ABSTRACT

Title of Dissertation: CLASSICAL AND QUANTUM ALGORITHMIC DEVELOPMENTS IN LATTICE FIELD THEORY

Hyunwoo Oh
Doctor of Philosophy, 2026

Dissertation Directed by: Professor Paulo F. Bedaque
Department of Physics

Lattice field theory provides a powerful framework for studying non-perturbative aspects of quantum chromodynamics (QCD), enabling first-principles numerical calculations of hadron properties and high-precision theoretical predictions of the Standard Model. Despite its successes, several longstanding computational challenges hinder progress in frontier regimes. Among these are the sign problem, which makes simulations exponentially costly in certain settings; the signal-to-noise problem, which limits precision for key observables; and the infinite variance problem, which can render estimators ill-defined. Addressing these issues is essential for advancing Monte Carlo simulations.

This thesis explores new strategies for overcoming these obstacles using classical and quantum computational methods. On the classical side, it develops and analyzes algorithms for variance reduction in Monte Carlo simulations. Approaches include reweighting schemes to eliminate the infinite variance problem; contour deformations and complex normalizing flows to mitigate the sign problem; and control variates—constructed to respect symmetry and using machine learning—to reduce

the signal-to-noise problem. These methods are benchmarked on representative lattice field theory models.

On the quantum side, the thesis investigates algorithms for efficient eigenstate preparation, a central requirement for simulating lattice field theory on quantum devices. Methods based on Hamiltonian paths are studied, including the rodeo projection algorithm, projection-based schemes inspired by the quantum Zeno effect, and adiabatic evolution. Their computational costs are analyzed in terms of path length and volume of the system, with proposed improvements that reduce fluctuations and demonstrate a potential quantum advantage in simulating quantum field theory on quantum computers.

Together, these developments contribute to tackling the numerical barriers that arise in simulating strongly interacting quantum many-body systems. The results presented here highlight crucial advancements in both classical variance reduction and quantum computational methodologies, laying the groundwork for more accurate simulations of non-perturbative QCD in the future.

CLASSICAL AND QUANTUM ALGORITHMIC DEVELOPMENTS IN LATTICE FIELD THEORY

by

Hyunwoo Oh

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2026

Advisory Committee:
Professor Paulo F. Bedaque, Chair/Advisor
Professor Thomas D. Cohen, Co-chair/Co-Advisor
Professor Zackaria Chacko
Professor Manuel Franco Sevilla
Professor Maria Cameron

Acknowledgments

I have been fortunate to study physics in Maryland and I owe much of this experience to the support of many wonderful people.

First and foremost, I am deeply grateful to my advisor, Paulo Bedaque, for giving me the opportunity to join the nuclear theory group. His guidance and mentorship have taught me how to approach research and have been fundamental to my academic growth.

I also would like to sincerely thank my co-advisor, Tom Cohen, for not only teaching me physics but for always being there to listen to my thoughts and concerns. I am proud to have been able to produce more publications together than I could have ever imagined.

I have been lucky to collaborate with an incredible group of people along the way. In chronological order, my thanks go to Andrei Alexandru, Andrea Carosso, Scott Lawrence, Yukari Yamauchi, Srijit Paul, Maneesha Pradeep, Veronica Wang, and Andrew Li. Each of you played a crucial role in making this journey both productive and enjoyable.

Living in Maryland is an unforgettable experience, and much of that was thanks to my fellow graduate students. I thank Edward Broadberry, Michael Cervia, Yonatan Gazit, Yanda Geng, Emily Jiang, Edison Murairi, Andy Sheng, Lucas Smith, and Dan Sokratov for the fun lunch breaks on the third floor of PSC and off-campus gatherings. I am especially grateful to Jaehong Choi, Donghee Yeum, and Youngjoon Lee for their support and camaraderie.

None of this would have been possible without the constant encouragement of my family. Their love and support kept me going through the toughest moments, and I will always be grateful.

Lastly, I want to express my deepest gratitude to my new family, Christin Lee. I cannot imagine this journey without your unwavering support. I am truly here because of you, and I am beyond grateful to have you by my side.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview

Quantum chromodynamics (QCD) is the fundamental theory that describes the strong interaction—one of the four known fundamental forces in nature. It governs the behavior of quarks and gluons, the elementary constituents of hadrons such as protons and neutrons, and thereby underlies the structure of nuclei and the forces that bind them.

A defining feature of QCD is asymptotic freedom, whereby the strength of the interaction between quarks becomes weaker at high energies or short distances, making perturbation theory a useful tool in that regime. However, at low energies where the rich phenomenology of hadronic and nuclear physics emerges, the coupling becomes strong, and perturbative techniques break down. As a result, understanding hadron structure, confinement, chiral symmetry breaking, and nuclear forces from first principles requires fully non-perturbative methods.

One of the most powerful tools for exploring non-perturbative QCD is lattice field theory, where spacetime is discretized into a finite lattice, and the infinite-dimensional path integral is replaced by a finite-dimensional integral. This formulation regularizes the theory while maintaining gauge invariance, making it suitable for numerical computations using classical computers. Over the past decades, lattice QCD has led to major insights into the mass spectrum of hadrons, the internal structure of nucleons, and weak matrix elements relevant for testing the Standard Model and searching for new physics.

This thesis focuses on the numerical challenges that arise in performing first-principles calculations of QCD on the lattice. Specifically, it addresses these challenges both in the Euclidean path integral formalism commonly used in classical Monte Carlo simulations and in the Hamiltonian for-

mulation that provides a natural framework for future quantum simulations.

While Monte Carlo methods have proven indispensable for evaluating the high-dimensional integrals encountered in lattice QCD, they are not without significant limitations. Three major issues are the sign problem, which makes simulations exponentially costly in regimes such as finite chemical potential or real-time evolution; the signal-to-noise problem, which limits the precision of observables such as baryon correlation functions; and the infinite variance problem, which can render certain estimators ill-defined even in principle. These issues pose serious barriers to progress in several frontier areas of QCD and nuclear theory.

This thesis systematically explores both classical and quantum computational strategies for overcoming the numerical challenges in simulating lattice field theories. It proposes new algorithms, analyzes their performance, and connects them to ongoing theoretical and computational developments in lattice field theory.

To lay the groundwork for these investigations, Section 1.2 provides a brief overview of QCD, summarizing its essential features and phenomenological implications. Section 1.3 introduces lattice field theory and explains how it provides a framework for regularizing and simulating quantum field theories numerically on a computer.

Chapter 2 addresses several fundamental numerical challenges encountered in lattice field theory calculations. Section 2.2 begins with an overview of Monte Carlo methods, which remain the primary tool for evaluating high-dimensional path integrals in lattice formulations. In Section 2.3, we examine the infinite variance problem that often arises in the computation of fermionic observables and present a strategy to eliminate this pathology. Section 2.4 focuses on the sign problem, a major obstacle in many lattice simulations, and explores approaches for mitigating its severity, including contour deformation and the use of complex normalizing flows. Finally, Section 2.5 analyzes the signal-to-noise problem,

a common issue that limits the precision of observables at large time separations, and introduces the control variates method, an unbiased technique to reduce variance and enhance statistical efficiency.

Chapter 2 draws upon the following publications:

- Andrei Alexandru, Paulo F. Bedaque, Andrea Carosso, and Hyunwoo Oh. Infinite variance problem in fermion models. Phys. Rev. D **107**, 094502 (2023).
- Hyunwoo Oh, Andrei Alexandru, Paulo F. Bedaque, and Andrea Carosso. A solution for infinite variance problem of fermionic observables. PoS, LATTICE2023 021.
- Scott Lawrence, Hyunwoo Oh, and Yukari Yamauchi. Lattice scalar field theory at complex coupling. Phys. Rev. D **106**, 114503 (2022).
- Paulo F. Bedaque and Hyunwoo Oh. Leveraging neural control variates for enhanced precision in lattice field theory. Phys. Rev. D **109**, 094519 (2024).
- Hyunwoo Oh. Control variates with neural networks. PoS, LATTICE2023 051.
- Hyunwoo Oh. Training neural control variates using correlated configurations. Phys. Rev. D **112**, 074501 (2025).

Chapter 3 focuses on the problem of quantum state preparation for simulating lattice field theories on quantum computers—a crucial step for realizing accurate and efficient quantum simulations. The chapter begins in Section 3.1 with a review of quantum simulation approaches for lattice gauge theories, setting the stage for subsequent developments. In Section 3.2, we analyze the rodeo projection algorithm, highlighting its limitations and presenting an improved method that reduces fluctuations in excited-state suppression. Section 3.3 introduces the notion of a Hamiltonian path and defines the path length as a metric for quantifying the complexity of state preparation methods. Section 3.4 explores a

method inspired by the quantum Zeno effect, which offers an alternative projection-based mechanism for preparing ground states. Section 3.5 provides a comprehensive discussion of adiabatic state preparation, including a review of the adiabatic theorem, the switching theorem, and an analysis of how errors behave in different regimes.

Chapter 3 is based on the following publications:

- Thomas D. Cohen and Hyunwoo Oh. Optimizing the rodeo projection algorithm. Phys. Rev. A **108**, 032422 (2023).
- Thomas D. Cohen and Hyunwoo Oh. Efficient vacuum-state preparation for quantum simulation of strongly interacting local quantum field theories. Phys. Rev. A **109**, L020402 (2024).
- Thomas D. Cohen and Hyunwoo Oh. Asymptotic errors in adiabatic evolution. Phys. Rev. A **111**, 042612 (2025).
- Thomas D. Cohen, Andrew Li, Hyunwoo Oh, and Maneesha Sushama Pradeep. Practical limitations of the switching theorem for adiabatic state preparation. Eur. Phys. J. D **80**, 41 (2026).
- Thomas D. Cohen and Hyunwoo Oh. Corrections to adiabatic behavior for long paths. Phys. Rev. A **110**, 062601 (2024).
- Thomas D. Cohen, Hyunwoo Oh, and Veronica Wang. Numerical study of computational cost of maintaining adiabaticity for long paths. *arXiv*:2412.08626.

This thesis concludes in Chapter 4 with a discussion of its broader implications and outlines potential directions for future research building on the results presented.

1.2 Quantum chromodynamics

Quantum chromodynamics (QCD) [4] is the quantum field theory (QFT) that describes the strong interaction—the fundamental force responsible for binding quarks and gluons into hadrons and nuclei. It is a non-Abelian gauge theory with the $SU(3)$ gauge group, which represents the local symmetry of color charge. The degrees of freedom in QCD include fermion quark fields, which transform in the fundamental representation of $SU(3)$, and gluon fields which transform in the adjoint representation and mediate the strong force.

The classical Lagrangian density of QCD for N_f quark flavors is given by

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} F^a_{\mu\nu} F^{a\,\mu\nu} + \sum_{f=1}^{N_f} \bar{\psi}_f (i\gamma^\mu D_\mu - m_f)\psi_f, \tag{1.1}$$

where

- $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$ is the gluon field strength tensor,
- A^a_μ are the gluon gauge fields,
- f^{abc} are the structure constants of the $SU(3)$ Lie algebra,
- ψ_f is the Dirac spinor for the fth quark flavor,
- $D_\mu = \partial_\mu - i g A^a_\mu T^a$ is the covariant derivative acting on quark fields,
- T^a are the generators of $SU(3)$ in the fundamental representation.

The gauge fields couple to the quark fields via the covariant derivative, ensuring local $SU(3)$ gauge invariance of the Lagrangian.

The QCD Lagrangian is invariant under local $SU(3)$ gauge transformations:

$$\psi(x) \to U(x)\psi(x), \quad A_\mu(x) \to U(x)A_\mu(x)U^\dagger(x) + \frac{i}{g}U(x)\partial_\mu U^\dagger(x), \tag{1.2}$$

where $A_\mu \equiv \sum_a A_\mu^a T^a$ and $U(x) \in SU(3)$. This invariance requires the introduction of eight gluon fields A_μ^a corresponding to the eight generators of the gauge group. The non-Abelian nature of $SU(3)$ leads to self-interactions among the gluons, encoded in the nonlinear terms of the field strength tensor.

While gauge invariance is often referred to as a symmetry, it is fundamentally distinct from global symmetries that correspond to conserved physical quantities via Noether's theorem [5]. Local gauge invariance does not correspond to a physical symmetry of the system in the usual sense; rather it reflects a redundancy in our description of the degrees of freedom. It allows us to write the theory in a local way, introducing gauge fields that mediate interactions while encoding the redundancy in the field representation. Despite this redundancy, gauge invariance plays a crucial role: it dictates the structure of the interactions and ensures consistency with the principles of quantum field theory. In particular, it leads to the self-interacting gauge bosons in non-Abelian theories, enforces the form of covariant derivatives, and constrains the allowable counterterms in the renormalization procedure.

One of the most profound features of QCD is asymptotic freedom [6, 7]: the effective interaction between quarks becomes weaker at short distances or high energies. This behavior is quantified through the renormalization group and the QCD beta function, which governs the running of the strong coupling constant $\alpha_s(\mu) \equiv \frac{g^2(\mu)}{4\pi}$ with respect to the renormalization scale μ.

At one-loop order, the beta function in $SU(N_c)$ gauge theory is

$$\beta(g) = \mu\frac{dg}{d\mu} = -\frac{g^3}{16\pi^2}\left(\frac{11}{3}C_A - \frac{4}{3}N_f T_F\right), \tag{1.3}$$

where

- $C_A = N_c$ is the Casimir invariant of the adjoint representation for $SU(N_c)$,
- $T_F = \frac{1}{2}$ is the trace convention of the fundamental representation,
- N_f is the number of quark flavors.

For QCD, this yields

$$\beta(g) = -\frac{g^3}{16\pi^2}\left(11 - \frac{2}{3}N_f\right). \tag{1.4}$$

For $N_f < 16.5$, the parenthesis is positive, making the beta function negative, i.e., the coupling decreases at high energy scales. Integrating the renormalization group equation gives the running of α_s:

$$\alpha_s(\mu) \approx \frac{2\pi}{\beta_0 \ln(\mu/\Lambda_{\text{QCD}})}, \tag{1.5}$$

with $\beta_0 = 11 - \frac{2}{3}N_f$, and Λ_{QCD} is the QCD scale parameter. This expression shows the logarithmic weakening of the coupling at high energies—a behavior confirmed experimentally in deep inelastic scattering and jet physics [8].

While asymptotic freedom enables perturbative calculations at high energies, at low energies the coupling becomes large, and perturbation theory breaks down. This regime is characterized by confinement, where quarks and gluons are bound into color-neutral hadrons. The non-perturbative dynamics of confinement, spontaneous chiral symmetry breaking, and hadron mass generation are central phenomena in QCD and are typically studied using lattice formulations.

1.3 Lattice field theory

Quantum field theories often require ultraviolet (UV) regularization to handle the short-distance divergences that arise in loop integrals. Traditional methods such as dimensional regularization [9] and Pauli-Villars regularization [10] are analytically elegant and widely used in perturbative treatments. However, for strongly coupled regimes where perturbation theory fails, such as low-energy QCD, a non-perturbative regularization scheme is essential. Lattice field theory provides such a framework by discretizing spacetime into a finite grid with lattice spacing a, effectively imposing a UV cutoff at momentum scales of order π/a. Additionally, this approach makes the path integral well-defined.

In lattice regularization, the continuum spacetime is replaced by a hypercubic grid. The continuum limit corresponds to taking $a \to 0$, while keeping physical quantities such as the volume $V = (aL)^4$ fixed. In practice, this means performing calculations at multiple values of a and extrapolating to zero lattice spacing. This discretization makes the theory finite-dimensional and amenable to numerical computation, albeit at the cost of breaking some continuum symmetries, most notably Lorentz invariance.

Gauge symmetry, however, can be preserved exactly on the lattice. This was a central insight in Wilson's formulation of lattice gauge theory [11], where the fundamental gauge degrees of freedom are represented not by the gauge fields $A_\mu(x) \in \mathfrak{su}(N)$, where $\mathfrak{su}(N)$ represents the Lie algebra of the Lie group $SU(N)$, but by link variables $U_\mu(x) \in SU(N)$, which live on the links between adjacent lattice sites and represent parallel transporters. The Wilson action is constructed from these link variables and ensures exact local gauge invariance even at finite lattice spacing.

For fermions, discretization introduces additional challenges. A naive lattice transcription of the fermion action in Eq. (1.1) leads to the fermion doubling problem, in which 2^d fermion species (where

d is the number of spacetime dimensions) emerge in the continuum limit. Wilson's solution [11] involves adding a second derivative term, the Wilson term, to lift the doublers by giving them masses of order $1/a$, though this explicitly breaks chiral symmetry. Alternative fermion formulations—such as staggered, domain wall, and overlap fermions—have been developed to address the trade-offs between preserving chiral symmetry, computational cost, and continuum behavior.

To systematically understand and reduce discretization effects, the Symanzik effective theory [12, 13] provides a powerful framework. In this approach, the lattice theory is viewed as a deformation of the continuum QFT by higher-dimensional operators:

$$\mathcal{L}_{\text{lattice}} = \mathcal{L}_{\text{continuum}} + a \cdot \mathcal{L}_1 + a^2 \cdot \mathcal{L}_2 + \cdots \tag{1.6}$$

where each correction $\mathcal{L}_n$ contains higher-dimensional operators consistent with the symmetries of the lattice action. These corrections vanish in the continuum limit, but at finite lattice spacing they contribute systematic errors that can be mitigated using improved actions.

In addition to UV regularization, practical lattice calculations must also regulate the infrared divergence by simulating in a finite spacetime volume. This introduces finite-volume effects, particularly relevant for light hadrons like the pion. To suppress these artifacts, it is standard to require $m_\pi L \gtrsim 4$, ensuring that the Compton wavelength of the lightest hadron is well-contained within the box.

Over the past 50 years, lattice field theory has undergone tremendous development—both in theoretical formulations and numerical methodologies—enabling precise calculations of hadron masses, matrix elements, and thermodynamic properties from first principles. Nonetheless, open challenges remain that call for both theoretical and numerical advancements, particularly in the study of multi-particle systems, resonances, and real-time dynamics. While theoretical developments continue to

address some of these issues, this thesis focuses on the numerical aspects, which will be discussed in detail in the next chapter.

Chapter 2: Numerical problems in lattice field theory

2.1 Overview

Lattice methods have played a pivotal role in uncovering the non-perturbative dynamics of quantum chromodynamics (QCD), providing reliable insights into hadron structure, confinement, chiral symmetry breaking, and other low-energy phenomena. By discretizing spacetime and formulating QCD on a Euclidean lattice, one can rigorously define the path integral and evaluate it using Markov Chain Monte Carlo methods. This approach has been extraordinarily successful in calculating static properties of hadrons, such as masses, decay constants, and parton distributions, with increasing precision. However, despite these accomplishments, lattice field theory faces significant numerical challenges that restrict its applicability to some of the most pressing questions in nuclear and particle physics.

Foremost among these challenges is the sign problem, which arises in a variety of contexts including finite baryon density, real-time evolution, and systems with a topological theta term. In such cases, the path integral measure becomes complex, rendering probabilistic interpretation—and thus importance sampling—invalid. This leads to exponential degradation in the signal-to-noise ratio as the system size or time increases, making direct simulations prohibitively expensive and completely infeasible. Although some limited progress has been made through analytic continuation, reweighting, and Lefschetz thimble techniques, a general solution to the sign problem remains elusive and is widely regarded as one of the grand challenges in computational physics.

Compounding these difficulties is the signal-to-noise problem, which can persist even when the sign problem is absent. This issue arises when physical correlators decay exponentially with Euclidean time, while the associated statistical noise either decays more slowly or remains constant. As a result,

the signal-to-noise ratio deteriorates exponentially with time, making it increasingly difficult to extract long-distance physics. For example, in baryonic correlation functions, the signal decays e^{-Mt}, where M is the mass of the state of interest, whereas the noise decays more slowly, often governed by a different spectrum. This exponential degradation severely limits the ability of classical lattice simulations to reliably extract quantities such as mass, matrix elements, or form factors at large time separations.

A third, less widely appreciated, but equally problematic issue is the infinite variance problem, particularly relevant in fermionic systems and reweighting schemes. This occurs when the observable being sampled has a heavy-tailed distribution due to near-zeros in the reweighting factor or determinant, leading to divergence of higher moments and rendering standard error estimates unreliable. In such cases, even a large number of Monte Carlo samples may not suffice to produce a statistically stable result, and naive application of statistical methods can produce misleading conclusions.

These challenges are not merely technical inconveniences; they impose fundamental limitations on our ability to probe important aspects of strongly interacting systems. They hinder the study of QCD at finite baryon density—a regime relevant to the physics of neutron stars and heavy-ion collisions—as well as the extraction of transport coefficients, spectral functions, and real-time observables necessary for understanding the dynamical response of quantum systems.

This chapter is devoted to a detailed exploration of these numerical challenges and the methods designed to address them.

- Section 2.2 reviews the standard Monte Carlo approach where the problems arise.
- Section 2.3 examines the infinite variance problem originating from the fermionic path integral and presents a potential remedy based on reweighting using an extra time-slice probability distribution [14, 15].

- Section 2.4 addresses the sign problem in lattice field theory, with a focus on recent developments using contour deformation and complex normalizing flows—tools adapted from modern machine learning [16].

- Section 2.5 discusses the signal-to-noise problem in lattice field theory and demonstrates how control variates—a standard statistical variance reduction technique—can be integrated with machine learning to systematically improve the statistical efficiency of Monte Carlo estimators [17–19].

2.2 Monte Carlo method

In quantum physics, measurable quantities are encoded in observables, which are represented mathematically by Hermitian operators. These observables encode all accessible information about a quantum system and their expectation values are directly related to experimental outcomes. In the path integral formulation of quantum field theory, the expectation value of an observable O is expressed as a weighted integral over all possible field configurations. Specifically, for gauge theories such as QCD, this takes the form:

$$\langle O \rangle = \frac{1}{Z} \int [DU]\, O(U) \mathrm{e}^{-S[U]}. \tag{2.1}$$

where $[DU]$ is the integration measure over gauge field configurations U, $S[U]$ is the Euclidean action, and $Z = \int [DU]\, \mathrm{e}^{-S[U]}$ is the partition function.

Despite its utility in theoretical physics, the path integral is not rigorously defined in the continuum, due to the infinite-dimensional nature of the integration space and the lack of a well-defined measure. However, by introducing a lattice regularization, one can render the theory mathematically well-defined and amenable to numerical computation. This approach discretizes spacetime into a fi-

nite hypercubic grid with lattice spacing a and finite volume $V = T \times L^3$, replacing the continuum fields defined on lattice sites and links. The continuum limit is approached by taking $a \to 0$ while maintaining a fixed physical volume or performing controlled extrapolations, which requires that the parameters of the action depend on a.

In QCD, the gauge group is $SU(3)$, corresponding to the symmetry group of the strong interactions. Consider a relatively modest lattice volume of 10^4 sites (far smaller than those used in state-of-the-art simulations). At each lattice site, there are 4 gauge links (one in each spacetime direction), and each link is an $SU(3)$ matrix, which can be parameterized by 8 real degrees of freedom (the number of generators of the group). The resulting dimension of the integration space is then approximately

$$\dim[DU] = 4 \times 8 \times 10^4 = 3.2 \times 10^5, \tag{2.2}$$

which is already well beyond the reach of standard quadrature or deterministic integration techniques. In practice, modern lattice QCD simulations often operate on significantly larger lattices, such as 64^4, pushing this dimensionality into the hundreds of millions or more.

To tackle such high-dimensional integrals, Monte Carlo integration is the only viable and efficient method. The essential idea of Monte Carlo techniques is to estimate expectation values via statistical sampling. Crucially, the exponential weighting $\mathrm{e}^{-S[U]}$ defines a natural probability distribution over field configurations, allowing one to preferentially sample those configurations that dominate the path integral, which is known as importance sampling. By generating an ensemble of contributions $\{U_i\}$ distributed according to the Boltzmann weight $\mathrm{e}^{-S[U]}$, the expectation value of an observable is approximated as

$$\langle O \rangle \approx \frac{1}{N} \sum_{i=1}^{N} O(U_i), \tag{2.3}$$

with the Markov Chain ergodic theorem ensuring convergence to the true expectation value in the limit of a large sample size N.

The success of lattice gauge theory relies heavily on the efficiency and reliability of these importance-sampling-based Monte Carlo methods, which are capable of navigating extremely high-dimensional spaces. Algorithms such as Metropolis [20, 21], Hybrid Monte Carlo [22], and heat bath updates [23] (Gibbs sampler) have been developed and optimized to enable accurate calculations of non-perturbative quantities such as hadron masses, decay constants, and form factors.

2.2.1 Markov Chain Monte Carlo and error estimation

Markov Chain Monte Carlo (MCMC) methods form a powerful set of algorithms designed to approximate complex probability distributions by generating samples through iterative procedures. These methods are especially useful in situations where analytical integration or direct sampling is not feasible. The core idea is to construct a Markov chain whose stationary distribution coincides with the target distribution. After an initial burn-in phase, the chain is assumed to have equilibrated, and samples drawn thereafter can be used to estimate statistical properties such as means, variances, and higher moments.

One important issue in MCMC is the autocorrelation between successive samples. Since each sample is generated conditionally from the previous one, strong correlations can persist along the chain, especially if mixing is poor. This autocorrelation reduces the number of effectively independent samples, thereby degrading the precision of statistical estimates.

To quantify the uncertainty of an observable from a finite sample, the jackknife method [24, 25] is often employed. It is a resampling-based approach for estimating the variance of an estimator.

Given a dataset $x_1, x_2, \ldots, x_n$ and an estimator $\hat{\theta}$ computed from the full sample, jackknife replicates are formed by leaving out one data point at a time:

$$\hat{\theta}_{(i)} = \hat{\theta}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n). \tag{2.4}$$

The jackknife estimate of the variance is then

$$\mathrm{Var}_{\mathrm{jack}}(\hat{\theta}) = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}_{(i)} - \bar{\theta} \right)^2, \tag{2.5}$$

where $\bar{\theta}$ is the average over all jackknife replicates, and the corresponding error (standard deviation) is $\mathrm{err}_{\mathrm{jack}}(\hat{\theta}) = \sqrt{\mathrm{Var}_{\mathrm{jack}}(\hat{\theta})}$. This method assumes that the data points are statistically independent, an assumption that fails in the presence of autocorrelation, such as in MCMC samples.

To address this, the blocked jackknife (or jackknife with binning) is used. Here, the data is grouped into B non-overlapping bins, each of size b, so that the total number of samples is $n = B \cdot b$. Each bin j contains the subset $x_{(j-1)b+1}, \ldots, x_{jb}$, and a jackknife replicate is formed by omitting one entire bin at a time:

$$\hat{\theta}_{(j)} = \hat{\theta}(\text{data excluding bin } j). \tag{2.6}$$

The variance estimate then becomes

$$\mathrm{Var}_{\mathrm{jack,b}}(\hat{\theta}) = \frac{B-1}{B} \sum_{j=1}^{B} \left(\hat{\theta}_{(j)} - \bar{\theta} \right)^2. \tag{2.7}$$

Choosing an appropriate bin size helps to mitigate the impact of autocorrelation within bins, leading to more accurate error estimates. In practice, the bin size is increased until the estimated variance

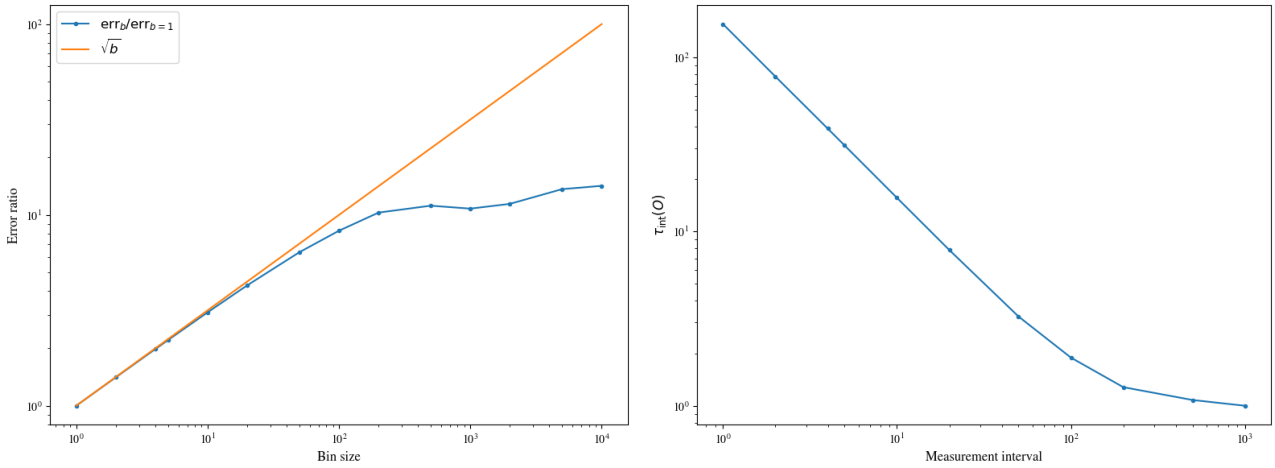

Figure 2.1: Estimation of blocked jackknife errors from 10^5 correlated configurations. *Left*: Ratio of the jackknife error at each bin size to the error at bin size 1, highlighting the growth of the estimated error due to autocorrelations. *Right*: Integrated autocorrelation time across different measurement intervals.

plateaus, indicating that autocorrelations are adequately captured.

