete lattice with $N \le \theta a/16$, 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 $\mathbb{E}\left\|K_f\right\| \ge \frac{a}{16}$, therefore

$$\mathbb{P}\left[\|K_u\| > \theta\frac{a}{16}\right] > \mathbb{P}\left[\left\|K_f\right\| > \theta\,\mathbb{E}\left\|K_f\right\|\right] > (1-\theta)^2\frac{(a/16)^2}{2ea^2} = (1-\theta)^2\frac{1}{512e}. \quad (2.51)$$

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

$$\widehat{g}[k] = \alpha \begin{cases} \sum_{p\in\mathbb{Z}^d} \widehat{f}[k + p(2N+1)], & \text{if } k \in [-N..N]^d, \\ 0, & \text{otherwise.} \end{cases} \quad (2.52)$$

Here α is a normalization constant chosen such that $\mathbb{E}_{X\sim\mathrm{Unif}}[g(X)^2] = C^2$. 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 2.2.1 and that the total variation distance between the distributions whose densities are proportional to $|f|^2$ and $|g|^2$ is at least $\mathbb{P}\left[\left\|K_f\right\| > N\right]$, which is itself lower bounded by $(1-\theta)^2\frac{1}{512e}$ due to $N < \theta\frac{a}{16}$ and (2.51). 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. □

2.2.2 Fourier differentiation method

Assume we are given function values at a set of points $x_1 < x_2 < \cdots < x_n$. How can we estimate the derivatives at these points? By denoting the function as u, the most straightforward approach involves employing the following finite difference

$$u'(x_i) \approx (D_1 u)_i := \frac{u(x_i) - u(x_{i-1})}{x_i - x_{i-1}}. \quad (2.53)$$

One can show that the aforementioned method has an accuracy of $O(h)$, where $h = \max_i x_i - x_{i-1}$ is the largest spacing between consecutive points. If x_i's are structured nicely, one can have better estimators. For instance, in case x_is are equidistant, we can deploy a central approximation

$$(D_2 u)_i := \frac{u(x_{i+1}) - u(x_{i-1})}{x_{i+1} - x_{i-1}}. \quad (2.54)$$

One can show that D_2 provides a more accurate estimation than D_1, as its error is bounded by $O(h^2)$. Using more values of $u(x_j)$ for approximating the derivatives at

a point can improve the approximation accuracy. A survey of this topic is presented in [STW11]. Assuming that the function is periodic, one can approximate derivatives exponentially accurately if the points $\{x_i\}_{i=1}^n$ are equidistant. This is the topic of this section.

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:

$$\widetilde{\partial}_j u_N[n] := F_N^{-1}\left(\frac{i2\pi k_j}{\ell}\left(F_N\overrightarrow{u_N}\right)[k]\right). \tag{2.55}$$

A straightforward calculation yields the following convolution relation

$$\widetilde{\partial}_j u_N[n] = \sum_{m\in[-N..N]} u_N[n_1,\cdots,n_{j-1},m,n_{j+1},\cdots,n_d]\, a[n_j-m] \tag{2.56}$$

where

$$a[m] = \begin{cases} 0, & \text{if } m = 0, \\ \frac{\pi(-1)^{m+1}}{\ell \sin\left(\frac{\pi m}{2N+1}\right)}, & \text{otherwise.} \end{cases} \tag{2.57}$$

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

$$\begin{aligned} \widetilde{\partial}_j^r u_N[n] :=& F_N^{-1}\left[\left(\frac{i2\pi k_j}{\ell}\right)^r\left(F_N\overrightarrow{u_N}\right)[k]\right] \\ =& \sum_{m\in[-N..N]} u_N[n_1,\cdots,n_{j-1},m,n_{j+1},\cdots,n_d]\, a^{(r)}[n_j-m], \end{aligned} \tag{2.58}$$

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 [STW11]) in which case this composability property may not hold. Note that $\left\|\widetilde{\partial}_j u\right\|_2 \leq 2\pi N/\ell \|u\|_2$ 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 2.2.4. *Let u and v be two ℓ-periodic functions in all dimensions. The Fourier derivatives $\widetilde{\partial}_j$ have the following properties:*

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

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

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

4. $\widetilde{\partial}_j^{\,2}$ *is a symmetric operator.*

Proof. 1. It suffices to show that the Fourier transforms of the two sides coincide.

$$\begin{aligned}
(2N+1)^{d/2}\left\{F_N\left(\widetilde{\partial}_j(u\cdot v)\right)\right\}[k] &\overset{(1)}{=} \sum_{q\in[-N..N]^d}\frac{i2\pi k_j}{L}\,\widehat{u}[q]\,\widehat{v}[k-q]\\
&= \sum_{q\in[-N..N]^d}\frac{i2\pi(q_j+k_j-q_j)}{L}\,\widehat{u}[q]\,\widehat{v}[k-q]\\
&= \sum_{q\in[-N..N]^d}\left(\frac{i2\pi q_j}{L}\widehat{u}[q]\right)\widehat{v}[k-q]+\left(\frac{i2\pi(k_j-q_j)}{L}\widehat{v}[k-q]\right)\widehat{u}[q]\\
&\overset{(2)}{=} (2N+1)^{d/2}\left[F_N\left((\widetilde{\partial}_j u)\cdot v\right)+F_N\left((\widetilde{\partial}_j v)\cdot u\right)\right]
\end{aligned} \tag{2.59}$$

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 $(2N+1)^{d/2}$).

2. To show this, we make use of equation (2.56):

$$\begin{aligned}
\left|(\widetilde{\partial}_j u)[n]\right| &= \left|\sum_{m\in[-N..N]} u[n_1,n_2,\cdots,m,\cdots,n_d]\,a[n_j-m]\right|\\
&\overset{(1)}{\leq} \frac{2\pi}{\ell}\,\|u\|_\infty \sum_{m=1}^{N}\frac{1}{\sin\left(\frac{\pi m}{2N+1}\right)}\\
&\leq \frac{2\pi}{\ell}\,\|u\|_\infty\left(\int_{x=1}^{N}\frac{dx}{\sin\left(\frac{\pi x}{2N+1}\right)}+\frac{1}{\sin\left(\pi/(2N+1)\right)}\right)\\
&\overset{(2)}{\leq} \frac{2\pi}{\ell}\,\|u\|_\infty\left(-\frac{(2N+1)}{\pi}\ln\left(\tan\left(\frac{\pi}{4N+2}\right)\right)+\frac{2N+1}{2}\right)\\
&\overset{(3)}{\leq} \frac{2\pi}{\ell}\,\|u\|_\infty(2N+1)\left[\frac{1}{\pi}\ln\left(\frac{4N+2}{\pi}\right)+\frac{1}{2}\right]
\end{aligned} \tag{2.60}$$

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

$$\left|(\widetilde{\partial}_j u)[n]\right| \leq \frac{48}{\ell}\|u\|_\infty N\ln N. \tag{2.61}$$

3. Note that for any vector v defined on the discrete lattice, one has

$$\sum_{n\in[-N..N]^d} v[n] = \left(\sqrt{2N+1}\right)^d \hat{v}[0] \tag{2.62}$$

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.

4. (2.56) and (2.57) 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 [STW11]. Fig. 2.3 depicts an example of the Fourier interpolation and this derivative estimation method.

Proposition 2.2.5. *Let u be (C,a)-semi-analytic and periodic, and let $N \geq 4ad$. It is the case that*

$$\sqrt{\sum_{j=1}^{d}\left\|\overrightarrow{\partial_j u_N} - \overrightarrow{\widetilde{\partial}_j u_N}\right\|^2} \leq \frac{40\sqrt{2}\pi e^3\, a}{\ell}\, C\,(2N+1)^{d/2}\; e^{-\frac{N}{2a}}, \text{ and} \tag{2.63}$$

$$\left\|\overrightarrow{\nabla^2 u_N} - \overrightarrow{\widetilde{\nabla^2} u_N}\right\| \leq \frac{200\sqrt{2}\pi^2 e^3\, a^2}{\ell}\, C^2\,(2N+1)^{d/2}\; e^{-0.4\frac{N}{a}}. \tag{2.64}$$

Proof. We first prove (2.63). We have

$$\partial_j u[n] = \sum_{k\in\mathbb{Z}^d} \frac{i2\pi k_j}{\ell}\widehat{u}[k]\, e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \tag{2.65}$$

$$= \sum_{k\in[-N..N]^d} e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \sum_{p\in\mathbb{Z}^d} \frac{i2\pi(k_j + p_j(2N+1))}{\ell}\widehat{u}[k+(2N+1)p]. \tag{2.66}$$

And similarly,

$$\widetilde{\partial}_j u[n] = \frac{1}{(2N+1)^{d/2}} \sum_{k\in[-N..N]^d} \frac{i2\pi k_j}{\ell}\widetilde{u}[k]\, e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \tag{2.67}$$

$$= \sum_{k\in[-N..N]^d} e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \sum_{p\in\mathbb{Z}^d} \frac{i2\pi k_j}{\ell}\widehat{u}[k+(2N+1)p] \tag{2.68}$$

where (2.68) follows from (2.22). Hence

$$\partial_j u[n] - \widetilde{\partial}_j u[n] = \sum_{k\in[-N..N]^d} e^{i\frac{2\pi\langle k,n\rangle}{2N+1}} \sum_{p\in\mathbb{Z}^d\setminus\{0\}} \frac{i2\pi p_j(2N+1)}{\ell}\widehat{u}[k+(2N+1)p] \tag{2.69}$$

which using Parseval's theorem gives

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

Then, using inequality (2.24) together with Lemma A.0.5, 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

$$\sum_{j=1}^{d}\left\|\overrightarrow{\partial_j u_N}-\overrightarrow{\widetilde{\partial}_j u_N}\right\|^2\leq\frac{32\pi^2}{\ell^2}C^2\,(2N+1)^d\,2^d\,N^{-2m}\,[(m+1)!]^2a^{2m+2}\tag{2.70}$$
$$\leq\frac{32\pi^2}{\ell^2}\,C^2\,(2N+1)^d\,2^{N/2a}\,N^{-2m}\,[(m+1)!]^2\,a^{2m+2}\tag{2.71}$$

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

$$\sqrt{\sum_{j=1}^{d}\left\|\overrightarrow{\partial_j u_N}-\overrightarrow{\widetilde{\partial}_j u_N}\right\|^2}\leq\frac{8\sqrt{2}\pi e^3}{\ell}C\,(2N+1)^{d/2}\,N\,e^{-0.6\frac{N}{a}}.\tag{2.72}$$

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 (2.64). As in above, we start by writing the Fourier transform of the Laplacians:

$$\widetilde{\nabla^2}u_N[n]=\sum_{k\in[-N..N]^d}\frac{-4\pi^2}{\ell^2}\|k\|^2\sum_{p\in\mathbb{Z}^d}\widehat{u}[k+(2N+1)p]\tag{2.73}$$
$$\nabla^2 u_N[n]=\sum_{k\in[-N..N]^d}\frac{-4\pi^2}{\ell^2}\sum_{p\in\mathbb{Z}^d}\|k+(2N+1)p\|^2\,\widehat{u}[k].\tag{2.74}$$

Therefore,

$$\left\|\overrightarrow{\widetilde{\nabla^2}u_N}-\overrightarrow{\nabla^2 u_N}\right\|^2=\frac{16\pi^4}{\ell^4}(2N+1)^d\sum_{k\in[-N..N]^d}\left|\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\left(\|k+(2N+1)p\|^2-\|k\|^2\right)\widehat{u}_N[k]\right|^2$$
$$\overset{(a)}{\leq}\frac{16\pi^4}{\ell^4}(2N+1)^d\sum_{k\in[-N..N]^d}\left\{\left(\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+p(2N+1)\|^{-2m}\right)\right.\tag{2.75}$$
$$\left.\times\left(\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|k+(2N+1)p\|^4\|k+(2N+1)p\|^{2m}\,|\widehat{u}_N[k]|^2\right)\right\}$$

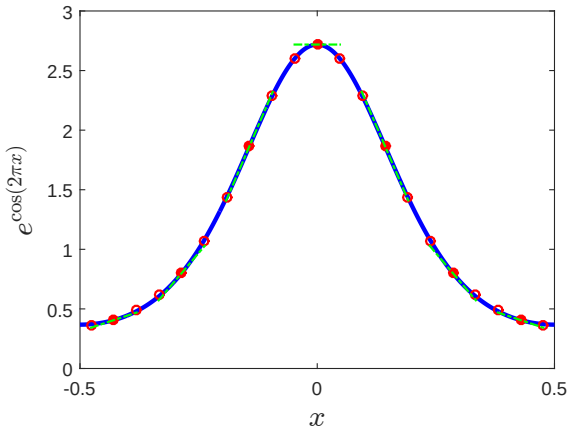

Figure 2.3: Applying the Fourier interpolation of Theorem 2.2.1 on the input function $u(x) = e^{\cos 2\pi x}$ of Example 2.2.2. The plot shows the interpolation results with $N = 3$ and $M = 10$. Filled circles correspond to the initial samples, and the hollow circles represent the interpolation output. The solid blue line represents the graph of the underlying function u. And the dashed green lines show the Fourier derivative estimations.

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 A.0.5 and set $m = \lfloor N/a \rfloor - 3$ (which is guaranteed to be a natural number since $N \geq 4ad$) to conclude the proof. □

2.2.3 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 [Rom80, KP02], according to which we have

$$\frac{d^m}{dx^m} f\left(g(x)\right) = \sum_{\substack{i_1,i_2,\cdots,i_m \in \{0,\cdots,m\} \\ i_1+2i_2+\cdots+mi_m=m}} \frac{m!}{i_1!\cdots i_m!} f^{(i_1+\cdots+i_m)}(g(x)) \left(\frac{dg/dx}{1!}\right)^{i_1} \cdots \left(\frac{d^m g/dx^m}{m!}\right)^{i_m} \tag{2.76}$$

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 2.2.6. *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*

$$D^\alpha (f \circ g)(x) = \alpha! \sum_{\lambda=1}^{|\alpha|} f^{(\lambda)}(g(x)) \sum_{s=1}^{|\alpha|} \sum_{p_s(\lambda,\alpha)} \prod_{j=1}^{s} \frac{1}{k_j!} \left(\frac{D^{\ell_j} g}{\ell_j!}\right)^{k_j}, \tag{2.77}$$

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 [CS96]. □

Reference [KP02, Lemma 1.4.1] proves that the coefficients in (2.76) follow

$$\sum_{\substack{i_1,i_2,\cdots,i_m\in\{0,\cdots,m\}\\ i_1+2i_2+\cdots+mi_m=m}} \frac{(i_1+\cdots+i_m)!}{i_1!\cdots i_m!} R^{i_1+\cdots+i_m} = \frac{R}{R+1}(R+1)^m, \tag{2.78}$$

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

Lemma 2.2.5. *Let α be a d-dimensional vector of non-negative integers. We have*

$$\sum_{\lambda=1}^{|\alpha|} \lambda!\, R^\lambda \sum_{s=1}^{|\alpha|} \sum_{p_s(\lambda,\alpha)} \prod_{j=1}^{s} \frac{1}{k_j!} = \frac{R}{R+1}(R+1)^{|\alpha|}. \tag{2.79}$$

Proof. The proof is a generalization of [KP02, 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 (2.77) 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

$$D^\alpha (x_1+\cdots+x_d)^n\big|_{x=0} = \alpha!\, \mathbf{1}_{\{|\alpha|=n\}} \tag{2.80}$$

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

$$\left(D^\alpha g\right)(0) = \alpha!. \tag{2.81}$$

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 $x = 0$.

$$(f\circ g)(x) = 1 + \frac{R}{R+1}\sum_{n=1}^{\infty}(R+1)^n \left(\sum_{i=1}^d x_i\right)^n \tag{2.82}$$

Combining (2.80) and (2.82) 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 (2.77). □

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

$$\sup_{x \in U} D^\alpha f(x) \leq C\, a^{|\alpha|}\, \alpha!, \tag{2.83}$$

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

$$\sqrt{\mathop{\mathbb{E}}_{X \sim \mathrm{Unif}(U)} \left(\sum_{\alpha : |\alpha| = m} |D^\alpha g(X)|^2 \right)} \leq C\, a^m\, m!. \tag{2.84}$$

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 2.2.2. *Using the same argument as in Proposition 2.2.1, for any open set U, we have $A_U(C, a) \subseteq S_U(3^{d-1}C, a)$.*