The integrated autocorrelation time, τ_{int}, is a standard way to quantify autocorrelation in a Markov Chain. It is defined as

$$\tau_{\text{int}}(\theta) \equiv 1 + 2\sum_{i=1}^{\infty} \frac{\text{Cov}(\theta_k, \theta_{k+i})}{\sigma^2}, \tag{2.8}$$

with $\sigma^2 = \text{Var}(\theta)$ denoting the variance of the observable. This quantity characterizes how strongly each configuration is correlated with later ones and effectively tells us how many correlated measurements correspond to a single independent sample. Accordingly, if $\bar{\theta}$ is estimated from n samples, the variance is increased by τ_{int}:

$$\text{Var}(\bar{\theta}) = \frac{\sigma^2}{n}\tau_{\text{int}}. \tag{2.9}$$

Put differently, autocorrelation decreases the effective number of independent measurements by a factor of τ_{int}, which must be included when assessing statistical errors.

The left panel of Figure 2.1 illustrates error estimation using the blocked jackknife method for different bin sizes. When the bin size is too small to capture the full extent of correlations, the estimated error grows approximately as $\sqrt{b}$, where b is the bin size. This occurs because residual correlations remain within bins, leading the jackknife method to underestimate the true variance in this regime.

The right panel of Figure 2.1 displays how τ_{int} varies with the measurement interval. As the interval increases, autocorrelations between successive samples diminish, and τ_{int} approaches 1, corresponding to statistically independent configurations.

2.2.2 Improved estimators

It is important to emphasize that Monte Carlo estimation, when implemented correctly, provides an unbiased estimator for path integrals: the sampling procedure does not introduce a systematic bias into the estimated observable. In other words, for a finite ensemble of configurations drawn according to a properly normalized probability distribution, the expectation value of the Monte Carlo estimate equals its exact value $\langle O \rangle$; as the number of samples increases, the estimate converges to this value due to the law of large numbers. The only form of error in such estimations is statistical, stemming from the finite number of the samples.

However, even when estimators are unbiased, high variance can significantly hinder the practical usefulness of the results. Observables with large variances require a correspondingly larger number of samples to achieve a desired level of precision. Thus, much of the effort in improving Monte Carlo methods is directed toward variance reduction—that is, improving the efficiency of estimators while trying to avoid altering their expectation values. Two broad strategies exist for achieving this.

The first strategy involves replacing a high-variance observable O with another observable $\tilde{O}$

that shares the same expectation value but has a lower variance:

$$\langle \tilde{O} \rangle = \langle O \rangle \quad \text{and} \quad \mathrm{Var}(\tilde{O}) \leq \mathrm{Var}(O). \tag{2.10}$$

This idea lies at the heart of the control variates method, which will be discussed in detail in Section 2.5. Control variates exploit correlations between the observable of interest and other quantities whose expectation values are either known exactly or can be computed with higher precision. By subtracting a suitably scaled correlated quantities, one can reduce the fluctuations of the target observable while keeping its expectation value intact.

The second approach involves modifying the probability distribution used to generate samples. Instead of sampling configurations according to the given distribution $P(\phi)$, one may choose a different distribution $\tilde{P}(\phi)$, and reweight the observable accordingly:

$$\langle O \rangle_P = \int [D\phi]\, O(\phi) P(\phi) = \frac{\int [D\phi]\, O(\phi) \frac{P(\phi)}{\tilde{P}(\phi)} \tilde{P}(\phi)}{\int [D\phi]\, \frac{P(\phi)}{\tilde{P}(\phi)} \tilde{P}(\phi)} = \frac{\langle OW \rangle_{\tilde{P}}}{\langle W \rangle_{\tilde{P}}}. \tag{2.11}$$

Here, one samples from $\tilde{P}(\phi)$, a possibly cheaper or more stable distribution, and corrects for the mismatch by reweighting each observable using the ratio $W(\phi) \equiv P(\phi)/\tilde{P}(\phi)$. More explicitly, if $\tilde{O} \equiv \frac{\sum_{i=1}^{N} O_i W_i}{\sum_{i=1}^{N} W_i}$ is the new estimator using reweighting,

$$\mathrm{Var}(\tilde{O}) \approx \frac{1}{N \langle W \rangle_{\tilde{P}}^2} \langle W^2 (O - \langle O \rangle_P)^2 \rangle_{\tilde{P}}{}^{[1]}. \tag{2.12}$$

Therefore, if

$$\frac{1}{\langle W \rangle_{\tilde{P}}^2} \langle W^2 (O - \langle O \rangle_P)^2 \rangle_{\tilde{P}} \lesssim \langle (O - \langle O \rangle_P)^2 \rangle_P, \tag{2.13}$$

[1] It can be derived using the delta method in the first order.

one can achieve variance reduction using a new observable. This technique is widely used in lattice field theory, especially in cases where sampling directly from $P(\phi)$ is difficult or numerically unstable, such as in finite-density QCD or simulations near phase transitions. Additionally, this technique can be exploited when samples $\phi \sim P(\phi)$ are already available but one would like to study physics with $\tilde{P}(\phi)$ instead.

While this reweighting formula in Eq. (2.11) is formally unbiased, its practical implementation with finite sample sizes introduces an important subtlety. The problem lies in the estimation of the denominator in Eq. (2.11). To see this explicitly, consider a simplified case in which we estimate the reciprocal of the mean of a positive observable A. The expectation value of the inverse of its finite-sample average $\bar{A}$ can be expanded around the mean:

$$\left\langle \frac{1}{\overline{A}} \right\rangle = \frac{1}{\langle A \rangle_\infty} - \left\langle \frac{\overline{A} - \langle A \rangle_\infty}{\langle A \rangle_\infty^2} \right\rangle + \left\langle \frac{\left(\overline{A} - \langle A \rangle_\infty\right)^2}{\langle A \rangle_\infty^3} \right\rangle - \cdots , \tag{2.14}$$

where $\langle A \rangle_\infty$ is the exact (infinite-sample) mean, and the larger brackets without subscripts denote an average over repeated trials with finite samples. The third term shows that $1/\bar{A}$ is biased, i.e., the average of the inverse is not equal to the inverse of the average. This bias originates from the nonlinear nature of inversion and becomes more severe when the variance of A is large or the number of samples is small.

In the context of reweighting, this issue becomes particularly noticeable when the same ensemble is used to estimate both the numerator and the denominator of Eq. (2.11). The correlation between the two estimators further complicates the statistical structure. Explicitly, if one estimates a ratio $R = \bar{B}/\bar{A}$, the full bias to $O(1/N)$ must account for the covariance between the numerator and

denominator:

$$\langle R\rangle = \frac{\langle B\rangle_\infty}{\langle A\rangle_\infty} + \frac{1}{N}\left[\frac{\langle B\rangle_\infty \mathrm{Var}(A)}{\langle A\rangle_\infty^3} - \frac{\mathrm{Cov}(A,B)}{\langle A\rangle_\infty^2}\right] + O(\frac{1}{N^2}), \tag{2.15}$$

which requires a large number of samples to sufficiently suppress higher-order terms[2]. This is especially dangerous in systems with sign problems, where the denominator itself—corresponding to the average sign—may be exponentially small and dominated by rare events. In such situations, reweighting becomes not only noisy but also fundamentally unreliable.

Therefore, one must be cautious when applying the reweighting method in practice. It is essential to assess the potential bias introduced by finite-sample estimations of denominators and, where possible, to design estimators that are explicitly unbiased.

2.3 Infinite variance problem

The efficiency of Monte Carlo calculations relies heavily on the variance of observables; observables with a variance that is small compared to their expectation value require fewer samples to achieve a target precision. However, in the extreme case, some observables can exhibit infinite variance. While the first moment may be finite, the divergence of the second moment means that the standard Central Limit Theorem fails. Consequently, the empirical sample variance diverges as the number of Monte Carlo samples increases, rendering standard statistical error estimation fundamentally unreliable.

In fermionic systems, the origin of the infinite variance problem is explicitly clear when examining the role of the fermion determinant. Because fermionic observables depend on the inverse of the fermion matrix M, their variance depends on the square of this inverse. Consider an observable defined as $O(\phi) = \frac{1}{\det M(\phi)} f(\phi)$, where f is an analytic function of the bosonic field ϕ. Upon integrat-

[2]Note that the systematic bias is not exactly removed for finite samples.

ing out the fermionic degrees of freedom, the expectation values for the first and second moments are given by

$$\langle O \rangle = \frac{1}{Z} \int [D\phi]\, f(\phi) \mathrm{e}^{-S_B(\phi)}\,, \tag{2.16}$$

$$\langle O^2 \rangle = \frac{1}{Z} \int [D\phi]\, \frac{1}{\det M} f(\phi)^2 \mathrm{e}^{-S_B(\phi)}\,. \tag{2.17}$$

For the first moment, the $\det M$ in the observable cancels the fermion determinant in the probability measure, yielding a well-defined, finite integral. However, for the second moment, an overall factor of $(\det M)^{-1}$ remains. If the phase space includes configurations where the fermion determinant approaches zero, the integral for $\langle O^2 \rangle$ can diverge. Thus, the expectation value of the observable remains finite, but its theoretical variance becomes infinite.

This issue has been known as *exceptional configurations* in lattice QCD [26]. These configurations arise due to near-zero eigenvalues of the Wilson-Dirac operator, leading to large fluctuations in quark propagators and numerical instabilities [27, 28]. While first observed in quenched simulations of Wilson fermions, similar divergences also appear in dynamical simulations [29], where they can cause instabilities in the Hybrid Monte Carlo (HMC) algorithm [22]. Various approaches have been proposed to mitigate these instabilities, such as modifying the quark propagator to account for zero mode poles [27, 28] or introducing twisted mass fermions to regulate infrared behavior [30]. The approach we investigate here is similar to previous strategies that utilize sampling with a positive weight, ensuring non-zero values across the field space, combined with reweighting [31–33].

In this thesis, we will use the Hubbard model to illustrate this problem and consider one possible solution. The Hubbard model [34] is a simplified representation of interacting electrons in a lattice, capturing the competition between electron hopping (kinetic energy) and on-site Coulomb repulsion

(interaction energy). The model has been extensively studied as a potential framework for understanding high-temperature superconductivity observed in cuprate materials [35, 36]. The electron creation operator at site x is denoted by $\hat{\psi}^{\dagger}_{a,x}$, where the spin index a takes the values $\uparrow$, or $\downarrow$. With this the Hamiltonian is written as follows:

$$\begin{aligned} H = &- \kappa \sum_{\langle x,y \rangle} (\hat{\psi}^{\dagger}_{\uparrow,x} \hat{\psi}_{\uparrow,y} + \hat{\psi}^{\dagger}_{\downarrow,x} \hat{\psi}_{\downarrow,y}) + U \sum_{x} (\hat{\psi}^{\dagger}_{\uparrow,x} \hat{\psi}_{\uparrow,x} - \frac{1}{2})(\hat{\psi}^{\dagger}_{\downarrow,x} \hat{\psi}_{\downarrow,x} - \frac{1}{2}) \\ &- \mu \sum_{x} (\hat{\psi}^{\dagger}_{\uparrow,x} \hat{\psi}_{\uparrow,x} + \hat{\psi}^{\dagger}_{\downarrow,x} \hat{\psi}_{\downarrow,x} - 1). \end{aligned} \tag{2.18}$$

where κ represents the hopping parameter, U denotes the strength of the on-site Coulomb interaction, and μ is the chemical potential. We will use $\kappa \equiv 1$ units. Note that the Hamiltonian is shifted so that $\mu = 0$ corresponds to the half-filling, i.e., there are the same numbers (on average) of electrons with spin-up and spin-down on the lattice. It is widely known that the Hubbard model has the sign problem except the half-filling, and in order to study the infinite variance problem, we will focus on the half-filling case.

On a bipartite lattice, it is useful to use the particle-hole basis for removing the sign problem at the half-filling: by redefining $\hat{\psi}_{1,x} \equiv \hat{\psi}_{\uparrow,x}$, $\hat{\psi}_{2,x} \equiv (-1)^x \hat{\psi}^{\dagger}_{\downarrow,x}$, the Hamiltonian becomes

$$\begin{aligned} H = &- \kappa \sum_{\langle x,y \rangle} (\hat{\psi}^{\dagger}_{1,x} \hat{\psi}_{1,y} + \hat{\psi}^{\dagger}_{2,x} \hat{\psi}_{2,y}) + \frac{U}{2} \sum_{x} (\hat{\psi}^{\dagger}_{1,x} \hat{\psi}_{1,x} - \hat{\psi}^{\dagger}_{2,x} \hat{\psi}_{2,x})^2 \\ &- \mu \sum_{x} (\hat{\psi}^{\dagger}_{1,x} \hat{\psi}_{1,x} - \hat{\psi}^{\dagger}_{2,x} \hat{\psi}_{2,x}). \end{aligned} \tag{2.19}$$

Note that the factor $(-1)^x$ in the redefinition of operators is $+1$ for even sites -1 for odd sites, which can be defined for bipartite lattices.

Using the path integral formulation with the Hubbard-Stratonovich transformation, which is

derived in Appendix A, one can derive an action

$$S \equiv S_B + S_F = -\beta \sum_{x,t} \cos \phi_{t,x} - \log \det M_1(\phi) - \log \det M_2(\phi), \tag{2.20}$$

where S_B and S_F denote the bosonic and fermionic parts of the action, respectively. The fermion matrices M_i are given as

$$M_a(\phi) = \mathbb{I} + B_a(\phi_N) \cdots B_a(\phi_1), \tag{2.21}$$

where $\phi_t \equiv \{\phi_{t,x}\}$ denotes the auxiliary field on the time-slice t, and $B_a(\phi)$ are given by

$$B_a(\phi_t) = \mathrm{e}^{-H_2} \mathrm{e}^{-\tilde{H}_4(\phi_t)}, \tag{2.22}$$

with

$$(H_2)_{x,y} = \kappa\epsilon\delta_{\langle x,y\rangle} + \varepsilon_a \mu\epsilon\delta_{x,y}, \tag{2.23}$$

$$\tilde{H}_4(\phi_t)_{x,y} = -i\varepsilon_a \sin \phi_{t,x}\delta_{x,y}. \tag{2.24}$$

Here, $\delta_{\langle x,y\rangle}$ is the nearest-neighbor hopping matrix, and $\varepsilon_{1,2} = \pm 1$. The parameter β is related to ϵU by

$$\mathrm{e}^{-\epsilon U/2} = \frac{I_0(\sqrt{\beta^2 - 1})}{I_0(\beta)}, \tag{2.25}$$

which is derived in Appendix A.3. It is obvious from the form of the matrix that when $\mu = 0$, $\det M_1 = (\det M_2)^*$, and therefore the model is free from the sign problem.

The Hubbard model is particularly useful to illustrate the infinite variance problem originated from fermion determinants since it has two determinants from spin-up and spin-down electrons. To

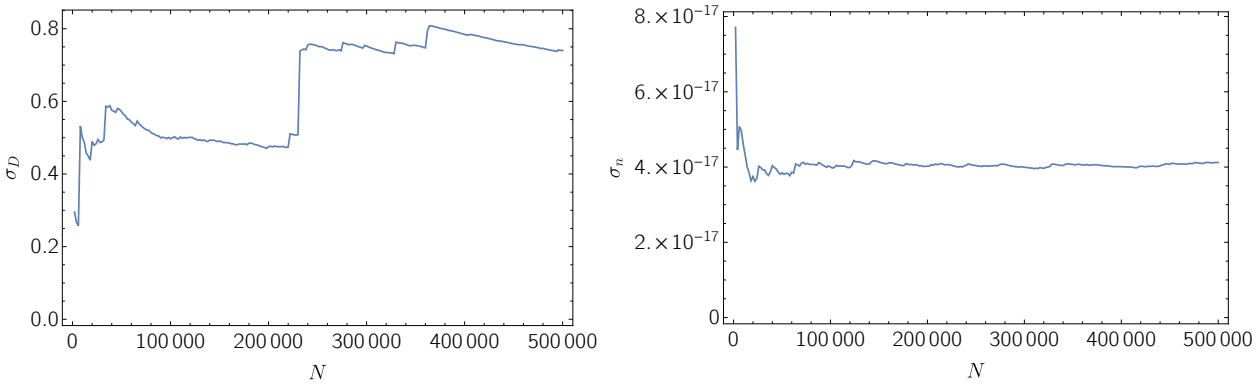

Figure 2.2: Cumulative estimates of standard deviations obtained from standard Monte Carlo simulations of the Hubbard model on a 4×4 lattice, with $U = 8$, $\mu = 0$, $T = 0.5$, and $\epsilon = 2.0$. The left plot shows the standard deviation, error $\times \sqrt{N}$, of the double occupancy, while the right plot shows that of the density.

exploit this property, we can consider the density n and the double occupancy D [15]. If we write them in terms of the fermion determinants, we find:

$$D(\phi) = \frac{1}{V}\sum_x \langle n_\uparrow(x) n_\downarrow(x)\rangle = \frac{1}{V}\sum_x M_2^{-1}(\phi)_{x,x}\left(1 - M_1^{-1}(\phi)_{x,x}\right), \tag{2.26}$$

$$n(\phi) = \frac{1}{V}\sum_x \langle n_\uparrow(x) + n_\downarrow(x)\rangle = \frac{1}{V}\sum_x \left(M_2^{-1}(\phi)_{x,x} - M_1^{-1}(\phi)_{x,x}\right). \tag{2.27}$$

Figure 2.2 shows the estimates of the standard deviations (error $\times\sqrt{N}$) for each observable as the number of configurations increases. The double occupancy exhibits an infinite variance problem, whereas the density does not. This discrepancy arises because the double occupancy includes terms proportional to the inverse of fermion matrices, specifically $(M_1 M_2)^{-1}$ in Eq. (2.3), while the density involves only a single inverse, such as M_1^{-1} or M_2^{-1}.

2.3.1 Extra time-slice method

Different methods have been suggested to deal with infinite variance problems [37–39] and we will utilize the extra time-slice method [39]. It exploits the reweighting method by using the probability distribution from an extra time-slice.

In order to introduce the new probability distribution, we define a function F as

$$F(\phi) \equiv \int [d\phi^*]\, \mathrm{e}^{-S_B(\phi^*)} \det M_{N+1}(\phi, \phi^*), \tag{2.28}$$

where M_{N+1} is the fermion matrix with $N+1$ time-slices and ϕ^* is the auxiliary field arising from the $(N+1)$th time-slice. Then we can rewrite the partition function as

$$Z = \int [d\phi]_N\, \mathrm{e}^{-S_B(\phi)} \det M_N(\phi) \frac{F(\phi)}{F(\phi)} = \int [d\phi]_N [d\phi^*]\, R(\phi)\, \mathrm{e}^{-S_B(\phi,\phi^*)} \det M_{N+1}(\phi, \phi^*), \tag{2.29}$$

where $R(\phi) = \det M_N(\phi)/F(\phi)$, $[\mathrm{d}\phi]_N$ denotes the path integral measure with N time-slices, and $[d\phi^*]$ is the additional measure from the extra time-slice. More specifically,

$$\begin{aligned} [d\phi]_N &= \prod_{i=1}^{N} \prod_{x=1}^{V} d\phi_{i,x}, \\ [d\phi^*] &= \prod_{x=1}^{V} d\phi_{N+1,x}. \end{aligned} \tag{2.30}$$

Therefore, one can use the new probability distribution and observables can be estimated using the reweighting procedure:

$$\langle O \rangle_N = \frac{\langle O(\phi) R(\phi) \rangle_{N+1}}{\langle R(\phi) \rangle_{N+1}}, \tag{2.31}$$

where the subscript $N+1$ denotes the new probability distribution:

$$p_{N+1}(\phi,\phi^*) = \frac{\mathrm{e}^{-S_0(\phi,\phi^*)}\det M_{N+1}(\phi,\phi^*)}{\int[\mathrm{d}\phi]_N[\mathrm{d}\phi^*]\mathrm{e}^{-S_0(\phi,\phi^*)}\det M_{N+1}(\phi,\phi^*)}. \tag{2.32}$$

Through this reweighting process, the variance of the observable O is determined by the behavior of OR under the new probability distribution. Importantly, any potential singularities arising from zeros of the fermion determinant in O, such as factors of $1/\det M_N$, are canceled by corresponding factors in the reweighting factor R. As a result, the infinite variance problem caused by the fermion determinant is resolved.

2.3.2 Sub-Monte Carlo method

Reference [39] suggested that one can integrate $F(\phi)$ analytically using the BSS formula [40] and expand it in $\epsilon \equiv \beta/N$:

$$\begin{aligned} F(\phi) &\equiv \int [d\phi^*]\,\mathrm{e}^{-S_0(\phi^*)}\det M_{N+1}(\phi,\phi^*) = \mathrm{Tr}\left[\mathrm{e}^{-\tilde{H}_2}\mathrm{e}^{-\epsilon H_4}B(\phi_N)\cdots B(\phi_1)\right] \\ &= \mathrm{Tr}\left[(1-\epsilon H)B(\phi_N)\cdots B(\phi_1)\right] + O(\epsilon^2) = (1-\epsilon H(\phi))\det M_N(\phi) + O(\epsilon^2). \end{aligned} \tag{2.33}$$

The main advantage of this approach is that it incurs no extra cost beyond the increased expense from introducing a new auxiliary field. However, it also has several drawbacks. First, truncation introduces an error of order ϵ^2, and the number of Wick contractions grows exponentially with higher orders. Additionally, a small ϵ is required not only to minimize Trotterization errors but because $F(\phi)$ can vanish when $1-\epsilon H(\phi) = 0$, potentially leading to another infinite variance problem.

Instead of using the perturbative approach, we can rather estimate $F(\phi)$ using Monte Carlo

calculations using the new auxiliary field ϕ^* [14] (which we call sub-Monte Carlo method):

$$F(\phi) = Z' \langle \det M_{N+1}(\phi, \phi^*) \rangle_{P_B}, \tag{2.34}$$

where the subscript P_B represents the Monte Carlo samplings using the bosonic part of the action $\mathrm{e}^{-S_B(\phi^*)}$ and Z' is a normalization constant different from the original partition function Z.

However, from $R(\phi) = \det M_N(\phi)/F(\phi)$, what one needs to estimate is $1/F(\phi)$, and taking an inverse of $\bar{F}(\phi)$ is a biased way to estimate $1/F(\phi)$ as discussed in Section 2.2.2. Therefore, we need to estimate $1/F(\phi)$ in an unbiased way. The authors in [41] suggested an unbiased estimator of $1/\langle A \rangle$:

$$\hat{\xi}_A \equiv \frac{w}{q_n} \prod_{i=1}^{n} (1 - wA_i)\,. \tag{2.35}$$

Here, q_n can be an arbitrary discrete probability distribution and w should be chosen to be smaller than $1/\langle A \rangle$. The variance minimizing choice of q_n and w [42] is

$$\begin{aligned} w &= \min\left\{ \frac{1}{k\overline{A}}, \frac{\overline{A}}{\overline{A^2}}, \frac{1}{A_{\max}} \right\}, \\ p &= 1 - \left[1 - 2w\overline{A} + w^2 \overline{A^2} \right]^{\frac{1}{2}}, \\ q_n &= p(1-p)^n. \end{aligned} \tag{2.36}$$

Note that the estimation of $1/F(\phi)$ using the sub-Monte Carlo (sub-MC) method is more expensive than the perturbative approach since it involves another Monte Carlo sampling during the measurement of the original Monte Carlo process. However, it is not as expensive as the original Monte Carlo process. First of all, the sub-MC is only done for measurements, not during the Monte Carlo updates performed between measurements. Furthermore, applying the action of Eq. (2.20) reduces

Eq. (2.33) to the following form:

$$
\begin{aligned}
F(\phi) &\equiv \int [d\phi^*]\, \mathrm{e}^{-S_B(\phi^*)} \det M_{N+1}(\phi, \phi^*) \\
&= \int [d\phi^*] \det \left(\mathbb{I} + B_1(\phi^*) B_1(\phi_N) \cdots B_1(\phi_1)\right) \mathrm{e}^{\beta \sum_x \cos \phi_x^*}.
\end{aligned}
\tag{2.37}
$$

During the sub-MC process, the background configurations $\phi_1, \cdots \phi_N$ are held fixed. Because the bosonic action S_B takes the form of a von Mises distribution, the proposed field ϕ^* can be sampled directly and independently, circumventing the autocorrelation inherent to standard Markov Chain methods. Furthermore, the computational bottleneck of evaluating the fermion determinant is heavily mitigated. The configurations are updated using the purely bosonic measure, with the fermion determinant evaluated only as a reweighting factor during the measurement phase even if MCMC methods are utilized. More crucially, exploiting the factorization properties of the fermion matrices [43] reduces the determinant to $\det\left(\mathbb{I} + B_1(\phi^*) B(\phi)\right)$. Since the bulk product $B(\phi) \equiv B_1(\phi_N) \cdots B_1(\phi_1)$ is strictly fixed during the sub-MC samplings, the computational cost of evaluating the determinant for each ϕ^* is drastically reduced compared to a full lattice calculation.

2.3.3 Results

Figure 2.3 compares the standard Monte Carlo and the sub-MC method. It is evident from the right figure that the variance of sample mean approaches constant while the variance of the standard method diverges.

In Figure 2.4, the left panel compares two approaches to the $\epsilon \to 0$ extrapolation: the sub-MC method and the expansion method from [39], which introduces ϵ^2 systematic errors. The sub-MC method exhibits a very mild dependence on ϵ. Although this is due to the method being unbiased and

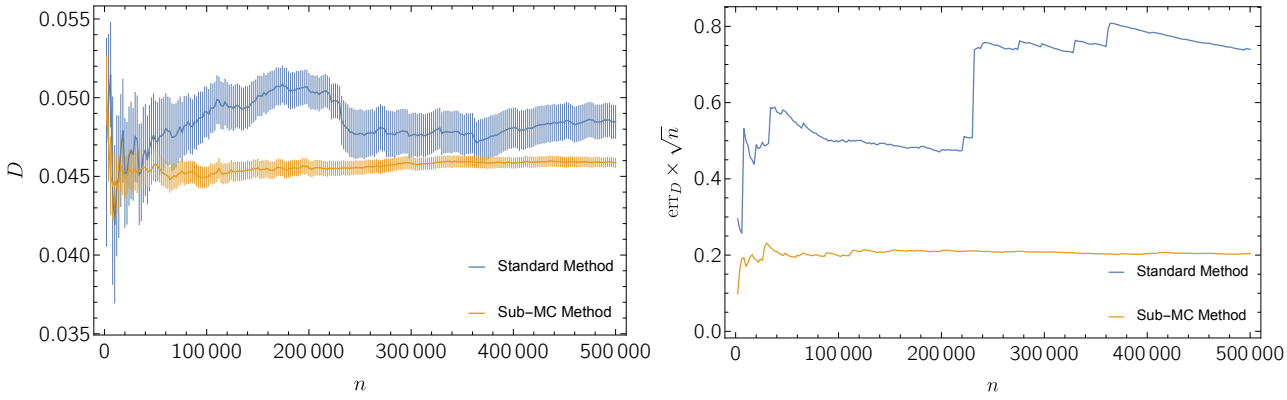

Figure 2.3: The double occupancy and its error estimate on a 4×4 lattice, with $U = 8$, $\mu = 0$, $T = 0.5$, and $\epsilon = 2.0$. The left plot is the Monte-Carlo history of the cumulative mean and error for the double occupancy as a function of number of updates, and the right plot represents the standard deviations which should be constant if the error estimate is reliable. Sudden jumps in the error are a symptom of the infinite variance problem.

the action being accurate up to ϵ^2, the small magnitude of the ϵ^2 coefficient is coincidental. However, the expansion method introduces a significant systematic error of order ϵ. Note that the extrapolation for the expansion method uses the first four points due to the increasing error at large ϵ. This indicates that a small ϵ is necessary for applying the expansion method to extrapolate to $\epsilon \to 0$. As a result, despite the extra Monte Carlo sampling step, the sub-MC method is more cost-effective in the long run.