Proposition 2.2.7. *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.*

1. *$f_1 + f_2 \in A_{U \cap V}(C_1 + C_2, \max\{a_1, a_2\})$ if $U, V \subseteq \mathbb{R}^d$ are in the same Euclidean domains. Also, the same property holds for the semi-analytic families.*
2. *$f_1 \cdot f_2 \in A_{U \cap V}(C_1 C_2, a_1 + a_2)$.*
3. *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

$$D^\alpha (f_1 \cdot f_2)(x) = \sum_{\beta \in \prod_{i=1}^d \{0, \cdots, \alpha_i\}} \binom{\alpha}{\beta} D^{\alpha - \beta} f_1(x) \cdot D^\beta f_2(x), \tag{2.85}$$

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

$$\begin{aligned} \sup_{x \in U \cap V} \left| D^\alpha (f_1 \cdot f_2)(x) \right| &\leq C_1 C_2 a_1 \prod_{i=1}^d \sum_{\beta_i = 0}^{\alpha_i} a_1^{\alpha_i - \beta_i} a_2^{\beta_i} \\ &\leq C_1 C_2 \prod_{i=1}^d (a_1 + a_2)^{\alpha_i} = C_1 C_2 (a_1 + a_2)^{|\alpha|}. \end{aligned} \tag{2.86}$$

(c) Using Proposition 2.2.6, we have

$$\sup_{x\in U} D^\alpha(f_2\circ f_1) = \sup_{x\in U} \alpha! \sum_{\lambda=1}^{|\alpha|} f_2^{(\lambda)} \sum_{s=1}^{\alpha} \sum_{p_s(\lambda,\alpha)} \prod_{j=1}^{s} \frac{1}{k_j!}\left(\frac{D^{\ell_j} f_1}{\ell j!}\right)^{k_j} \tag{2.87}$$

$$\leq \alpha! C_2 a_1^{|\alpha|} \sum_{\lambda=1}^{\alpha} \lambda! a_2^\lambda C_1^\lambda \sum_{s=1}^{|\alpha|} \sum_{p_s(\lambda,\alpha)} \prod_{j=1}^{s} \frac{1}{k_j!} \tag{2.88}$$

$$\overset{(1)}{=} \alpha! \frac{C_2 C_1 a_2}{1+a_2 C_1}\left[a_1(1+a_2 C_1)\right]^{|\alpha|}, \tag{2.89}$$

where (1) follows from Lemma 2.2.5. □

Corollary 2.2.3. *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 2.2.2).*

Proof. This follows from Proposition 2.2.7(c) and the fact that $g(x) = e^x$ is in $A_\Delta(e^\Delta, 1)$. □

Corollary 2.2.4. *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 2.2.7(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 2.2.5. *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*

$$\widetilde{a} \leq 2^D \exp\left(\sum_{k=1}^{D} 2\left\|W^{(k)}\right\|_\infty + \left\|b^{(k)}\right\|_\infty\right) \tag{2.90}$$

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 2.2.7(a) we get $f_i^{(1)} \in A(\left\|w_i^{(1)}\right\|_1, 1)$. Also, note that $\sup_{x\in[-1,1]^d} \left|f_i^{(1)}(x)\right| \leq \left\|w_i^{(1)}\right\|_1 + \left|b_i^{(1)}\right|$. Applying Corollary 2.2.4 yields $g_i^{(1)} \in A_1(1, a_i)$

with $a_i = 2(1 + \|w_i\|_1)e^{\left\|w_i^{(1)}\right\|_1 + \left|b_i^{(1)}\right|} \leq 2e^{2\left\|w_i^{(1)}\right\|_1 + \left|b_i^{(1)}\right|}$. 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 2.2.7(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 2.2.4 once more to complete our induction.

2.3 Extension to squared-integrable functions

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.

$$u(x) = \frac{1}{(2\pi)^{d/2}} \int_{\omega\in\mathbb{R}^d} e^{i\langle\omega,x\rangle}\widehat{u}(\omega)\,\mathrm{d}\omega \tag{2.91}$$

$$\widehat{u}(\omega) = \frac{1}{(2\pi)^{d/2}} \int_{x\in\mathbb{R}^d} e^{-i\langle\omega,x\rangle}u(x)\,\mathrm{d}x \tag{2.92}$$

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

$$\widetilde{u}(\omega) := \left(\frac{H}{\sqrt{2\pi}}\right)^d \sum_{j\in\mathbb{Z}^d} e^{-i\langle\omega,jH\rangle}u(jH). \tag{2.93}$$

The coefficients of (2.91), (2.92), and (2.93) are chosen so that

$$\int_{x\in\mathbb{R}^d} |u(x)|^2\,\mathrm{d}x = \int_{\omega\in\mathbb{R}^d} |\widehat{u}(\omega)|^2\,\mathrm{d}\omega, \tag{2.94}$$

$$H^d \sum_{j\in\mathbb{Z}^d} |u(jH)|^2 = \int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widetilde{u}(\omega)|^2\,\mathrm{d}\omega. \tag{2.95}$$

Note the similarities between (2.92) and (2.93), 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

$$\widetilde{u}(\omega) = \sum_{k \in \mathbb{Z}^d} \widehat{u}\left(\omega + \frac{2\pi k}{H}\right). \tag{2.96}$$

Note that $\widetilde{u}$ depends only on the values of u at the lattice points $H\mathbb{Z}^d$. Moreover, if $\widehat{u}$ has a bounded support circumscribed within a fundamental domain of $\widetilde{u}$ then (2.96) 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 2.3.1. *A function $u \in L^2(\mathbb{R}^d)$ is said to be semi-analytic if*

$$\left\| \sum_{\alpha : |\alpha| = m} D^\alpha u \right\| \leq C\, a^m\, m! \tag{2.97}$$

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

We note that Definition 2.3.1 is equivalent to

$$\sqrt{\int_{\omega \in \mathbb{R}^d} \|\omega\|^{2m} |\widehat{u}(\omega)|^2 \, \mathrm{d}\omega} \leq C\, a^m\, m!. \tag{2.98}$$

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

Example 2.3.1. *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 2.3.2. *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*

$$\sqrt{\int_{\omega \in \mathbb{R}} |\omega|^{2m} |\widehat{u}(\omega)|^2 \, \mathrm{d}\omega} = \sqrt{\mathop{\mathbb{E}}_{X \sim \mathcal{N}(0,a)}[X^{2m}]} = \left(\frac{a}{2}\right)^m \sqrt{(2m-1)!!} \leq a^m\, m!. \tag{2.99}$$

Example 2.3.3. *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,*

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

We may now show the interpolation result of this section.

Theorem 2.3.1. *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*

$$V : |j\rangle \mapsto \sum_{k\in\mathbb{Z}} V_{jk} |k\rangle \tag{2.100}$$

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

$$\left\| V^{\otimes d} |u_H\rangle - |u_h\rangle \right\| \leq \frac{8\sqrt{2}e^3\, C}{\|u\|} e^{-0.6\,\frac{\pi}{aH}} \tag{2.101}$$

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

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

$$\begin{aligned} u(hj) &= \frac{1}{(2\pi)^{d/2}} \int_{\omega\in\mathbb{R}^d} e^{i\langle jh,\omega\rangle} \widehat{u}(\omega)\, \mathrm{d}\omega \\ &\approx \frac{1}{(2\pi)^{d/2}} \int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} e^{i\langle jh,\omega\rangle} \widetilde{u}(\omega)\, \mathrm{d}\omega \\ &= \left(\frac{H}{2\pi}\right)^d \int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} e^{i\langle jh,\omega\rangle} \sum_{k\in\mathbb{Z}^d} e^{-i\langle Hk,\omega\rangle} u(kH)\, \mathrm{d}\omega \\ &= \sum_{k\in\mathbb{Z}^d} u(kH) \prod_{a=1}^{d} W_{j_a k_a}, \end{aligned} \tag{2.102}$$

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 (2.102) is small. With a similar argument as the one used in Lemma 2.2.1, we can show that

$$\sqrt{\int_{\omega\notin[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widehat{u}(\omega)|\, \mathrm{d}\omega} \leq 2e^3 C\, e^{-\frac{0.6\pi}{Ha}}, \tag{2.103}$$

and also, following the proof of Lemma 2.2.2, we obtain

$$\sqrt{\int_{\omega\in[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widehat{u}(\omega) - \widetilde{u}(\omega)|^2\, \mathrm{d}\omega + \int_{\omega\notin[-\frac{\pi}{H},\frac{\pi}{H}]^d} |\widehat{u}(\omega)|^2\, \mathrm{d}\omega} \leq 2\sqrt{2}e^3\, C e^{-\frac{3\pi}{5aH}}. \tag{2.104}$$

The rest of the proof follows directly from the arguments made in Corollary 2.2.2, and Proposition 2.2.2. □

We also note that the concentration results of Section 2.2.1 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).

Chapter 3

Sampling

> Random numbers should not be generated with a method chosen at random.
>
> Donald Knuth

Sampling has been a longstanding task of interest. In this section, we are interested in the following task.

Design an algorithm that draws samples from a given distribution over a set.

In quantum settings, randomness arises naturally due to the probabilistic nature of quantum mechanics. However, in classical settings, the generation of randomness is not as straightforward. Therefore, it is commonly assumed that the generation of uniformly random numbers, say in the interval $[0,1]$, is easily achievable. In practice, algorithms like the middle-square method, introduced by von Neumann [VN+51], are utilized to generate pseudo-random numbers, which emulate samples from $\mathrm{Unif}([0,1])$ in the real world.

The question we presented above lacks clarity as we did not specify how the distribution is provided to us. In a straightforward scenario, we assume that we are given a distribution on the real line, characterized by an efficiently computable CDF denoted as $F : \mathbb{R} \to [0,1]$. In this case, it is easy to show that generating a random variable X from a uniform distribution on the interval $[0,1]$, say $U \sim \mathrm{Unif}([0,1])$, and then using $F^{-1}(U)$ as a sample, yields a value drawn from the desired distribution. For instance, $X = \lambda^{-1} \ln\left(\frac{1}{1-U}\right)$ is distributed according to the exponential distribution $\mathrm{d}\mu = \lambda e^{-\lambda x}\mathrm{d}x$. In many cases, such an inverse does not have an algebraically closed form, such as the normal distribution $\mathcal{N}(0,1)$. For the specific example of normal distribution, one might use the so-called Box-Muller algorithm [BM58]. The algorithm utilizes two independent random variables, denoted as U and V, drawn from a uniform distribution over the interval $[0,1]$. From these, the algorithm generates two independent random Gaussian variables, represented by X and Y, following a standard normal distribution with mean 0 and variance 1. This algorithm leverages the insight that the random variable $X^2 + Y^2$ follows an exponential distribution. Additionally,

the angle formed between the vector (X, Y) and the horizontal axis $(1, 0)$ is uniformly distributed. To see this note that

$$\mathrm{d}\mu = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \mathrm{d}x\, \mathrm{d}y$$

Hence, letting $R^2 := X^2 + Y^2$ and $\Theta := \angle((X, Y), (1, 0))$, we have

$$\mathrm{d}\mu = \frac{1}{2\pi} e^{-\frac{r^2}{2}} r\mathrm{d}r\, \mathrm{d}\theta = \frac{1}{2} e^{-\frac{r^2}{2}} \mathrm{d}(r^2)\, \frac{\mathrm{d}\theta}{2\pi}$$

which indicates that R^2 has an exponential distribution with $\lambda = \frac{1}{2}$, and $\Theta \sim \mathrm{Unif}([0, 2\pi))$. As a result, $X = \sqrt{2 \ln\left(\frac{1}{1-U}\right)} \cos(2\pi V)$, with indpendent and identically distributed (iid) random variables $U, V \sim \mathrm{Unif}([0, 1])$ has a normal distribution.

Another setting of interest is a uniform sampling from a given subset of $\mathbb{R}^d$. For instance, uniform sampling from the surface of the unit ball in $\mathbb{S}^{d-1}$. This can be readily done by noting that a normalized Gaussian i.e., $\widehat{X} = \frac{X}{\|X\|}$ for $X \sim \mathcal{N}(0, \mathbb{I}_d)$, is uniformly distributed over the unit ball $\mathbb{S}^{d-1}$. Hence, using d iid standard Gaussians we can generate the desired samples. Furthermore, one can use the fact that the area of the ball of radius r is proportional to r^{d-1}, we conclude that $Z := \widehat{X} \cdot R$, where R follows the distribution $\mathrm{d}\mu = d\, r^{d-1}\, \mathrm{d}r$ for $r \in [0, 1]$, has a uniform distribution inside the unit ball i.e., $Z \sim \mathrm{Unif}(\mathbb{B}^d)$. Note that the random variable R here can be generated using the inverse CDF method described earlier.

The aforementioned problem can be extended to the more general task of uniformly sampling from a bounded set $K \subset \mathbb{R}^d$. In this context, it is common to assume the presence of a membership oracle, denoted as O_K, which determines if a given point x belongs to K i.e., $O_K(x) = \mathbb{1}_{\{x \in K\}}$. The complexity of algorithms addressing this problem is typically assessed based on the number of queries made to O_K and the required arithmetic operations [Vem05]. Notably, classical algorithms have been developed that can generate samples uniformly at random from K within polynomial time (in terms of both the dimensionality and desired sampling accuracy) [Vem05, LV04].

In the broadest form of sampling, known as Gibbs sampling, we assume the availability of an oracle that computes the logarithm (with a shift) of our target distribution. Specifically, we aim to sample from the distribution $\mu(\mathrm{d}x) = \frac{e^{-f(x)}}{Z} \mathrm{d}x$, where $f(x)$ is accessible through the oracle. It is worth noting that when f is defined on the positive half of the real line as $f(x) = \lambda x$ for $x \geq 0$, we retrieve the problem of sampling from the exponential distribution. Similarly, setting $f(x) = \frac{1}{2}x^2$ corresponds to the standard Gaussian sampling problem. There has been extensive research focused on cases where f is a convex function (e.g., see [DCWY18, CLL+22]). Note that this scenario encompasses the problem of uniform sampling from a convex set K. Notably, the state-of-the-art sampling algorithms involve devising stochastic processes that reach the desired distribution as their (often unique) stationary state. The complexities of the algorithms heavily depend on the mixing times of such processes.

So far, we discussed classical approaches. However, it is worth exploring the potential of quantum algorithms in addressing the problem at hand. In quantum sampling, we aim to prepare the state $|\psi\rangle = \sum_i \sqrt{p_i}\, |i\rangle$. Measuring this state in the

computational basis allows us to obtain samples that follow the distribution determined by p_i. This quantum approach offers a level of generality beyond what classical algorithms can achieve. Note that classical algorithms prepare a mixed state $\rho = \sum_i p_i |i\rangle \langle i|$. Quantum algorithms provide additional capabilities by working directly with the superposition state $|\psi\rangle$. For example, specific quantum algorithms, like the mean estimation algorithm presented in [KO23], can be applied to $|\psi\rangle$. Leveraging this feature enables the development of faster algorithms for mean estimation. This feature will indeed help us devise faster algorithms for mean estimation in section Section 3.2.2. The article [GR02] demonstrates that in cases where the distribution is efficiently integrable, it is possible to efficiently prepare the state $|\psi\rangle$. The algorithm described in the paper achieves this by refining the discretization through the addition of auxiliary qubits and the application of controlled gates. To elaborate further, let us assume that we have already prepared the state:

$$|\psi_m\rangle = \sum_{i=0}^{2^m-1} \sqrt{p_i^{(m)}} |i\rangle .$$

Here, p_i represents the probability of the random variable x falling within the interval $[x_L^{(i,m)}, x_R^{(i,m)})$, which corresponds to an equidistant discretization of the domain of the random variable X. To approximate a continuous distribution determined by a measure, denoted as μ, we assume that $X \in [0,1]$ almost surely. Consequently, we have $x_L^{(i,m)} = \frac{i-1}{2^m-1}$ and $x_R^{(i,m)} = \frac{i}{2^m-1}$, resulting in

$$p_i^{(m)} = \int_{\frac{i-1}{2^m-1}}^{\frac{i}{2^m-1}} \mathrm{d}\mu.$$

The goal is to demonstrate that if $\int_a^b \mathrm{d}\mu$ is efficiently computable for any $0 \leq a < b \leq 1$, whether on a quantum or classical computer, then we can efficiently prepare $|\psi_n\rangle$ for any n. For instance, if integrating f over any interval takes at most time T, then we can prepare $|\psi_n\rangle$ in time $O(Tn)$. To achieve a finer discretization, we introduce an additional auxiliary qubit in the state $|0\rangle$. Next, we define

$$f(i) := \frac{\int_{x_L^{(i,m)}}^{\frac{x_R^{(i,m)} - x_L^{(i,m)}}{2}} \mathrm{d}\mu}{\int_{x_L^{(i,m)}}^{x_R^{(i,m)}} \mathrm{d}\mu}.$$

The goal of $f(i)$ is to represent the probability of X lying in the left half of the i-th interval, conditioned on X being on this interval. Therefore, we have

$$|\psi_{m+1}\rangle = \sum_{i=0}^{2^m-1} \sqrt{p_i^{(m)} f(i)} |i\rangle |0\rangle + \sum_{i=0}^{2^m-1} \sqrt{p_i^{(m)} (1 - f(i))} |i\rangle |1\rangle .$$

In transforming $|\psi_m\rangle$ into $|\psi_{m+1}\rangle$, we implement a circuit that performs the following transformations:

$$|i\rangle|0\rangle \mapsto |i\rangle|0\rangle|\theta_i\rangle \tag{3.1}$$
$$\mapsto |i\rangle\left(\sqrt{f(i)}|0\rangle + \sqrt{1-f(i)}|1\rangle\right)|\theta_i\rangle \tag{3.2}$$
$$\mapsto |i\rangle\left(\sqrt{f(i)}|0\rangle + \sqrt{1-f(i)}|1\rangle\right), \tag{3.3}$$

where in (3.1), we have $\cos\theta_i = \sqrt{f(i)}$, which can be implemented by adding auxiliary qubits and deploying the integration subroutine for the computation of $f(i)$. Equation (3.2) is performed using controlled rotations controlled by the $|\cos\theta_i\rangle$ register. Finally, in Equation (3.3), we uncompute $|\theta_i\rangle$ and discard the resulting $|0\rangle$ register. Using this recursive approach $n-1$ times (while starting with $|\psi_1\rangle = \sqrt{\mu(X \le 0.5)}|0\rangle + \sqrt{\mu(X > 0.5)}|1\rangle$), yields the desired state $|\psi_n\rangle$.