The right plot in Figure 2.4 illustrates the effect of the the unbiased estimation. The biased estimator uses $1/\overline{F}$ as an estimator for $1/F(\phi)$. The bias scales as $1/n_{\text{sub}}$, where n_{sub} is the number of sub-MC samples. Since the number of sub-MC samples for the unbiased estimator is randomly selected from q_n in Eq. (2.36), the average value $n_{\text{sub}} \approx 2k$ is used for the figure which accounts for the cost k associated with estimating w and p in Eq. (2.36). The stochastic errors in the final results arise from both the configuration-to-configuration variance of the observable and the variance of the estimator. The blue band represents the $n_{\text{sub}} \to \infty$ extrapolation of the biased estimator. Notably, the number of sub-MC samplings needed to reach the asymptotic regime is quite modest. Moreover,

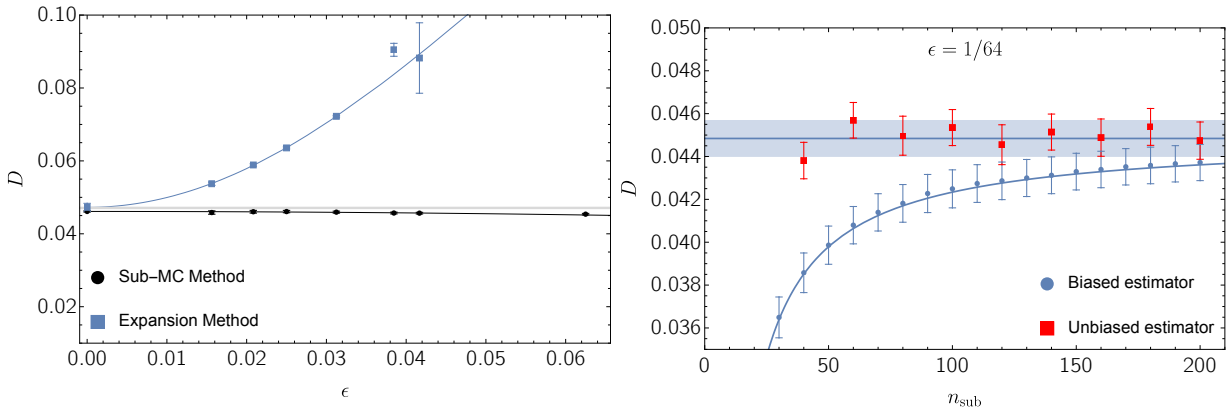

Figure 2.4: Double occupancy computed with the perturbative and sub-MC methods on a 4×4 lattice, with $U = 8$, and $\beta = 2$. The left plot compares the perturbative and sub-MC methods. For extrapolation, the fit for expansion uses the first four points, whereas the sub-MC method utilizes all points, with the gray band representing the benchmark result from [1]. The right plot contrasts the biased and unbiased estimators. The solid line illustrates the $1/n_{\text{sub}}$ fit for the biased estimator, while the blue band shows the $n_{\text{sub}} \to \infty$ limit of the biased estimator along with its associated error.

since the sub-MC process is performed only on configurations used for measurements, the overhead of applying sub-MC method is minimal.

2.4 Sign problem

Monte Carlo methods struggle with oscillating integrals because frequent sign changes of the integrand cause significant cancellations between positive and negative contributions. This cancellation results in a high variance, making the statistical error very large and convergence extremely slow. Consequently, achieving an accurate estimate requires an impractically large number of samples, rendering the method inefficient for such integrals. This problem is called the (numerical) sign problem.

In lattice field theory, such oscillating integrals can arise when the action can have complex values. While for a given physical theory, the Minkowski action is always real, its Euclidean action can be complex. Note that while its action can be complex for different configurations, the partition

function is real.

One of the cases where the sign problem appears is when there is a finite baryon chemical potential. Let us consider the Euclidean Dirac matrix

$$\det D(\mu) = \det(\not{\partial} + m - \mu\gamma_0), \tag{2.38}$$

where $\not{\partial} \equiv \gamma^\mu \partial_\mu$ where γ^μ are the gamma matrices which are Hermitian. One can check that D satisfies the following γ_5-hermiticity:

$$\gamma_5 D(\mu)\gamma_5 = \gamma_5(\not{\partial} + m - \mu\gamma_0)\gamma_5 = \not{\partial}^\dagger + m + \mu\gamma_0 = \left(\not{\partial} + m - \mu^*\gamma_0\right)^\dagger = D(-\mu^*)^\dagger, \tag{2.39}$$

which implies $\det D(\mu) = (\det D(-\mu))^*$ for real μ. Therefore, if there are even flavors (such as u and d quarks) and their masses are the same, the fermion determinant is positive when $\mu = 0$, while this fact does not hold in general when $\mu \neq 0$.

Another source of sign problems is the real-time simulation of quantum field theories. While the Euclidean path integral is used for thermodynamics of the system, discussed in Appendix A.1, in order to study the time-dependent observables, one needs to consider the Minkowski action:

$$\langle O \rangle = \frac{1}{Z} \int D\phi \, O(\phi) \mathrm{e}^{iS_M[\phi]}, \tag{2.40}$$

where the Minkowski action S_M is real for every configuration, the factor i in the Boltzmann factor makes it purely a phase and the integrand highly oscillatory. In general, real-time sign problems are more severe than Euclidean path integrals with finite baryon chemical potentials.

A typical way to deal with the sign problem is by reweighting with the probability distribution

from the real part of the action:

$$\langle O\rangle = \frac{1}{Z}\int D\phi\, O(\phi)\mathrm{e}^{-S[\phi]} = \frac{\frac{1}{Z_Q}\int D\phi\, O(\phi)\mathrm{e}^{-iS_I[\phi]}\mathrm{e}^{-S_R[\phi]}}{\frac{1}{Z_Q}\int D\phi\, \mathrm{e}^{-iS_I[\phi]}\mathrm{e}^{-S_R[\phi]}}, \tag{2.41}$$

where Z and Z_Q are the partition function and the quenched partition function, respectively.

The measure of the sign problem is the denominator in Eq. (2.41), which is called "average sign":

$$\langle\sigma\rangle \equiv \frac{1}{Z_Q}\int D\phi\, \mathrm{e}^{-iS_I[\phi]}\mathrm{e}^{-S_R[\phi]} = \frac{Z}{Z_Q} \text{ and } |\langle\sigma\rangle| \leq 1, \tag{2.42}$$

and therefore

$$\langle O\rangle = \frac{\frac{1}{Z_Q}\int D\phi\, O(\phi)\mathrm{e}^{-iS_I[\phi]}\mathrm{e}^{-S_R[\phi]}}{\langle\sigma\rangle}. \tag{2.43}$$

Thus, as the average sign goes to zero, the variance of O diverges:

$$\mathrm{Var}(\bar{O}) \approx \frac{1}{N|\langle\sigma\rangle|^2}\left(\frac{1}{Z_Q}\int D\phi\, |O(\phi)|^2\mathrm{e}^{-S_R[\phi]} - \left|\frac{1}{Z_Q}\int D\phi\, O(\phi)\mathrm{e}^{-S_R[\phi]}\right|^2\right), \tag{2.44}$$

where $\bar{O}$ denotes the estimation of the observable with a finite number of samples, N. This implies that more Monte Carlo samples are needed to estimate observables within a fixed error.

It is easy to check that the sign problem gets exponentially worse as the volume increases. In the thermodynamic limit, both Z and Z_Q scale exponentially with the volume V, since the free energy $F = -\log Z$ is extensive. Writing $Z = \mathrm{e}^{-Vf}$ and $Z_Q = \mathrm{e}^{-Vf_Q}$, where f and f_Q are the free energy densities of the full and phase-quenched systems respectively, we obtain [44]

$$\langle\sigma\rangle = \mathrm{e}^{-V(f-f_Q)}. \tag{2.45}$$

This expression shows that the average sign decreases exponentially with volume. As a result, the variance of any reweighted observable grows exponentially, and an exponentially large number of samples is required to maintain a fixed signal-to-noise ratio.

2.4.1 Contour deformation

One of the general approaches to alleviate the sign problem is contour deformation, which leverages the multi-dimensional version of Cauchy's theorem from complex analysis. Specifically, if $f(\phi)$ is an analytic function in $\mathbb{C}^n$ (which embeds $\mathbb{R}^n$),

$$\int_{\mathbb{R}^n} d\phi_1 \wedge \cdots \wedge d\phi_N f(\phi) = \int_{\gamma} d\phi_1 \wedge \cdots \wedge d\phi_N f(\phi), \tag{2.46}$$

where γ is any contour embedded in $\mathbb{C}^N$ space and lies in the same homology class as $\mathbb{R}^n$. A detailed discussion of the mathematical framework underlying contour deformation is provided in Appendix B.

This flexibility allows one to deform the original integration domain to a more favorable contour where the sign problem is milder. The improvement stems from the following observation: the average sign can be written as

$$\langle \sigma \rangle = \frac{Z}{Z_Q} = \frac{\int D\phi \, \mathrm{e}^{-S[\phi]}}{\int D\phi \, |\mathrm{e}^{-S[\phi]}|}. \tag{2.47}$$

If the action $S[\phi]$ is holomorphic (which is the case in field theory), the Boltzmann weight $\exp(-S)$ is holomorphic and the partition function Z is invariant under contour deformations within the same homology class. However, the denominator Z_Q is non-holomorphic and thus can be modified by contour deformation. By deforming the contour to reduce Z_Q, the average sign improves. A comprehensive review of this method is given in Ref. [45].

Historically, contour deformation techniques trace back to the Lefschetz thimble method [46–61], initially introduced in physics for analytic continuation in Chern-Simons theory [62, 63]. In this framework, the integral is rewritten as a sum over thimbles, which are manifolds defined by the downward holomorphic flow equation:

$$\left.\frac{\partial S}{\partial \phi_i}\right|_{\phi=\phi_c} = 0\,. \tag{2.48}$$

Each thimble is tied with a critical point of the action, satisfying

$$\frac{d\phi}{dt} = -\overline{\frac{\partial S}{\partial \phi}}. \tag{2.49}$$

The thimble $\mathcal{T}_c$ is the set of all points that flow to ϕ_c as $t \to \infty$. Importantly, along each thimble, the imaginary part of the action is constant, meaning the sign problem is absent. If a single thimble dominates, the sign problem is completely eliminated.

However, practical challenges arise because finding all relevant thimbles is nontrivial, and interference between thimbles with different phases can reintroduce a sign problem. To address this, the generalized thimble method [64, 65] was proposed. Instead of the downward flow, it utilizes the upward holomorphic flow,

$$\frac{d\phi}{dt} = +\overline{\frac{\partial S}{\partial \phi}}, \tag{2.50}$$

which has the properties

$$\begin{aligned} \frac{dS_R}{dt} &= \sum_i \left(\frac{\partial S}{\partial \phi_i}\overline{\frac{\partial S}{\partial \phi_i}}\right) \geq 0, \\ \frac{dS_I}{dt} &= 0. \end{aligned} \tag{2.51}$$

Thus, the upward flow increases the real part of the action while keeping the imaginary part constant, reducing Z_Q and improving the average sign.

Despite their theoretical appeal, Lefschetz thimble-based methods have drawbacks:

- Integrating the flow equations is computationally costly.
- The flow introduces a nontrivial Jacobian determinant, which includes solving N-coupled ordinary differential equations and computing a determinant of a $N \times N$ matrix, further complicating calculations [60].
- As the flow time increases, the resulting probability distribution becomes multimodal, making sampling challenging. Solutions like parallel tempering [66, 67] alleviate ergodicity issues but at additional computational cost.

To circumvent these difficulties, an alternative approach is to directly parameterize the contour to optimize the average sign, rather than relying on the holomorphic flow [68–70]. Specifically, one can exploit the following fact:

$$\nabla_\lambda |\langle\sigma\rangle| = |\langle\sigma\rangle| \frac{\int_{\mathbb{R}^N} D\phi\, \nabla S_R(\phi;\lambda) \mathrm{e}^{-S_R(\phi;\lambda)}}{\int_{\mathbb{R}^N} D\phi\, \mathrm{e}^{-S_R(\phi;\lambda)}}. \tag{2.52}$$

This gradient depends only on the real part of the action, meaning that the optimization process itself is free from the sign problem. By parameterizing the integration contour (e.g., via neural networks) and using gradient descent, one can systematically search for manifolds that reduce the severity of the sign problem.

It is important to note that this parameterization only allows for deformations of the form $\phi \to f(\phi) + i\, g(\phi)$, which are homeomorphic to $\mathbb{R}^N$. As a result, this approach excludes any integration contours that are not topologically equivalent to $\mathbb{R}^N$. For instance, one could imagine a contour defined by a manifold γ that is topologically the product of an interval $[0, 1]$ and a $(N-1)$-dimensional sphere,

with the ends of the interval identified with disjoint regions in $\mathbb{R}^N$. Whether such nontrivial topologies could offer advantages in addressing the sign problem remains an open question.

2.4.2 Complex normalizing flow

A normalizing flow is a bijective map $\tilde{\phi}(\phi)$ between two probability distributions:

$$\left(\det \frac{\partial \tilde{\phi}}{\partial \phi}\right) P_f\left(\tilde{\phi}(\phi)\right) = P_i\left(\phi\right), \tag{2.53}$$

where P_i is the initial probability distribution and P_f is the final or desired probability distribution. In applications for lattice field theories, P_i is often chosen as the Gaussian distribution and P_f is the Boltzmann factor $\exp(-S[\phi])$. For the scalar field theory discussed in this thesis, we will use P_i as the Gaussian distribution from this point forward.

The merit of normalizing flows for lattice calculations is that one can compute expectation values of observables without sampling through MCMC. Uncorrelated samples can be efficiently generated by transforming from the easily sampled distribution P_i. However, in many cases it is hard to find the exact map; for such an approximate normalizing flow, we define the induced action as

$$\left(\det \frac{\partial \tilde{\phi}}{\partial \phi}\right) e^{-S_{\text{induced}}[\tilde{\phi}(\phi)]} = \mathcal{N} e^{-\phi^2/2}, \tag{2.54}$$

where $\mathcal{N}$ is a normalization constant. The expectation value with respect to the desired action can be calculated via reweighting:

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O}\ \mathrm{e}^{S_{\text{induced}} - S} \rangle_n}{\langle \mathrm{e}^{S_{\text{induced}} - S} \rangle_n}. \tag{2.55}$$

Above, the subscript n denotes the expectation value with respect to the normal distribution.

Normalizing flows have recently attracted attention as an effective tool for accelerating Monte Carlo simulations in lattice field theory [71–73]. As established in Ref. [74], contour deformation inherently connects to the framework of normalizing flows. For systems governed by a complex action S, the flow equation defined in Eq. (2.53) parametrizes a deformed integration manifold designed to mitigate the sign problem. Consequently, the training of normalizing flows in this context is conceptually equivalent to established contour deformation methodologies.

For two smooth probability distributions, normalizing flows are known to always exist [75]. However, the conditions for the existence of complex normalizing flows are not fully understood.

It is useful to parameterize the space of normalizing flows using neural networks. The next step is to define a suitable loss function, where smaller values indicate better model performance. Here, we define a loss function composed of two terms: one that accounts for the logarithm of the magnitude of the Boltzmann factor and another that captures the phase:

$$\begin{aligned} L_{\mathrm{nf}}(\lambda) = &\left\langle \left| \underbrace{\frac{\phi^2}{2} + \operatorname{Re}\left[\log\left|\frac{\partial\tilde{\phi}_\lambda}{\partial\phi}\right| - \log\mathcal{N} - S(\tilde{\phi}_\lambda(\phi))\right]}_{\operatorname{Re}(S_{\text{induced}}-S)} \right|^2 \right\rangle_n \\ &+ \left\langle \left| 1 - \underbrace{\left(\operatorname{sgn}\left|\frac{\partial\tilde{\phi}_\lambda}{\partial\phi}\right|\right) \mathrm{e}^{-i\operatorname{Im}\left(S(\tilde{\phi}_\lambda(\phi))+\log\mathcal{N}\right)}}_{e^{i\operatorname{Im}(S_{\text{induced}}-S)}} \right|^2 \right\rangle_n , \end{aligned} \tag{2.56}$$

where $\operatorname{sgn} z = \frac{z}{|z|}$ denotes the sign function. This cost function ensures that the transformed distribution aligns with the target distribution by minimizing deviations in both the real and imaginary parts of the induced action. Although a different relative weighting could, in principle, be beneficial, we have not investigated this possibility. During training, the real and imaginary parts of the normalization factor $\mathcal{N}$ are treated as additional parameters to be optimized.

Training a complex normalizing flow is more challenging than training a real normalizing flow or an integration contour. In particular, ensuring that the trained normalizing flow exhibits the correct asymptotic behavior necessary for Cauchy's integral theorem to hold is nontrivial. A useful approach in this context is *adiabatic training*, as described in [76]. In this method, a one-parameter family of actions S_α is introduced, where S_0 corresponds to a well-understood action for which a normalizing flow can be exactly prepared, and S_1 represents the target action. The action is gradually evolved in small steps of $\Delta\alpha$, retraining the normalizing flow at each stage. Various adiabatic paths can be chosen to interpolate between S_0 and S_1. For simplicity, unless stated otherwise, we adopt the following path:

$$S_\alpha(\tilde{\phi}) = \alpha S(\tilde{\phi}) + \frac{(1-\alpha)}{2}\tilde{\phi}^2. \tag{2.57}$$

We perform adiabatic training with step size $\Delta\alpha = 0.05$. However, it is worth noting that this particular path is unsuitable when the original path integral diverges along the real plane, requiring an alternative approach for analytic continuation, as discussed in Section 2.4.3.

2.4.3 Scalar field theory at complex coupling

As a test ground of the contour deformation and the complex normalizing flow approaches to the sign problem, we study scalar field theories on an isotropic lattice with periodic boundary conditions. The action is

$$S = \frac{1}{2}\sum_{x,\mu}(\phi_{x+\mu} - \phi_x)^2 + \frac{m^2}{2}\sum_x \phi_x^2 + \lambda\sum_x \phi_x^4, \tag{2.58}$$

where m^2 and λ are allowed to be complex numbers. $0+1$D case corresponds to the anharmonic oscillator:

$$H_{\text{aho}} = \frac{1}{2}p^2 + \frac{m^2}{2}x^2 + \lambda x^4. \tag{2.59}$$

In this thesis, the anharmonic oscillator will be studied through the complex normalizing flow and the $1+1$D case will be studied with contour deformation.

For physical theories, the partition function is real. In contrast, the theory studied here is not physical in that the partition function can acquire a complex value. This behavior is reflected in the average sign, as the quenched partition function remains real by definition, while the full partition function may develop a nontrivial phase.

However, our approach connects naturally to the study of $\mathcal{PT}$-symmetric theories on the lattice. $\mathcal{PT}$-symmetric theories, though non-Hermitian, are invariant under the combined action of parity ($\mathcal{P}$) and time-reversal ($\mathcal{T}$) symmetries [77]. While such Hamiltonians can possess entirely real spectra when $\mathcal{PT}$-symmetry is unbroken, spontaneous $\mathcal{PT}$-symmetry breaking can result in complex eigenvalues. Such theories extend conventional quantum mechanics and appear in various fields, including high-energy physics [78, 79], statistical mechanics [80–83], and optics [84–87]. A well-known example is the Hamiltonian $H = p^2 + ix^3$, which is $\mathcal{PT}$-symmetric but not Hermitian. $\mathcal{PT}$-symmetric systems are particularly relevant in open quantum systems and non-equilibrium physics [88–92].

This subtle non-Hermiticity leads to significant computational challenges, particularly manifesting as severe sign problems in lattice formulations. In path integrals, even when the underlying theory has real physical observables, the integrand may acquire a rapidly fluctuating phase, rendering standard Monte Carlo techniques inefficient. One effective strategy to address this issue is contour deformation, which was successfully applied to $\mathcal{PT}$-symmetric quantum mechanical systems in [93].

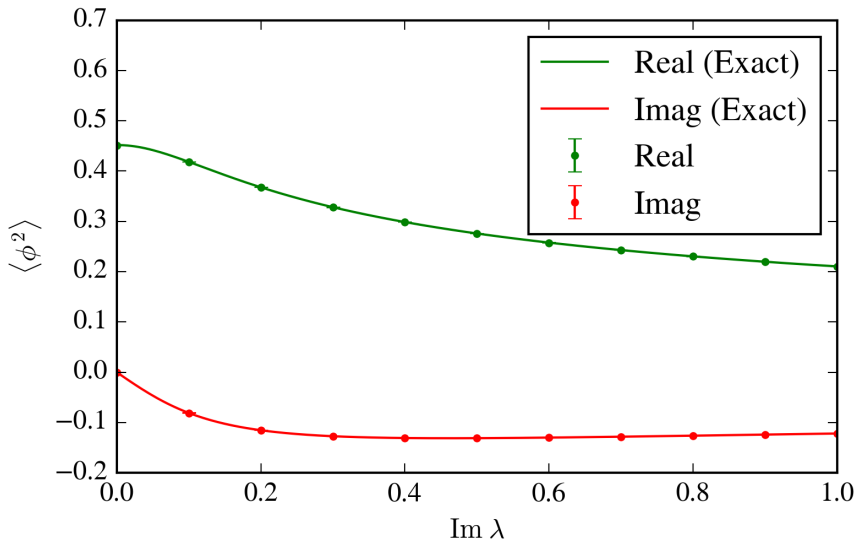

Figure 2.5: The expectation value $\langle \phi^2 \rangle$ is plotted against $\mathrm{Im}\,\lambda$ for a one-dimensional 10-site lattice with parameters $m^2 = 0.5$ and $\mathrm{Re}\,\lambda = 0.1$. For reference, the exact result is obtained by directly diagonalizing the non-Hermitian Hamiltonian.

Regardless of the physical interpretation of this theory, the model is closely connected to real-time lattice simulations formulated via the Schwinger-Keldysh contour [94, 95]. This formalism is widely used to compute real-time observables on the lattice [96, 97].

First, we construct normalizing flows $\tilde{\phi}(\phi)$ from neural networks. The precise functional form used is

$$\tilde{\phi}(\phi) = (M_R + iM_I)\phi + (L_R + iL_I)f(\phi). \tag{2.60}$$

The function $f(\phi)$ is a fully connected network with a sigmoid activation function, taking V inputs, featuring $2V$ nodes per internal layer, and producing $2V$ outputs. The trainable parameters include four real-valued matrices M_R, M_I, L_R, and L_I, along with the weights and biases of f. Note that the network ends with the activation function without an additional affine transformation. This convention is used to indicate that the complex contour is generated by a single network with real inputs.

At first sight, the separation of the M_* terms from the definition of f may seem unnecessary and

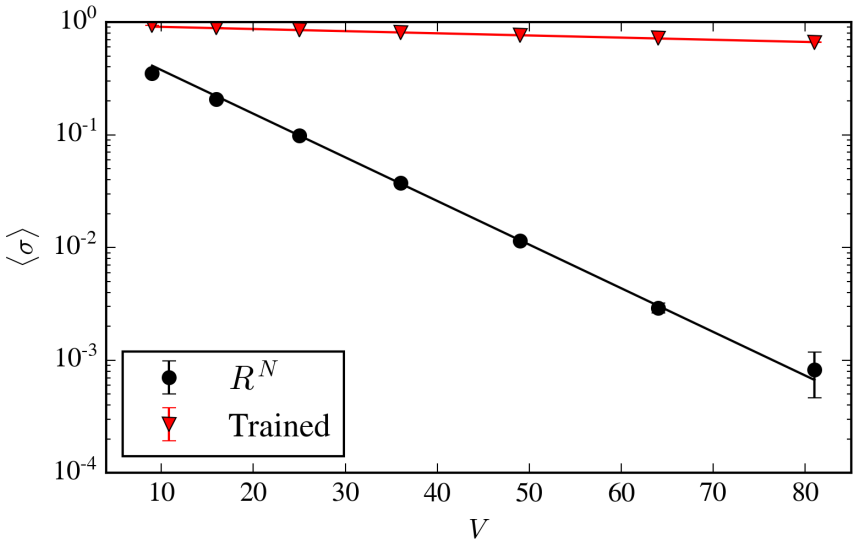

Figure 2.6: The average sign plotted as a function of volume for two-dimensional scalar field theory with parameters $m^2 = 0.5$ and $\lambda = 1.0i$. The integration contour is defined by an affine transformation, corresponding to a zero-layer network. For the simulations, 5×10^6 configurations were sampled on the real plane, while only 5×10^5 configurations were required on the trained contours, due to the improved average sign. Statistical uncertainties, estimated via bootstrapping with blocking, are indicated but remain smaller than the symbol size.

redundant. However, due to the choice of activation function, $\sigma(x) = \frac{1}{1+e^{-x}}$, f is inherently bounded above and below. Consequently, a normalizing flow relying solely on f cannot map the real plane onto a contour satisfying Cauchy's integral theorem. The inclusion of simple linear terms restores this capability.

To validate our approach and implementation, Figure 2.5 presents $\langle\phi^2\rangle$ for various complex λ. The exact results, obtained through direct matrix manipulations of the Trotterized imaginary-time evolution operator, show excellent agreement with Monte Carlo computations.

To investigate the $1+1$D scalar theory, we train integration contours using the algorithm outlined in Section 2.4.1. For the moderate lattice sizes considered, linear contours of the form

$$\tilde{\phi}(\phi) = \phi + M_R\phi + iM_I\phi \tag{2.61}$$

are sufficient to mitigate the sign problem. Here, M_R and M_I are $V \times V$ real matrices treated as trainable parameters, where V represents the number of lattice sites. Figure 2.6 illustrates the performance of these trained integration contours as a function of the lattice size, ranging from 3×3 to 9×9. A fully connected network is employed, with optimization performed using the *ADAM* optimizer [98]. The results demonstrate that the contour deformation method yields an exponential improvement in the average sign with respect to the lattice volume.

2.4.3.1 Studying partition functions

In this subsection, we will extract information of the partition functions with the two methods: contour deformation and the complex normalizing flow.

Partition functions are fundamental objects in quantum field theory and statistical mechanics, as they contain complete information about the system. One particularly important aspect of partition functions is their analytic structure, especially the location of their zeros in the complex parameter space. These zeros, known as Lee-Yang zeros [99, 100], are crucial because they offer insight into phase transitions and symmetry-breaking phenomena. In the thermodynamic limit, the accumulation of Lee-Yang zeros near the real axis signals the onset of phase transitions. Thus, understanding the zeros of the partition function is not only of theoretical interest but also essential for characterizing critical phenomena in both Hermitian and non-Hermitian systems.

In conventional Monte Carlo methods, extracting the partition function is typically challenging, as these methods yield only the probability distribution $\frac{1}{Z}\exp(-S[\phi])$, leaving the normalization factor Z unknown. One approach to determining the partition function utilizes annealed importance sampling [101], which involves constructing a discrete sequence of intermediate actions, $S_1, S_2, \ldots, S_N$.

This sequence bridges a well-understood base distribution (e.g., a Gaussian action for S_1) and the target system governed by S_N. The partition function is then computed via

$$\frac{Z_N}{Z_1} = \frac{Z_2}{Z_1}\frac{Z_3}{Z_2}\cdots\frac{Z_N}{Z_{N-1}}, \tag{2.62}$$

using reweighting techniques, though this process is nontrivial and computationally intensive.

In contrast, if an exact normalizing flow were available, the partition function would be immediately determined by the trained normalization factor $\mathcal{N}$. For approximate flows, the partition function can still be estimated by averaging over samples:

$$\frac{Z}{Z_n} = \langle e^{S_{\text{induced}}-S}\rangle. \tag{2.63}$$

The scalar field theory is well-defined only for $\operatorname{Re}\lambda \geq 0$, as the path integral $\int \mathcal{D}\phi\, e^{-S}$ becomes divergent when $\operatorname{Re}\lambda$ is negative due to the instability introduced by the ϕ^4 interaction term. However, it is possible to analytically continue the theory beyond this region of convergence. The analytically continued partition function remains well-defined and analytic everywhere except at the branch point $\lambda = 0$. One intuitive perspective on this continuation is through a rotation of the field variables, $\phi \to e^{i\theta}\phi$. Under such a transformation, the coupling constant transforms as $\lambda \to e^{-4i\theta}\lambda$, allowing one to choose θ to ensure convergence of the path integral. This comes at the cost of introducing complex phases into the m^2 and kinetic terms.