Note that the above algorithm of [GR02] is very limited as our distribution might not be efficiently integrable. For instance, when considering the general task of Gibbs sampling, it is very inefficient. Hence, in more generic cases, quantum random walks are often exploited. In short, quantum walks provide a powerful framework for quantum algorithms in sampling, analogous to classical algorithms that employ stochastic processes. The key idea involves transforming a series of Markov chains into a sequence of Grover's search operations, initially introduced in [WA08]. Building upon this foundation, subsequent works such as [CCH+23, CLL+22] have demonstrated quantum algorithms capable of preparing uniform distributions over convex bodies and also log-concave Gibbs distributions, outperforming their classical counterparts.

The rest of this chapter is organized as follows. In Section 3.1.1 we discuss the mixing times of Markov chains. The mixing time of a Markov chain that is constructed to reach a desired distribution is at the heart of the analysis of the complexity of a sampling algorithm. In Section 3.1.3 we describe the work of [WA08] and how can it be used to achieve quantum speedups. We discuss some of the prior classical and quantum algorithms proposed for Gibbs sampling in Section 3.1.2 and Section 3.1.4. Finally, in Section 3.2 we describe our Gibbs sampling algorithm, which uses a novel idea based on solving PDEs. We analyze the complexity of our algorithm and compare it with the state-of-the-art.

3.1 Background and prior works

3.1.1 Finite Markov chains

Here we aim to overview the important results of the mixing times of finite, discrete-time Markov chains. For a detailed analysis of the topic see [LP17, CLRS22]. Consider a homogeneous time Markov chain over the state space $[n]$, with transition matrix $P_{j\leftarrow i} := \mathbb{P}[X_{n+1} = j | X_n = i]$. Let us denote the set of probability vectors over $[n]$ by $\mathcal{P}_n$ that is $\mathcal{P}_n := \{q \in \mathbb{R}^n : q_i \ge 0, \sum_{i=1}^n q_i = 1\}$. We say that $\pi \in \mathcal{P}_n$ is a stationary state of the chain

P, if $\pi P = \pi$. A Markov chain that is irreducible[1] and aperiodic[2] has a unique stationary state. In addition to irreducibility and aperiodicity, we assume that P is reversible, that is

$$\pi_i P_{j \leftarrow i} = \pi_j P_{i \leftarrow j}. \tag{3.4}$$

The reversibility condition (3.4) serves as a valuable tool for analyzing the spectral gap of P and its relationship to the mixing time, as outlined below. Note that (3.4) concludes

$$\sqrt{\frac{\pi_i}{\pi_j}} P_{j \leftarrow i} = \sqrt{\frac{\pi_j}{\pi_i}} P_{i \leftarrow j} \tag{3.5}$$

which concludes that $\mathrm{diag}(\sqrt{\pi})\, P\, \mathrm{diag}(\sqrt{\pi})^{-1}$ is a symmetric matrix, and hence, its eigenvalues are all real. Furthermore, as this symmetric matrix is related to P via a similarity transform, we have that all eigenvalues of P are real as well.

Symmetric Markov chains

Let us first consider the case where the stationary distribution of P is uniform. This is the case, for instance, when P is a random walk on a regular graph. In such cases, P is itself symmetric since $\pi_i = \frac{1}{n}$ for all $i \in [n]$. Note that all symmetric Markov chains are reversible. Furthermore, let

$$\lambda(P) := \max_{q \in \mathcal{P}_n} \frac{\|qP - \pi\|}{\|q - \pi\|} = \max_{x \in \mathbb{R}^n : x \perp \pi} \frac{\|xP\|}{\|x\|}. \tag{3.6}$$

To see the second equality above, note that for any $q \in \mathcal{P}_n$, the vector $x = q - \pi$ is orthogonal to π, and hence, the left hand side is less than or equal to the expression on the right. To prove the converse, note that for any $x \in \mathbb{R}^n$ that is orthogonal to π, there exists a small enough number $c > 0$ such that $\pi + cx \in \mathcal{P}_n$.

Note that from the first equality in Eq. (3.6) we have

$$\|q \cdot P^n - \pi\| \leq \|q - \pi\| \left(\lambda(P)\right)^n \tag{3.7}$$

Furthermore, let us denote the eigen-values of P by $\lambda_1, \lambda_2, \cdots, \lambda_n$ orderd so that $1 = \lambda_1 > |\lambda_2| > \cdots |\lambda_n|$.[3] From the second equality in (3.6), we have $\lambda(G) = |\lambda_2|$, and hence setting $n = O\left(\frac{1}{\gamma} \log\left(\frac{1}{\varepsilon}\right)\right)$ guarantees ε-closeness to the distribution π, where $\gamma := 1 - |\lambda_2|$ is the spectal gap of P. Notably, the graphs with large spectral gaps allow fast convergence. Extensive research has been conducted on classical algorithms that leverage Markov chains to achieve high efficiency, as demonstrated by their

[1] An irreducible Markov chain is a chain in which any state can be reached from any other state with non-zero probability, after some finite number of steps. Mathematically, P is irreducible, if for all $i \neq j$, there exists some $n \geq 1$ such that $(P^n)_{ij} \neq 0$.

[2] Let $\mathcal{N}(i) := \{n \in \mathbb{Z}_{\geq 0} : (P^n)_{ii} \neq 0\}$. The Markov chain P is said to be aperiodic, if for all $i \in [n]$ we have $\gcd(\mathcal{N}(i)) = 1$.

[3] By the Perron–Frobenius theorem, all eigen-values of any transition matrix P are in the unit disc. Furthermore, the eigen-value $\lambda_1 = 1$, has unit geometric and algebraic multiplicity if the chain is irreducible and aperiodic. Under the latter conditions, we also have $|\lambda_i| < 1$ for all $i = 2, \cdots, n$.

proven convergence rates. Notable examples include efficient algorithms for estimating the permanent of matrices with non-negative entries [JSV04], the widely used page rank algorithm for search engines [PBMW99], and algorithms for counting problems [KLM89]. These algorithms stand as remarkable achievements in leveraging Markov chains to tackle complex computational tasks effectively.

The convergence rate for random walks on a graph is also sometimes bounded via a quantity called *conductance*. Let us consider an undirected d-regular graph $G = (V, E)$ with the adjacency matrix A. In this case, we have

$$P = \frac{A}{d}, \tag{3.8}$$

which is related to the Laplacian of G by $\mathcal{L}_G = \mathbb{I} - P$. Note that by the second equality in (3.6), we have that the gap $\gamma := 1 - \lambda(P)$ can be written as

$$\gamma = \frac{1}{d} \min_{x \in \mathbb{R}^n : x \perp \pi} \frac{\langle x, \mathcal{L}_G x \rangle}{\langle x, x \rangle} = \frac{1}{d} \min_{x \in \mathbb{R}^n : c \perp \pi} \frac{\sum_{(uv) \in E} (x_u - x_v)^2}{\sum_{v \in V} x_v^2} \tag{3.9}$$

To see the second equality above, note that to each $(u, v) \in E$, we can assign the terms $x_u^2 + x_v^2 - 2x_u x_v$ in the expansion of $\langle x, \mathcal{L}_G x \rangle$ which is equal to $(x_u - x_v)^2$. We can relate the spectral gap of our random walk to a geometrical property of the underlying graph. Let us now define this geometrical feature (see [MT+06] for a more detailed account of this topic).

Definition 3.1.1. *Consider a cut $(S, \overline{S})$ where $|S| \leq \frac{n}{2}$. We define the conductance of this cut to be*

$$\phi(S) := \frac{|E(S, \overline{S})|}{d|S|}. \tag{3.10}$$

Accordingly, the conductance of the graph is defined as

$$\Phi(G) := \min_{S : |S| \leq \frac{n}{2}} \phi(S). \tag{3.11}$$

We notice that $\Phi(G)$ shows the connectivity of each set of vertices to the rest of the graph. Furthermore, note that for a d-regular graph we have $|E(S, \overline{S})| \leq d \cdot |S|$, and hence $\phi(S) \leq 1$ for all $S \subset V$. Also, notice that setting x as $x_v = \frac{\mathbf{1}_{\{v \in S\}}}{|S|} - \frac{\mathbf{1}_{\{v \in \overline{S}\}}}{|\overline{S}|}$ in (3.9) gives $\gamma \leq 2\phi(S)$, which by minimizing over S results in

$$\gamma/2 \leq \Phi(G). \tag{3.12}$$

We note that the inequality is tight, as for a hypercube of ℓ bits, we have $\gamma = \frac{2}{\ell}$ and $\Phi = \frac{1}{\ell}$. Interestingly, conductance is upper bounded by the square root of the spectral gap by Cheeger's inequality.

Theorem 3.1.1 (Cheeger's inequality, Theorem 5.7 of [MT+06])**.** *It is the case that*

$$\gamma/2 \leq \Phi(G) \leq \sqrt{2\gamma}. \tag{3.13}$$

We note that the second inequality in (3.13) is also tight, as for the cycle C_n, we have $\Phi \approx \frac{2}{n}$, while $\gamma \approx \frac{\pi^2}{n^2}$).

Therefore, conductance, as a geometric measure, can also be used to upper bound the mixing time. Indeed, Cheeger's inequality indicates that $O(\Phi^{-2})$ many steps are sufficient to get close to the stationary state.

Generic reversible walks

Let us discuss the mixing time of more generic random walks. This is helpful in understanding the mixing time of Langevin dynamics. We have the following result about the mixing time.

Theorem 3.1.2 (Theorem 3.4 of [Sou20]). *Let P be a reversible Markov chain with stationary distribution π. Then, for all starting distributions q, we have*

$$\left\| \frac{q \cdot P^t}{\pi} - J \right\|_\pi \leq \left\| \frac{q}{\pi} - J \right\|_\pi (1 - |\lambda_2|)^t, \tag{3.14}$$

where J denotes the all-one vector; i.e., $J_v = 1, \forall v \in V$.

In the above theorem, the distance is defined as $\|x\|_\pi^2 = \sqrt{\sum_{v \in V} \pi_v x_v^2}$. As before, $|\lambda_2|$ corresponds to the second largest eigenvalue, in magnitude, of the matrix P.

3.1.2 Stochastic differential equations

In this section, we discuss a few properties of the Itô integral, which provides us with insights into the meaning of a SDE. The goal here is not to present a comprehensive and rigorous review of the topic, but rather to develop a better intuition of the fundamental aspects of such stochastic processes. For a more detailed and rigorous treatment, we recommend referring to [KKSS91, Eva23].

Firstly, let us recall that a stochastic process is a collection of random variables. In our context, we can focus on a sequence of the form $\{X(t) : t \geq 0\}$. We define the Wiener process, denoted as $W(\cdot)$, which satisfies the following properties: (i) $W(0) = 0$ almost surely: The Wiener process starts at zero with probability one. (ii) $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ for all $t \geq s \geq 0$: The increments of the Wiener process between any two time points t and s follow a normal distribution with mean zero and variance proportional to the time difference $t - s$. (iii) For any ordered sequence of $0 < t_1 < \cdots < t_n$, the random variables $W(t_1)$, $W(t_2) - W(t_1)$, $\cdots$, $W(t_n) - W(t_{n-1})$ are mutually independent: The increments of the Wiener process at different time intervals are mutually independent random variables.

There are constructions for a Wiener process (which proves its existence), and furthermore, one can prove that any two Wiener processes $W(\cdot)$ and $W'(\cdot)$ have the same distributions (i.e., for any sequence $0 < t_1 < \cdots < t_n$ the joint distributions of $\{W(t_i)\}_{i=1}^n$ and $\{W'(t_i)\}_{i=1}^n$ coincide).[4] For a given $\omega \in \Omega$, the map $t \mapsto W(t, \omega)$ is referred to as a sample path. It has been demonstrated that a Wiener process possesses

[4] This has a straightforward proof, which follows directly by property (iii) and (ii) of a Wiener process.

continuous sample paths almost surely. However, it is important to note that the sample path is nowhere differentiable almost surely. To gain a better understanding, we can examine the concept of Hölder continuity. Recall that a function f is said to be locally Hölder continuous with an exponent γ if, for any x, there exits $\varepsilon > 0$ such that $|f(x) - f(y)| \leq C\,|x-y|^{\gamma}$ for a constant $C > 0$ and all $y \in (x-\varepsilon, x+\varepsilon)$. It is known that a Wiener process is not locally Hölder continuous (almost surely) for any $\frac{1}{2} < \gamma \leq 1$, but it is locally Hölder continuous (almost surely) for all $\gamma < \frac{1}{2}$.

Let us now discuss Itô integrals. Our objective is to provide meaning to the following integrals

$$\int W\,\mathrm{d}t, \quad \int g\,\mathrm{d}W, \quad \int G\,\mathrm{d}W, \tag{3.15}$$

where we denote a function as g and a stochastic process as G. It is worth noting that the first integral ($\int W\,\mathrm{d}t$) is well-defined since $W(\cdot,\omega)$ is a continuous function almost surely. This is indeed extendable to any function that is continuous almost surely. However, we must establish a definition for integration with respect to $\mathrm{d}W$. It is crucial to understand that integrals involving $\mathrm{d}W$ do not possess any inherent meaning until they are properly defined. In this discussion, we adopt the convention of Itô integration, which is widely employed in stochastic calculus. Nevertheless, it is worth mentioning that alternative approaches such as Stratonovich integration exist as well [Oks13].

When performing Itô integration, we can interpret the integration of a function using the following identity

$$\int_0^1 g\,\mathrm{d}W = -\int_0^1 g' W\,\mathrm{d}t. \tag{3.16}$$

It is important to note that the right-hand side of Equation (3.16) is well-defined, thus leaving us with the task of defining $\int G\,\mathrm{d}W$. Let us first define such integrals when $G(\cdot)$ has specific features. We say G is a step process[5] if there is an ordered sequence $0 = t_0 < t_1 < \cdots < t_n = 1$ such that $G(t) = G(t_i)$ for all $i \in \{0, 1, \cdots, n\}$. For such a step process, we let

$$\int_0^1 G\,\mathrm{d}W := \sum_{i=1}^{n} G(t_{i-1})\,\left(W(t_i - W(t_{i-1})\right). \tag{3.17}$$

An intriguing observation is that, for a set of stochastic processes considered to be "nice", any stochastic process can be effectively approximated by a sequence of step processes. Furthermore, the previously defined Itô integral exhibits convergence for any convergent sequence within this "nice" set. As a result, a more general definition for Itô integrals emerges. The precise characterization of this "nice" set involves technical measure-theoretic assumptions, which are beyond the scope of our current discussion. However, it is worth emphasizing that the aforementioned "nice" set encompasses all stochastic processes G for which $\mathbb{E}\left[\int_0^1 G^2\,\mathrm{d}t\right] < \infty$. This quantity serves as a norm, which also defines the convergence of stochastic processes. To demonstrate the

[5]We omit a few technical conditions for the sake of readability and conciseness. In particular, we emphasize that the stochastic process $G(\cdot)$ that is to be integrated satisfies the following: for any $t > 0$, $G(t)$ depends only on the past history ($t' \leq t$) of the Wiener process.

eccentricity of Itô calculus, we highlight that $\int_0^T W\,\mathrm{d}W = \frac{1}{2}W(T)^2 - \frac{1}{2}T$.