Figure 2.7 illustrates these results. In this example, we utilized a normalizing flow built from a neural network with a single hidden layer. The adiabatic training process used for the normalizing flows does not depend on whether the path integral converges on the real axis; rather, it effectively

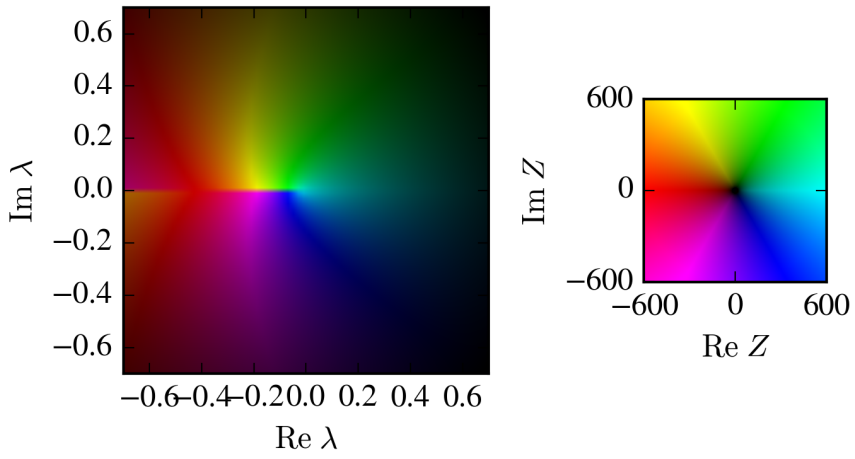

Figure 2.7: The partition function for lattice scalar field theory in $0+1$ dimensions for a lattice with 10 sites and $m^2 = 0.5$, evaluated at complex values of the coupling λ. For negative λ, the partition function is defined via analytic continuation, with the branch cut placed along the negative real axis. The color scheme, displayed to the right, corresponds to the mapping of $f(\lambda) = \lambda$ under the same scale for reference.

performs an analytic continuation following the chosen path. Specifically, the normalizing flows in Figure 2.7 were first trained at a set of radii $(0.1, 0.2, \ldots, 1.0)$ along the real λ-axis, and then retrained while rotating λ in the complex plane. This rotation was confined to angles between $-\pi$ and π, producing the branch cut seen along the negative real axis.

Contour deformation methods are basically the Monte Carlo method dealing with sign problems. Therefore, one cannot directly estimate the partition function. However, one can study the zeros of the partition function from the following quantity.

Reference [74] argued that theories with polynomial actions often admit perfect integration contours, supporting this with several examples of one-dimensional integrals. In the preceding subsection, we provided numerical evidence that quantum mechanical path integrals can also have perfect contours. However, in this subsection, we highlight an important source of counterexamples: the zeros of a partition function.

To illustrate, consider the one-dimensional integral:

$$I(\alpha) = \int_{-\infty}^{\infty} dz\, e^{-\alpha z^2 - z^4}. \tag{2.64}$$

While specific values of α in [74] allowed for perfect contours, there are other values where the average phase can be made arbitrarily small. This can be demonstrated by showing two key points:

1. For any $\epsilon > 0$, $|I(\alpha)| < \epsilon$.

2. For any fixed α, the quenched partition function $I_Q(\alpha) = \int_{-\infty}^{\infty} dz\, |e^{-\alpha z^2 - z^4}|$ has a positive lower bound.

Together, these imply that the average phase $\langle \sigma \rangle = \frac{I}{I_Q}$ can be bounded by $\frac{\epsilon}{B}$, where B is the lower bound of the quenched partition function.

The first point follows from the great Picard theorem: since $I : \mathbb{C} \to \mathbb{C}$ has an essential singularity at infinity, it attains nearly all complex values, and zeros typically cluster near certain α values. For the second point, consider contours in the complex plane parameterized as $z = re^{i\theta}$. For any fiducial radius R, a contour either passes within radius R or remains outside. In the former case, the contour must traverse an annulus where the modulus of the integrand is bounded below by some constant C_R, contributing at least C_R to I_Q. In the latter case, in unstable angular regions, the integrand grows exponentially, and its contribution can be bounded from below by a term proportional to R.

Thus, combining both cases yields a non-zero lower bound on the quenched partition function, albeit a loose one. While in physical theories it may be challenging to compute this bound tightly, the existence of zeros of a partition function is a general obstacle to perfect contours.

While normalizing flows offer a way to compute the partition function directly via Monte Carlo

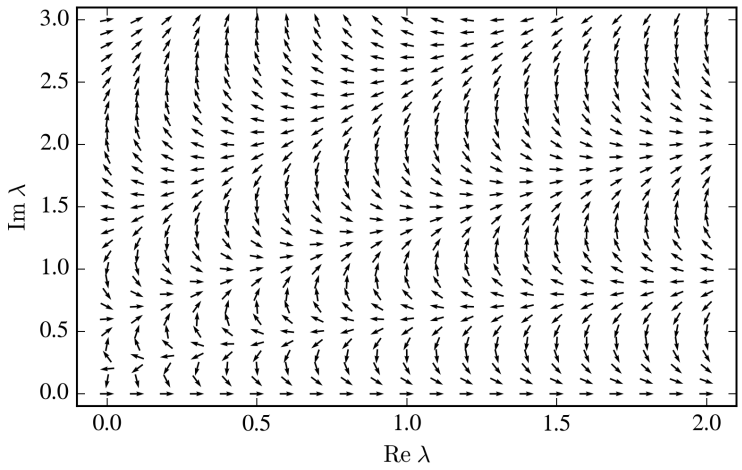

Figure 2.8: Phases of the partition function for $m^2 = 0.5$ on an 8×8 lattice, plotted across the complex λ-plane. As the partition function is analytic everywhere except at the branch point $\lambda = 0$, the number of zeros within any enclosed region can be identified by evaluating the winding number of the phase along a closed path encircling that region.

sampling, they are not immediately practical for identifying zeros of the partition function. This is apparent from Figure 2.7, where the existence of zeros cannot be easily inferred. The difficulty arises because the average phase becomes vanishingly small in the vicinity of the zeros of the partition function, making it impossible to distinguish between an actual zero and a value that is merely very close to zero.

Turning to scalar field theory, zeros of the partition function are expected, for instance when external sources like a $J\phi$ term are introduced—this is well-established [102]. Crucially, however, the presence of these zeros can be detected without evaluating the partition function precisely. Specifically, by integrating the logarithmic derivative of the partition function along a closed contour in the complex λ-plane:

$$I_\gamma = \frac{1}{2\pi i} \oint_\gamma d\lambda \, \frac{Z'(\lambda)}{Z(\lambda)}, \tag{2.65}$$

one obtains an integer counting the zeros enclosed by the contour. This interpretation links the zeros

to the winding number of the phase of $Z(\lambda)$. Therefore, even when direct Monte Carlo sampling near zeros is impractical, the locations of zeros can be inferred from phase behavior. In the two-dimensional case of this subsection, while directly determining Z is challenging, the phase (imaginary part of the free energy) is accessible via the improved average sign, allowing zeros to be identified through their topological signature.

Figure 2.8 shows the phase of the partition function across the first quadrant of the complex λ-plane. By inspecting the winding number of the phase along closed loops, one can determine whether zeros are present within a given region. From the plot, it is evident that no zeros appear in the first quadrant for $|\lambda| \lesssim 2$. Additionally, since the partition function satisfies $Z(\bar{\lambda}) = \overline{Z(\lambda)}$, the absence of zeros extends to the corresponding region in the fourth quadrant.

2.5 Signal-to-noise problem

Even in the absence of a complex Boltzmann factor, signal-to-noise problems can still arise when the observable itself exhibits significant fluctuations. In such cases, the issue does not stem from the sampling measure but rather from the observable's intrinsic structure—effectively a sign problem encoded in the observable. In the lattice field theory community, this type of difficulty is typically referred to as a signal-to-noise problem to distinguish it from sign problems associated with the path integral weight.

Mathematically, the signal-to-noise ratio for an observable O is defined as

$$\mathrm{SNR}(O) = \frac{\langle O \rangle}{\sqrt{\mathrm{Var}(O)}}. \tag{2.66}$$

This ratio quantifies the strength of the average signal relative to the statistical noise. A low SNR

indicates that large fluctuations dominate over the mean, making accurate estimation difficult.

A canonical example of this phenomenon is found in Euclidean correlation functions, such as

$$C(t) = \langle O_1(t) O_2(0) \rangle. \tag{2.67}$$

As the Euclidean time separation t increases, the correlator $C(t)$ typically decays exponentially, $C(t) \sim \mathrm{e}^{-mt}$, where m is the mass of the lightest state coupling to the operators. However, the variance of $C(t)$ often remains roughly constant or decays more slowly, resulting in an exponentially worsening signal-to-noise ratio with increasing t.

This behavior is elegantly captured by the Parisi-Lepage argument [103, 104], which describes the degradation of the signal-to-noise ratio in terms of the spectral content of the observable and its fluctuations. Consider the nucleon two-point correlation function: the signal behaves as $\exp(-m_N t)$, while the variance is dominated by multi-pion states and scales as $\exp(-3m_\pi t)$, where m_N and m_π are the nucleon and pion masses, respectively. This leads to a signal-to-noise ratio of the form

$$\mathrm{SNR}(C(t)) \propto \exp\left(-\left(m_N - \frac{3}{2} m_\pi\right) t\right), \tag{2.68}$$

which decays exponentially with time, making precise measurements at large t increasingly costly.

In practical computations, we estimate observables using a finite number of Monte Carlo samples:

$$\bar{O} = \frac{1}{N} \sum_{i=1}^{N} O_i, \tag{2.69}$$

with the associated statistical error (standard deviation)

$$\text{err}(\bar{O}) = \sqrt{\frac{\text{Var}(O)}{N}}. \tag{2.70}$$

From this, it follows that the signal-to-noise ratio of the estimator improves only as $\sqrt{N}$. Consequently, when the signal decays exponentially while the noise remains significant, one requires an exponentially large number of samples in t to maintain a fixed precision. This exponential scaling severely limits the accessible time separations in correlator studies and is a central challenge in lattice QCD and other applications of MCMC to quantum field theory.

2.5.1 Control variates

To address signal-to-noise problems in lattice field theory, the control variates method was recently proposed as a promising variance reduction technique [105]. While this method is well-established in the statistics and Monte Carlo literature, its prior applications have largely been limited to low-dimensional integration problems—typically in spaces with fewer than ten dimensions [106–112].

The essential idea behind control variates is straightforward. Given an observable O, suppose one can construct a function f, called a control variate, such that $\langle f \rangle = 0$. Then the modified observable

$$\tilde{O} = O - f \tag{2.71}$$

has the same expectation value as the original:

$$\langle \tilde{O} \rangle = \langle O \rangle. \tag{2.72}$$

However, the variance of $\tilde{O}$ may differ from that of O:

$$\mathrm{Var}(\tilde{O}) = \mathrm{Var}(O) + \langle f^2 \rangle - 2\langle Of \rangle. \tag{2.73}$$

If f is chosen such that it is highly correlated with O and its variance is small, then $\mathrm{Var}(\tilde{O})$ can be significantly reduced. Thus, effective control variates reduce statistical noise without altering the mean.

In practice, constructing a zero-mean control variate is challenging. To this end, a constructive framework was developed using Stein's method [113] or the lattice Schwinger-Dyson equations [105, 114]. These rely on the identity:

$$\int D\phi \, \frac{\delta}{\delta \phi} \left(g(\phi) \mathrm{e}^{-S(\phi)} \right) = 0, \tag{2.74}$$

which holds under suitable boundary conditions for any appropriately smooth function $g(\phi)$. This is the vanishing of the integral of a total derivative, and its use in physics dates back to the zero-variance principle of quantum Monte Carlo [115, 116].

Applying Eq. (2.74), one obtains general forms of a control variate: for a scalar-valued function $g : M^N \to \mathbb{R}$,

$$f = \sum_i \left(\frac{\partial g}{\partial \phi_i} - g \frac{\partial S}{\partial \phi_i} \right), \tag{2.75}$$

where S is the action, M is the field target space, and N is the number of degrees of freedom[3].

[3]For example, if we consider $U(1)$ gauge theory on 4-dimension, $N = 4V$ where V is the number of sites.

Similarly, for a vector-valued function $g : M^N \to \mathbb{R}^N$,

$$f = \sum_i \left(\frac{\partial g_i}{\partial \phi_i} - g_i \frac{\partial S}{\partial \phi_i} \right). \tag{2.76}$$

These are called Stein control variates, and the choice of g determines their effectiveness.

Importantly, using control variates cannot make the variance worse. If a particular choice of f increases the variance of $\tilde{O}$, it can simply be discarded. More flexibly, one can introduce a rescaling factor a and define

$$\tilde{O}_a = O - af, \tag{2.77}$$

with corresponding variance

$$\mathrm{Var}(\tilde{O}_a) = \mathrm{Var}(O) + a^2 \langle f^2 \rangle - 2a \langle Of \rangle. \tag{2.78}$$

This is minimized when

$$a = \langle Of \rangle / \langle f^2 \rangle, \tag{2.79}$$

yielding the optimal variance

$$\min_a \mathrm{Var}(O_a) = \mathrm{Var}(O) - \frac{\langle Of \rangle^2}{\langle f^2 \rangle} \leq \mathrm{Var}(O). \tag{2.80}$$

This ensures that variance is never increased, and the method guarantees improvement.

A remarkable property of the control variate framework is the existence of a perfect control

variate, which achieves zero variance. This is given by

$$f_P = O - \langle O \rangle. \tag{2.81}$$

Although f_P is not practical in computations (since $\langle O \rangle$ is the quantity being estimated), it serves as an ideal benchmark and motivates the search for good approximations.

It is important to note that the Stein control variates constructed via Eq. (2.76) are universal in the sense that, with a suitably chosen auxiliary function g, they can represent arbitrary control variates. In contrast, the special case in Eq. (2.75) lacks this universality; a concrete counterexample will be presented in Section 2.5.3.2. The universality of the Stein construction was first rigorously established in [117], where the framework was formulated using tools from functional analysis. In addition, alternative theoretical justifications based on algebraic topology [114] and differential equations have been developed to further support and generalize this approach. These derivations are detailed explicitly in Appendix C.

2.5.1.1 Contour deformation and control variates

The contour deformation method can be leveraged to mitigate signal-to-noise problems. When the Euclidean action is real but the observable may take complex values, the expectation value can be expressed as

$$\int_{\mathbb{R}^n} D\phi\, O(\phi) \exp(-S(\phi)) = \int_{\gamma} D\tilde{\phi}\, O(\tilde{\phi}) \exp(-S(\tilde{\phi})), \tag{2.82}$$

where the contour γ lies in the same homology class as $\mathbb{R}^n$. If γ can be parametrized as a deformation of the original real manifold, then the integral can be written as

$$\begin{aligned}\int_\gamma D\tilde{\phi}\, O(\tilde{\phi}) \exp(-S(\tilde{\phi})) &= \int_{\mathbb{R}^n} D\phi \left|\frac{\partial\tilde{\phi}}{\partial\phi}\right| O(\tilde{\phi}(\phi)) \exp(-S(\tilde{\phi}(\phi))) \\ &= \int_{\mathbb{R}^n} D\phi \left(\left|\frac{\partial\tilde{\phi}}{\partial\phi}\right| O(\tilde{\phi}(\phi)) \exp(-S(\tilde{\phi}(\phi)) + S(\phi))\right) \exp(-S(\phi)) \\ &\equiv \int_{\mathbb{R}^n} D\phi\, \tilde{O}(\phi) \exp(-S(\phi)).\end{aligned} \tag{2.83}$$

Thus, the original observable O and the modified observable $\tilde{O}$ have identical expectation values but potentially different variances. This property has been exploited to enhance signal-to-noise ratios in lattice calculations [118–120].

Moreover, this technique can be reinterpreted through the lens of control variates by defining

$$f_c(\phi) = O(\phi) - \tilde{O}(\phi). \tag{2.84}$$

However, it has been argued that a perfect contour yielding zero variance might not exist in general [121], which limits the scope of contour deformation for more intricate models.

In addition to addressing signal-to-noise issues, control variates have also been utilized to alleviate sign problems [122, 123] by subtracting fluctuations in the partition function:

$$Z = \int D\phi\, \mathrm{e}^{-S(\phi)} = \int D\phi\, \left(\mathrm{e}^{-S(\phi)} - f(\phi)\right), \tag{2.85}$$

as long as $\int D\phi\, f(\phi) = 0$. Within this framework, the contour deformation method can be understood

as a special case of control variates, where

$$f(\phi) = \mathrm{e}^{-S(\phi)} - \left|\frac{\partial\tilde{\phi}}{\partial\phi}\right| \mathrm{e}^{-S(\tilde{\phi}(\phi))}, \tag{2.86}$$

with $\tilde{\phi}$ representing the deformed contour.

Although contour deformation may not completely eliminate sign or signal-to-noise problems, it can yield substantial improvements with even simple parameterizations—for instance, the constant shift along imaginary axes of the integration variables [124].

Guidance for identifying effective contours can also be drawn from the theory of perfect control variates. Since perfect control variates (as in Eq. (2.81)) respect the symmetries of the observable and action, it is natural to expect that optimal deformed contours should exhibit the same symmetries. For example, if the action is invariant under spacetime translations, then the deformed contour should preserve translational symmetry. This perspective can significantly reduce the space of candidate deformations and guide the search for efficient variance-reducing contours.

2.5.2 Neural control variates

The construction of control variates hinges on identifying a function f with zero expectation value that is strongly correlated with the observable of interest. In the Stein-based approach, this control variate f is derived from an auxiliary function g, as in Eq. (2.76). To proceed, one must specify a functional form or parametrization for g. A natural and highly expressive choice is to employ neural networks.

Neural networks are, under mild assumptions, universal function approximators [125,126], offer a flexible framework for learning complex, high-dimensional functions directly from data. This makes

them particularly well-suited for constructing effective control variates in settings where analytic intuition or hand-crafted ansätze are unavailable or insufficient.

This strategy gives rise to neural control variates (NCVs), wherein a neural network is trained to produce a control variate f correlated with the target observable, while ensuring that its expectation value remains zero. Training is typically guided by minimizing the variance of the modified observable $\tilde{O} = O - f$, often using Monte Carlo estimates of the relevant expectations.

The NCV framework has proven versatile and powerful, showing promising results in a variety of domains. In lattice field theory, neural networks have been used to construct control variates for difficult observables with large variance [17–19]. In the context of Bayesian inference, NCVs have been applied to improve the efficiency of posterior sampling and model evidence estimation [127–131]. More broadly, applications in probabilistic machine learning [132, 133] demonstrate that NCVs can generalize across domains, offering a data-driven route to variance reduction in high-dimensional integration tasks.

This neural approach effectively bridges statistical variance reduction techniques with modern function approximation, enabling scalable and adaptive solutions to signal-to-noise and sign problems in computational physics and beyond.

2.5.2.1 Imposing symmetry

In many physical systems, both the action and observables are constrained by symmetries. While general classes of neural networks are universal approximators, incorporating symmetry directly into the construction of control variates is often more efficient than relying on the network to learn it from data.

A primary example is translational invariance. To ensure that the control variate f respects this symmetry, it suffices to construct the auxiliary function g to be translationally covariant. A general and practical way to achieve this is to define g through a scalar function $g_0 : M^V \to \mathbb{R}$, where M^V denotes the field space over the lattice, such that

$$g(T_y\phi)_x = g(\phi)_{x+y}, \tag{2.87}$$

where T_y denotes a translated field configuration by vector y, and $g(\phi)_x$ is the value of $g(\phi)$ at site x.

This construction is the most general form of translationally covariant functions. To see this, note that translational invariance of the control variate f implies

$$f(T_y\phi) = f(\phi) \tag{2.88}$$

for all translations y. Using the definition of f in terms of the auxiliary function g,

$$\begin{aligned} f(T_y\phi) &= \sum_x \left(\frac{\partial g_x(T_y\phi)}{\partial \phi_{x+y}} - g_x(T_y\phi) \frac{\partial S(T_y\phi)}{\partial \phi_{x+y}} \right) \\ &= \sum_x \left(\frac{\partial g_{x+y}(\phi)}{\partial \phi_{x+y}} - g_{x+y}(\phi) \frac{\partial S(\phi)}{\partial \phi_{x+y}} \right) = f(\phi). \end{aligned} \tag{2.89}$$

This equality holds for all y if and only if

$$g(T_y\phi)_x = g(\phi)_{x+y}, \tag{2.90}$$

which defines the condition for translational covariance of g. Additionally this g can be defined by its

first component: $g(x)_0 \equiv g_0(x)$. Then $g_0(x) : M^V \to \mathbb{R}$ and then

$$g(\phi)_x = g(T_x\phi)_0 = g_0(T_x\phi). \tag{2.91}$$

A similar construction applies to other symmetries. For example, under a Z_2 symmetry $\phi \to -\phi$, one can enforce that g is an odd function. In the case of feed-forward neural networks, this can be achieved by removing biases in all layers and using odd activation functions (e.g., $\mathrm{arcsinh}$). Such symmetry constraints reduce the effective parameter space and improve training efficiency while ensuring that the learned control variates respect the underlying physical symmetries.

By incorporating symmetry at the level of function design, one can greatly enhance the efficiency and generalization of neural control variates, particularly in structured physical problems such as lattice field theory.

2.5.2.2 Loss function

The natural loss function for training g is the variance of $O - f$:

$$L(w) = \langle (O - f_w)^2 \rangle - \langle O - f_w \rangle^2, \tag{2.92}$$

where f_w is a Stein control variate, parametrized by w of the network defining g. Note that $\langle O - f_w \rangle^2$ can be omitted since it does not affect the new variance.

Since the number of configurations are small (typically of the order of hundreds in lattice QCD) while the neural network can easily contain a much larger number of parameters, the risk of overfitting exists. In that case, even though $\langle f \rangle = 0$ by construction, Eq. (2.92) is minimized by having

$f(\phi_i) \approx O(\phi_i)$ for every ϕ_i in the sample, leading to the sample average $\overline{f} \approx \overline{O}$. This is not a good approximation to the ideal control variate f. This problem was recognized in [127] where the authors suggest to minimize instead the quantity:

$$L(w) = \langle O - f_w - \mu_0 \rangle^2, \tag{2.93}$$

where μ_0 is the sample average of the observable $\mu_0 = \overline{O}$.

In this case, overfitting leads to $f_w(\phi_i) \approx O(\phi_i) - \mu_0$ for every ϕ_i in the sample and $\overline{f_w} \approx 0$, a much better approximation to the ideal control variate. In fact, Refs. [117, 127, 134] suggested to use, instead of a fixed value μ_0, a variable μ that is updated during the training process just like the parameters w of f_w. We verified that in our examples this last procedure was significantly better and all results presented below are obtained with this last procedure[4]. Finally, to further avoid overfitting, we applied the L_2 regularization:

$$L(w, \mu) = \langle (O - f_w - \mu)^2 \rangle + \delta \sum w^2, \tag{2.94}$$

where $\sum w^2$ is the sum of the squares of all neural network parameters and δ is the regularization parameter[5].

2.5.2.3 Training with correlated configurations

Since the estimation of statistical errors for observables requires uncorrelated data, it is standard practice to use decorrelated configurations when computing final estimates. This methodology has

[4]Note that $\langle O \rangle$ is still estimated by $\overline{(O - f_w)}$, not $\overline{(O - f_w - \mu)}$.

[5]If one uses improved optimizers such as *ADAM* [98], *AdamW* [135] can be for proper weight decay.

carried over to recent developments in the construction of control variates using Monte Carlo samples, where it is widely assumed that independent configurations are employed for training [17, 105, 136]. The rationale is straightforward: autocorrelated configurations contain no genuinely new information compared to decorrelated ones, so training on them should, in principle, yield similar results to training on a smaller set of independent samples.

However, while correlated samples are generally regarded as statistically redundant for the purposes of error estimation, they are far from meaningless. In fact, correlations arise precisely because the Markov chain explores the probability distribution through local updates. As a result, successive configurations reflect smooth transitions through the configuration space, which encode structural information about the distribution beyond what is captured by a sparse set of decorrelated samples[6]. This additional information—arising from correlations between successive configurations in the Markov chain—can be valuable when training control variates or machine learning models. By utilizing correlated samples, one can potentially capture more detailed features of the observable's fluctuation patterns, leading to more effective variance reduction for a given computational budget.

Furthermore, it is important to recognize that the autocorrelation properties of the original observable O may differ from those of the improved observable $\tilde{O}$ obtained via control variates. While the design of control variates often aims to produce an improved estimator that fluctuates in tandem with O, it does not guarantee that the autocorrelation time will be preserved. If $\tilde{O}$ happens to decorrelate more rapidly than O, then discarding correlated configurations—based on the autocorrelation time of O—may be counterproductive. In such a scenario, the data considered "correlated" for O may in fact be decorrelated, or at least weakly correlated, for $\tilde{O}$, leading to the unnecessary exclusion of valuable training data. This misalignment underscores the importance of evaluating autocorrelation

[6]The same reasoning applies even when global updates are used in the Markov chain.

properties separately for each observable involved in the control variate construction.

Note that the above argument does not mean that one will use the same number of correlated configurations for training rather than the same number of decorrelated configurations. If we have a fixed length of Monte Carlo chains, for example, 10^3 decorrelated configurations, we can measure 100 times more often in the same chain to train the neural control variates with 10^5 correlated configurations, while they are statistically the same for measuring uncertainties.

In principle, a good control variate should exhibit fluctuations similar to those of O, differing only in that its expectation value is zero. This similarity suggests that f might retain comparable autocorrelation properties. Nevertheless, the relationship does not guarantee that $\tilde{O}$ has the same autocorrelation behavior. Consider, for instance, the perfect control variate defined by $f_P = O - \langle O \rangle$. This choice eliminates all stochastic fluctuations, yielding an estimator with zero variance and, consequently, zero autocorrelation time. While such a perfect control variate is rarely attainable, it highlights that variance reduction can significantly alter the autocorrelation structure, and this should be considered when interpreting results.

Given these considerations, a safer and more flexible strategy is to retain and utilize correlated samples when training control variates. By doing so, one avoids discarding potentially informative data and makes more efficient use of available computational resources—especially in scenarios where generating additional independent configurations is expensive. Importantly, this approach does not conflict with standard practices for error estimation. While training may proceed on correlated data, the final evaluation of observables and their uncertainties should still rely on decorrelated configurations or techniques such as the blocked jackknife or autocorrelation-aware estimators. In this way, one can exploit the full informational content of MCMC samples while maintaining statistical rigor in uncertainty quantification.

2.5.3 Results

2.5.3.1 Scalar field theory

As a testbed for variance reduction techniques in quantum field theory, we consider the scalar field theory with a ϕ^4 interaction—a widely used toy model in this context [16, 17, 105, 137]. We study this theory in two, three, and four spacetime dimensions. The discretized lattice action takes the form:

$$S = \frac{1}{2}\sum_{x,\mu}(\phi_{x+\mu} - \phi_x)^2 + \frac{m^2}{2}\sum_x \phi_x^2 + \frac{\lambda}{4!}\sum_x \phi_x^4, \tag{2.95}$$

where $x = (t, \vec{x})$ denotes the spacetime lattice site, and $\mu = 1, \ldots, d$ indexes the spacetime directions.

The observable of interest is the zero-momentum two-point correlation function, defined as

$$O(t) = \sum_{t'}\left(\frac{1}{V}\sum_{\vec{x}}\phi_{t+t',\vec{x}}\right)\left(\frac{1}{V}\sum_{\vec{x}}\phi_{t',\vec{x}}\right), \tag{2.96}$$

where V is the spatial volume of the system.

In Figure 2.9, we present the training history of neural control variates on a 20×20 lattice for both weak and strong coupling regimes, $\lambda = 0.5$ and $\lambda = 24.0$, using neural networks with and without hidden layers. The target observable is the mid-lattice two-point correlator, Eq. (2.96), evaluated at $t = L_0/2 = 10$, where the signal-to-noise ratio is the worst.

The plots show the ratio of standard deviations—indicating the improvement in statistical uncertainty—as a function of training step. A dataset of 10^4 fully decorrelated configurations is utilized for training the neural network, while an independent test set of 10^3 configurations is served to evaluate the variance of the estimator. The results demonstrate significant variance reduction, though the

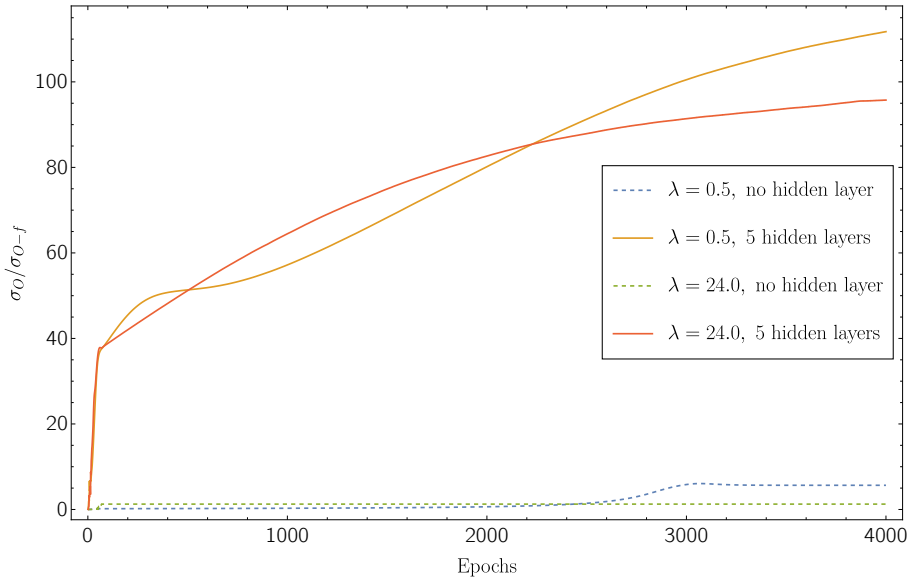

Figure 2.9: Training history of neural control variates on a 20×20 lattice at weak ($\lambda = 0.5$) and strong ($\lambda = 24.0$) coupling. The curves show the ratio $\sigma_{\text{Raw}}/\sigma_{\text{CV}}$, quantifying the improvement in standard deviation due to control variates. Dashed lines correspond to networks without hidden layers (i.e., linear control variates), while solid lines represent networks with 5 hidden layers of 4 neurons each. 10^4 decorrelated configurations are used for training, and 10^3 independent samples are used to estimate the variance.

performance depends sensitively on the size of the training set.

These findings can be compared to those in [105], which used a different, more direct construction of control variates. The method in [105] is exact in the free-field limit and performs well at weak coupling. Neural control variates also excel in this regime, particularly when a linear model (no hidden layers) is used, consistent with the sufficiency of linear control variates in free theories. However, at strong coupling, the linear ansatz offers limited improvement, as shown by the dashed lines in Figure 2.10. By contrast, introducing nonlinearity via hidden layers yields significantly better variance reduction, highlighting the advantage of flexible neural architectures in strongly interacting systems.

Reducing the uncertainty of a single time-slice does not necessarily lead to improved estimates of derived quantities such as the mass. To explore this, we applied our method to anisotropic 40×10 lattices with couplings near the continuum limit. Ideally, one would construct separate control variates

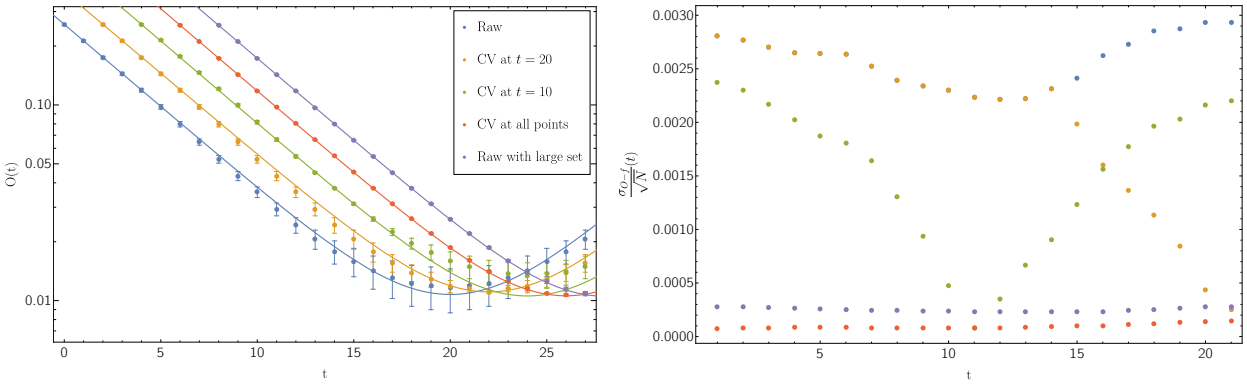

Figure 2.10: Correlation functions on a 40×10 lattice with parameters $m^2 = 0.01$ and $\lambda = 0.1$. Both raw and control variate (CV) results are shown. A total of 2×10^3 samples were used, with 10^3 reserved for training the neural control variate and all samples used for evaluating observables. For comparison, high-statistics raw results computed from 2×10^5 samples are also included. The left panel shows the correlation functions along with their fits; horizontal shifts have been applied for clarity. The right panel displays the corresponding standard deviations of the correlators from the left panel.

for each time-slice, but this is computationally prohibitive. Instead, we trained NCVs to optimize the variance at $t = L_0/2 = 20$ or $t = L_0/4 = 10$. The resulting correlators are shown in Figure 2.10, and mass fits using the functional form

$$C(t) = A(\mathrm{e}^{-mt} + \mathrm{e}^{-m(L_0-t)}) \tag{2.97}$$

are summarized in Table 2.1. The fits incorporate correlations between time-slices, and a regularization strength $\delta = 0.1$ was used during training.

	A	m	χ^2/dof
Raw, $N = 2 \times 10^3$	0.000112(18)	0.1935(40)	0.55
CV at $t = 20$	0.0001191(51)	0.1920(13)	0.29
CV at $t = 10$	0.0001083(44)	0.1944(13)	0.55
CV at all points	0.0001096(7)	0.1938(2)	0.91
Raw, $N = 2 \times 10^5$	0.0001088(18)	0.1940(4)	0.34

Table 2.1: The fitted values of correlators in Figure 2.10.

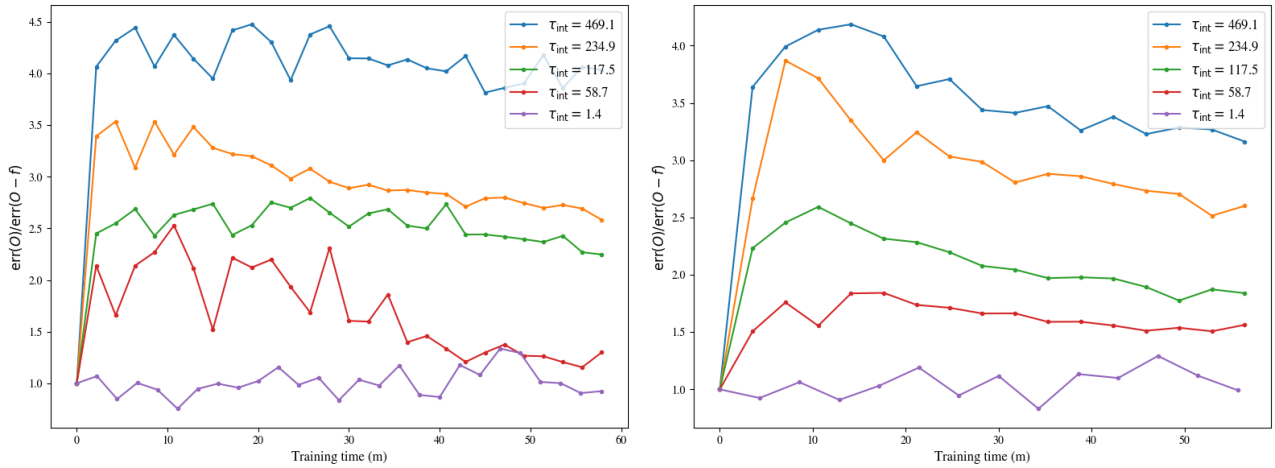

Figure 2.11: Training history for the three-dimensional scalar field theory on a $32 \times 8 \times 8$ lattice with parameters $m^2 = 0.01$ and $\lambda = 0.1$. All configurations are drawn from the same Monte Carlo chain, but with different autocorrelation times. The left panel shows the standard deviation reduction ratio, $\sigma_{\mathrm{Raw}}/\sigma_{\mathrm{CV}}$, using a neural network with two hidden layers of 32 neurons each. The right panel shows the same quantity for a deeper architecture with four hidden layers of 64 neurons each.

At the optimized time-slice ($t = 20$ or $t = 10$), the variance is reduced by a factor of ~ 20. Due to correlations between time-slices, neighboring points also benefit, though to a lesser extent. In this setup, the uncertainty in the extracted mass is reduced by a factor of ~ 3, equivalent to using a dataset nine times larger.

While targeting a single time-slice yields substantial local improvements, the global impact on mass estimation remains moderate. To address this, we employed *transfer training*—using the trained model from one time-slice as the initialization for the next. This strategy enables us to construct effective control variates across all time-slices, achieving ~ 20 times variance reduction uniformly. As shown in Table 2.1, the resulting mass uncertainty matches that obtained from high-statistics datasets, confirming the reliability and consistency of the neural control variate approach even with limited data.

We now turn to training with correlated Monte Carlo samples. Configurations are generated using the Metropolis-Hastings algorithm, with one measurement per sweep. The training dataset includes 10^4 samples, while the test set consists of 10^3 fully decorrelated configurations, obtained by

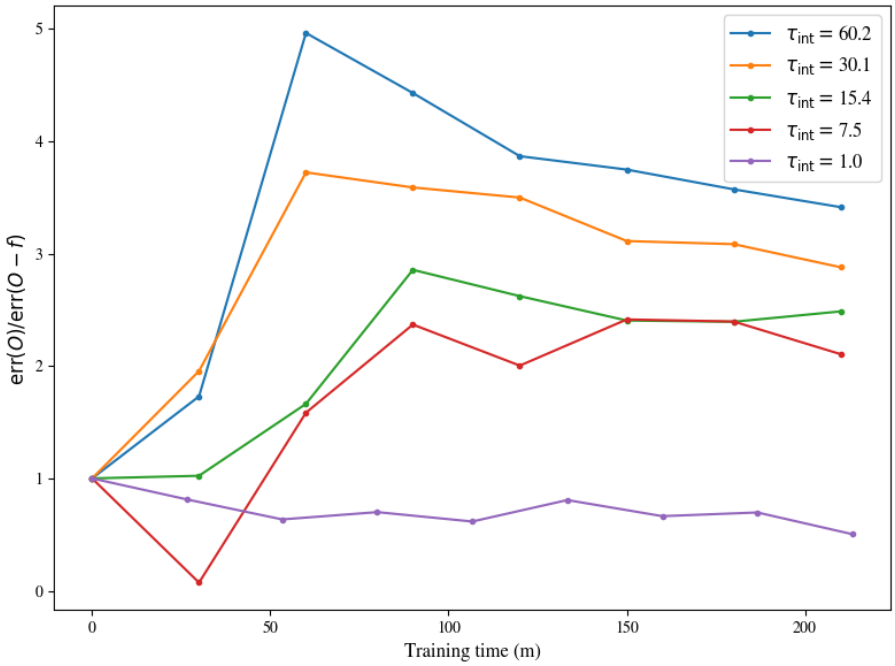

Figure 2.12: Training history for the four-dimensional scalar field theory on a $16 \times 8 \times 8 \times 8$ lattice with parameters $m^2 = 0.1$ and $\lambda = 0.5$. Configurations are generated from a single Monte Carlo chain, with different measurement intervals (in units of sweeps) used to control the degree of sample correlation.

inserting sufficiently long intervals between measurements.

Training is performed using the *ADAM* optimizer with a fixed learning rate of 10^{-4} and a mini-batch size of 32. To prevent exploding gradients, we employ gradient clipping [138] with a threshold of 1.0. To focus solely on the impact of sample correlation, L_2 regularization is omitted, as it primarily mitigates overfitting and is not expected to qualitatively affect the results.

Figure 2.11 shows results in three dimensions on a $32 \times 8 \times 8$ lattice. Two network architectures are compared: the left panel uses two hidden layers with 32 neurons each, while the right panel uses four hidden layers with 64 neurons each. In both setups, increasing the number of training samples leads to greater error reduction—even when the samples are correlated.

Results in four dimensions are displayed in Figure 2.12. As before, we observe that adding more training data, even if the samples are not fully decorrelated, enhances the effectiveness of variance

reduction when the total MCMC and training budget is held fixed.

2.5.3.2 2D $U(1)$ gauge theory

We now turn to numerical experiments involving a simplified model: two-dimensional $U(1)$ gauge theory with open boundary conditions. While gauge theories are typically formulated in terms of link variables, in two dimensions with open boundaries, it is possible to fix the gauge such that the path integral can be expressed directly in terms of plaquette variables [118]. In this formulation, the Euclidean action takes the form

$$S = -\beta \sum_x \left(1 - \cos \theta_x^P\right), \tag{2.98}$$

with a corresponding partition function

$$Z = \int \prod_x d\theta_x^P \, \mathrm{e}^{-\beta \sum_x \left(1 - \cos \theta_x^P\right)}. \tag{2.99}$$

As an observable, we consider the Wilson loop over a region A:

$$O(A) = \prod_{x \in A} \mathrm{e}^{\mathrm{i}\theta_{\mathrm{P}}}, \tag{2.100}$$

which is known to suffer from a severe signal-to-noise problem. For all results shown here, the coupling is fixed at $\beta = 5.55$, corresponding to a fine lattice spacing close to the continuum limit [118]. As shown in Figure 2.13, the signal-to-noise ratio of this observable decreases exponentially with increasing area.

To better isolate the effect of correlation in training, we employ unusually high acceptance rates in the Metropolis-Hastings algorithm, ranging from 0.9 to 0.95—substantially higher than typical val-

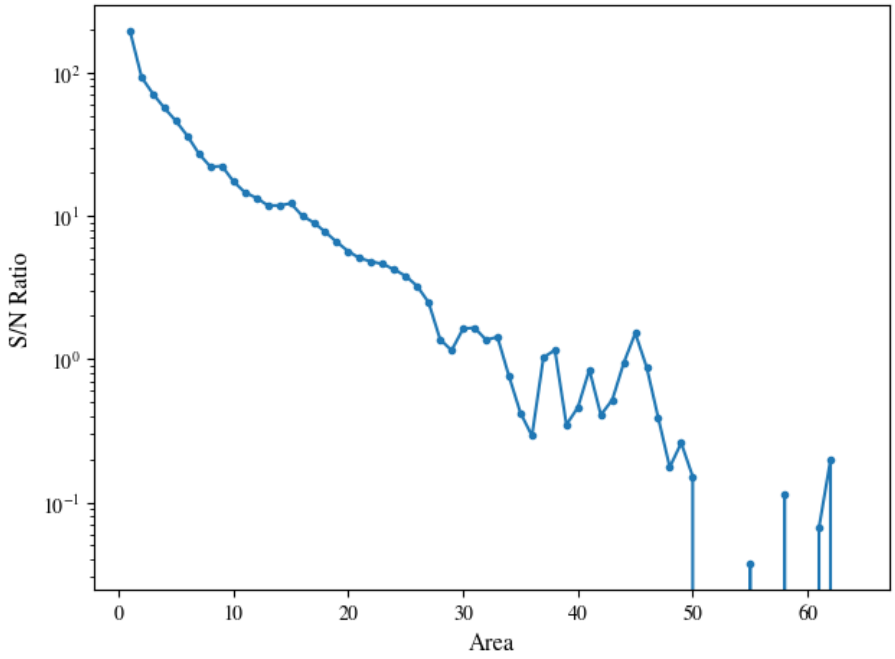

Figure 2.13: Signal-to-noise ratio of Wilson loops as a function of loop area at coupling $\beta = 5.55$, estimated from 10^3 decorrelated samples.

ues (0.3–0.7). The training dataset consists of 10^4 correlated samples, while the test set comprises 10^3 fully decorrelated configurations, generated by increasing the interval between measurements in the Markov chain.

For this toy model, we do not impose symmetry constraints on the neural network. Specifically, the network is not designed to enforce translational invariance or Z_2 symmetry, as discussed in Section 2.5.2.1. Instead, we use a standard feedforward architecture to parametrize g, with input features given by $\cos\theta^P$ and $\sin\theta^P$, representing the real and imaginary parts of the plaquettes. The activation function used is the Continuously Differentiable Exponential Linear Unit (CELU) [139], defined as

$$\mathrm{CELU}(x,\alpha) = \begin{cases} x & \text{if } x \geq 0, \\ \alpha\left(\exp\left(x/\alpha\right) - 1\right) & \text{otherwise,} \end{cases} \tag{2.101}$$

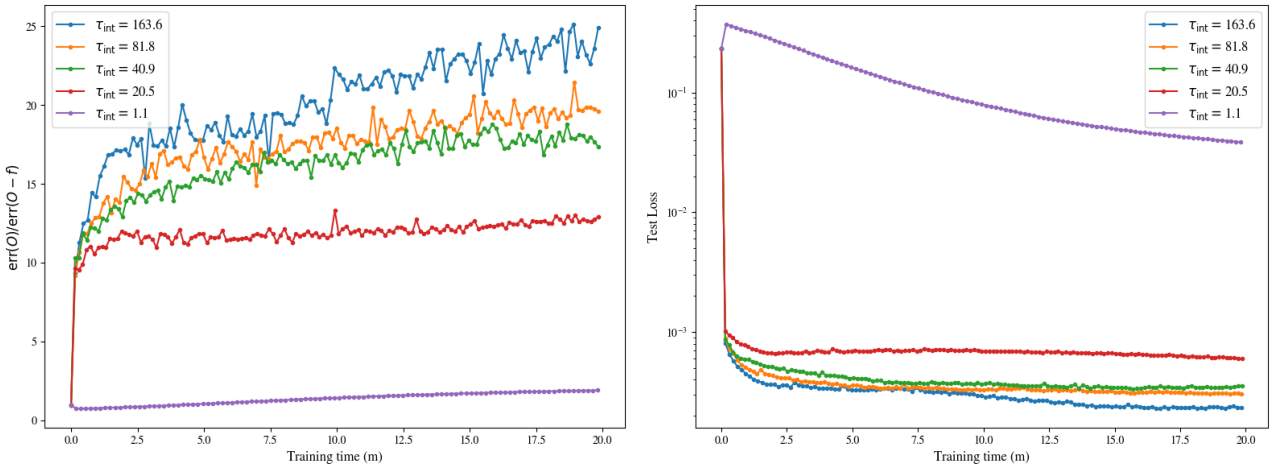

Figure 2.14: Training dynamics for the Wilson loop with area $A = 4$ at coupling $\beta = 5.55$, based on a single Monte Carlo chain sampled with integrated autocorrelation times. The left panel illustrates the reduction in standard deviation compared to the original observable O. The right panel shows how the loss function evolves during training, evaluated on an independently decorrelated test set.

with $\alpha = 1.0$ throughout this work.

Training is performed using the *ADAM* optimizer with a fixed learning rate of 10^{-4} and a mini-batch size of 32. To mitigate the risk of exploding gradients, gradient clipping is applied with a threshold of 1.0.

Figure 2.14 illustrates the effect of using varying degrees of autocorrelation in training samples. The left panel shows the variance reduction achieved by the neural control variate relative to the raw observable O, while the right panel tracks the test loss, Eq. (2.94). A network with three hidden layers of 32 nodes each is used. The integrated autocorrelation time is varied, controlling the correlation among training configurations, and the results for fully decorrelated samples are presented for comparison.

To study scaling behavior, we increase the area of the Wilson loop to $A = 16$. The corresponding results are presented in Figure 2.15. In this case, we use a smaller network with two hidden layers of 16 nodes each. The overall trend persists: incorporating a larger number of (partially correlated)

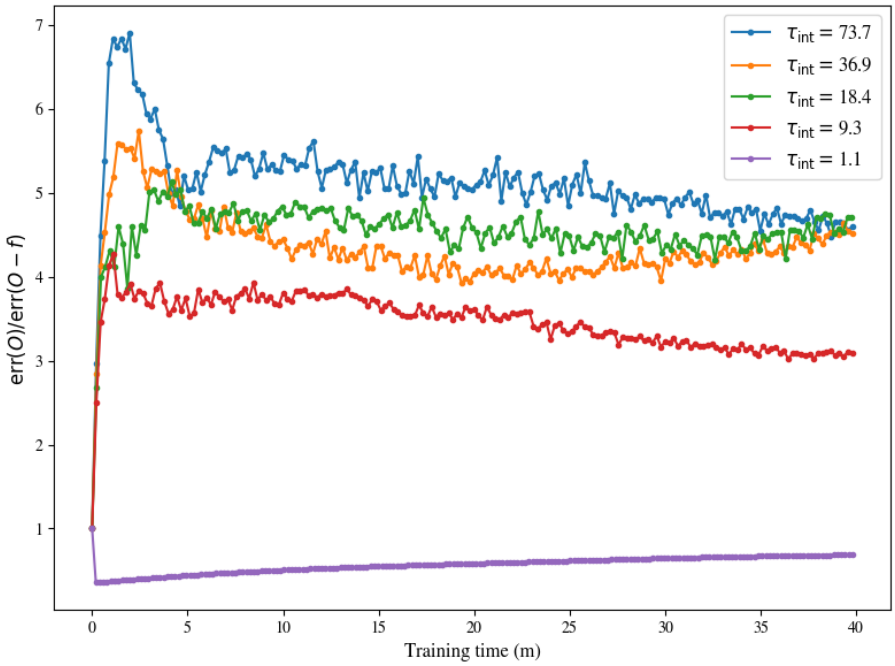

Figure 2.15: Training history for the Wilson loop with area $A = 16$ at coupling $\beta = 5.55$, using a single Monte Carlo chain sampled at various integrated autocorrelation times.

configurations in the training data leads to better variance reduction, assuming a fixed total length of MCMC chain. However, due to the more intricate loss landscape arising from the larger number of variables, the reduction in variance is less pronounced compared to the $A = 4$ case.

While Eq. (2.76) provides a general framework for constructing control variates, one can leverage model-specific structures to design more effective variants in practice. In this toy model, the factorized form of both the action and the observable allows for a tailored control variate of the form:

$$f = \prod_{x \in A} \mathrm{e}^{i\theta_x^P} - \prod_{x \in A} \left(\mathrm{e}^{i\theta_x^P} - f_1(\theta_x^P) \right), \quad \text{where} \quad f_1(\theta^P) = \frac{dg}{d\theta^P} - g \frac{dS_1}{d\theta^P}. \tag{2.102}$$

Here, $S_1(\theta^P) = \beta(1 - \cos\theta^P)$ denotes the single-plaquette action. Since the observable in Eq. (2.100) is complex-valued while its expectation value is real, f_1 is generally complex. Accordingly, the function g is modeled using two separate neural networks, g_r and g_i, such that $g(\theta^P) = g_r(\theta^P) + i\, g_i(\theta^P)$.

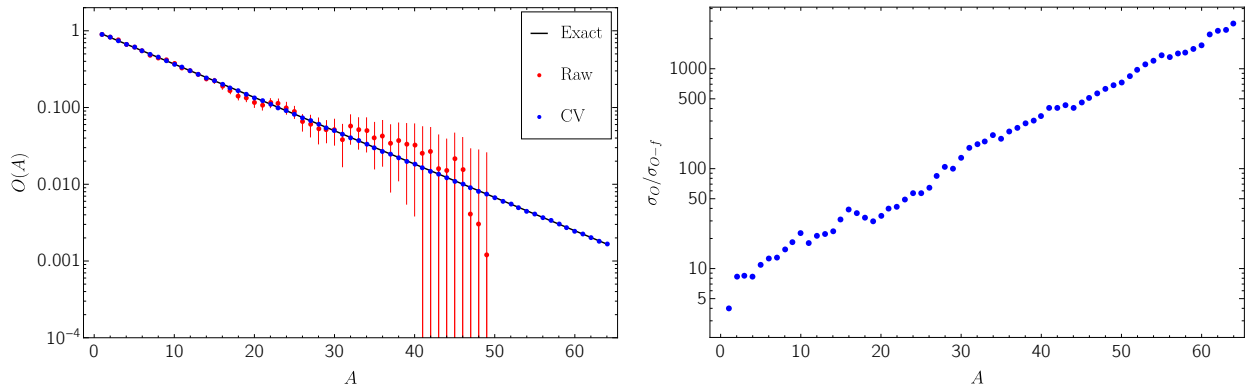

Figure 2.16: Wilson loops computed on an 8×8 lattice at coupling $\beta = 5.555$, using the control variate ansatz from Eq. (2.102). The left panel shows the expectation values of the Wilson loops with 1σ error bars. The right panel displays the achieved variance reduction, quantified by the ratio of standard deviations $\sigma_{\text{Raw}}/\sigma_{\text{CV}}$ between the original observable and the control variate.

Figure 2.16 shows the error reduction achieved using these neural control variates. Because two-dimensional gauge theories are exactly solvable, the exact expectation value $(I_1(\beta)/I_0(\beta))^A$ is also shown for comparison. For each area, both g_r and g_i are represented by feedforward networks with four hidden layers and 16 neurons per layer. To ensure that the control variate f satisfies the periodic boundary condition required for Stein's identity, Eq. (2.74), the input to g is constructed using $\cos\theta^P$ and $\sin\theta^P$—the real and imaginary parts of the plaquette angle. As illustrated in the right panel of Figure 2.16, although the raw observable suffers from an exponential signal-to-noise degradation with increasing area, the neural control variates effectively mitigate this problem, yielding exponential error suppression.

This example also illustrates a limitation of the generic control variate form in Eq. (2.75): it may not encompass the "perfect" control variate. Consider a Wilson loop with area $A = 2$, parameterized by two plaquette angles, x and y. Finding the perfect control variate corresponds to solving the partial

differential equation:

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} - \beta \sin(x) g - \beta \sin(y) g = \exp(i(x+y)) - \left(\frac{I_1(\beta)}{I_0(\beta)}\right)^2. \tag{2.103}$$

This equation admits a general solution using a change of variables to $x+y$ and $x-y$, along with the method of integrating factors:

$$\begin{aligned} g(x,y) = & \, \mathrm{e}^{-2\beta \cos\left(\frac{x+y}{2}\right)\cos\left(\frac{x-y}{2}\right)} \\ & \times \left(\int^{\frac{x+y}{2}} du \, \mathrm{e}^{2\beta \cos(u)\cos\left(\frac{x-y}{2}\right)} \left(\mathrm{e}^{2iu} - \left(\frac{I_1(\beta)}{I_0(\beta)}\right)^2 \right) + C\left(\frac{x-y}{2}\right) \right), \end{aligned} \tag{2.104}$$

where C is a function determined by boundary conditions. However, it is straightforward to verify that this general solution lacks periodicity, implying that the perfect control variate falls outside the function class defined by Eq. (2.75).

3.1 Overview

Quantum computers, originally proposed by Richard Feynman [140], represent a fundamentally different paradigm for simulating quantum systems—one that aligns naturally with the structure of the problems they are designed to solve. Unlike classical computers, which face exponential resource requirements when simulating large or entangled quantum systems, quantum computers can in principle perform these simulations using polynomial resources. This promise is particularly compelling in areas where classical simulations are known to be intractable.

A canonical example illustrating the power of quantum algorithms is Shor's factoring algorithm [141], which solves the problem of factoring large integers exponentially faster than the best-known classical approaches [142–145]. However, the advantages of quantum computing extend far beyond cryptographic applications. Quantum simulations have the potential to address some of the most severe computational bottlenecks in physics, including the infamous sign problem [44]. This problem arises in various contexts—such as real-time dynamics [96, 146, 147] and systems at finite chemical potential [148, 149]—where classical Monte Carlo techniques fail due to exponentially suppressed signal-to-noise ratios stemming from complex or oscillatory weights in the path integral.

With the rapid development of quantum hardware, including the emergence of noisy intermediate-scale quantum (NISQ) devices, the use of quantum simulations to circumvent—not merely mitigate—these problems has transitioned from a theoretical aspiration to an increasingly viable strategy. In tandem with advances in quantum error mitigation and the construction of Hamiltonian formulations suitable for gauge theories, researchers have begun to design algorithms that can potentially outperform classical techniques in extracting real-time and finite-density physics from first principles.

The standard procedure of a quantum simulation can be divided into three essential components:

1. The preparation of an appropriate quantum state, typically a ground state of the target Hamiltonian.

2. The time evolution of this state under a given Hamiltonian.

3. The measurement of observables of interest.

Each of these stages poses unique challenges, but among them, state preparation remains a particularly demanding and central problem.

Indeed, high-fidelity state preparation is not only technically difficult, but also crucial for the success of a simulation. In many physical applications, such as those involving lattice gauge theories or strongly correlated systems, the quantum simulation must begin in a well-defined and physically relevant quantum state—often the vacuum, a thermal state, or a low-energy eigenstate. The quality of this initial state directly impacts the reliability of downstream observables and the feasibility of subsequent computational steps. Because of this, state preparation has been the subject of intensive research for nearly three decades [150–178].

Unlike certain problems in quantum information theory—such as factoring or Grover's unstructured search [179]—where a single-shot high-probability success suffices, quantum simulations of physical systems typically require the prepared state to be generated many times. Observables are often estimated via statistical sampling, and the reliability of the result depends on repeated measurements of a well-prepared quantum state. For this reason, it can be worthwhile to expend significant resources upfront to generate a high-fidelity initial state if doing so enables more efficient or scalable reuse of that state in repeated runs.