Let us now discuss SDEs. We say that a stochastic process $X(\cdot)$ satisfies the following SDE

$$\mathrm{d}X = F\,\mathrm{d}t + G\,\mathrm{d}W, \tag{3.18}$$

for all $t \in [0, T]$, if it satisfies the following equation (in the Itô sense)

$$X(t) = X(t_0) + \int_{t_0}^{t} F(X(t'), t')\,\mathrm{d}t + \int_{t_0}^{t} G(X(t'), t')\,\mathrm{d}W \tag{3.19}$$

for all $0 \leq t_0 \leq t \leq T$. We remark that the existence and uniqueness of solutions to an SDE are well-studied [KKSS91, Eva23].

We emphasize that in the context of SDEs, one can interpret $\mathrm{d}W$ as being almost equal to $\sqrt{dt}$. This notion can be formalized using the Itô formula. Consider a function $u : \mathbb{R} \times [0, T] \to \mathbb{R}$ that is continuous and possesses continuous partial derivatives of first and second order. If $X(t)$ is the solution to the SDE $\mathrm{d}X = F\mathrm{d}t + G\mathrm{d}W$, then according to the Itô formula, we have

$$\mathrm{d}u(X(t), t) = \partial_t u\,\mathrm{d}t + \partial_x u\,\mathrm{d}X(t) + \frac{1}{2}\partial_x^2 u\,G^2\,\mathrm{d}t$$

As an example, when we set $F = 0$, $G = 1$, and $u(x) = x^2$, we obtain $\mathrm{d}(W^2) = 2W(t)\,\mathrm{d}W + \mathrm{d}t$. This confirms the expression we previously discussed for the integration of $W\,\mathrm{d}W$.

Langevin dynamics

The definition of SDEs can be extended to higher dimensions. In such cases, F and W are higher dimensional stochastic processes. A specific instance of these equations is known as Langevin dynamics, which corresponds to the case where $F = -\boldsymbol{\nabla} f$ and $G = \sqrt{2}$, with f being a potential function. Therefore, the SDE of our interest is

$$\mathrm{d}X(t) = -\boldsymbol{\nabla} f(X(t))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W(t)$$

It is well-known that if $\lim_{\|x\|\to\infty} f(x) = \infty$, then Langevin dynamics possesses a unique solution. Furthermore, its unique stationary state corresponds to the Gibbs measure associated with the potential function f [Pav14]. As a result, extensive research has been conducted on the development of algorithms for preparing Gibbs states through the simulation of Langevin dynamics.

Classical algorithms

There have been several studies focused on the design of algorithms for sampling using discretizations of Langevin dynamics [RT96, DCWY18, LST20, CEL+21]. Often in these works, the underlying function is assumed to be log-concave. This is due to the fact that the mixing times of log-concave distributions, can be small i.e., having an upper bound independent of dimension and temperature.

Let the desired distribution be described by the density $\mathrm{d}\mu \propto e^{-f(x)}\,\mathrm{d}x$, where f is a convex function. If we assume $\mathrm{Hess}(f) \geq \mu\mathbb{I}$ at all points, in other words if f is μ-strongly convex, it has been shown that the inverse spectral gap of the operator $\mathcal{L}$ is at most $\frac{1}{\mu}$ [Pav14, BGL+14]. This result is known as the Bakry–Emery criteria.

There have been various proposals for simulating Langevin dynamics using different algorithms. One of the most basic approaches is the unadjusted Langevin dynamics, which applies the following update rule:

$$x_{i+1} = x_i - h\nabla f + \sqrt{2h}\, n_i, \tag{3.20}$$

where $n_i \sim \mathcal{N}(0,1)$ are iid random Gaussian variables. The unadjusted Langevin dynamics method, which was used at least as early as [Par81], has limitations highlighted by [RT96]. Specifically, certain choices of the step size h can result in a transient Markov chain, leading to a lack of a unique limiting distribution. Although some of these issues have been addressed [CB18], the discretization error often hinders the achievement of fast algorithms.

On the other hand, Metropolized algorithms offer speed ups by adjusting the discretized transition rules. Metropolis adjusted Langevin algorithm (MALA), which incorporates a probabilistic rejection mechanism and yields a reversible Markov chain that converges to the desired limiting distribution $\mathrm{d}\mu \propto e^{-f}\,\mathrm{d}x$. In particular, let (see e.g., [DCWY18])

$$y_i := x_i - h\nabla f + \sqrt{2h}\, n_i, \tag{3.21}$$

$$\alpha := \min\left[1, \frac{\exp\left(f(x_i) + \|y_i - x_i + h\nabla f(x_i)\|\right)}{\exp\left(f(y_i) + \|x_i - y_i + h\nabla f(y_i)\|\right)}\right] \tag{3.22}$$

We then perform the update $x_i \leftarrow y_i$ with probability α. For a μ-strongly convex function f, with L-Lipschitz gradients ∇f, it is shown that the mixing time of MALA is $O(\kappa d)$, where d is the dimension of the domain and $\kappa = \frac{L}{\mu}$ [LST20].

3.1.3 From classical to quantum random walks

Szegedy [Sze04] developed an approach to quantizing any random walk on a regular graph. This was later generalized to reversible Markov chains [MNRS07]. Their approach essentially prepares a unitary matrix that has a phase gap, which is quadratically larger than the gap of its corresponding classical random walk. This gap was used to achieve quantum speed ups through phase estimation.

Let us review how quantum walks perform. Given a reversible Markov chain P over $[n]$, we consider the Hilbert space $\mathbb{C}^n$, and define the following quantum states

$$|P_v\rangle := \sum_{y\in V} \sqrt{P_{u\leftarrow v}}\,|v\rangle\,. \tag{3.23}$$

Accordingly, we define the following subspaces

$$\mathcal{A} := \mathrm{span}\{|v\rangle \otimes |P_v\rangle : v \in V\}, \tag{3.24}$$

$$\mathcal{B} := \mathrm{span}\{|P_v\rangle \otimes |v\rangle : v \in V\}. \tag{3.25}$$

For a subspace $\mathcal{K}$, we let $\Pi_{\mathcal{K}}$ be the projector onto $\mathcal{K}$. The quantum random walk operator corresponding to the classical walk is defined as

$$W(P) := (2\Pi_{\mathcal{B}} - \mathbb{I})(2\Pi_{\mathcal{A}} - \mathbb{I}). \tag{3.26}$$

Furthermore, denoting the stationary distribution by π, we define

$$|\pi\rangle := \sum_v \sqrt{\pi_v} |v\rangle |P_v\rangle . \tag{3.27}$$

It is transparent that $|\pi\rangle \in \mathcal{A}$. Furthermore, note that

$$\sum_v \sqrt{\pi_v} |v\rangle |P_v\rangle = \sum_{v,u} \sqrt{\pi_v P_{u\leftarrow v}} |v\rangle |u\rangle = \sum_{v,u} \sqrt{\pi_u P_{v\leftarrow u}} |v\rangle |u\rangle = \sum_u \sqrt{\pi_u} |P_u\rangle |u\rangle \tag{3.28}$$

where in the second last step, we use the fact that P is reversible. Hence, by (3.28), we have $|\pi\rangle \in \mathcal{B}$. As a result

$$W(P) |\pi\rangle = |\pi\rangle . \tag{3.29}$$

The next theorem helps us better understand the spectral properties of W. In the following, we denote by $\lambda_0 = 1$ the largest (in absolute value) eigen-value of a reversible Markov chain.

Theorem 3.1.3 (Adaption of Theorem 1 of [WA08]). *Let P be a reversible Markov chain. Let $\lambda_1, \cdots, \lambda_m$ be the non-zero eigen-values of P (with possible repititions). We have*

1. *On $\mathcal{A} \cap \mathcal{B}$, the operator $W(P)$ acts as identity. Furthermore, $\mathcal{A} \cap \mathcal{B}$ is a one-dimensional subspace spanned by $|\pi\rangle$.*
2. *On $\mathcal{A}^\perp \cap \mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}^\perp$, the operator $W(P)$ acts as $-\mathbb{I}$. Furthermore, $\dim(\mathrm{Ker}(P)) = \dim(\mathcal{A}^\perp \cap \mathcal{B}) = \dim(\mathcal{A} \cap \mathcal{B}^\perp)$.*
3. *On $\mathcal{A} + \mathcal{B}$, those eigenvalues of $W(P)$ that have non-zero imaginary part are $e^{\pm 2i\theta_1}, e^{\pm 2i\theta_2}, \cdots, e^{\pm 2i\theta_m}$, where $\theta_i \in (0, \frac{\pi}{2})$ are defined such that $\cos\theta_i = \lambda_i$.*
4. *$W(P)$ has no other eigenvalues on $\mathcal{A} + \mathcal{B}$.*

We note that the algorithms that we would work with, ensure that we are always in the subspace $\mathcal{A} + \mathcal{B}$, and hence, the spectral properties above, would be significantly useful for us. In particular, the fact that $|\pi\rangle$ is the unique eigen-vector with eigen-value $+1$ on $\mathcal{A} + \mathcal{B}$ is very important to us.

Let us now recall Grover's fixed-point search algorithm [Gro05].

Lemma 3.1.1 (Lemma 1 of [WA08]). *Let $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^d$ be two quantum states. Also, let R_1 and R_2 be defined so that*

$$R_i := \omega |\psi_i\rangle \langle\psi_i| + (\mathbb{I} - |\psi_i\rangle \langle\psi_i|) , \tag{3.30}$$

where $\omega = e^{i\pi/3}$. There is an algorithm, that given $m \in \mathbb{N}$, uses $M = 3^m$ many R_1 and R_2 and prepares a unitary U with the following property

$$|\langle\psi_2, U\psi_1\rangle|^2 \geq 1 - (1-p)^M \tag{3.31}$$

where $p \leq |\langle \psi_2, \psi_1 \rangle|^2$.

As a result of the latter lemma, we can get ε-close (up to a global phase) to $|\psi_2\rangle$, by $O(\frac{1}{p} \log \frac{1}{\varepsilon})$ many applications of R_1 and R_2. The interesting fact about the fixed-point algorithms is that they do not overshoot. We highlight that this fixed-point search algorithm can be improved (in terms of dependence on p), as demonstrated in [YLC14, GSLW19]. Next, we show that given a random walk operator W, one can prepare such an operator R (on a restricted subspace).

Lemma 3.1.2 (Adaption of Corollary 2 of [WA08])**.** *Let W be a unitary acting on $\mathbb{C}^d$ with a unique eigen-vector $|\psi_0\rangle$ with eigen-value $\lambda_0 = 1$. Furthermore, let $\lambda_j = e^{2\pi i \phi_j}$ for $j = 1, \cdots, d-1$, be the rest of the eigen-values of W. Let $\Delta = \min_{j=1,\ldots,d-1} |\phi_j|$ be the phase gap of W. Let $R = \omega |\psi_0\rangle \langle \psi_0| + (\mathbb{I} - |\psi_0\rangle \langle \psi_0|)$. There is an algorithm to prepare the unitary $\widetilde{R}$ such that $\left\| \widetilde{R} - R \right\| \leq \varepsilon$, with $O(\frac{1}{\Delta} \log \frac{1}{\varepsilon})$ many uses of controlled W.*

Proof. We sketch the proof here. One can use the phase estimation algorithm with precision Δ. We then have a unitary U that acts approximately as

$$U : |\psi_j\rangle |0\rangle \mapsto |\psi_j\rangle |\phi_j\rangle . \tag{3.32}$$

Then, we apply the unitary $\mathbb{I} \otimes V$ with $V = \omega |0\rangle \langle 0| + (\mathbb{I} - |0\rangle \langle 0|)$. □

Combining all the above provides us with the following.

Theorem 3.1.4 (Adaption of Theorem 2 of [WA08])**.** *Let $P_0, P_1, \cdots, P_r$ be classical reversible Markov chains with stationary distributions $\pi_0, \pi_1, \cdots, \pi_r$ and spectral gaps larger than δ. Assume π_i and π_{i+1} are close in fidelity i.e., for some $p > 0$, we have*

$$\left(\sum_v \sqrt{\pi_i(v) \pi_{i+1}(v)} \right)^2 = \langle \pi, \pi_{i+1} \rangle^2 \geq p. \tag{3.33}$$

Let $(W_i)_{i=1}^r$ be the random walk operators corresponding to the Markov chains. Assume we are given $|\pi_0\rangle$. There a quantum algorithm that makes

$$O\left(\frac{r}{p\sqrt{\delta}} \log\left(\frac{r}{\varepsilon}\right) \operatorname{poly}\left(\log \frac{r \log \varepsilon^{-1}}{p} \right) \right) \tag{3.34}$$

many calls to controlled-W_is overall, and samples from a distribution ε-close (in total variation distance) to π_r.

The fact that the complexity in Theorem 3.1.4 depends on $\frac{1}{\sqrt{\delta}}$ (as opposed to $\frac{1}{\delta}$) allows quantum algorithms that achieve quadratic speed ups. In the following subsection, we show how [CLL+22] uses the aforementioned techniques to develop a quantum analog of MALA (see Section 3.1.2 for a review of the classical method).

3.1.4 Quantum MALA

Let us consider a convex potential function f associated with the target Gibbs measure π. We can efficiently determine the minimum of f, and we assume that this minimum occurs at the origin and has a value of zero. Additionally, we assume that f is μ-strongly convex and possesses L-Lipschitz gradients. Our objective is to construct a sequence of Markov chains and employ the annealing algorithm proposed in [WA08], as discussed in Section 3.1.3. Let $f_0 = \frac{\|x\|^2}{2\sigma_1^2}$. Hence, $\pi_0(x) \propto e^{-f_0}$ represents a multivariate Gaussian. Furthermore, consider $f_1, f_2, \cdots, f_M$ such that $f_i = f(x) + \frac{\|x\|^2}{2\sigma_i^2}$, where $\sigma_{i+1}^2 = (1+\alpha)\sigma_i^2$ for all $i \in [M-1]$. We denote the corresponding Gibbs measures by $\pi_1, \cdots, \pi_M$. Intuitively, π_i and π_{i+1} have large fidelities, and moreover, for a large enough M, we expect π_M to be close to π. The following lemma formalizes this intuition

Lemma 3.1.3 (Adaption of Lemma B.6 of [CLL+22]). *For a Gibbs distribution π_i, let*

$$|\pi_i\rangle := \int \mathrm{d}x \sqrt{\pi_i(x)}\,|x\rangle\,. \tag{3.35}$$

We have

1. $\langle \pi_0, \pi_1 \rangle \geq e^{-\sigma_1^2 dL/4}$,
2. $\langle \pi_i, \pi_{i+1} \rangle \geq e^{-2(\frac{\alpha}{\alpha+2})^2 d}$, *for all* $i \in [M-1]$, *and*
3. $\langle \pi_M, \pi \rangle \geq e^{-\frac{d}{2\mu\sigma_M^4}}$.

As a result of the above lemma, choosing $\alpha = \frac{1}{\sqrt{d}}$, $M = O(\sqrt{d})$ and $\sigma_1^2 = \frac{\varepsilon}{2Ld}$ guarantees that all consecutive inner products are lower bounded by a constant. Furthermore, [CLL+22] shows one can implement the quantum walk corresponding to MALA with constantly many queries to an oracle for the derivatives of f. Putting all this together with the fact that the mixing time of MALA is $O(\kappa d)$ and utilizing Theorem 3.1.4, we conclude that with $\widetilde{O}(d\sqrt{\kappa})$ many queries to an oracle for the derivatives of f, we can sample according to the desired Gibbs distribution. Notice that this provides a Grover-type speed up in the condition number κ.