This chapter is dedicated to addressing this foundational issue: the efficient and scalable preparation of physical states for quantum simulations, with particular emphasis on applications to lattice gauge theories. Our goal is to propose, analyze, and compare strategies that can prepare quantum states with high accuracy while maintaining feasibility in both current and future quantum computing architectures. We are especially interested in methods that offer robustness against errors, exhibit favorable scaling with system size, and maintain compatibility with near-term devices.

The chapter is organized as follows.

- Section 3.2 examines the rodeo algorithm proposed in [2], where we observe that iterations with randomly selected times lead to exponentially large fluctuations in the suppression of excited states, and improve the algorithm both in removing fluctuations and improving the suppression of excited states [180].

- Section 3.3 introduces the concept of path length, defining it as a quantitative measure of the difficulty in preparing eigenstates along a given path. The section also explores how the path length scales with system volume in local quantum field theories [181].

- Section 3.4 explores an efficient projection-based algorithm for traversing Hamiltonian paths [181].

- Section 3.5 analyzes various aspects of errors in adiabatic state preparation.

  - Section 3.5.1 and Section 3.5.2 review the adiabatic theorem and the switching theorem, introducing key concepts relevant to understanding and quantifying errors.

  - Section 3.5.3 presents the concept of typical error, providing a framework for error estimation using the switching theorem [182].

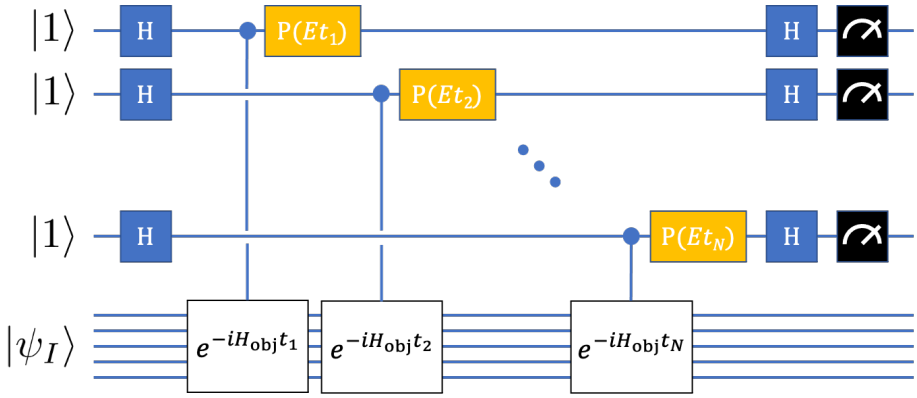

Figure 3.1: The circuit diagram of the rodeo algorithm. This figure is adapted from [2].

– Section 3.5.4 explores how the switching theorem can be applied to reduce errors in practice [183].

– Section 3.5.5 investigates how errors scale in the limit of long path lengths [184, 185].

3.2 Rodeo algorithm

The rodeo algorithm [2, 186, 187] is a cost-effective method for quantum state preparation, designed to project an initial state onto an energy eigenstate in systems with discrete spectra. Its success probability scales with the square of the overlap between the initial and target states, while the time required grows only logarithmically with the desired accuracy. The rodeo algorithm can be used on its own for direct state preparation or combined with techniques such as adiabatic quantum computing [188–192], particularly in cases where the overlap between the initial state and the target state is small. It operates through quantum interferometry and requires a series of iterations to achieve high accuracy, with each iteration involving a specific evolution time. The circuit diagram for the rodeo algorithm is shown in Figure 3.1.

It is worth noting that when implemented on a gate-based digital quantum computer, the "time" in the rodeo algorithm corresponds to a discretized evolution time, such as that appearing in a Trotter-

ized approximation [193, 194]. However, the rodeo algorithm is also well-suited to analog quantum simulators, provided there is a mechanism to control the simulator—effectively turning it on and off—via quantum gates. This concept of gate-controlled analog quantum simulation presents a promising avenue where the rodeo algorithm might demonstrate a decisive advantage. For the purposes of analyzing the rodeo algorithm, we will focus on such idealized scenarios and set aside concerns about Trotterization errors.

While the rodeo algorithm is applicable to projecting onto any discrete eigenstate of a Hermitian operator, for clarity and simplicity, this paper will focus on projecting onto the ground state of a physically relevant Hamiltonian—a likely common use case. Furthermore, to streamline the discussion, we will assume that a constant shift has been applied to the Hamiltonian so that the ground state energy is zero. Nevertheless, the essential points of the analysis remain valid for projections onto arbitrary discrete eigenstates and Hamiltonians with general ground state energies.

Before analyzing the full rodeo algorithm as discussed in [2], it is useful to first examine the effect of a single iteration using the framework of density matrices and reduced density matrices. The Hilbert space of the total system is enlarged by introducing an ancilla qubit, initialized in the state $|\uparrow\rangle$ (in the z basis), while the physical system begins in a pure quantum state. Since the ancilla qubit becomes entangled with the physical system during the iteration, a natural and convenient description of the system is through the use of density matrices, allowing us to easily track both the total system and its reduced subsystems.

Prior to the jth iteration of the algorithm, the physical system is in the state

$$|\psi\rangle_{\text{phys}}^{j,\text{initial}} = \alpha_g^j|\psi_g\rangle + \sum_c \alpha_c^j|\psi_c\rangle \text{ with } |\alpha_g^j|^2 + \sum_c |\alpha_c^j|^2 = 1, \tag{3.1}$$

where $|\psi_g\rangle$ denotes the target eigenstate (typically the ground state), and $|\psi_c\rangle$ represent other eigenstates. The corresponding full density matrix in the enlarged Hilbert space—including the ancilla qubit—is given by

$$\hat{\rho}_{\text{total}}^{j,\,\text{initial}} = |\uparrow\rangle\langle\uparrow| \otimes \hat{\rho}_{\text{phys}}^{j,\,\text{initial}}\,, \tag{3.2}$$

where the subscripts "phys" and "total" refer to the Hilbert spaces of the physical system and the combined system (physical system plus ancilla qubit), respectively. The superscript j indicates the iteration number.

The total system then undergoes unitary evolution described by

$$\hat{\rho}_{\text{total}}^{j,\,\text{final}} = \hat{U}_j\,\hat{\rho}_{\text{total}}^{j,\,\text{initial}}\,\hat{U}_j^{\dagger}, \tag{3.3}$$

where

$$\hat{U}_j = \exp\left(-i\left(\hat{p}_x^{\uparrow} \otimes \hat{H}_{\text{phys}} T_j\right)\right) \text{ with } \hat{p}_x^{\uparrow} \equiv \frac{1}{2}(1+\hat{\sigma}_x)\,. \tag{3.4}$$

Here, $\hat{p}_x^{\uparrow}$ acts on the ancilla qubit and serves as a projector onto the up state in the x basis. This evolution is implemented by first applying a Hadamard gate to the ancilla qubit, then allowing the physical system to evolve under its Hamiltonian for a time T_j conditioned on the ancilla qubit being in the "up" state, followed by another Hadamard gate on the ancilla. The core idea of the rodeo algorithm is the quantum interference between the Hamiltonian-evolved components and the non-evolved components, enabling selective amplification of desired eigenstates across iterations.

The full density matrix after the time evolution is

$$\hat{\rho}_{\text{total}}^{j,\,\text{final}} = |\uparrow\rangle\langle\uparrow| \otimes \hat{\rho}_{\uparrow\uparrow}^{j,\,\text{final}} + |\downarrow\rangle\langle\downarrow| \otimes \hat{\rho}_{\downarrow\downarrow}^{j,\,\text{final}} + |\downarrow\rangle\langle\uparrow| \otimes \hat{\rho}_{\downarrow\uparrow}^{j,\,\text{final}} + |\uparrow\rangle\langle\downarrow| \otimes \hat{\rho}_{\uparrow\downarrow}^{j,\,\text{final}}\,, \tag{3.5}$$

where

$$\begin{aligned}\hat{\rho}^{j,\,\mathrm{final}}_{\uparrow\uparrow} &\equiv |\alpha^j_g|^2 |\psi_g\rangle\langle\psi_g| \\ &+ \sum_{c,c'} \cos\left(\frac{\omega_c T_j}{2}\right) \cos\left(\frac{\omega'_c T_j}{2}\right) \left(\alpha_c \alpha^*_c \,|\psi_c\rangle\langle\psi_{c'}| + \alpha^*_c \alpha_{c'} \,|\psi_{c'}\rangle\langle\psi_c|\right) \\ &+ \sum_{c} \cos\left(\frac{\omega_c T_j}{2}\right) \left(\alpha_g \alpha^*_c e^{-i\omega_c T_j/2} \,|\psi_g\rangle\langle\psi_c| + \alpha^*_g \alpha_c e^{i\omega_c T_j/2} \,|\psi_c\rangle\langle\psi_g|\right),\end{aligned} \tag{3.6}$$

$$\begin{aligned}\hat{\rho}^{j,\,\mathrm{final}}_{\downarrow\downarrow} &\equiv \sum_{c} |\alpha^j_c|^2 \sin^2\left(\frac{\omega_c T_j}{2}\right) |\psi_c\rangle\langle\psi_c| \\ &+ \sum_{c\neq c'} \sin\left(\frac{\omega_c T_j}{2}\right) \sin\left(\frac{\omega_{c'} T_j}{2}\right) \left(\alpha_c \alpha^*_c \,|\psi_c\rangle\langle\psi_{c'}| + \alpha^*_c \alpha_{c'} \,|\psi_{c'}\rangle\langle\psi_c|\right),\end{aligned} \tag{3.7}$$

$$\hat{\rho}^{j,\,\mathrm{final}}_{\downarrow\uparrow} \equiv \left(\sum_c \frac{1}{2}\left(e^{-i\omega_c T_j} - 1\right) \alpha_c |\psi_c\rangle\right) \left(\alpha^*_g \langle\psi_g| + \sum_{c'} \frac{1}{2}\left(e^{i\omega_{c'} T_j} + 1\right) \alpha^*_{c'} \langle\psi_{c'}|\right), \tag{3.8}$$

$$\hat{\rho}^{j,\,\mathrm{final}}_{\uparrow\downarrow} \equiv \left(\alpha^*_g |\psi_g\rangle + \sum_c \frac{1}{2}\left(e^{-i\omega_c T_j} + 1\right) \alpha^*_c |\psi_{c'}\rangle\right) \left(\sum_c \frac{1}{2}\left(e^{i\omega_{c'} T_j} - 1\right) \alpha^*_{c'} \langle\psi_c|\right), \tag{3.9}$$

with $\omega_a = E_a/\hbar$.

The last step is to measure the state of the ancilla qubit. This corresponds to the reduced density matrix for the physical system at the end of the iteration, obtained from Eq. (3.5), in the following form:

$$\begin{aligned}\hat{\rho}^{j,\,\mathrm{final}}_{\mathrm{phys}} &= \hat{\rho}^{j,\,\mathrm{final}}_{\uparrow\uparrow} + \hat{\rho}^{j,\,\mathrm{final}}_{\downarrow\downarrow} \\ &= |\alpha^j_g|^2 |\psi_g\rangle\langle\psi_g| + \sum_c |\alpha^j_c|^2 |\psi_c\rangle\langle\psi_c| \\ &+ \sum_c \cos\left(\frac{\omega_c T_j}{2}\right) \left(\alpha_g \alpha^*_c e^{-i\omega_c T_j/2} \,|\psi_g\rangle\langle\psi_c| + \alpha^*_g \alpha_c e^{i\omega_c T_j/2} \,|\psi_c\rangle\langle\psi_g|\right) \\ &+ \sum_{c\neq c'} \Bigg(\left(\cos\left(\frac{\omega_c T_j}{2}\right) \cos\left(\frac{\omega'_c T_j}{2}\right) + \sin\left(\frac{\omega_c T_j}{2}\right) \sin\left(\frac{\omega'_c T_j}{2}\right)\right) \\ &\qquad \times \left(\alpha_c \alpha^*_c \,|\psi_c\rangle\langle\psi_{c'}| + \alpha^*_c \alpha_{c'} \,|\psi_{c'}\rangle\langle\psi_c|\right) \Bigg) \\ &= P^{j,\,\mathrm{s}} \hat{\rho}^{j,\,\mathrm{s}} + P^{j,\,\mathrm{u}} \hat{\rho}^{j,\,\mathrm{u}},\end{aligned} \tag{3.10}$$

where

$$
\begin{aligned}
P^{j,\,\mathrm{s}} &= |\alpha_g|^2 + \sum_c |\alpha_c^j|^2 \,\cos^2\left(\frac{\omega_c T_j}{2}\right), \\
P^{j,\,\mathrm{u}} &= \sum_c |\alpha_c^j|^2 \sin^2\left(\frac{\omega_c T_j}{2}\right), \\
\hat{\rho}^{j,\,\mathrm{s}} &= \frac{\left(\alpha_g|\psi_g\rangle + \sum_c \alpha_c^j \frac{1}{2}\left(1+e^{-i\omega_c T_j}\right)|\psi_c\rangle\right)\left(\alpha_g^*\langle\psi_g| + \sum_c' \alpha_{c'}^{*j}\frac{1}{2}\left(1+e^{i\omega_{c'}T_j}\right)|\langle\psi_{c'}|\right)}{|\alpha_g^j|^2 + \sum_c |\alpha_c^j|^2\cos^2\left(\frac{\omega_c T_j}{2}\right)}, \\
\hat{\rho}^{j,\,\mathrm{u}} &= \frac{\left(\sum_c \alpha_c^j \sin\left(\frac{\omega_c T_j}{2}\right)|\psi_c\rangle\right)\left(\sum_{c'} \alpha_{c'}^{*j}\sin\left(\frac{\omega_{c'}T_j}{2}\right)|\langle\psi_{c'}|\right)}{\sum_c |\alpha_c^j|^2\sin^2\left(\frac{\omega_c T_j}{2}\right)}.
\end{aligned}
\tag{3.11}
$$

Here, $P^{j,\,\mathrm{s}}$ is the probability that the auxiliary qubit is measured in the up state (along the z axis), corresponding to a "successful" iteration, while $P^{j,\,\mathrm{u}}$ is the probability of finding it in the down state, denoting an "unsuccessful" iteration.

The terms "successful" and "unsuccessful" correspond to the measurement outcome of the ancilla qubit. Measuring the qubit in the down state projects the physical system onto $\hat{\rho}^{j,\,\mathrm{u}}$, which lacks ground state components, making the iteration unsuccessful. Measuring the up state projects the system onto $\hat{\rho}^{j,\,\mathrm{s}}$ enhancing the ground state component and marking a successful iteration. It can be summarized as

$$
\begin{aligned}
&|\psi\rangle_{\mathrm{phys}}^{j,\,\mathrm{initial}} \xrightarrow[\text{iteration}]{\text{successful}} |\psi\rangle_{\mathrm{phys}}^{j,\,\mathrm{final}} \text{ with} \\
&|\psi\rangle_{\mathrm{phys}}^{j,\,\mathrm{initial}} = \alpha_g|\psi_g\rangle + \sum_c \alpha_c|\psi_c\rangle \text{ and } |\psi\rangle_{\mathrm{phys}}^{j,\,\mathrm{final}} = \frac{\alpha_g|\psi_g\rangle + \sum_c \frac{1}{2}\left(1+e^{-i\omega_c T_j}\right)\alpha_c^j|\psi_c\rangle}{\sqrt{|\alpha_g^j|^2 + \sum_c |\alpha_c^j|^2\cos^2\left(\frac{\omega_c T_j}{2}\right)}}\,.
\end{aligned}
\tag{3.12}
$$

3.2.1 Analysis of the rodeo algorithm

Reference [2] primarily emphasized maximizing the suppression factor per iteration of the rodeo algorithm. This might suggest that the optimal strategy is always to maximize suppression in each step and simply run enough iterations to achieve the desired accuracy. However, this approach is not always the most efficient. Greater suppression typically requires longer evolution times, which may cancel out the benefits. Performing more iterations with shorter evolution times may reduce overall computational cost.

The rodeo algorithm suppresses excited-state components based on their energies. Since the method targets discrete spectra, there is a minimal excitation energy, denoted Δ. Knowledge of Δ is crucial for optimization: investing effort to suppress components below this threshold is unnecessary, as they are absent by construction.

This minimum excitation energy defines a natural timescale:

$$T_0 = \frac{2\pi\hbar}{\Delta}, \tag{3.13}$$

corresponding to one full phase period of the lowest excited state.

While time provides a useful proxy for computational cost, it is a dimensionful quantity and can be rescaled alongside energies without altering the problem's difficulty. Additionally, most analyses of the rodeo algorithm involve products of energies and times. For convenience, we introduce dimensionless variables:

$$\zeta = \frac{ET}{2\pi\hbar} = \frac{E}{\Delta}\frac{T}{T_0}, \quad \zeta_{\text{tot}} = \frac{ET_{\text{tot}}}{2\pi\hbar} = \frac{E}{\Delta}\frac{T_{\text{tot}}}{T_0}, \tag{3.14}$$

where T is the time for a single iteration, and T_{tot} is the total runtime. The factor of 2π ensures that ζ naturally counts periods of the phase.

The rodeo algorithm proceeds iteratively, taking as input at the jth step the physical state $|\psi\rangle_{\text{phys}}^{j,\,\text{initial}}$ and, upon a successful iteration, producing the state $|\psi\rangle_{\text{phys}}^{j,\,\text{final}}$. The evolution time for each iteration is denoted T_j.

The key features of the rodeo algorithm are:

1. **Suppression of excited states**:

   After n consecutive successful iterations, the cth excited state is suppressed relative to the ground state by

$$\begin{gathered} s_c = \prod_{j=1}^{n} s_c^j \text{ with } s_c^j = \cos^2\left(\frac{\omega_c T_j}{2}\right) = \cos^2\left(\pi\zeta_c^j\right), \\ \text{where } \omega_c \equiv \frac{E_c}{\hbar} \text{ and } \zeta_c^j \equiv \frac{\omega_c T^j}{2\pi}. \end{gathered} \tag{3.15}$$

   The total suppression factor after n iterations for any component can be expressed in terms of ζ_{tot} (from Eq. (3.14)) as

$$s(\zeta_{\text{tot}}) = s\left(\frac{E_c}{\Delta}\frac{T_{\text{tot}}}{T_0}\right) = \prod_{j=1}^{n} \cos^2\left(\frac{\omega_c T_j}{2}\right) \text{ where } T_{\text{tot}} = \sum_{k=1}^{n} T_n \,. \tag{3.16}$$

2. **Enhancement of the ground state probability**:

   The probability that the system is in the ground state after n successful iterations is

$$\begin{aligned} P_g^n &= \frac{|\alpha_g|^2}{|\alpha_g|^2 + \sum_c \left(\prod_{j=1}^{n} \cos^2\left(\pi\zeta_c^j\right)\right)|\alpha_c|^2} = \frac{P_g^i}{P_g^i + \int_\Delta^\infty dE\; s\left(\frac{E}{\Delta}\frac{T_{\text{tot}}}{T_0}\right)\rho(E)} \\ &= \frac{P_g^i}{P_g^i + (1 - P_g^i)S_E}. \end{aligned} \tag{3.17}$$

Here, $\rho(E)$ is the spectral density of the initial state:

$$\rho(E) \equiv {}^{\text{initial}}_{\text{phys}}\langle\psi|\delta(\hat{H}-E)|\psi\rangle^{\text{initial}}_{\text{phys}}, \tag{3.18}$$

and S_E, the overall suppression factor for excited states, is

$$S_E \equiv \frac{\int_{\Delta-\epsilon}^{\infty} dE\ s\left(\frac{E}{\Delta}\frac{T_{\text{tot}}}{T_0}\right)\rho(E)}{\int_{\Delta-\epsilon}^{\infty} dE\,\rho(E)}, \tag{3.19}$$

where $0 < \epsilon < \Delta$ ensures inclusion of the first excited state. As $n \to \infty$, $S_E \to 0$ and $P_g^n \to 1$.

3. **Success probability**:

   Assuming the ground state energy is precisely known and there are no implementation errors, the probability of n consecutive successful iterations is

$$P^{n\,\text{s}} = |\alpha_g|^2 + \sum_c \left(\prod_{j=1}^{n} \cos^2\left(\pi\zeta_c^j\right)\right)|\alpha_c|^2. \tag{3.20}$$

   In the limit $n \to \infty$,

$$\lim_{n\to\infty} P^{n\,\text{s}} \to |\alpha_g|^2 = |\langle\psi_g|\psi\rangle^{\text{initial}}_{\text{phys}}|^2. \tag{3.21}$$

4. **Restarts after failure**:

   If an iteration fails before the desired ground state accuracy is reached, the initial state must be recreated. On average, the initial state must be prepared $1/P^{n\,\text{s}}$ times to obtain n successful iterations.

5. **Expected total runtime**:

   The expected total time to achieve a projected state with accuracy corresponding to n successful iterations is

$$T^{\text{expected}} = \frac{T^{n\,\text{s}}}{|\langle\psi_g|\psi\rangle_{\text{phys}}^{\text{initial}}|^2}, \tag{3.22}$$

   where $T^{n\,\text{s}}$ is the runtime for n successful iterations. This expression assumes that contributions from excited states are negligible after n iterations.

To understand these features, it is instructive to examine the details of a single iteration. From Eq. (3.1), the probability that the system is initially in the ground state at the start of iteration j is

$$P_g^{j,\,\text{initial}} = |\alpha_g^j|^2. \tag{3.23}$$

According to Eq. (3.12), after a successful iteration, the state becomes:

$$|\psi\rangle_{\text{phys}}^{j,\,\text{final}} = \alpha_g'^j|\psi_g\rangle + \sum_c \alpha_c'^j|\psi_c\rangle, \tag{3.24}$$

with

$$\alpha_g'^j = \frac{\alpha_g^j}{\sqrt{|\alpha_g^j|^2 + \sum_c |\alpha_c^j|^2 \cos^2\left(\pi\zeta_c^j\right)}} \text{ and } \alpha_c'^j = \frac{\alpha_c^j \cos\left(\pi\zeta_c^j\right) e^{-i\pi\zeta_c^j}}{\sqrt{|\alpha_g^j|^2 + \sum_c |\alpha_c^j|^2 \cos^2\left(\pi\zeta_c^j\right)}}\,. \tag{3.25}$$

Thus, each successful iteration increases the ground state probability:

$$P_g^{j,\,\text{s}} = |\alpha_g'^j|^2 = \frac{P_g^{j,\,\text{initial}}}{|\alpha_g^j|^2 + \sum_c |\alpha_c^j|^2 \cos^2\left(\pi\zeta_c^j\right)} \geq P_g^{j,\,\text{initial}}. \tag{3.26}$$

The inequality holds because the denominator is at most unity and is saturated only if all excited components satisfy either $\alpha_c^j = 0$ or $\cos^2(\pi\zeta_c^j) = 0$—a highly unlikely scenario. Therefore, each successful iteration reliably enhances the ground-state component.

From Eq. (3.25), it is clear that the enhancement of the ground state probability comes through the relative suppression of components other than the ground state, which is given by

$$s_c^j \equiv \left(\frac{|\alpha_c'^j|}{|\alpha_g'^j|} \Big/ \frac{|\alpha_c^j|}{|\alpha_g^j|}\right)^2 = \cos^2\left(\pi\zeta_c^j\right). \tag{3.27}$$

Since this suppression happens for each iteration, the net suppression for component c is the product $\prod_{j=1}^{n} s_c^j$ which establishes feature 1 above; feature 2 follows consequently.

The reduced density matrix for the physical system is given in Eq. (3.10). Because the reduced density matrix traces over the ancilla qubit, it accounts for both successful and unsuccessful iterations. From the explicit form of Eq. (3.10), it is evident that the total probability for the physical system to be in the ground state—including contributions from both successful and unsuccessful iterations—is simply $|a_g^j|^2$. As noted earlier, this is precisely the initial probability of the system being in the ground state at the start of the jth iteration. Therefore, when considering the full evolution, including both successful and unsuccessful outcomes, the total probability for the ground state remains unchanged by each iteration. Since this property holds for all iterations, it follows that the overall ground state probability is preserved throughout, remaining equal to its initial value before any iterations have been applied. Feature 3 follows immediately from this observation, while features 4 and 5 are natural consequences of feature 3.

Once the number of iterations and the times assigned to each iteration are fixed, the cumulative suppression factor for components with any given energy is fully determined. However, the algo-

rithm as described thus far remains incomplete: explicit criteria for selecting the iteration times and determining the total number of iterations have not yet been specified.

Note that the suppression factor for each iteration depends solely on $\zeta_c^j = E_c T^j/(2\pi\hbar)$. Since the objective is to suppress contributions from various excited-state energies present in the initial state, the key is to select iteration times that effectively target these components. It is clear that using the same time for each iteration is suboptimal: for energies satisfying $\omega_c T^j$ equal to an odd integer (with $\omega_c = E_c/\hbar$), the suppression factor is unity—that is, no suppression occurs—regardless of the number of iterations performed. Moreover, components with energies near these points will experience only weak suppression. The remainder of this section is devoted to exploring various strategies for selecting the iteration times in order to overcome these limitations and ensure efficient suppression of unwanted components.

3.2.2 Optimizing the random rodeo algorithm

In its original formulation, the algorithm selected iteration times randomly from a Gaussian distribution with a predetermined standard deviation [2]. We will refer to this version of the rodeo algorithm as the random rodeo algorithm (RRA). The underlying logic of RRA is simple: the iteration times are uncorrelated, except insofar as they share the same overall standard deviation. Given this lack of correlation, one might reasonably expect that, after many iterations, all excited-state components will experience significant suppression, as the random variation in times prevents the formation of coherent regions that remain largely unsuppressed.

This section focuses on how to optimize the RRA to achieve maximal suppression with minimal computational cost. In subsequent sections, we will explore alternative time-selection strategies that

generally yield stronger suppression at lower computational expense than the optimized RRA. However, in order to appreciate the relative advantages of these alternative approaches, it is essential first to understand the limits of what can be achieved within the RRA framework.

At present, the optimization problem is not fully specified. Different components of the initial state may experience varying levels of suppression, and the appropriate optimization strategy depends on the information available about the initial state's spectral composition. For instance, if it is known that the excited-state contributions are concentrated within a narrow energy band—say, between Δ and 2Δ—then the algorithm can be tailored specifically to suppress components within this range. In contrast, if the initial state contains substantial contributions spanning many orders of magnitude in excitation energies, a qualitatively different strategy may be required to ensure broad and effective suppression.

Let us consider the case where no detailed information about the initial state is available beyond the fact that it is normalized and that all excited-state components have energies greater than or equal to Δ. In this situation, a conservative optimization strategy is appropriate—one aimed at maximizing suppression of the components with $E \geq \Delta$, which are, on average, the least suppressed (i.e., have the largest expectation value of the suppression factor s) for a fixed average total runtime. Since the random rodeo algorithm selects iteration times randomly, the only control parameter is the average total runtime. While one cannot guarantee that individual components will not experience smaller-than-average suppression, one can ensure that, statistically, the average suppression for all components exceeds a desired threshold. Equivalently, one can ask: how much average total time is required to ensure that the expected suppression for all components with energies above Δ is at least a specified value?

This optimization problem is straightforward to analyze. From Eq. (3.15), the suppression factor

for a component with energy $E = \hbar\omega$ after n iterations is

$$s(\omega) = \prod_{j=1}^{n} \cos^2\left(\frac{\omega T_j}{2}\right), \tag{3.28}$$

where the T_j are randomly and independently drawn for each iteration. Thus, the ensemble average of the suppression factor is:

$$\begin{aligned} s^{\mathrm{Avg}}(\zeta, n) &= \int \left(\prod_{j=1}^{n} dT_j \; p(T_j, T)\right) \cos^2\left(\frac{\omega T_1}{2}\right) \cos^2\left(\frac{\omega T_2}{2}\right) \cdots \cos^2\left(\frac{\omega T_n}{2}\right) \\ &= \left(\int dT' p(T', T) \cos^2\left(\frac{\omega T'}{2}\right)\right)^n \\ &= \left(\frac{1}{2}\right)^n \left(1 + \exp\left(-\frac{\pi\omega^2 T^2}{4}\right)\right)^n = \left(\frac{1}{2}\right)^n \left(1 + \exp\left(-\pi^3\zeta^2\right)\right)^n, \end{aligned} \tag{3.29}$$

where n is the number of iterations, $p(T_j, T)$ is the probability distribution for the iteration times (specifically, the positive half of a normal distribution with mean T such that $T = \sqrt{\frac{2}{\pi}} T_{\mathrm{RMS}}$), and in the final line we have introduced $\zeta \equiv \frac{\omega T}{2\pi}$.