3.1.5 Quantum algorithms for solving differential equations

Our algorithm benefits from the use of the high precision quantum linear differential equation solver developed by Berry et al. [BCOW17]. Krovi later improved this algorithm in [Kro23]. The solver is designed to tackle the problem of interest, which is solving the following Ordinary Differential Equation (ODE) at time $T > 0$:

$$\frac{d\vec{x}}{dt} = A\vec{x} + \vec{b}\,, \quad \vec{x}(0) = \vec{x}_{\text{init}}, \tag{3.36}$$

where $\vec{x}$, $\vec{b} \in \mathbb{C}^n$ 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 ([Kro23, Lemma 6])

$$\overrightarrow{x}(T) = e^{At}\overrightarrow{x}(0) + \left(e^{At} - \mathbb{I}\right) A^{-1} h \overrightarrow{b}. \tag{3.37}$$

More specifically, letting

$$T_k(z) := \sum_{j=0}^{k} \frac{z^j}{j!}$$
$$S_k(z) := \sum_{j=0}^{k-1} \frac{z^j}{(j+1)!},$$

we have $T_\infty(z) = e^z$ and $S_\infty(z) = (e^z - 1)z^{-1}$. Notably the sequence $\{T_k(z)\}_k$ and $\{S_k(z)\}_k$ converge superexpontially fast to $T_\infty(z)$ and $S_\infty(z)$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \leq 0$ [BCOW17, Lemma 10, 12]. Hence, they propose an algorithm that updates $\overrightarrow{x}$ according to

$$\overrightarrow{x}(nh) \leftarrow T_k(Ah)\ \overrightarrow{x}((n-1)h) + S_k(Ah)\, b. \tag{3.38}$$

To implement (3.38) on a quantum computer, one can define the following matrix

$$N_1(A) := \mathbb{I} - \sum_{i=0}^{m-1} |i\rangle\langle i| \otimes \left(\sum_{j=0}^{k-1} |j\rangle\langle j-1| \otimes \frac{A}{j}\right) - \sum_{i=0}^{m-1} |i+1\rangle\langle i| \otimes \left(\sum_{j=0}^{k} |0\rangle\langle k|\right) \otimes \mathbb{I} \tag{3.39}$$

As we will see shortly, the first register (indexed by i in summations) counts the time step, while the second register (indexed by j) counts the terms in Taylor expansion. Consider the equation

$$N_1(A)\,|x\rangle = |0\rangle\,|0\rangle\,|x_{\text{init}}\rangle + h \sum_{i=0}^{m-1} |i+1\rangle\,|1\rangle\,|b\rangle$$

Letting $|x\rangle = \sum_{i,j\in[m]\times[k]} |i\rangle\,|j\rangle\,|x_{i,j}\rangle$ we have (note that $|x_{i,j}\rangle$ might not be normalized to 1)

$$\begin{aligned} |x_{0,0}\rangle &= |x_{\text{init}}\rangle \\ |x_{i,j}\rangle &= |x_{i,j-1}\rangle \frac{A}{j} + h\,|b\rangle\,\delta_{j,1} \quad \forall i \in [m-1], j \in [k] \\ |x_{i+1,0}\rangle &= \sum_{j=0}^{k} |x_{i,j}\rangle \qquad\qquad\qquad \forall i \in \{0, \cdots, m-1\}. \end{aligned} \tag{3.40}$$

As a result of (3.40) we have

$$|x_{i+1,0}\rangle = T_k(Ah)\,|x_{i,0}\rangle + S_k(Ah)\,h\,|b\rangle\,. \tag{3.41}$$

Therefore, we can consider the linear system (3.39) and solve it via a quantum linear solver [CKS17, HHL09]. Note that we are interested in the state $|x_{m,0}\rangle$. However, it has a small amplitude. To overcome this issue, we repeat $|x_{m,0}\rangle$ many times, say q times. This can be performed by changing N_1 as

$$N(A) := N_1(A) - \sum_{i=m}^{m+q} |i+1\rangle\langle i| \otimes |0\rangle\langle 0| \otimes \mathbb{I}, \tag{3.42}$$

and then utilizing amplitude amplification in the end. Hence, we solve for

$$C(A)|x\rangle = |0\rangle|0\rangle|x_{\text{init}}\rangle + h\sum_{i=0}^{m-1} |i+1\rangle|1\rangle|b\rangle$$

We refer the reader to [BCOW17, Kro23] for the computations of the condition number of $C(A)$. As shown in [BCOW17], κ_C is upper bounded by $m+q$ (i.e., the number of time steps) multiplied by the condition number of the matrix V whose columns are eigenvectors of A. [Kro23] shows that this upper bound can be made tighter via replacing κ_V by the log-norm of A. Here, we just provide the main result of [Kro23] and eliminate the complexity and correctness proofs.

Theorem 3.1.5 (Adoption of Theorem 7 of [Kro23])**.** *Suppose $A = VDV^{-1}$ is an $N \times N$ diagonalizable matrix, where $D = diag(\lambda_1, \cdots, \lambda_N)$ satisfies $\text{Re}(\lambda_i) \leq 0$ 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*

$$\frac{d}{dt}\overrightarrow{x} = A\overrightarrow{x} + \overrightarrow{b} \tag{3.43}$$

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

$$g = \max_{t\in[0,T]} \left\|\overrightarrow{x}(t)\right\| / \left\|\overrightarrow{x}(T)\right\|. \tag{3.44}$$

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

$$O\left(gT\|A\|\kappa_V \text{ poly}\left(s, \log d, \log\left(1 + \frac{Te^2\|b\|}{\|x_T\|}\right), \log\left(\frac{1}{\varepsilon}\right), \log(T\|A\|\kappa_V)\right)\right), \tag{3.45}$$

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

$$O(\text{polylog}(1 + \frac{Te^2\|b\|}{\|x_T\|}, 1/\varepsilon, T\|A\|)), \tag{3.46}$$

.

Proof. We can observe that the only difference in the complexity result between our theorem and the one presented in [Kro23, Theorem 7] is the substitution of $C(A)$ with

κ_V. It is worth recalling from [Kro23, 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 [Kro23, Lemma 4]. □

3.2 Gibbs Sampling

In this section we present and analyze our Gibbs sampling algorithm.

3.2.1 The algorithm

Our goal is to obtain the steady state solutions to the FPE

$$\partial_t \rho(x,t) = \nabla \cdot \left(e^{-E(x)} \nabla \left(e^{E(x)} \rho(x,t) \right) \right) \tag{3.47}$$

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

$$\mathrm{d}Y_t = -\nabla E(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t \tag{3.48}$$

on a high dimensional torus $\mathbb{T} = \mathbb{R}^d/\mathbb{Z}^d$. By this notation we mean to 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 will 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 (3.47) corresponds to the Gibbs state $\rho_s(x) \propto e^{-E(x)}$ of the potential. We intend to find this distribution by solving (3.47) 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 Section 3.2.1 for a pseudo-code of our algorithm.

By solving the FPE we mean preparing a quantum state that encodes the solution of the PDE (3.47) 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

$$\mathcal{L}(-) = \nabla \cdot \left(e^{-E} \nabla \left(e^{E} - \right) \right) \tag{3.49}$$

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

$$\frac{d}{dt}\overrightarrow{u(t)} = \mathbb{L}\overrightarrow{u(t)} \tag{3.50}$$

using the machinery of [BCOW17, Kro23] to find a high precision approximation of the solution, $u(T)$ (line 2 of Section 3.2.1).

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 2.2.1 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

$$|\rho_s\rangle \propto \sum_x e^{-E(x)} |x\rangle , \tag{3.51}$$

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

$$p(x) = \frac{1}{\bar{Z}_{\beta=2}} e^{-2E(x)} \simeq \frac{\ell^d}{(2n+1)^d Z_{\beta=2}} e^{-2E(x)}. \tag{3.52}$$

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 Section 3.2.1).

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 [OS75] (line 3 of Section 3.2.1).

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. C.1 in the appendix).
2: Deploy the algorithm of [BCOW17] 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 1.2.1).
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 $\tilde{x}$ 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 $\tilde{x}$.

To this end, in Chapter 2 we introduced *Fourier interpolation* and use this technique to upsample our quantum state in the Fourier domain. However, our interpolation technique is useful beyond the applications considered herein. For instance, quantum algorithms for solving partial differential equations (e.g., [CLO21]) 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. (2.56) in the appendix). In Appendix D.1 we show that assuming that the FPE generates a semi-analytic one-parameter family of probability measures $\{e^{\mathcal{L}t}\rho_0 : t \geq 0\}$, Section 3.2.1 samples from a distribution ε-close to the Gibbs distribution (in total variation distance) by making $\widetilde{\mathcal{O}}(d^7 \frac{e^{\Delta/2}}{\ell^2} \kappa_{E/2} \operatorname{polylog}(1/\varepsilon))$ queries to the oracle (1.6) 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.

3.2.2 Algorithm's analysis

We are now ready to state the computational complexity of our Gibbs sampler (Section 3.2.1).

Theorem 3.2.1 (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. Section 3.2.1 samples from a distribution ε-close to the Gibbs distribution (in total variation distance), by making*

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

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 $\operatorname{polylog}(Cade^{\Delta}(1+\ell L))/\varepsilon)$.

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 $\nabla E(x) = 0$. In [MBM16] this condition is quantified 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. [LE20] generalizes 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 [LE20, 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.0.2. This is a generalization of the Bakry-Emery criterion [BGL+14, Proposition 4.8.1] well beyond strong convexity. We have,

Corollary 3.2.1. *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 [LE20, 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

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

This follows from the observation that E satisfies a Poincaré inequality with $\kappa_E = \mathcal{O}\left(\frac{1}{\lambda^2}\right)$ as per [LE20, 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.[6]

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 [KO23].

Corollary 3.2.2 (Mean estimation)**.** *Let E be an energy function, satisfying the assumptions made in Theorem 3.2.1. Furthermore, let f be an L_f-Lipschitz ℓ-periodic function with diameter*

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

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*

$$\mathcal{O}\left(d^7 a^4 e^{\Delta/2} \frac{\kappa_{E/2}}{\ell^2} \frac{\Delta_f}{\varepsilon} \log\left(\frac{1}{\delta}\right) \operatorname{polylog}\left(C, a, \frac{1}{\varepsilon}, \Delta_f, L_f \ell, L\ell\right)\right) \tag{3.53}$$

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 [LE20] 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. In both cases, 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.

Chapter 4

Conclusion

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 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. Ergo, the exponentially poor dependence on Δ indicates the unsurprising difficulty of low-temperature Gibbs sampling. 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).

Another interesting avenue for future research is further investigation into the requirement of compactness of the domain of definition of the potential. In this paper, we considered periodic boundary conditions, hence tori as the domains of definition of the potential. However, other compact domains, or compactly supported Gibbs measures are left for further investigations. Perhaps such variations would require other choices of spectral methods; e.g., by deploying Chebyshev polynomials.

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. It is worth noting that our upsampling results of Chapter 2 are generalizable to non-periodic functions as shown in Section 2.3. The measure concentration results are also straightforward to generalize. However, establishing connections between semi-analyticity and analyticity for non-periodic functions, as well as developing efficient algorithms for upsampling them, are left as future areas of exploration.

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, especially 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.

References

[AHS85] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.

[AS64] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1964.

[BCOW17] Dominic W Berry, Andrew M Childs, Aaron Ostrander, and Guoming Wang. Quantum algorithm for linear differential equations with exponentially improved dependence on precision. *Communications in Mathematical Physics*, 356(3):1057–1081, 2017.

[Ber11] Nils Berglund. Kramers' law: Validity, derivations and generalisations. *arXiv preprint arXiv:1106.5799*, 2011.

[BGL+14] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

[BHMT02] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[BM58] George EP Box and Mervin E Muller. A note on the generation of random normal deviates. *The annals of mathematical statistics*, 29(2):610–611, 1958.

[CB18] Xiang Cheng and Peter Bartlett. Convergence of langevin mcmc in kl-divergence. In *Algorithmic Learning Theory*, pages 186–211. PMLR, 2018.

[CCBJ17] Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped langevin mcmc: A non-asymptotic analysis. arxiv e-prints, page. *arXiv preprint arXiv:1707.03663*, 2017.

[CCH+23] Shouvanik Chakrabarti, Andrew M Childs, Shih-Han Hung, Tongyang Li, Chunhao Wang, and Xiaodi Wu. Quantum algorithm for estimating volumes of convex bodies. *ACM Transactions on Quantum Computing*, 4(3):1–60, 2023.

[CEL+21] Sinho Chewi, Murat A Erdogdu, Mufan Bill Li, Ruoqi Shen, and Matthew Zhang. Analysis of Langevin Monte Carlo from Poincaré to Log-Sobolev. *arXiv preprint arXiv:2112.12662*, 2021.

[CKS17] Andrew M Childs, Robin Kothari, and Rolando D Somma. Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM Journal on Computing*, 46(6):1920–1950, 2017.

[CL20] Andrew M Childs and Jin-Peng Liu. Quantum spectral methods for differential equations. *Communications in Mathematical Physics*, 375(2):1427–1457, 2020.

[CLL+22] Andrew M Childs, Tongyang Li, Jin-Peng Liu, Chunhao Wang, and Ruizhe Zhang. Quantum algorithms for sampling log-concave distributions and estimating normalizing constants. *arXiv preprint arXiv:2210.06539*, 2022.

[CLO21] Andrew M Childs, Jin-Peng Liu, and Aaron Ostrander. High-precision quantum algorithms for partial differential equations. *Quantum*, 5:574, 2021.

[CLRS22] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2022.

[CS96] G Constantine and T Savits. A multivariate faa di bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, 1996.

[CS16] Anirban Narayan Chowdhury and Rolando D Somma. Quantum algorithms for Gibbs sampling and hitting-time estimation. *arXiv preprint arXiv:1603.02940*, 2016.

[CvDHS15] Andrew M Childs, Wim van Dam, Shih-Han Hung, and Igor E Shparlinski. Optimal quantum algorithm for polynomial interpolation. *arXiv preprint arXiv:1509.09271*, 2015.

[DCWY18] Raaz Dwivedi, Yuansi Chen, Martin J Wainwright, and Bin Yu. Log-concave sampling: Metropolis-Hastings algorithms are fast! In *Conference on learning theory*, pages 793–797. PMLR, 2018.

[DM19] Yilun Du and Igor Mordatch. Implicit generation and generalization in energy-based models. *arXiv preprint arXiv:1903.08689*, 2019.

[Eva23] Lawrence C. Evans. An introduction to stochastic differential equations. https://www.cmor-faculty.rice.edu/~cox/stoch/SDE.course.pdf, 2023. Accessed: 2023-05-10.

[GPSMH19] Eduardo García-Portugués, Michael Sørensen, Kanti V Mardia, and Thomas Hamelryck. Langevin diffusions on the torus: estimation and applications. *Statistics and Computing*, 29(1):1–22, 2019.

[GR] IS Gradshteyn and IM Ryzhik. Tables of integrals, series, and products 2000 6th ed San Diego.

[GR02] Lov Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions. *arXiv preprint quant-ph/0208112*, 2002.

[Gro05] Lov K Grover. Fixed-point quantum search. *Physical Review Letters*, 95(15):150501, 2005.

[GSLW19] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 193–204, 2019.

[GWJ+19] Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. *arXiv preprint arXiv:1912.03263*, 2019.

[HHL09] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical review letters*, 103(15):150502, 2009.

[Hin07] Geoffrey E Hinton. Learning multiple layers of representation. *Trends in cognitive sciences*, 11(10):428–434, 2007.

[HJA20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[HM18] Yassine Hamoudi and Frédéric Magniez. Quantum chebyshev's inequality and applications. *arXiv preprint arXiv:1807.06456*, 2018.

[HMZ20] Mitch Hill, Jonathan Mitchell, and Song-Chun Zhu. Stochastic security: Adversarial defense using long-run dynamics of energy-based models. *arXiv preprint arXiv:2005.13525*, 2020.

[Hop82] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[HOT06] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[Hsu02] Elton P Hsu. *Stochastic analysis on manifolds*. Number 38. American Mathematical Soc., 2002.

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004.

[KHR22] Frederic Koehler, Alexander Heckett, and Andrej Risteski. Statistical efficiency of score matching: The view from isoperimetry. *arXiv preprint arXiv:2210.00726*, 2022.

[KK09] Daniel M Kane and Samuel A Kutin. Quantum interpolation of polynomials. *arXiv preprint arXiv:0909.5683*, 2009.

[KKSS91] Ioannis Karatzas, Ioannis Karatzas, Steven Shreve, and Steven E Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 1991.

[KLM89] Richard M Karp, Michael Luby, and Neal Madras. Monte-carlo approximation algorithms for enumeration problems. *Journal of algorithms*, 10(3):429–448, 1989.

[KO23] Robin Kothari and Ryan O'Donnell. 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)*, pages 1186–1215. SIAM, 2023.

[Kom60] Hikosaburo Komatsu. A characterization of real analytic functions. *Proceedings of the Japan Academy*, 36(3):90–93, 1960.

[KP02] Steven G Krantz and Harold R Parks. *A primer of real analytic functions*. Springer Science & Business Media, 2002.

[Kro23] Hari Krovi. Improved quantum algorithms for linear and nonlinear differential equations. *Quantum*, 7:913, 2023.

[LE20] Mufan Bill Li and Murat A Erdogdu. Riemannian langevin algorithm for solving semidefinite programs. *arXiv preprint arXiv:2010.11176*, 2020.

[LKK+21] Jin-Peng Liu, Herman Øie Kolden, Hari K Krovi, Nuno F Loureiro, Konstantina Trivisa, and Andrew M Childs. Efficient quantum algorithm for dissipative nonlinear differential equations. *Proceedings of the National Academy of Sciences*, 118(35):e2026805118, 2021.

[LP17] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[LST20] Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Logsmooth gradient concentration and tighter runtimes for Metropolized hamiltonian monte carlo. In *Conference on learning theory*, pages 2565–2597. PMLR, 2020.

[LV04] László Lovász and Santosh Vempala. Hit-and-run from a corner. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 310–314, 2004.

[LW18] Tongyang Li and Xiaodi Wu. Quantum query complexity of entropy estimation. *IEEE Transactions on Information Theory*, 65(5):2899–2921, 2018.

[MBM16] Song Mei, Yu Bai, and Andrea Montanari. The landscape of empirical risk for non-convex losses. *arXiv preprint arXiv:1607.06534*, 2016.

[MNRS07] Frédéric Magniez, Ashwin Nayak, Jérémie Roland, and Miklos Santha. Search via quantum walk. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 575–584, 2007.

[Mon15] Ashley Montanaro. Quantum speedup of monte carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015.

[MP11] David A Meyer and James Pommersheim. On the uselessness of quantum queries. *Theoretical Computer Science*, 412(51):7068–7074, 2011.

[MT$^+$06] Ravi Montenegro, Prasad Tetali, et al. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.

[MV00] Peter A Markowich and Cédric Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp*, 19:1–29, 2000.

[NC02] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[NHH$^+$20] Erik Nijkamp, Mitch Hill, Tian Han, Song-Chun Zhu, and Ying Nian Wu. On the anatomy of mcmc-based maximum likelihood learning of energy-based models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5272–5280, 2020.

[NHZW19] Erik Nijkamp, Mitch Hill, Song-Chun Zhu, and Ying Nian Wu. Learning non-convergent non-persistent short-run mcmc toward energy-based model. *Advances in Neural Information Processing Systems*, 32, 2019.

[Oks13] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[OS75] Alan V Oppenheim and Ronald W Schafer. Digital signal processing(book). *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.

[Par81] Giorgio Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180(3):378–384, 1981.

[Pav14] Grigorios A Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, volume 60. Springer, 2014.

[PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[Pet07] Valentin V Petrov. On lower bounds for tail probabilities. *Journal of statistical planning and inference*, 137(8):2703–2705, 2007.

[PW09] David Poulin and Pawel Wocjan. Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer. *Physical review letters*, 103(22):220502, 2009.

[PZ32] Raymond EAC Paley and Antoni Zygmund. A note on analytic functions in the unit circle. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 28, pages 266–272. Cambridge University Press, 1932.

[RC22] Sergi Ramos-Calderer. Efficient quantum interpolation of natural data. *arXiv preprint arXiv:2203.06196*, 2022.

[Rom80] Steven Roman. The formula of faa di bruno. *The American Mathematical Monthly*, 87(10):805–809, 1980.

[RT96] Gareth O Roberts and Richard L Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

[Sha79] Adi Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[SK21] Yang Song and Diederik P Kingma. How to train your energy-based models. *arXiv preprint arXiv:2101.03288*, 2021.

[Sou20] Perla Sousi. Mixing times of markov chains. https://www.math.ubc.ca/~jhermon/Mixing/mixing-notes.pdf, 2020. Accessed: 2023-06-12.

[SSDK+20] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[STW11] Jie Shen, Tao Tang, and Li-Lian Wang. *Spectral methods: algorithms, analysis and applications*, volume 41. Springer Science & Business Media, 2011.

[Sze04] Mario Szegedy. Quantum speed-up of markov chain based algorithms. In *45th Annual IEEE symposium on foundations of computer science*, pages 32–41. IEEE, 2004.

[TD00] Barbara M Terhal and David P DiVincenzo. Problem of equilibration and the computation of correlation functions on a quantum computer. *Physical Review A*, 61(2):022301, 2000.

[vAGGdW17] Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Quantum sdp-solvers: Better upper and lower bounds. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 403–414. IEEE, 2017.

[Vem05] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[vH16] Ramon van Handel. Probability in high dimension. https://web.math.princeton.edu/~rvan/APC550.pdf, 2016. Accessed: 2022-09-20.

[VN+51] John Von Neumann et al. Various techniques used in connection with random digits. *Applied Math Series*, 12(36-38):1, 1951.

[WA08] Pawel Wocjan and Anura Abeyesinghe. Speedup via quantum sampling. *Physical Review A*, 78(4):042336, 2008.

[YLC14] Theodore J Yoder, Guang Hao Low, and Isaac L Chuang. Fixed-point quantum search with an optimal number of queries. *Physical review letters*, 113(21):210501, 2014.

APPENDICES

Appendix A

Lemmas used in Section 2.2

Lemma A.0.1. *For any integer $m \geq 0$ and $z > 0$*

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

Proof. We note that

$$\sum_{k=0}^{\infty} \frac{z^k\, k^m}{k!} = \frac{\partial^m}{\partial x^m}\bigg|_{x=0} \sum_{k=0}^{\infty} \frac{z^k\, e^{kx}}{k!} = \frac{\partial^m}{\partial x^m}\bigg|_{x=0} e^{z\, e^x}. \tag{A.2}$$

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

$$\frac{\partial^s}{\partial x^s} f(x) = \sum_{r=1}^{s} C_s[r] \; (g(x))^r \; f(x) \tag{A.3}$$

Taking derivative of both sides yields the following recursive relations

$$C_{s+1}[r] = \begin{cases} C_s[1], & \text{if } r = 1, \\ r\, C_s[r] + C_s[r-1], & \text{if } 2 \leq r \leq s, \\ C_s[s], & \text{if } r = s+1. \end{cases} \tag{A.4}$$

Therefore

$$\sum_{r=1}^{s+1} C_{s+1}[r] = \sum_{r=1}^{s} C_s[r]\, (r+1) \leq (s+1) \sum_{r=1}^{s} C_s[r] \tag{A.5}$$

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

$$\sum_{r=1}^{m} C_m[r] \leq m!. \tag{A.6}$$

Lastly, we note that

$$\left.\frac{\partial^m}{\partial x^m}\right|_{x=0} e^{z\,e^x} \leq e^z \max\{z^m, 1\} \sum_{r=1}^{m} C_m[r] \leq e^z\, z^m\, m! \tag{A.7}$$

since $g(0) = z$, and $f(0) = e^z$, and the last inequality follows from (A.6). Combined with (A.2) the result follows. □

Lemma A.0.2. *For any $z > 1$ and $m \in \mathbb{N}$, the following holds*

$$\sum_{k=0}^{\infty} k^m\, z^{-k} \leq \frac{1}{1 - z^{-1}} \max\left(2, \frac{2}{z-1}\right)^m m!. \tag{A.8}$$

Proof. First, we note that the function we are upper bounding is a special case of the Lerch transcendents [GR], for which we also provide a lower bound in Example 2.2.3. 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

$$\sum_{k\geq 0} k^m z^{-k} = \frac{\partial^m}{\partial \alpha^m} \sum_{k\geq 0} e^{\alpha k} = \frac{\partial^m}{\partial \alpha^m} \frac{1}{1 - e^{\alpha}}. \tag{A.9}$$

By a simple induction, we arrive at the following form

$$\frac{\partial^m}{\partial \alpha^m} \frac{1}{1 - e^{\alpha}} = \sum_{r=1}^{m} C_m[r] \frac{e^{r\alpha}}{(1 - e^{\alpha})^{r+1}}, \tag{A.10}$$

with the following recursive relation for the coefficients

$$C_{s+1}[r] = \begin{cases} C_s[1], & \text{if } r = 1, \\ r\,(C_s[r] + C_s[r-1]) & \text{if } 2 \leq r \leq s, \\ r C_s[s], & \text{if } r = s+1. \end{cases} \tag{A.11}$$

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,

$$\frac{\partial^m}{\partial \alpha^m} \frac{1}{1 - e^{\alpha}} \leq \frac{1}{1 - z^{-1}} \max\{1, \frac{1}{z-1}\}^m 2^m\, m!. \tag{A.12}$$

□

Lemma A.0.3. *For any $m \geq 1$*

$$\sum_{k=0}^{m} k!(m-k)! \leq 3\, m!. \tag{A.13}$$

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

$$\sum_{k=0}^{m} k!(m-k)! = m! \sum_{k=0}^{m} \frac{1}{\binom{m}{k}} \tag{A.14}$$

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

$$\sum_{k=0}^{m} \frac{1}{\binom{m}{k}} = 2 + \frac{2}{m} + \sum_{k=2}^{m-2} \frac{1}{\binom{m}{k}} \leq 2 + \frac{2}{m} + \frac{m-3}{m} \leq 3. \tag{A.15}$$

□

Lemma A.0.4. *For all $d \geq 1$*

$$\sum_{\substack{i_1,i_2,\cdots,i_d \in \{0,\cdots,m\} \\ i_1+i_2+\cdots+i_d=m}} i_1!\, i_2! \cdots i_d! \leq 3^{d-1}\, m!. \tag{A.16}$$

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

$$\begin{aligned} \sum_{i_1+i_2+\cdots+i_d=m} i_1!\, i_2! \cdots i_{d+1}! &= \sum_{i_1=0}^{m} i_1! \left(\sum_{i_2+\cdots+i_{d+1}=m-i_1} i_2! \cdots i_{d+1}! \right) \\ &\leq 3^{d-1} \sum_{i_1=0}^{m} i_1\, (m-i_1)! \\ &\leq 3^d\, m! \end{aligned} \tag{A.17}$$

where the last step is due to Lemma A.0.3. □

Lemma A.0.5. *Let m be an integer greater than d. We have*

$$\max_{x \in [-1,1]^d} \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|x + 2\,p\|^{-2m} \leq 2^{d+1} \tag{A.18}$$

for a universal constant ξ.

Proof. Firstly, we note that

$$\max_{x \in [-1,1]^d} \sum_{p \in \mathbb{Z}^d \setminus \{0\}} \|x + 2\,p\|^{-2m} \leq \sum_{j=1}^{d} \binom{d}{j} \max_{x \in [-1,1]^j} \sum_{p \in (\mathbb{Z} \setminus \{0\})^j} \frac{1}{\|x + 2p\|^{2m}} \tag{A.19}$$

$$\leq \sum_{j=1}^{d} \binom{d}{j} \max_{x \in [-1,1]^j} \sum_{p \in (\mathbb{Z} \setminus \{0\})^j} \frac{1}{\|x + 2p\|^{2j}}. \tag{A.20}$$

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:

$$\begin{aligned}\max_{x\in[-1,1]^d}\sum_{p\in\mathbb{Z}^d\setminus\{0\}}\|x+2\,p\|^{-2m} &\le \sum_{j=1}^{d}\binom{d}{j}\left(\frac{1}{\sqrt{j}}+\frac{1}{2^j}\int_{r=1}^{\infty}\frac{dV_j}{r^{2j}}\right)\\ &\le \sum_{j=1}^{d}\binom{d}{j}\left(\frac{1}{\sqrt{j}}+\frac{\pi^{j/2}}{2^j\Gamma(j/2+1)}\int_{r=1}^{\infty}r^{-1-j}\,dr\right)\\ &= \sum_{j=1}^{d}\binom{d}{j}\left(\frac{1}{\sqrt{j}}+\frac{\pi^{j/2}}{2^j\Gamma(j/2+1)}\frac{1}{j}\right)\\ &\le 2^d\sup_{j\in\mathbb{N}}\left(\frac{1}{\sqrt{j}}+\frac{\pi^{j/2}}{2^j\Gamma(j/2+1)}\frac{1}{j}\right)=2^{d+1}\end{aligned} \tag{A.21}$$

□

Lemma A.0.6 (Lemma 13 of [BCOW17]). *Let $\overrightarrow{a}$ and $\overrightarrow{b}$ be two vectors of the same vector space. It is the case that*

$$\left\|\frac{\overrightarrow{a}}{\|\overrightarrow{a}\|}-\frac{\overrightarrow{b}}{\left\|\overrightarrow{b}\right\|}\right\|\le\frac{2\left\|\overrightarrow{a}-\overrightarrow{b}\right\|}{\max\left[\|\overrightarrow{a}\|,\left\|\overrightarrow{b}\right\|\right]}. \tag{A.22}$$

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

$$\begin{aligned}\left\|\frac{\overrightarrow{a}}{\|\overrightarrow{a}\|}-\frac{\overrightarrow{b}}{\left\|\overrightarrow{b}\right\|}\right\| &= \left\|\frac{\overrightarrow{a}}{\|\overrightarrow{a}\|}-\frac{\overrightarrow{b}}{\|\overrightarrow{a}\|}+\frac{\overrightarrow{b}}{\|\overrightarrow{a}\|}-\frac{\overrightarrow{b}}{\left\|\overrightarrow{b}\right\|}\right\|\\ &\le \left\|\frac{\overrightarrow{a}}{\|\overrightarrow{a}\|}-\frac{\overrightarrow{b}}{\|\overrightarrow{a}\|}\right\|+\left\|\frac{\overrightarrow{b}}{\|\overrightarrow{a}\|}-\frac{\overrightarrow{b}}{\left\|\overrightarrow{b}\right\|}\right\|\\ &\le \frac{2\left\|\overrightarrow{a}-\overrightarrow{b}\right\|}{\|\overrightarrow{a}\|}\end{aligned} \tag{A.23}$$

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

Lemma A.0.7. *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*

$$\mathrm{TV}\left(P_\psi,P_\phi\right)\le\||\psi\rangle-|\phi\rangle\|. \tag{A.24}$$

Proof. We have

$$\begin{aligned}
\mathrm{TV}\left(P_\psi, P_\phi\right) &= \frac{1}{2}\sum_{i\in I}\left||\psi(i)|^2 - |\phi(i)|^2\right| \\
&= \frac{1}{2}\sum_{i\in I}\big|\,|\psi(i)\,| - |\,\phi(i)|\big|\cdot\big||\psi(i)| + |\phi(i)|\,\big| \\
&\le \frac{1}{2}\sqrt{\sum_{i\in I}\big|\,|\psi(i)\,| - |\,\phi(i)|\big|^2}\sqrt{\sum_{i\in I}\big|\,|\psi(i)\,| + |\,\phi(i)|\big|^2} \\
&\overset{(a)}{\le} \frac{1}{\sqrt{2}}\sqrt{\sum_{i\in I}|\psi(i)-\phi(i)|^2}\sqrt{\sum_{i\in I}|\psi(i)|^2+|\phi(i)|^2} \\
&= \||\psi\rangle - |\phi\rangle\|
\end{aligned} \tag{A.25}$$

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

Lemma A.0.8. *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*

$$M = \left\lceil \frac{1}{\delta}\frac{L\ell d/2 + 10/3\sqrt{2}\,ae^4C}{\mathcal{U}} \right\rceil \tag{A.26}$$

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

Proof. We have

$$\mathrm{TV}(\mu,\widehat{\mu}) = \frac{1}{2}\int dx\left|\frac{u^2}{\ell^d\mathcal{U}^2} - \sum_{n\in[-M..M]^d}\mathbf{1}_{\{x\in\mathbb{B}_n\}}\frac{u_M^2[n]}{\left\|\overrightarrow{u_M}\right\|^2}\left(\frac{2M+1}{\ell}\right)^d\right|. \tag{A.27}$$

Using the triangle inequality, we have

$$\begin{aligned}
\mathrm{TV}(\mu,\widehat{\mu}) &\le \frac{1}{2}\int dx\left|\frac{u^2}{\ell^d\mathcal{U}^2} - \sum_{n\in[-M..M]^d}\mathbf{1}_{\{x\in\mathbb{B}_n\}}\frac{u_M^2[n]}{\mathcal{U}^2\ell^d}\right| \\
&\quad + \frac{1}{2}\int\frac{dx}{\ell^d}\sum_{n\in[-M..M]^d}\mathbf{1}_{\{x\in\mathbb{B}_n\}}u_M^2[n]\cdot\left|\frac{1}{\mathcal{U}^2} - \frac{(2M+1)^d}{\left\|\overrightarrow{u_M}\right\|^2}\right| \\
&= \frac{1}{2\ell^d\mathcal{U}^2}\int dx\sum_{n\in[-M..M]^d}\mathbf{1}_{\{x\in\mathbb{B}_n\}}\left|u^2 - u_M^2[n]\right| + \frac{1}{2\mathcal{U}^2}\left|\frac{\left\|\overrightarrow{u_M}\right\|^2}{(2M+1)^d} - \mathcal{U}^2\right| \\
&\overset{(a)}{\le} \frac{L\ell d}{\mathcal{U}(2M+1)} + \frac{2\sqrt{2}e^3C}{\mathcal{U}}e^{-3M/5a}
\end{aligned} \tag{A.28}$$

where in (a) we use Lemma 2.2.2 with the choice of M in (A.26). Using the inequality

$e^{-x/e} \leq 1/x$, we obtain

$$\mathrm{TV}(\mu, \widehat{\mu}) \leq \frac{L\ell d/2 + 10/3\sqrt{2}\, a e^4 C}{\mathcal{U} M} \tag{A.29}$$

which concludes the proof. □

Appendix B

Langevin diffusion on a torus

Let $\{Y_t\}_{t\geq 0}$ be a continuous time stochastic process in $\mathbb{R}^d$ satisfying the overdamped Langevin dynamics at thermodynamic $\beta = 1$,

$$dY_t = -\nabla E(Y_t)\, dt + \sqrt{2}\, dW_t \tag{B.1}$$

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} [Pav14, 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 [GPSMH19] 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, [Hsu02] discusses stochastic calculus on manifolds.