It is clear that as $\zeta \to 0$, $s^{\mathrm{Avg}} \to 1$. This reflects the design of the algorithm, which ensures that the ground state (with $\omega = 0$) remains unsuppressed. Conversely, for large ζ, the suppression factor asymptotes to $1/2^n$. Notably, this asymptotic behavior is reached rapidly since $\exp(-\pi^3\zeta^2)$ becomes negligible for moderate values of ζ.

It is important to note that the total average runtime, T_{tot}, is simply the number of iterations times the average time per iteration: $T_{\mathrm{tot}} = nT$. Rewriting Eq. (3.29) in terms of T_{tot}, or equivalently $\zeta_{\mathrm{tot}} = \omega T_{\mathrm{tot}}/2\pi$, yields:

$$s^{\mathrm{Avg}}(\zeta_{\mathrm{total}}, n) = \left(\frac{1}{2}\right)^n \left(1 + \exp\left(\frac{-\pi^3\zeta_{\mathrm{tot}}^2}{n^2}\right)\right)^n. \tag{3.30}$$

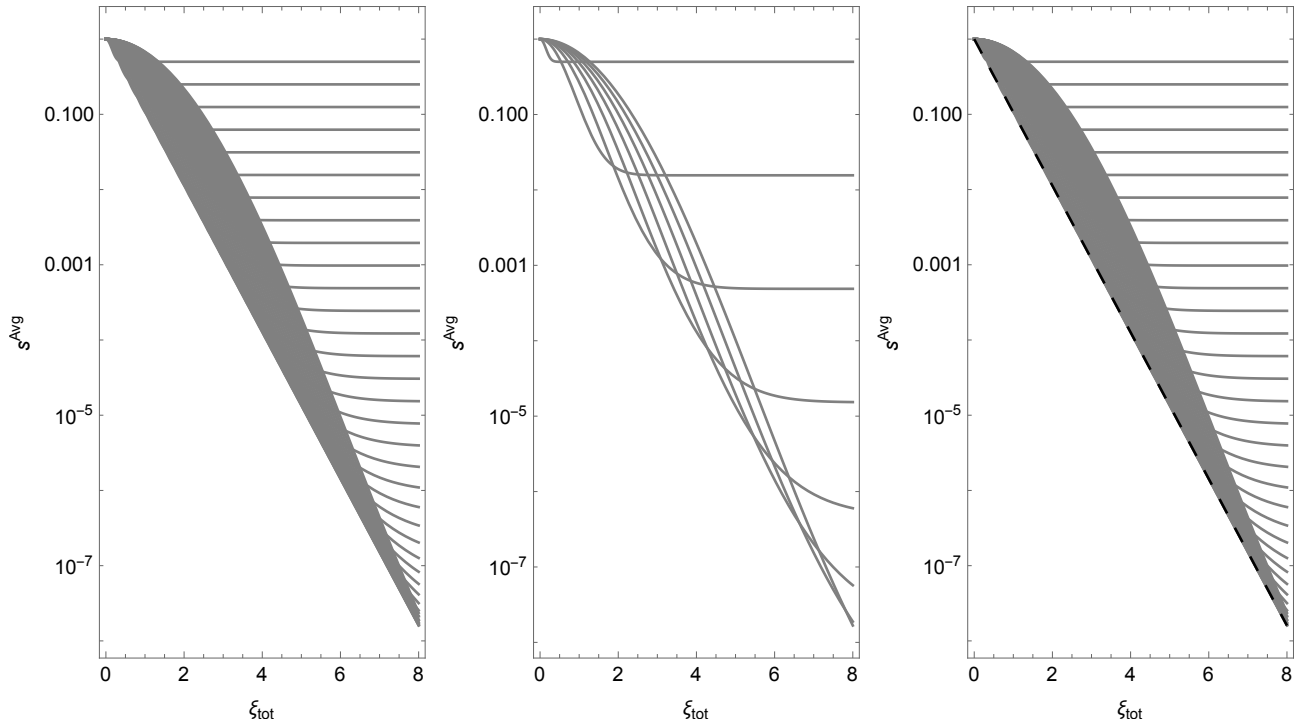

Figure 3.2: The average suppression factor is plotted as a function of ζ_{tot} for various values of n. The left panel displays all values of n from 1 to 40. As the curves become densely packed near the separatrix, making the features difficult to distinguish, the middle panel highlights every fifth value of n for clarity. The right panel includes a dashed line representing the exponential fit given in Eq. (3.34).

This form highlights the trade-off between the number of iterations and the total runtime required to achieve a given level of average suppression.

In Figure 3.2, s^{Avg} is plotted as a function of ζ_{tot} for various values of n. The leftmost panel shows curves for all n from 1 to 40. It is evident that the ζ_{tot}–s^{Avg} plane separates into two distinct regions. The lower left region corresponds to situations where no choice of n yields a given value of ζ_{tot} with the corresponding level of average suppression s^{Avg}; in this region, achieving such suppression with RRA is impossible. Conversely, in the upper right region, there exist values of n for which RRA achieves the desired suppression level at the specified ζ_{tot}. The boundary between these regions is marked by a well-defined separatrix.

The dense clustering of curves in the left panel makes it difficult to discern the detailed structure

of the separatrix. To better illustrate its behavior, the middle panel of Figure 3.2 shows only every fifth value of n. For $\zeta_{\rm tot} \gtrsim 1$, the separatrix is strikingly well-approximated by a simple exponential—manifesting as a straight line in the log-plot.

The optimal exponential fit is the one that intersects the exact separatrix at every n. This can be achieved by considering the functional form of Eq. (3.30) while allowing n to be a continuous real variable instead of restricting it to discrete integers. By fixing ζ_{tot} and minimizing $s^{\rm Avg}$, $\partial s^{\rm Avg}/\partial n = 0$ gives

$$n = \alpha \zeta_{\rm tot}, \tag{3.31}$$

where α satisfies

$$\frac{2\pi^3}{\alpha^2\left(1+\exp\left(\frac{\pi^3}{\alpha^2}\right)\right)} = \log\left(\frac{2}{1+\exp\left(-\frac{\pi^3}{\alpha^2}\right)}\right). \tag{3.32}$$

Numerically solving for α yields

$$\alpha \approx 4.271. \tag{3.33}$$

Substituting this result, the minimal suppression factor $S^{\rm Avg}$ for a given $\zeta_{\rm tot}$ takes the exponential form:

$$S^{\rm Avg}(\zeta_{\rm tot}) = \exp\left(-\beta\,\zeta_{\rm tot}\right) \text{ with } \beta = -\alpha \log\left(\frac{1+\exp\left(-\frac{\pi^3}{\alpha^2}\right)}{2}\right) \approx 2.244. \tag{3.34}$$

Eqs. (3.31)-(3.34) provide the minimum achievable value of $s^{\rm Avg}$ for any n. However, since the RRA is implemented with integer values of n, Eq. (3.34) serves as a lower bound:

$$s^{\rm Avg}(\zeta_{\rm tot}, n) \geq S^{\rm Avg}(\zeta_{\rm tot}) = \exp\left(-\beta\,\zeta_{\rm tot}\right) \text{ with } \beta \approx 2.244. \tag{3.35}$$

In cases where $\zeta_{\rm tot}$ yields integer n values from Eq. (3.31), the bound is saturated, and Eq. (3.35)

becomes exact. Moreover, since the curves for successive n values near the separatrix are densely packed for $\zeta_{\text{tot}} \gtrsim 1$, the continuous approximation remains highly accurate. This excellent agreement is clearly visible in the right panel of Figure 3.2.

Using this result, we return to the optimization problem of determining the required total time to ensure that all components with energy greater than or equal to Δ (the minimum excitation energy or gap) are suppressed on average by at least a predetermined factor S. Given that $\zeta_{\text{tot}} = \frac{ET_{\text{tot}}}{2\pi\hbar}$, we can rewrite Eq. (3.35) as

$$T_{\text{tot}} \geq -\frac{2\pi\hbar \log(S)}{\beta E}. \tag{3.36}$$

To guarantee suppression for all components, we substitute $E = \Delta$, yielding

$$T_{\text{tot}} \geq -\frac{2\pi\hbar \log(S)}{\beta \Delta}. \tag{3.37}$$

By using the natural timescale T_0, we obtain

$$\frac{T_{\text{tot}}}{T_0} \geq -\frac{\log(S)}{\beta} \approx 0.4456 \log(S). \tag{3.38}$$

3.2.3 Fluctuations in the random rodeo algorithm

It may seem reasonable to assume that maintaining the ensemble-averaged suppression factor below a fixed threshold S across all excited-state energies would guarantee effective suppression. In practice, however, this is not the case—the suppression factor exhibits exponentially large fluctuations, making it both challenging and computationally expensive to consistently achieve the desired level of suppression.

A clear indication of exceptionally large fluctuations can be found in the existence of multiple, qualitatively distinct "natural" expectations for the suppression after n iterations.

One intuitive expectation follows from the fact that the average of $\cos^2(\zeta)$ is $\frac{1}{2}$. Away from ζ=0, where no suppression occurs, one might assume that each iteration effectively samples $\cos^2(\zeta)$ randomly, leading to an average suppression of $\frac{1}{2}$ per iteration. Alternatively, one could argue that for large ζ, the phases in the cosine function become uniformly distributed due to the random time assignments, and as noted in Ref. [2], the geometric mean of $\cos^2(\zeta)$ suggests a suppression of $\frac{1}{4}$ per iteration.

Which of these expectations is correct? The surprising answer is that strong evidence supports both. Reference [2] provides numerical evidence for suppression scaling as 4^{-n}, showing that for $\zeta = 3.0$, the logarithm of suppression closely follows $-N \log 4$ (with significant fluctuations) and nearly does so for $\zeta = 2$. This suggests that, at least for sufficiently large ζ, typical suppression per iteration is close to $\frac{1}{4}$.

At the same time, Eq. (3.30) reveals that the ensemble-averaged suppression approaches 2^{-n} rather than 4^{-n} as ζ increases. This apparent contradiction can be resolved by recognizing that the average suppression is heavily influenced by rare, atypical cases where suppression factors are much larger than the typical ones. In other words, while most realizations of the algorithm exhibit suppression around 4^{-n}, the ensemble average is skewed by occasional runs with significantly less suppression, indicating the presence of exceptionally large fluctuations.

This can be easily verified. Table 3.1 presents the geometric mean and root-mean-square (RMS) values of the suppression factor for the ensemble used in RRA, alongside the previously computed arithmetic mean. These expressions are obtained using a similar approach to that of the arithmetic mean.

Statistical quantity	$s(\zeta_{\text{tot}}, n)$	$\zeta_{\text{tot}} \to \infty$	Exponential fit
Geometric mean	$\exp\left(n \int_{-\infty}^{\infty} dT \log\left(\cos^2(\pi T \zeta_{\text{tot}})\right) p(T,1)\right)$	4^{-n}	$\beta \approx 4.46$
Arithmetic mean	$\left(\frac{1}{2}\right)^n \left(1 + \exp\left(-\pi^3 \zeta_{\text{tot}}^2\right)\right)^n$	2^{-n}	$\beta \approx 2.244$
Root-mean-square	$\left(\frac{3}{8}\right)^{n/2} \left(1 + \frac{\exp(-4\pi^3\zeta_{\text{tot}}^2)}{3} + \frac{4\exp(-\pi^3\zeta_{\text{tot}}^2)}{3}\right)^{n/2}$	$\left(\frac{3}{8}\right)^{n/2}$	$\beta \approx 1.637$

Table 3.1: Three statistical quantities characterizing the suppression factor as a function of ζ_{tot} are shown, each defined with respect to the ensemble used in the random rodeo algorithm. The second column provides expressions for these quantities as functions of ζ_{tot}. The third column lists the asymptotic values these expressions approach for large ζ_{tot}; in all cases, the asymptotic regime is reached to excellent approximation when $\zeta_{\text{tot}} > 1$. The fourth column displays the best-fit exponential curves analogous to the separatrix in the $\zeta_{\text{tot}} - s$ plane. These were obtained following the same method used for Eqs. (3.31)–(3.35). As in that case, the exponential separatrix provides a lower bound on the time required to ensure that the ensemble-averaged suppression factor reaches (or surpasses) a target threshold for all energy modes.

Using the expressions for the root-mean-square (RMS) and arithmetic mean, one can readily compute the signal-to-noise ratio:

$$\begin{aligned} \frac{s^{\text{Avg}}}{\sigma} &= \left(\left(\frac{3}{2}\right)^n \left(\frac{\left(1 + \frac{\exp(-4\pi^3\zeta^2)}{3} + \frac{4\exp(-\pi^3\zeta^2)}{3}\right)}{\left(1 + \exp\left(-\pi^3\zeta^2\right)\right)^2} \right)^n - 1 \right)^{-\frac{1}{2}} \\ &\xrightarrow{\text{large } n} \left(\frac{3}{2}\right)^{-\frac{n}{2}} \left(\frac{\left(1 + \frac{\exp(-4\pi^3\zeta^2)}{3} + \frac{4\exp(-\pi^3\zeta^2)}{3}\right)}{\left(1 + \exp\left(-\pi^3\zeta^2\right)\right)^2} \right)^{-\frac{n}{2}} \xrightarrow{\text{large } \zeta} \left(\frac{3}{2}\right)^{-\frac{n}{2}}. \end{aligned} \tag{3.39}$$

At large n, the fluctuations in the distribution of s values grow exponentially with n, indicating increasingly large deviations from the mean.

The three statistical quantities, despite being sampled from the same distribution, exhibit significant differences. If the suppression factors were governed by a single characteristic scale, these quantities would align. However, as demonstrated in the third column, they not only differ but do so in an exponentially large manner as a function of n.

Previously, it was proposed that "typical" suppression factors scale approximately as 4^{-n} for sufficiently large ζ, aligning with the findings of Ref. [2]. This behavior can be understood intuitively: the suppression factor is a product of many independently chosen random variables, each sampled from the same distribution. In such cases, the central limit theorem suggests that, for large n, a log-normal distribution emerges when considering the logarithm of the suppression factors. Since the logarithms of these variables combine additively, their distribution tends toward a normal form.

An important consequence of this is that the geometric mean of a log-normal distribution corresponds to its median. Thus, at large n, the median suppression factor—a reasonable measure of a "typical" value—should be given by the geometric mean. In this case, a straightforward calculation confirms that the geometric mean is indeed 4^{-n}, explaining the observed scaling behavior.

However, it is worth noting a limitation: while the log-normal distribution provides an excellent approximation in most cases, its validity can be constrained by the bounded nature of the suppression factor, which is restricted to values $S \leq 1$. This affects quantities that receive significant contributions from the tail of the distribution extending beyond unity, such as the root-mean-square, which may not be fully captured by the log-normal approximation.

Up to this point, our focus has been on statistical fluctuations at fixed values of ζ_{tot} across the ensemble of RRA runs. However, it is natural to expect that the large fluctuations observed in the ensemble would also manifest as significant variations in ζ within a single run.

Figure 3.3 illustrates a representative suppression factor profile, providing numerical evidence supporting this expectation. It is after six iterations, with times drawn from a half-normal distribution with a mean of unity, plotted over the range $1 < \zeta < 10$. The choice of $n = 6$ ensures that the ensemble average suppression factor, $1/64$, remains visually distinguishable from zero. The presence of extreme fluctuations relative to the average is evident. The highest peak reaches 13 times the average value,

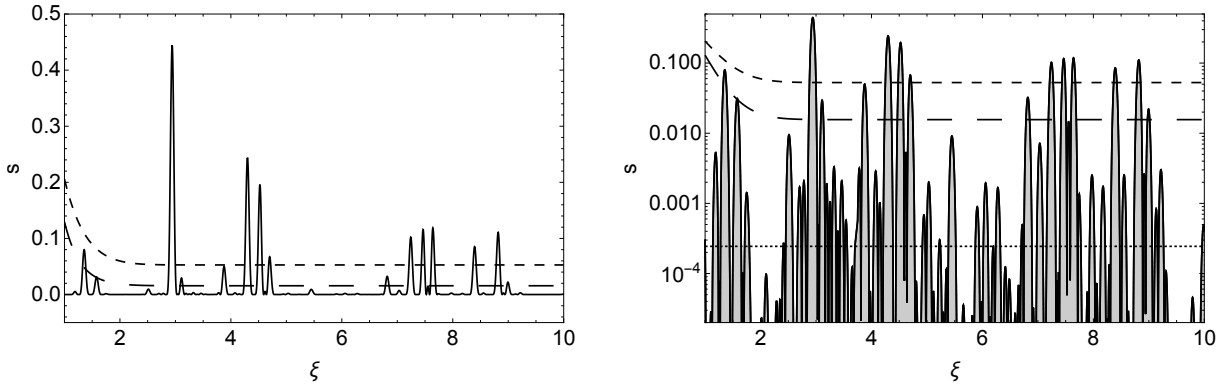

Figure 3.3: A single instance of the random rodeo algorithm with six iterations. The solid curve shows the suppression factor for one representative run, where each iteration time is drawn from the positive half of a normal distribution with a mean of one. The long-dashed line depicts the ensemble average suppression over many such runs using the same distribution. The short-dashed curve represents the root-mean-square (RMS) suppression, while the dotted line (shown in the log-scale plot) corresponds to the geometric mean.

while typical values between peaks are so small that they are only visible on a logarithmic scale.

The minimum suppression value observed in the log plot highlights the extreme suppression levels. Within the range $2 < \zeta < 10$, where the ensemble average has stabilized at approximately 2^{-6}, nearly half of the sampled ζ values exhibit suppression factors below 4^{-6}. A numerical integration of $\Theta(4^{-6} - s[\zeta])$ (where Θ is the Heaviside step function) estimates that roughly 58% of values in this interval satisfy $s(\zeta) < 4^{-6}$, reinforcing the notion that a "typical" suppression factor is of order 4^{-6}.

Additionally, the mean suppression factor in this range aligns closely with 2^{-n}. Specifically, for this run, $s(\zeta)$ averages to approximately 0.014, which is about 90% of the expected ensemble average value $2^{-6} \approx 0.0156$. Given the limited number of prominent peaks and the modest number of iterations ($n = 6$), this agreement is striking.

These large fluctuations pose a fundamental challenge for the efficient application of RRA. To ensure a high probability that all relevant components of the initial state reach a required suppression level, significantly more computational resources are needed than what would be expected based on

a "typical" suppression factor. In practical settings—especially for early quantum computers capable of implementing the rodeo algorithm—such fluctuations introduce substantial uncertainty, making it difficult to reliably estimate the degree to which the initial state has been projected onto the ground state.

3.2.4 Super iterations and optimizing the rodeo projections

There are two main motivations for exploring non-random (i.e., deterministic or structured) variants of the rodeo projection algorithm rather than relying on the random rodeo algorithm. The first motivation stems from the previous discussion: the random approach generates exponentially large fluctuations in suppression as the number of iterations n increases, making it inefficient for situations where reliable and predictable suppression is required. Ideally, one would want to deliberately select parameters that reduce or eliminate these fluctuations.

The second motivation is more general. For any well-defined quantity associated with suppression, there exists an optimal combination—specific values of n and ζ (or equivalently the times T_i used in each iteration)—that either maximizes or minimizes that quantity, depending on the goal. At best, a random approach can only approximate this optimal choice. In contrast, a carefully designed deterministic algorithm has the potential to perform strictly better, since it can directly target the optimal suppression characteristics.

A deeper understanding of the fluctuations suggests that they often originate from iterations where the suppression factor is close to unity. This happens when $\pi\zeta_c^j$ is an integer. On the other hand, suppression factors vanish when $\pi\zeta_c^j$ is a half-integer. This observation suggests a strategy: large fluctuations can be mitigated by ensuring that iterations with near-unity suppression are paired

with ones that force suppression to zero, so their product remains small. Remarkably, this can be done using a finite amount of time—specifically, just twice the time that would be used in a single iteration.

The key to this idea is a construction called a *super iteration*, which consists of an infinite sequence of standard iterations, each with half the time (or half of the ζ_c) of the one before it. Because this is a geometric series, the total time of a super iteration converges to exactly twice that of the initial iteration.

As shown in Eq. (3.15), the suppression factor from each sub-iteration depends on the cosine squared of a phase proportional to ζ. The second sub-iteration in a super iteration has a suppression factor of zero whenever ζ_c^j is twice a half-integer (i.e., an odd integer), which matches exactly where the first sub-iteration has a suppression factor of one. Thus, it cancels out these problematic cases. The third sub-iteration cancels out the next set (where ζ_c^j is four times a half-integer), and so on. In this way, the full super iteration cancels out all the points where the initial iteration alone would have had a suppression factor near unity—eliminating the peaks responsible for large fluctuations.

The total suppression factor from a super iteration at a given $\zeta_{c\,\mathrm{sup}}^j = 2\zeta_c^j$ is given by

$$s_{c\,\mathrm{sup}}^j = \prod_{k=0}^{\infty} \cos^2\left(\frac{\pi\zeta_c^j}{2^k}\right) = \prod_{k=1}^{\infty} \cos^2\left(\frac{\pi\zeta_{c\,\mathrm{sup}}^j}{2^k}\right) = j_0^2\left(\pi\zeta_{c\,\mathrm{sup}}^j\right) = \left(\frac{\sin(\pi\zeta_{c\,\mathrm{sup}})}{\pi\zeta_{c\,\mathrm{sup}}}\right)^2, \tag{3.40}$$

where $j_0(x)$ is the zeroth-order spherical Bessel function.

From Eq. (3.14), we have $\zeta = (E/\Delta)(T/T_0)$. If one allocates a total time of $T_0 = 2\pi\hbar/\Delta$—corresponding to one full period of phase evolution for the lowest excited state—and performs a single super iteration, the resulting suppression factor as a function of energy is given by $j_0^2(\pi E/\Delta)$, as shown in Figure 3.4. Since Δ is the minimum excitation energy, all excited-state components in the initial state satisfy $E \geq \Delta$, and Figure 3.4 plots s for $E/\Delta \geq 1$. The choice of $T = T_0$ ensures that

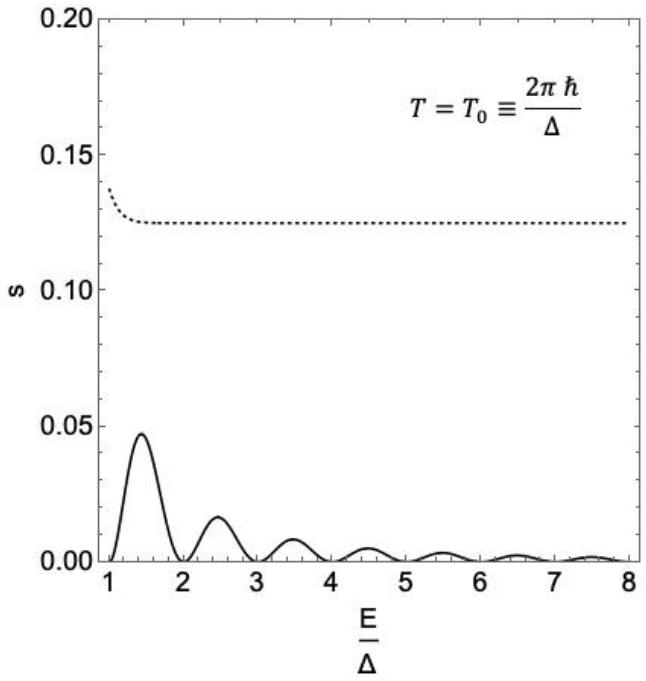

Figure 3.4: Comparison between a single super iteration and the ensemble-averaged suppression factor from the random rodeo algorithm. The solid curve shows the suppression factor resulting from a single super iteration as a function of excitation energy, expressed in units of Δ, the minimum excitation energy. The total time was fixed at $T_0 = 2\pi\hbar/\Delta$. For comparison, the dotted curve shows the ensemble-averaged suppression factor from the random rodeo algorithm, using three iterations with the same average total time—chosen to minimize the peak value of the average suppression.

the lowest excited state is completely suppressed, with $s = 0$ at $E = \Delta$. Two additional features stand out: the suppression factor remains below 0.05 for all $E > \Delta$, and it decreases rapidly with increasing E/Δ (or equivalently with ζ), indicating strong and efficient filtering of higher excitations within a modest runtime.

This level of suppression is particularly impressive given the modest total runtime—just a single period of the lightest excitation. For comparison, if the total runtime is similarly constrained to one excitation period, three RRA iterations are optimal for maximizing average suppression when $E > \Delta$, as shown in Figure 3.4. The ensemble average suppression from the RRA, represented by the dotted line, is consistently outperformed by the single super iteration across the entire range $E/\Delta > 1$, with suppression remaining at least a factor of 2.6 smaller than the RRA average and becoming orders of

magnitude more effective at higher values of E/Δ.

This rapid decay in suppression with increasing ζ highlights a significant advantage of super iterations. In designing optimized variants of the rodeo algorithm, one can focus on improving the suppression factor for ζ in the range $1 < \zeta < 2$, knowing that excellent suppression for $\zeta > 1$ will naturally follow. This feature, which is not shared by the random algorithm, underscores the advantage of selecting algorithm parameters deliberately rather than relying on randomness to achieve optimal suppression.

As discussed in Section 3.2.2, optimizing the rodeo projection algorithm requires a clear definition of the quantity to be optimized. Even if one limits the computational resources (or more precisely, the total time in units of $T_0 = 2\pi\hbar/\Delta$) and the goal is to optimize the suppression factor, the problem remains incomplete since the suppression factor depends on the excited energies.

If one has substantial knowledge of the spectral density of the initial state, the optimization can be tailored to focus on components with higher probabilities. However, merely knowing the spectrum of the theory, without further details about the initial state, can still impact optimization. For example, if the spectrum contains no states with energies between Δ and 2Δ, it would be inefficient to allocate computational resources to suppressing components in that energy range, as these states do not appear in the initial state.

A simple approach to optimization involves two steps: first, selecting a quantity $Q(s, T_{\text{tot}})$ depends on the suppression factor $s(\zeta_{\text{tot}})$ and the total time, and is designed to minimize and achieve strong suppression for the given state. Second, one seeks the suppression factor s produced by the rodeo projection under the constraint of total time, with the goal of minimizing Q.

It is important to note that, regardless of the choice for Q, there exists an optimal set of parameters that minimizes Q once T_{tot} is determined. However, finding this optimal choice might not always

be feasible. Instead, one can explore a range of possible parameters and select the one that minimizes Q most effectively.

We begin by considering a scenario where little is known about the initial state except that all unwanted components have energies greater than or equal to Δ. In this case, we adopt a conservative strategy that ensures a rigorous bound on the total suppression of excited states. The approach is straightforward: we define Q as the largest suppression value within the physically relevant range $E \geq \Delta$:

$$Q = \max_{E \geq \Delta} s\left(\frac{E}{\Delta}\frac{T_{\text{tot}}}{T_0}\right). \tag{3.41}$$

This choice of Q effectively represents the worst-case suppression in the possible physical range. By selecting Q in this way, we ensure that the suppression for all components of the state does not exceed this value, thus providing an upper bound for S_E, the total suppression of excited-state components:

$$S_E \equiv \frac{\int_{\Delta-\epsilon}^{\infty} dE\, s\left(\frac{E}{\Delta}\frac{T_{\text{tot}}}{T_0}\right)\rho(E)}{\int_{\Delta-\epsilon}^{\infty} dE\, \rho(E)} \leq Q. \tag{3.42}$$

This strategy does not require any specific knowledge about the initial state or the spectrum beyond the fact that the lowest excitation energy is Δ. Furthermore, Q depends solely on the suppression factor $s(\zeta_{\text{tot}})$ and the total time, independent of the initial state. Therefore, the optimization of Q needs to be performed only once for each T_{tot}, avoiding the need for repeated optimization for different initial states.

However, this approach is quite conservative. While it provides an upper bound for S_E, this bound may be much larger than the actual least upper bound. For example, if s is constructed from super iterations with energies several times Δ, the suppression factor becomes extremely small at

$\max_{E\geq\Delta} s\left(\frac{E}{\Delta}\frac{T_{\rm tot}}{T_0}\right)$	$T_{\rm tot}/T_0$	n	T_1/T_0	T_2/T_0	T_3/T_0	T_4/T_0	T_5/T_0	T_6/T_0	T_7/T_0	T_8/T_0
4.719×10^{-2}	.8129	1	.8129							
8.508×10^{-4}	1.5906	2	.9361	.6545						
2.421×10^{-5}	2.4222	3	.9494	.6638	.8090					
7.549×10^{-7}	3.0752	4	.9785	.6841	.8338	.5788				
7.385×10^{-9}	3.9865	5	.9764	.6827	.8320	.5776	.9180			
8.948×10^{-11}	4.7944	6	.9881	.6908	.8419	.5845	.9290	.7601		
5.689×10^{-12}	5.4500	7	.9925	.6939	.8457	.5871	.9331	.7634	.6343	
1.539×10^{-14}	6.4010	8	.9895	.6918	.8431	.5853	.9303	.7611	.6324	.9675

Table 3.2: Times for each super iteration and the corresponding maximum suppression factor for $E > \Delta$, as determined using the Whac-a-Mole scheme.

chosen energies, even with just a few super iterations. If the initial state is heavily weighted toward such energies, the actual suppression of excited components S_E will be much smaller than the bound given by Q. In this case, fixing the bound for S_E at the level needed for the problem may result in unnecessary computational resources being spent. After considering this conservative bound, we will explore ways to refine and improve it.