We start with a periodic energy function with period ℓ in all dimensions. It is shown in [GPSMH19, 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 (B.1) given by

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

The Fokker–Planck equation associated to (B.1) viewed as an Itô stochastic differential equation

$$\partial_t \sigma(y,t) = \nabla \cdot \left(e^{-E} \nabla \left(e^{E}\, \sigma(y,t) \right) \right) \tag{B.3}$$

describes the evolution of the probability density function $\sigma(-,t)$ of Y_t (see [Pav14] for

more details). The probability density $\rho_t = \rho(-,t)$ of the wrapped process X_t satisfies

$$\rho(x,t) = \sum_{k\in\mathbb{Z}^d} \sigma(y + k\,\ell) \tag{B.4}$$

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.

$$\partial_t \rho(x,t) = \nabla \cdot \left(e^{-E}\, \nabla \left(e^{E}\, \rho(x,t) \right) \right) \tag{B.5}$$

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

$$\mathcal{L}(-) = \nabla \cdot \left(e^{-E}\, \nabla \left(e^{E}\, - \right) \right). \tag{B.6}$$

Remark B.0.1. *It is worth mentioning that the Fokker–Planck generator is usually written in the form*

$$\mathcal{L}(-) = \nabla^2 E(-) + \nabla E \cdot \nabla(-) + \nabla^2(-) \tag{B.7}$$

however as we will see in Appendix B.1, the generator is better behaved under discretization when it is written in the form (B.6). Nevertheless, discretizing all the derivatives in the usual Fokker–Planck equation results in an operator of the same form

$$\widetilde{\nabla^2} E(-) + \widetilde{\nabla} E \cdot \widetilde{\nabla}(-) + \widetilde{\nabla^2}(-) = \widetilde{\nabla} \cdot \left(e^{-E}\, \widetilde{\nabla} \left(e^{E}\, - \right) \right) \tag{B.8}$$

where the tilde on top represents Fourier differentiation operators. This is due to the fact that discrete Fourier differentiation obeys Leibniz's rules (see Section 2.2.2 and Appendix B.1).

Remark B.0.2. *We can derive the uniqueness of the stationary state of the toroidal Fokker–Planck equation by considering the operator*

$$\mathcal{L}' := e^{E/2} \circ \mathcal{L} \circ e^{-E/2}. \tag{B.9}$$

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

$$\langle \rho, \mathcal{L}'\rho \rangle = -\int_{x\in[-L/2,L/2]^d} e^{-E} \left\| \nabla \left(e^{E/2} \rho \right) \right\|^2 \leq 0 \tag{B.10}$$

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 [MV00, Ber11, BGL+14, vH16]. A functional inequality known as the *Poincarè inequality* is equivalent to exponentially fast convergence of Langevin diffusion [BGL+14, 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.0.1. *Given the toroidal Fokker–Planck equation (B.5), 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*

$$\mathrm{Var}_{\rho_s}[f] \leq \lambda \, \mathbb{E}_{\rho_s}\left[\|\nabla f\|^2\right]. \tag{B.11}$$

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

$$\|\rho_t/\rho_s - 1\|_{L^2(\rho_s)} \leq e^{-t/\lambda}\|\rho_0/\rho_s - 1\|_{L^2(\rho_s)}. \tag{B.12}$$

Proof. Let us denote the ratio of the distributions by $h_t = h(-,t) = \rho_t/\rho_s$. Using the Fokker–Planck equation (B.3) one has $\partial_t h = \rho_s^{-1}\nabla\cdot(\rho_s\,\nabla h)$, which implies equality (a)

$$\frac{d}{dt}\int_{x\in[-\ell/2,\ell/2]^d}\rho_s\,(h-1)^2 = 2\int_{x\in[-\ell/2,\ell/2]^d}\rho_s\,(h-1)\partial_t h \tag{B.13}$$

$$\overset{(a)}{=} -2\int_{x\in[-\ell/2,\ell/2]^d}\rho_s\|\nabla h\|^2. \tag{B.14}$$

Note that $\mathbb{E}_{\rho_s}[h] = 1$ and hence, the left hand side of (B.13) is the time derivative of $\mathrm{Var}_{\rho_s}[h_t]$, while the right hand side of (B.14) 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.0.3. *We note that by Jensen's inequality* [1]

$$\mathbb{E}_{\rho_s}[|\rho_t/\rho_s - 1|] \leq \sqrt{\mathbb{E}_{\rho_s}(\rho_t/\rho_s - 1)^2} \tag{B.15}$$

which yields

$$\int_{x\in[-\ell/2,\ell/2]^d}|\rho_t(s) - \rho_s(x)| \leq \sqrt{\int_{x\in[-\ell/2,\ell/2]}\rho_s(x)\left(\frac{\rho_t}{\rho_s} - 1\right)^2}. \tag{B.16}$$

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

$$\mathrm{TV}(\rho_t,\rho_s) \leq \frac{1}{2}\sqrt{\mathrm{Var}_{\rho_s}[\rho_t/\rho_s]}. \tag{B.17}$$

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

Proposition B.0.2. *Let E be an ℓ-periodic energy potential with a bounded range Δ. Then for all ℓ-periodic $f \in L^2(\rho_s)$*

$$\mathrm{Var}_{\rho_s}[f(X)] \leq \frac{\ell^2\,e^{\Delta}}{4\pi^2}\,\mathbb{E}_{\rho_s}\left[\|\nabla f\|_2^2\right]. \tag{B.18}$$

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

Proof. We have

$$\begin{aligned}\mathrm{Var}_{\rho_s}\left[f(X)\right] &= \mathrm{Var}_{\rho_s}\left[f(X) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right] \\ &\leq \mathbb{E}_{\rho_s}\left[\left(f(X) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right)^2\right] \\ &\leq \frac{e^{-\min_x E(x)}}{Z} \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \left(f(x) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right)^2\end{aligned} \tag{B.19}$$

and also, due to Parseval's theorem

$$\int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \left(f(x) - \int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \frac{1}{\ell^d} f(x)\right)^2 = \ell^d \sum_{k\in\mathbb{Z}^d\setminus\{0\}} \left|\widehat{f}[k]\right|^2, \tag{B.20}$$

where $\widehat{f}$ is the Fourier transform of f. Note that the $k=0$ term is excluded in (B.20) 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

$$\int_{x\in[-\frac{\ell}{2},\frac{\ell}{2}]^d} \|\nabla f\|^2 = \ell^d \sum_{k\in\mathbb{Z}^d} \frac{4\pi^2}{\ell^2}\|k\|^2 \left|\widehat{f}[k]\right|^2 \geq \ell^d \sum_{k\in\mathbb{Z}^d\setminus\{0\}} \frac{4\pi^2}{\ell^2} \left|\widehat{f}[k]\right|^2. \tag{B.21}$$

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 (B.19), (B.20), and (B.21) prove the claim. □

Corollary B.0.1. *In Proposition B.0.1 we have $\lambda = \frac{\ell^2 e^{\Delta}}{4\pi^2}$ as the universal Poincaré constant.*

B.1 Discretization of the FPE

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

$$\begin{aligned}\mathbb{L} : \mathcal{V}_N &\to \mathcal{V}_N \\ \overrightarrow{f} &\mapsto \widetilde{\nabla}\cdot\left(e^{-E}\widetilde{\nabla}(e^{E}\overrightarrow{f})\right)\end{aligned} \tag{B.22}$$

where tilde on the top of derivative is used to represent Fourier derivative operators (see Section 2.2.2). Note that by the product rule of the Fourier derivative operator (1) we can rewrite $\mathbb{L}$ in terms of derivatives of e^{-E} and the function that $\mathbb{L}$ acts on as follows:

$$\mathbb{L}(-) = e^{E}\left(\left\|\widetilde{\nabla}e^{-E}\right\|^2 + \widetilde{\nabla^2}e^{-E}\right)(-) - e^{E}\,\widetilde{\nabla}e^{-E}\cdot\widetilde{\nabla}(-) + \widetilde{\nabla^2}(-). \tag{B.23}$$

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.1.1. *The discrete operator (B.22) has the following properties.*

1. *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.,* $D \leq 0$*).*

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

3. *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 1 we have

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

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 4). Hence, $\mathbb{L}'$ is symmetric. Note that this concludes $\mathbb{L}$ being diagonalizable, and moreover, since $\kappa_U \leq e^{\Delta/2}$, we also have $\kappa_V \leq e^{\Delta/2}$. Next, we show that $\mathbb{L}' \preccurlyeq 0$.

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

where (a) and (b) follow from 1 and 3, 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 $\left\|e^{-E}\right\|\left\|e^{E}\right\| \leq e^{\Delta}$, and $\left\|\widetilde{\nabla}\right\| \leq \sqrt{d}\,\frac{2\pi N}{\ell}$, and also $\left\|\left(\widetilde{\nabla}\cdot\right)\right\| \leq \sqrt{d}\frac{2\pi N}{\ell}$. Furthermore, using (B.8), the triangle inequality, and 2, we conclude that $\|\mathbb{L}\| \leq \frac{dN^2}{\ell^2}\left(4\pi^2 + 48^2(\ln N)^2\Delta + 96\pi\Delta\ln N\right)$, 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.1.2. *Consider the differential equation* $\frac{d}{dt}\overrightarrow{u} = \mathbb{L}\,\overrightarrow{u}$*, with initial condition* $\overrightarrow{u}(0) = \mathbf{1}$*. We have*

$$\sup_{t\geq 0} \left\| \overrightarrow{u}(t) \right\| \leq e^{\Delta/2} \left\| \mathbf{1} \right\|, \tag{B.24}$$

$$\inf_{t\geq 0} \left\| \overrightarrow{u}(t) \right\| = \left\| \mathbf{1} \right\| \tag{B.25}$$

Proof. We write the solution as $\overrightarrow{u}(t) = e^{\mathbb{L}t}\overrightarrow{u}(0)$. From 1 we have $\left\|e^{\mathbb{L}t}\right\| = \left\|V^{-1}\,e^{Dt}\,V\right\| \leq \|V\|\left\|V^{-1}\right\|\left\|e^{Dt}\right\|$. Therefore $\kappa_V = \|V\|\left\|V^{-1}\right\|$ and $\left\|e^{Dt}\right\| \leq 1$ imply that $\left\|e^{\mathbb{L}t}\right\| \leq \kappa_V \leq e^{\Delta/2}$. This proves (B.24).

For (B.25) we use the fact that $\langle \mathbf{1}|\,\mathbb{L} = 0$ (which follows from 3), to conclude that $\langle \mathbf{1}, u(t)\rangle = \langle \mathbf{1}, u(0)\rangle$. Using the Cauchy-Schwartz inequality one has

$$\|u(t)\|\,\|\mathbf{1}\| \geq \langle \mathbf{1}, u(0)\rangle \tag{B.26}$$

and given $\langle \mathbf{1}, u(0)\rangle = \|\mathbf{1}\|^2$, we have $\|u(t)\| \geq \|\mathbf{1}\|$. 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.2.1. *Let* $u(x,t)$ $(\forall t \geq 0)$ *be a solution to the Fokker–Planck equation* (B.3)*. 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 (B.3) v satisfies

$$\partial_t v = -\nabla E\cdot\nabla v + \nabla^2 v = \mathcal{L}^* v \tag{B.27}$$

which is the backward Kolmogorov equation with $\mathcal{L}^*$ the adjoint of the operator $\mathcal{L}$ (see for instance [Pav14] or [vH16, 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 (B.27) 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

$$v(x,t+s) = \mathbb{E}\left[v(X_{t+s},t)\middle|X_t = x\right] \quad (\forall s,t\geq 0) \tag{B.28}$$

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

$$v(x,t+s) \leq \max_{y\in\mathbb{T}} v(y,t) \quad (\forall x). \tag{B.29}$$

It now suffices to take the maximum of the left hand side of (B.29) 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 2.2.2.

Lemma B.2.2. *Let $\overrightarrow{u}$ denote the discretization of the solution to the Fokker–Planck equation ((B.27)), 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*

$$\left\| \frac{d}{dt}\overrightarrow{u(t)} - \mathbb{L}\overrightarrow{u(t)} \right\| \leq 8\times 10^5 \pi e^3 \, \frac{\sqrt{d}\, e^{\Delta}\, C\,(1+\ell L/48)(a^3+a^2)}{\ell^2}\, (2N+1)^{d/2}\, e^{-0.4N} \tag{B.30}$$

for every point $t\geq 0$ in time, where L is the Lipschitz constant of E. [2]

Proof. We write

$$\begin{aligned}\mathbb{L}\overrightarrow{u} - \frac{d}{dt}\overrightarrow{u} = {} & e^{E}\left(\left\|\widetilde{\nabla}e^{-E}\right\|^2 - \left\|\nabla e^{-E}\right\|^2\right)\overrightarrow{u} + e^{E}\left(\widetilde{\nabla^2}e^{-E} - \nabla^2 e^{-E}\right)\overrightarrow{u} \\ & + e^{E}\left(\nabla e^{-E}\cdot\nabla\overrightarrow{u} - \widetilde{\nabla}e^{-E}\cdot\widetilde{\nabla}u\right) + \left(\widetilde{\nabla^2}\overrightarrow{u} - \overrightarrow{\nabla^2 u}\right)\end{aligned} \tag{B.31}$$

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

$$\begin{aligned}&\sum_{n\in[-N..N]^d}\left| e^{E[n]}\,u[n]\,\left(\left\|\widetilde{\nabla}e^{-E}\right\|^2 - \left\|\nabla e^{-E}\right\|^2\right)_{[n]}\right|^2 \\ &\quad\leq \left(\max_x e^{E(x)}u(x)\right)^2 \sum_{n\in[-N..N]^d}\left|\left\|\widetilde{\nabla}e^{-E}\right\|^2 - \left\|\nabla e^{-E}\right\|^2\right|^2 \\ &\quad\overset{(a)}{\leq} e^{2\max_x E(x)}\left[\sum_{n\in[-N..N]^d}\left(\left\|\widetilde{\nabla}e^{-E}\right\| - \left\|\nabla e^{-E}\right\|\right)^2_{[n]}\left(\left\|\widetilde{\nabla}e^{-E}\right\| + \left\|\nabla e^{-E}\right\|\right)^2_{[n]}\right] \\ &\quad\overset{(b)}{\leq} d\,e^{2\Delta}\left(\frac{48}{\ell}N\ln N + L\right)^2 \sum_{n\in[-N..N]^d}\left\|\widetilde{\nabla}e^{-E} - \nabla e^{-E}\right\|^2_{[n]} \\ &\quad\overset{(c)}{\leq} A_1\frac{dC^2a^2\,e^{2\Delta}}{\ell^4}\left(N\ln N + \frac{\ell L}{48}\right)^2 (2N+1)^d\, e^{-N/a}\end{aligned} \tag{B.32}$$

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

where we have used Lemma B.2.1 in (a), a triangle inequality of ℓ_2 norms together with 2 in (b), and inequality (2.63) 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 (B.31), we use inequality (2.64) and Lemma B.2.1 to write

$$\sum_{n\in[-N.N]^d} \left| e^{E[n]}\, u[n] \left(\widetilde{\nabla^2} e^{-E} - \nabla^2 e^{-E} \right) \right|^2 \leq A_2 \frac{C^2 a^4\, e^{2\Delta}}{\ell^4}\, (2N+1)^d\, e^{-0.8N/a} \tag{B.33}$$

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

We rewrite the third term as

$$e^{E[n]} \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right)_{[n]} = e^{E[n]} \left[\left(\nabla e^{-E} - \widetilde{\nabla} e^{-E} \right) \cdot \widetilde{\nabla} u + \nabla e^{-E} \cdot \left(\nabla u - \widetilde{\nabla} u \right) \right] \tag{B.34}$$

which allows us to conclude that

$$\begin{aligned} \left| e^{E[n]} \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right)_{[n]} \right| &\leq e^{E[n]} \left| \left(\nabla e^{-E} - \widetilde{\nabla} e^{-E} \right) \cdot \widetilde{\nabla} u \right| + \left| \nabla E[n] \cdot \left(\nabla u - \widetilde{\nabla} u \right) \right| \\ &\leq e^{E[n]} \left\| \nabla e^{-E} - \widetilde{\nabla} e^{-E} \right\| \left\| \widetilde{\nabla} u \right\| + \|\nabla E\| \left\| \nabla u - \widetilde{\nabla} u \right\|. \end{aligned} \tag{B.35}$$

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

$$\begin{aligned} \left\| e^E \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right) \right\|^2 \leq\ & 2 \frac{d\, e^{2 \max_x E(x)}}{\ell^4} (48 N \log N)^2 \left\| \nabla e^{-E} - \widetilde{\nabla} e^{-E} \right\|^2 \\ & + 2dL^2 \left\| \nabla u - \widetilde{\nabla} u \right\|^2 \end{aligned} \tag{B.36}$$

which by applying (2.63) implies

$$\left\| e^E \left(\nabla e^{-E} \cdot \nabla u - \widetilde{\nabla} e^{-E} \cdot \widetilde{\nabla} u \right) \right\|^2 \leq 2A_1 \frac{d\, e^{2\Delta}\, C^2 a^2}{\ell^4} \left(N \ln N + \frac{\ell L}{48} \right)^2 (2N+1)^d\, e^{-N/a}. \tag{B.37}$$

Finally the last term of (B.31) is taken care of directly using (2.64):

$$\left\| \widetilde{\nabla^2} u - \nabla^2 u \right\| \leq \sqrt{A_2} \frac{C\, a^2}{\ell^2}\, (2N+1)^{d/2}\, e^{-0.4N} \tag{B.38}$$

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 $N \ln N \leq N^2 \leq 100 a^2 e^{0.1N/a}$, and combined with (B.32), (B.33), (B.37), (B.38), and (B.31) yield

$$\left\| \frac{d}{dt} \overrightarrow{u(t)} - \mathbb{L} \overrightarrow{u(t)} \right\| \leq A \frac{\sqrt{d} e^{\Delta} C\, (1 + \frac{\ell L}{48})(a^3 + a^2)}{\ell^2}\, (2N+1)^{d/2}\, e^{-0.4N} \tag{B.39}$$

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

Proposition B.2.1. *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,*

$$\left\| \overrightarrow{u}(T) - \overrightarrow{v}(T) \right\| \leq 1.6 \times 10^6 \pi e^3 \frac{T\, e^{3\Delta/2}\, C^2\, (a^3 + a^2)\, (1 + \frac{\ell L}{48})}{\ell^2} (2N+1)^{d/2}\, e^{-0.4\frac{N}{a}}. \tag{B.40}$$

Proof. For convenience we will denote the right hand side of (B.30) as $f[N]$ (which defers from the right hand side of (B.40) 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 $\left\|\overrightarrow{e}(t)\right\|$ is upper bounded in Lemma B.2.2. Note also that $\frac{d}{dt}\overrightarrow{v} = \mathbb{L}\overrightarrow{v}$ by definition. Hence, letting $\overrightarrow{\varepsilon} := \overrightarrow{u} - \overrightarrow{v}$ we get

$$\frac{d}{dt}\overrightarrow{\varepsilon}(t) = \mathbb{L}\overrightarrow{\varepsilon}(t) + \overrightarrow{e}(t). \tag{B.41}$$

By 1, $\mathbb{L} = V^{-1} D V$ 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 (B.41) from left by V, and let $\overrightarrow{\mathcal{E}} := V\overrightarrow{\varepsilon}$ and $\overrightarrow{b} := V\overrightarrow{e}$ to get

$$\frac{d}{dt}\overrightarrow{\mathcal{E}}(t) = D\,\overrightarrow{\mathcal{E}}(t) + \overrightarrow{b}(t). \tag{B.42}$$

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

$$\langle \overrightarrow{\mathcal{E}}, \frac{d}{dt}\overrightarrow{\mathcal{E}} \rangle = \langle \overrightarrow{\mathcal{E}}, D\,\overrightarrow{\mathcal{E}} \rangle + \langle \overrightarrow{\mathcal{E}}(t), \overrightarrow{b}(t) \rangle. \tag{B.43}$$

Since D is negative semi-definite

$$\mathrm{Re}\left(\langle \overrightarrow{\mathcal{E}}, \frac{d}{dt}\overrightarrow{\mathcal{E}} \rangle \right) \leq \mathrm{Re}\left(\langle \overrightarrow{\mathcal{E}}(t), \overrightarrow{b}(t) \rangle \right) \leq \left\| \overrightarrow{\mathcal{E}} \right\| \left\| \overrightarrow{b} \right\|. \tag{B.44}$$

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

$$\frac{d}{dt}\left\| \overrightarrow{\mathcal{E}}(t) \right\| \leq \left\| \overrightarrow{b}(t) \right\|. \tag{B.45}$$

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

$$\frac{d}{dt}\|V\overrightarrow{\varepsilon}\| \leq f[N]\|V|e\rangle\|. \tag{B.46}$$

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

$$\left\| V\,\overrightarrow{\varepsilon}(T) \right\| \leq T f[N] \sigma_{\max}(V). \tag{B.47}$$

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

Appendix C

Training EBMs

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

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

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 iid 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:

$$\text{KL}(p_{\text{data}}(x)||p_\theta(x)) = -\mathbb{E}_{x\sim p_{\text{data}}}[\log p_\theta(x)] + \text{constant}. \tag{C.2}$$

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

$$\mathbb{E}_{p_{\text{data}}}[\nabla_\theta \log p_\theta(x)] = -\beta \mathbb{E}_{p_{\text{data}}}[\nabla_\theta E_\theta(x)] + \beta \mathbb{E}_{p_\theta}[\nabla_\theta E_\theta(x)]. \tag{C.3}$$

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 ([SK21]).

The energy function is a composition of linear functions (affine transformations) with nonlinear ones such as the sigmoid function [1] for which we can build a quantum oracle as in Fig. 1.1. Our algorithm works by queries to an oracle for the discretization of the generator of the Fokker–Planck equation L (see (B.22)). In order to construct this oracle, we use the expression (B.8) restated here as

$$\mathbb{L}(-) = \widetilde{\nabla^2} E(-) + \widetilde{\nabla} E \cdot \widetilde{\nabla}(-) + \widetilde{\nabla^2}(-) \tag{C.4}$$

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

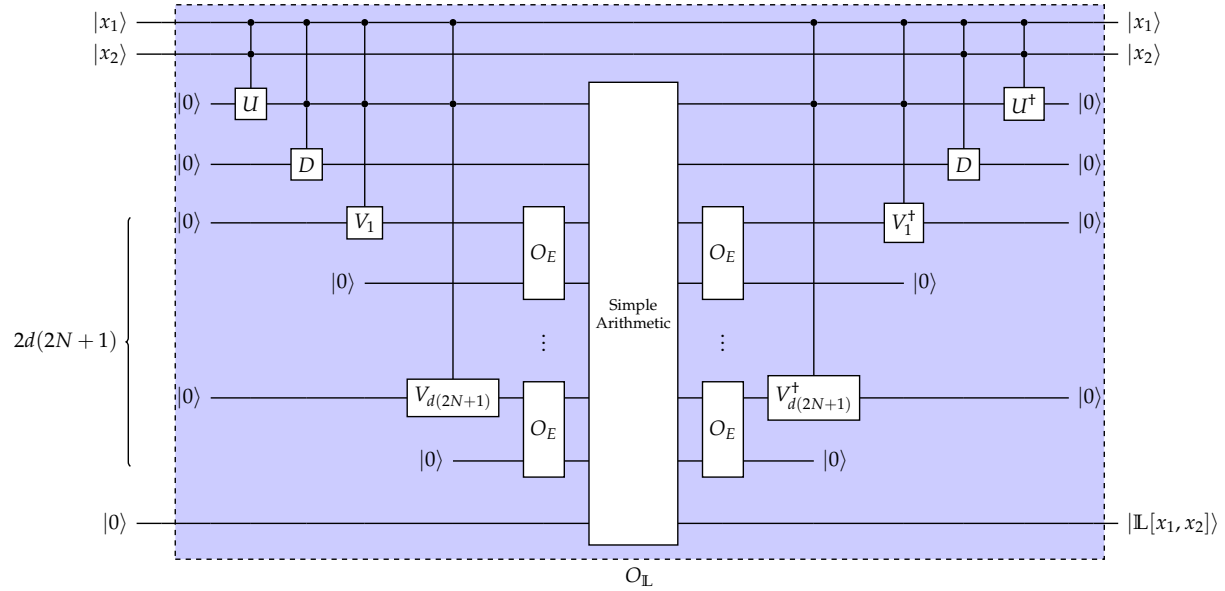

Figure C.1: The circuit for the oracle of discrete generator $\mathbb{L}$ comprising $2d(2N+1)$ copies of the energy potential oracle, O_E. To query $\mathbb{L}[x_1, x_2] = \langle x_2 | \mathbb{L} | x_1 \rangle$, first the controlled-$U$ gate checks for the difference between x_1 and x_2: the third 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 Section 2.2.2) and is applied via a circuit performing simple arithmetic.

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

After iterative queries to the oracle of $\mathbb{L}$, Section 3.2.1 returns a sample from the model distribution of the EBM. Repeated executions will then provide an approximation of the second term in (C.3). 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. C.1 in order to perform mean estimations of the components of the gradient $\nabla_\theta E_\theta(x)$ as per Corollary 3.2.2 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.

Appendix D

Proofs of the results in Section 3.2.2

D.1 Gibbs sampling

Theorem D.1.1 (Theorem 3.2.1 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). Section 3.2.1 samples from a distribution ε-close to the Gibbs distribution (in total variation distance), by making*

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

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 $\mathrm{polylog}(Cade^{\Delta}(1+\ell L))/\varepsilon)$.

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.2.1 we have

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

where we have also used Lemma A.0.6 and Lemma B.1.2. Hence, if we let $A = 3.2\times 10^6\pi e^3$, we may choose

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

to guarantee an at most ε distance between $\left|e^{\mathbb{L}T}\,\rho_0\right\rangle$ and $|\rho_T\rangle = \left|e^{\mathcal{L}T}\,\rho_0\right\rangle$. Now we apply Theorem 3.1.5 to obtain an output state $|\mathcal{A}\rangle$, for which we have $\||\mathcal{A}\rangle - |\rho_T\rangle\| \leq 2\varepsilon$. Hence, by Theorem 2.2.1, 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 D.3.1, together with Corollary B.0.1 implies that

choosing

$$T = \kappa_{E/2} \log\left(2e^{\Delta/2}/\varepsilon\right) \tag{D.3}$$

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 3.1.5 is now obtained by noting firstly that $\kappa_V \leq e^{\Delta/4}$ from 1 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.1.2. The norm of $\mathbb{L}$ is bounded by 3 as $\|\mathbb{L}\| = O(\Delta dN^2/\ell^2 \operatorname{polylog} N)$. Finally $\beta \leq 1$ for us also using Lemma B.1.2. This provides the complexity of every term in Theorem 3.1.5 with respect to N. We also note that

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

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 [KO23] 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. In [BHMT02], the authors consider having access to a unitary U, that acts as $U|0\rangle = \sqrt{p}|0\rangle + \sqrt{1-p}|0^\perp\rangle$, where $|0^\perp\rangle$ 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

$$|\psi\rangle = \frac{1}{2}e^{i\theta}\left(|0\rangle - i|0^\perp\rangle\right) + \frac{1}{2}e^{-i\theta}\left(|0\rangle + i|0^\perp\rangle\right) \tag{D.5}$$

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, [Mon15, LW18] extended the above algorithm and obtained mean estimation algorithms for more generic cases. [HM18] combines the latter algorithms and obtains a desired complexity of $\mathcal{O}^*(\frac{1}{\varepsilon})$. The recent work of [KO23] 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 [KO23]. We restate the following theorem from their work.

Theorem D.2.1 (Theorem 1.3 of [KO23])**.** *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*

$$\mathbb{P}\left[|\widehat{\mu} - \mathbb{E}[f(X)]| \geq \frac{\mathrm{Var}[f(X)]}{n}\right] \leq \delta. \tag{D.6}$$

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.2.1 (Corollary 3.2.2 in the manuscript)**.** *Let E be an energy function, satisfying the assumptions made in Theorem D.1.1. 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*

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

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 Section 3.2.1, 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[1] [Gro05]. 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

$$\frac{d/2(\ell L_f + \ell L\Delta_f + 10\sqrt{2}/3e^4\Delta_f aC)}{M} + 2\Delta_f\varepsilon_1, \tag{D.8}$$

[1]The $\frac{\pi}{3}$-amplitude amplification algorithm of [Gro05] has a dependence on the success probability of the algorithm, which is later improved in the works of [YLC14, GSLW19]. However, as the success probability of our circuit is $\Omega(1)$, we do not need to utilize the more complex algorithms.

which follows from the total variation bounds obtained above, and further, that of Lemma A.0.8. Hence, implementing line 3 of Section 3.2.1 with $M = \frac{\text{poly}(C,a,\Delta_f,L_f\ell,L\ell)}{\varepsilon}$, 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.2.1 concludes the result. □

D.3 Lemmas used in Appendix D.1

Lemma D.3.1. *Let $u(x)$ be an ℓ-periodic real-valued function satisfying*

$$\sqrt{\int_{\mathbb{T}} \rho_s \left(\frac{u}{\rho_s} - 1\right)^2} \leq \delta \sqrt{\int_{\mathbb{T}} \rho_s \left(\frac{1}{V\rho_s} - 1\right)^2} \tag{D.9}$$

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

$$\frac{1}{2}\int_{\mathbb{T}} \left| \frac{u^2}{\int_{\mathbb{T}} u^2} - \frac{\rho_s^2}{\int_{\mathbb{T}} \rho_s^2} \right| \leq 2\,\delta\, e^{\Delta/2}. \tag{D.10}$$

Proof. Note that from the assumption

$$\sqrt{\int_{\mathbb{T}} (u - \rho_s)^2} \leq e^{\Delta/2}\,\delta \sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}. \tag{D.11}$$

By a similar argument as in Lemma A.0.7 we have

$$\frac{1}{2}\int_{\mathbb{T}} \left| \frac{u^2}{\int u^2} - \frac{\rho_s^2}{\int_{\mathbb{T}} \rho_s^2} \right| \leq \sqrt{\int_{\mathbb{T}} \left(\frac{u}{\sqrt{\int_{\mathbb{T}} u^2}} - \frac{\rho_s}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} \right)^2}. \tag{D.12}$$

Furthermore, using the triangle inequality (c.f., Lemma A.0.6)

$$\sqrt{\int_{\mathbb{T}} \left(\frac{u}{\sqrt{\int_{\mathbb{T}} u^2}} - \frac{\rho_s}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} \right)^2} \leq 2\,\frac{\sqrt{\int_{\mathbb{T}} (u - \rho_s)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}}. \tag{D.13}$$

Now, putting equations (D.11), (D.12), and (D.13) together, we have

$$\frac{1}{2}\int_{\mathbb{T}} \left| \frac{u^2}{\int u^2} - \frac{\rho_s^2}{\int_{\mathbb{T}} \rho_s^2} \right| \leq 2\delta e^{\Delta/2} \frac{\sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}}. \tag{D.14}$$

Now consider a random variable X drawn uniformly at random from $\mathbb{T}$, and define

$Y := \rho_s(X)$. It is clear that

$$\frac{\sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} = \sqrt{\frac{\mathbb{E}\left[\left(Y - \frac{1}{V}\right)^2\right]}{\mathbb{E}\left[Y^2\right]}}. \tag{D.15}$$

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

$$\frac{\sqrt{\int_{\mathbb{T}} \left(\frac{1}{V} - \rho_s\right)^2}}{\sqrt{\int_{\mathbb{T}} \rho_s^2}} = \sqrt{\frac{\mathrm{Var}\left[Y\right]}{\mathbb{E}\left[Y^2\right]}} \leq 1. \tag{D.16}$$

Combining the latter equation with (D.14) completes the proof. □