Even with this conservative approach, the mathematical problem of finding s that minimizes $\max_{E\geq\Delta} s\left(\frac{E}{\Delta}\frac{T_{\rm tot}}{T_0}\right)$, subject to the constraint that s is obtained through a fixed number of rodeo iterations with total time, remains unresolved. However, one can resort to ad hoc strategies and test their performance. A natural choice is to consider prescriptions based on super iterations, given their apparent efficiency.

Table 3.2 illustrates a simple ad hoc prescription, primarily intended to show that there are schemes that significantly outperform the average results of RRA without suffering from the exponentially large fluctuations. However, it is important to note, though, that this strategy does not guarantee that the prescription is near the optimal solution.

We call this method the Whac-a-Mole prescription, named after the arcade game where "moles" appear randomly and must be whacked with a mallet before new ones pop up in different locations. Similarly, in this strategy, the "moles" represent the energies to be suppressed.

The prescription follows the structure of super iterations. Initially, one selects a super iteration with a time equal to T_0, yielding a suppression factor with zeros at all integer multiples of E/Δ. The focus is on energies grater than Δ. The largest suppression factor for $E > \Delta$ is identified, and the corresponding energy is treated as the "mole" to be whacked. The time for the next super iteration is chosen to place a zero at this energy maximum, ensuring that the smallest zero remains at $E = \Delta$. A third iteration follows the same process: the maximum suppression factor for $E > \Delta$ is found, and the time for the next super iteration is chosen to place a zero at that energy. This cycle can be repeated for as many iterations as desired. Each cycle results in progressively narrower suppression peaks at increasingly lower scales, with maxima appearing between $1 \leq E/\Delta \leq 2$.

The scheme thus far ensures that $E = \Delta$ remains the smallest zero. However, this constraint can be relaxed to optimize time by using a scaling factor less than one. This scaling does not affect the maximum suppression factor but shifts the suppression function so that energies previously below Δ move into the relevant physical range $E \geq \Delta$. Since the suppression factor at $E = \Delta$ was initially zero, applying a scaling factor slightly smaller than one pushes small suppression values into the physical region. The scaling is chosen to ensure that the suppression at $E = \Delta$ equals to the maximum suppression factor, as a smaller value would increase the maximum. This scaling leads to a modest time reduction after the first few cycles, but it can provide up to a 20% reduction when only a single super iteration is used.

It is useful to compare the results obtained from the Whac-a-Mole scheme with those from RRA. In Figure 3.5, we compare the *upper* bound of the suppression factor with the *lower* bound for the

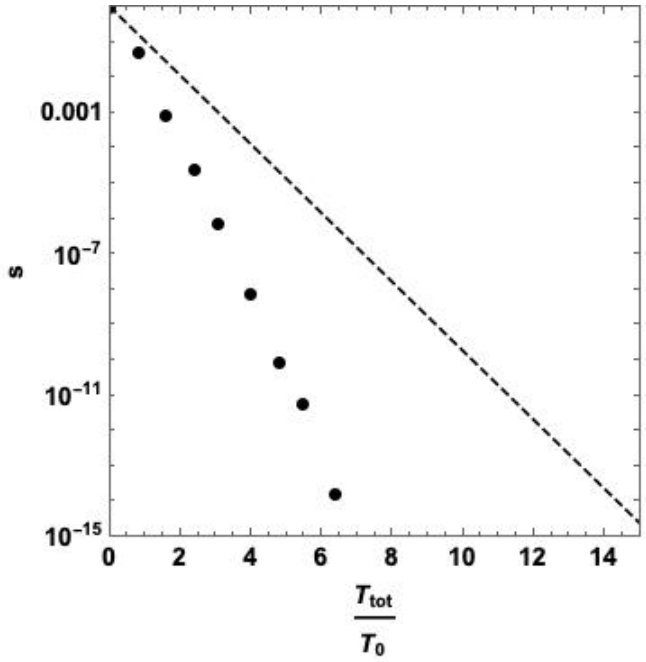

Figure 3.5: Comparison between super iterations using the Whac-a-Mole strategy and the random rodeo algorithm. The solid points represent an absolute upper bound on the suppression factor achieved with intentionally chosen times following the Whac-a-Mole prescription, plotted against total runtime for various numbers of super iterations. For reference, the dashed line shows a lower bound on the runtime required for the ensemble average suppression factor S to exceed a fixed threshold using the random rodeo algorithm from Ref. [2].

time required to achieve a fixed ensemble average of S. The intentional approach clearly outperforms the random approach by several orders of magnitude for fixed total times, especially when T is a few multiples of T_0 or higher. For example, at $T/T_0 = 6.4$, the intentional method produces an upper bound for the total suppression that is a factor of 3.77×10^7 smaller than the lower bound of the random method. If calculations are constrained to shorter times—likely the case when quantum computers are first large enough to implement the rodeo algorithm with sufficient coherence—the significant performance improvement at fixed times becomes particularly noticeable.

A more moderate view of the improvement can be seen by considering the total time needed to reach a certain suppression level, where the gain appears to be about a factor of two. Regardless of how one looks at it, the comparison emphasizes the significant advantages of using the intentional

method. The random approach, with its exponentially large fluctuations, complicates achieving consistent suppression. In contrast, by selecting times intentionally, this problem is completely avoided, providing a reliable upper bound for suppression.

The Whac-a-Mole method is entirely ad hoc, created as an example of a scheme that clearly outperforms the random approach. It is reasonable to think that a more thorough search of the parameter space could lead to even better results, which could be explored in future work.

For practical implementation, a helpful step would be to create a table similar to Table 3.2, but with a denser set of total times and ideally, improved suppression factors. This would make it straightforward to either select the suppression level needed for a specific problem and find the corresponding time or set a maximum allowable time and determine the achievable suppression. In either case, one could then use the times from the relevant row of the table, whether from super iterations or regular iterations, to efficiently implement the projection. Practical considerations for implementing rodeo projections are discussed in Appendix D.

3.3 Hamiltonian path

A general strategy for state preparation on quantum simulators is to begin with a system whose vacuum (ground) state is known and can be prepared explicitly, and then evolve the system by varying a parameter s that defines the Hamiltonian $H(s)$. We will call this the Hamiltonian path from the initial Hamiltonian to the final Hamiltonian. If this evolution is performed carefully, the system remains approximately in its instantaneous ground state $|g(s)\rangle$ throughout the process. This principle underlies several classes of algorithms, including adiabatic state preparation [170, 191, 192, 195–200] and approaches based on the quantum Zeno effect (QZE), which utilize projective measurements to

inhibit transitions out of the ground state [201–205]. The parameter s can always be chosen to be dimensionless without loss of generality.

3.3.1 Path length

Consider a family of Hamiltonians $H(s)$ parameterized by a continuous variable s, such that $H(s_i)$ and $H(s_f)$ correspond to the initial and final Hamiltonians, respectively. We assume that $H(s)$ has a discrete and non-degenerate spectrum for all $s \in [s_i, s_f]$,[1] and that $H(s)$ is infinitely differentiable with respect to s. The instantaneous ground state of $H(s)$ is denoted $|g(s)\rangle$.

The length of the path between two points a and b along this family, where $s_a \leq s_b$, is defined as [201–203]

$$L_{a,b} \equiv \int_{s_a}^{s_b} ds \, \| |g'(s)\rangle - |g(s)\rangle\langle g(s)|g'(s)\rangle \|, \text{ with } |g'(s)\rangle \equiv \frac{d}{ds}|g(s)\rangle. \tag{3.43}$$

This definition of path length provides an intuitive measure of the difficulty of preparing the ground state along the path from $H(s_i)$ to $H(s_f)$: it quantifies the rate of change of the ground state in the direction orthogonal to itself.

Importantly, this definition is invariant under reparameterization: any smooth, invertible change of variables in s leaves $L_{a,b}$ unchanged for all a and b.

Since the ground state $|g(s)\rangle$ is defined only up to an overall phase, one may exploit this freedom and fix the phase such that $\langle g(s)|g'(s)\rangle = 0$ along the entire path. In this gauge, the path length simplifies to

$$L_{a,b} = \int_{s_a}^{s_b} ds \, \| |g'(s)\rangle \|. \tag{3.44}$$

[1] These results extend to weaker conditions, but the assumptions stated here simplify the analysis.

While this phase choice is not necessary for practical calculations, it simplifies formal analysis and will be adopted throughout the remainder of this thesis. One implication of this choice is that if the path passes through the same Hamiltonian more than once (i.e., $H(s_a) = H(s_b)$ for $s_a \neq s_b$), the corresponding ground states $|g(s_a)\rangle$ and $|g(s_b)\rangle$ may differ by a Berry phase [206].

For formal clarity, we similarly fix the phases of excited states such that $\langle e_i(s)|e_i'(s)\rangle = 0$, where $|e_i(s)\rangle$ denotes an arbitrary excited state. Furthermore, it is convenient to reparameterize the trajectory in terms of parameter $\lambda(s)$ defined by

$$\lambda(s) \equiv \lambda_i + \int_{s_i}^{s} d\tilde{s} \, \| |g'(\tilde{s})\rangle \| \text{ with } L_{a,b} = \lambda_b - \lambda_a. \tag{3.45}$$

This reparameterization tracks the distance traveled along the path and allows one to describe motion in parameter space as a function $\lambda(t)$. The total path length is given by $L \equiv \lambda_f - \lambda_i$.

3.3.2 Scaling of path length in volume for local field theories

We will demonstrate that the path length scales with the square root of the volume of the system:

$$\frac{L_{V_2}}{L_{V_1}} = \sqrt{\frac{V_2}{V_1}} \, . \tag{3.46}$$

To derive Eq. (3.46), we rely on the assumption of local interactions and finite correlation lengths. At large volumes, these properties imply:

$$\langle g'(\lambda_0)|g'(\lambda_0)\rangle = - \left. \frac{\partial^2 \langle g(\lambda_1)|g(\lambda_0)\rangle}{\partial(\lambda_1)^2} \right|_{\lambda_1=\lambda_0} = \left(\frac{V}{\xi^d}\right) h(\lambda_0) \times \left(1 + O\left(\frac{\xi^d}{V}\right)\right), \tag{3.47}$$

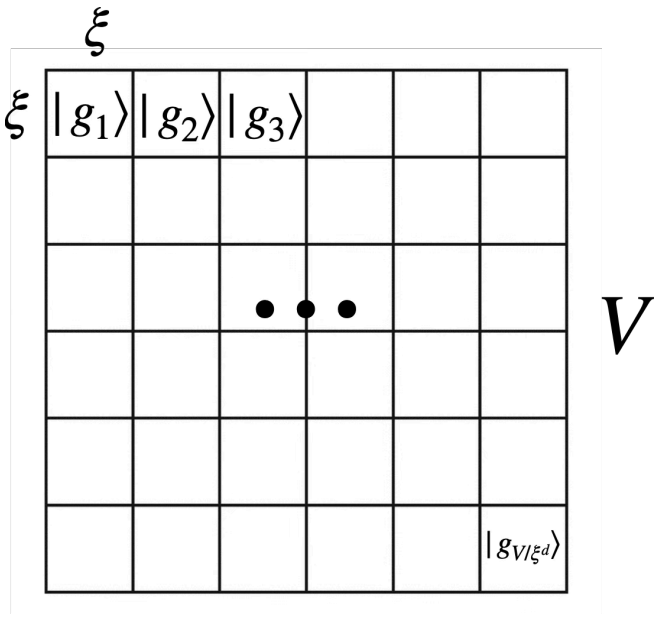

Figure 3.6: Schematic illustration of a ground state governed by local interactions, highlighting finite correlations.

where $h(\lambda_0)$ is a volume-independent function at large V, ξ is the correlation length, and d is the number of spatial dimensions. Intuitively, since the correlation length ξ is finite, the full volume can be thought of as being divided into many approximately independent subvolumes of size ξ^d, which is shown in Figure 3.6. More generally, the nth derivative of the overlap satisfies:

$$\left.\frac{\partial^n \langle g(\lambda_0+\Delta\lambda/2)|g(\lambda_0-\Delta\lambda/2)\rangle}{\partial(\Delta\lambda)^n}\right|_{\Delta\lambda=0} = -\frac{1+(-1)^n}{2}\frac{\langle g'(\lambda_0)|g'(\lambda_0)\rangle^{\frac{n}{2}} n!}{2^{n/2}\,(n/2)!}\left(1+O\left(\frac{\xi^d}{V}\right)\right), \tag{3.48}$$

which leads to the approximate expressions:

$$\begin{aligned}\langle g(\lambda_1)|g(\lambda_0)\rangle &= e^{-L^2_{\lambda_0,\lambda_1}/2}\left(1+O\left(\frac{\xi^d}{V}\right)\right),\\ \||g'(\lambda_0+\Delta\lambda)\rangle\| &= \||g'(\lambda_0)\rangle\| \times \left(1+O\left(\frac{\xi^d}{V}\right)^{\frac{1}{2}}\right).\end{aligned} \tag{3.49}$$

These results collectively imply that the path length L scales as $L \sim \left(\frac{V}{\xi^d}\right)^{\frac{1}{2}}$ in the large-volume limit.

3.4 Zeno-inspired state preparation

There are two key issues in optimizing the traversal of the Hamiltonian path. An obvious one is the choice of path through parameter space, since clearly this affects how efficiently a ground state can be produced. The other is finding an efficient manner to traverse the path to find the ground state of interest. This thesis addresses the second issue.

There are important technical restrictions when using the approach outlined here: for the algorithm to be efficient the theory must remains local for all values of λ along the path and that no phase transitions (in the infinite volume theory) are encountered. We recognize that the restriction that no phase transitions are encountered along the path may reduce the applicability of this approach, but also note that this restriction applies generally to all methods based on traversing paths in parameter space including ASP and methods based on the QZE.

Even in the absence of phase transitions, there is a challenge faced by such methods for field theories with large volumes: the natural path through the various ground states is necessarily long—with path lengths scaling with the square root of the volume. It is important to note that while the size of the minimum spectral gap has received significant attention particularly in the context of adiabatic quantum computing [3,188,207,208], the total cost of traversing a path in parameter space is also quite sensitive to the path length. This should be clear if one imagines a Hamiltonian trajectory that has been rescaled so the spectral gap is constant along the trajectory. Clearly, longer paths are more expensive to traverse. The approach developed here is designed to be efficient for long paths and should be useful for a variety of problems with long paths but it is particularly useful for QFTs with large volumes.

In this section, we will review the state preparation method based on the quantum Zeno effect and suggest a method to travel the path more efficiently than the QZE-based methods.

3.4.1 Quantum Zeno effect

One approach to preparing eigenstates is through the quantum Zeno effect, which exploits the idea that frequent measurements can inhibit the evolution of a quantum state. If a system is continuously—or sufficiently frequently—measured in a way that projects onto the instantaneous eigenstate of the Hamiltonian, the system is effectively "frozen" in that eigenstate throughout the evolution. This can be used to maintain the system in the ground state as the Hamiltonian changes.

Why does this work? The key lies in the dynamics of quantum transitions: the instantaneous change of a state due to a small perturbation is linear in time, whereas the probability of making a transition to an orthogonal state scales quadratically with time. This suppression of transitions is the essence of the quantum Zeno effect.

A helpful classical analogy comes from optics. Consider a series of polarizers acting as projectors or measurements. If two polarizers are oriented orthogonally, no light passes through both. However, if we insert many intermediate polarizers between them, each rotated slightly with respect to the previous one, a significant portion of the light can make it through.

Mathematically, suppose that we place $n-1$ polarizers between two orthogonal ones such that the angle between the neighboring polarizers is $\frac{\pi}{2n}$. Then the transmitted intensity I is given by

$$I = \left(\cos^2\left(\frac{\pi}{2n}\right)\right)^n \approx \left(1 - \frac{1}{2}\left(\frac{\pi}{2n}\right)^2\right)^{2n} \approx \left(1 - \frac{\pi^2}{4n}\right) \xrightarrow{\text{large n}} 1. \tag{3.50}$$

Thus, as the number of polarizers increases, the transmitted intensity approaches that of the initial light—albeit with a rotated polarization. In the quantum setting, this analogy reflects how repeated projections (measurements) can preserve the quantum state. The transmitted intensity corresponds to

the squared overlap between the final ground state and the state evolved under frequent measurements.

In realistic scenarios, performing infinitely many measurements is not feasible. The number of measurements must be optimized. It can be shown that the cost of implementing this method scales quadratically with the path length L:

$$C \sim T \gtrsim \frac{L^2}{\Delta}, \tag{3.51}$$

where Δ is the spectral gap.

Before deriving the cost scaling, it is useful to recast the definition of the path length in terms of the overlap between nearby ground states:

$$L = \int_0^\lambda d\lambda \sqrt{-\frac{1}{2}\frac{d^2}{d\Lambda^2}|\langle g(\lambda+\Lambda)|g(\lambda)\rangle|^2\Big|_{\Lambda=0}}. \tag{3.52}$$

This expression is equivalent to the original definition in Eq. (3.44) and is motivated by the interpretation of path length as a measure of how rapidly the ground state changes to excited states along the path. For small $\delta\lambda$, the squared overlap between neighboring ground states can be expanded as

$$\begin{aligned} |\langle g(\lambda+\delta\lambda)|g(\lambda)\rangle|^2 &= 1 + \frac{1}{2}(\delta\lambda)^2 \frac{d}{d\Lambda}|\langle g(\lambda+\Lambda)|g(\lambda)\rangle|^2\Big|_{\Lambda=0} + \cdots \\ &\equiv 1 - \Omega(\lambda)(\delta\lambda)^2 + O\left((\delta\lambda)^3\right), \end{aligned} \tag{3.53}$$

where we have defined

$$\Omega(\lambda) \equiv -\frac{1}{2}\frac{d^2}{d\Lambda^2}|\langle g(\lambda+\Lambda)|g(\lambda)\rangle|^2\Big|_{\Lambda=0}. \tag{3.54}$$

Physically, $\Omega(\lambda)$ quantifies the probability that the evolved state deviates from the instantaneous ground state over an infinitesimal segment of the path. Using the orthogonality condition $\langle g'|g\rangle = 0$,

it is straightforward to show that

$$\langle g''|g\rangle + \langle g|g''\rangle = -2\langle g'|g'\rangle, \tag{3.55}$$

which further confirms that Eq. (3.52) matches the standard definition of path length.

In the Zeno-based approach, the total cost is proportional to the number of measurements required to successfully traverse the Hamiltonian path. Let $N(\lambda)$ denote the expected number of measurements to reach a point λ along the path (assuming $\lambda_i = 0$), and let $n(\lambda)$ be the local density of measurements per unit path length. Then, over a small increment $\delta\lambda$, the expected number of measurements becomes:

$$N(\lambda + \delta\lambda) = N(\lambda) + n(\lambda)\delta\lambda + \left(P_{\text{fail}} \cdot N(\lambda) + P_{\text{fail}}^2 \cdot 2N(\lambda) + \cdots\right), \tag{3.56}$$

where P_{fail} is the failure probability for a segment of length $\delta\lambda$, and the subsequent terms account for retries due to projection failures (each of which restarts from $\lambda = 0$).

The success rate over $\delta\lambda$ with $n(\lambda)\delta\lambda$ measurements is given by

$$P_{\text{success}} = \left(1 - \Omega(\lambda)\left(\frac{\delta\lambda}{n(\lambda)\delta\lambda}\right)^2\right)^{n(\lambda)\delta\lambda} \approx 1 - \Omega(\lambda)\frac{\delta\lambda}{n(\lambda)}, \tag{3.57}$$

mirroring the polarizer analogy (where $\delta\lambda/(n(\lambda)\delta\lambda)$ corresponds to the angular spacing $\pi/2n$). Thus,

$$P_{\text{fail}} = \Omega(\lambda)\frac{\delta\lambda}{n(\lambda)}. \tag{3.58}$$

Substituting this into the recurrence equation gives

$$\begin{aligned} N(\lambda+\delta\lambda) &= N(\lambda) + n(\lambda)\delta\lambda + \frac{\Omega(\lambda)\frac{\delta\lambda}{n(\lambda)}}{(1-\Omega(\lambda)\frac{\delta\lambda}{n(\lambda)})^2} N(\lambda) \\ &\approx N(\lambda) + n(\lambda)\delta\lambda + \Omega(\lambda)\frac{\delta\lambda}{n(\lambda)} N(\lambda), \end{aligned} \tag{3.59}$$

leading to the differential equation:

$$\frac{dN}{d\lambda} = n(\lambda) + \Omega(\lambda)\frac{N(\lambda)}{n(\lambda)}. \tag{3.60}$$

To minimize the cost, the right-hand side of Eq. (3.60) must be minimized with respect to $n(\lambda)$. The optimal choice satisfies

$$n(\lambda) = \sqrt{\Omega(\lambda)N(\lambda)}, \tag{3.61}$$

yielding the inequality

$$\frac{dN}{d\lambda} \geq 2\sqrt{\Omega(\lambda)N(\lambda)}. \tag{3.62}$$

Solving this gives

$$N \geq \left(\int_0^\lambda d\lambda \sqrt{\Omega(\lambda)}\right)^2 = L^2, \tag{3.63}$$

Therefore, the total cost of Zeno-based state preparation scales as the square of the path length, $C = O(L^2)$.

This quadratic scaling arises because any failure in the projection process necessitates a full restart from the beginning. As the path length increases, so does the cumulative risk of failure, leading to quadratic scaling. Importantly, this derivation assumes idealized, perfect projective measurements. In practice, physical realizations of such projections may incur additional costs and limitations, which

are discussed in Appendix D.

In conclusion, while the Zeno method guarantees accurate ground state preparation and offers a known cost scaling, its quadratic dependence on path length means that it can become expensive for long paths. Careful consideration is required to assess when it offers advantages over other strategies such as adiabatic evolution.

3.4.2 A new scheme using projections and reflections

The scheme proposed here is different fundamentally from adiabatic state preparation, in which the system evolves continuously through every ground state along the path from the initial state $|g(\lambda_i)\rangle$ to the final state $|g(\lambda_f)\rangle$. It is also different from QZE-based schemes which typically consider only a discrete set of points λ_i along the path, chosen such that neighboring ground states have large overlaps: $|\langle g(\lambda_{i+1})|g(\lambda_i)\rangle|^2 \approx 1$.

Like QZE-based approaches, our method samples the path at discrete points. However, it crucially differs in the choice of these points. Rather than maximizing the overlap between neighboring states, we ideally select the points so that

$$|\langle g(\lambda_{i+1})|g_i(\lambda_i)\rangle|^2 = 1/2,\ L_{\lambda_i,\lambda_{i+1}} = \sqrt{\log(2)}, \tag{3.64}$$

where the second expression holds in the large-volume limit. We assume that the set of points along the path satisfying Eq. (3.64) have been determined to good approximation from the initial study of the path.

An idealized version of the algorithm makes use of an oracle, $\hat{U}(\lambda)_\gamma$, which depends on a parameter λ and a label $\gamma \in a, b$. This oracle acts on the combined state of the system $|\psi\rangle$ and two ancillary

qubits, $|\chi\rangle_a$ and $|\chi\rangle_b$. he action of $\hat{U}(\lambda)\gamma$ is defined as:

$$\begin{aligned}\hat{U}(\lambda)_a &= |g(\lambda)\rangle\langle g(\lambda)| \otimes \hat{I}_a \otimes \hat{I}_b + \left(\hat{I}_\psi - |g(\lambda)\rangle\langle g(\lambda)|\right) \otimes \hat{\sigma}_a^x \otimes \hat{I}_b, \\ \hat{U}(\lambda)_b &= |g(\lambda)\rangle\langle g(\lambda)| \otimes \hat{I}_a \otimes \hat{I}_b + \left(\hat{I}_\psi - |g(\lambda)\rangle\langle g(\lambda)|\right) \otimes \hat{I}_a \otimes \hat{\sigma}_b^x,\end{aligned} \tag{3.65}$$

where $\hat{\sigma}_{a,b}^x$ are Pauli-X operators acting on the ancillas, $\hat{I}_{a,b}$ are identity operators on the ancillas, and $\hat{I}_\psi$ is the identity on the state of the system.

Intuitively, $\hat{U}(\lambda)_a$ flips $|\chi\rangle_a$ if the system is not in the ground state of $H(\lambda)$, and leaves it unchanged otherwise. The same holds for $\hat{U}(\lambda)_b$ with respect to $|\chi\rangle_b$. Accurate approximations of these oracles can be constructed using projection algorithms including the rodeo algorithm [2, 150, 180, 186, 187, 209–213].

It is straightforward to show that if $|e\rangle$ is a superposition of excited states of $H(\lambda)$,

$$\hat{U}(\lambda)_b \left((\alpha|g(\lambda)\rangle + \beta|e\rangle) \otimes |\chi\rangle_a \otimes \hat{H}^{\mathrm{ad}} |\downarrow\rangle_b\right) = (\alpha|g(\lambda)\rangle - \beta|e\rangle) \otimes |\chi\rangle_a \otimes \hat{H}^{\mathrm{ad}} |\downarrow\rangle_b , \tag{3.66}$$

where $\hat{H}^{\mathrm{ad}}$ is the Hadamard gate. That is, $\hat{U}(\lambda)_b$ reflects the state of the physical system across the ground state $g(\lambda)\rangle$.

Now consider an idealized scenario in which the oracle can be implemented exactly, and the discrete points λ_i satisfying Eq. (3.64) are known precisely. A simple algorithm can then propagate a system from the ground state at λ_i to the ground state at λ_{i+1}. The scheme begins with the system in the state $|g(\lambda_i)\rangle$, the ancillary qubit a initialized in $|\uparrow\rangle_a$, and qubit b in $\hat{H}^{\mathrm{ad}} |\downarrow\rangle_b$. The full initial state is thus:

$$|\psi_i\rangle = |g(\lambda_i)\rangle \otimes |\uparrow\rangle_a \otimes \hat{H}^{\mathrm{ad}} |\downarrow\rangle_b \ . \tag{3.67}$$

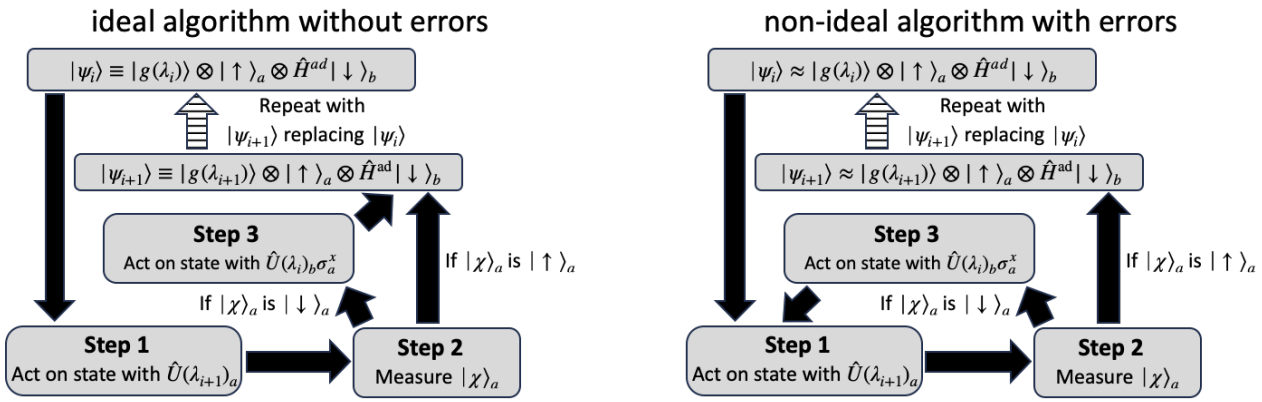

Figure 3.7: Flowcharts illustrating a single cycle of the algorithm in two settings: the idealized version assuming perfect operations (left), and a realistic, error-resilient implementation capable of handling imperfections (right).

The left panel of Figure 3.7 presents a flowchart of the idealized version of the algorithm, which assumes perfect operations. The process begins by applying the oracle $\hat{U}(\lambda_{i+1})_a$ to the initial state $|\psi_i\rangle$. The ancillary qubit $|\chi\rangle_a$ is then measured. In approximately half of the cases, the measurement yields $|\uparrow\rangle_a$, signaling that the system has successfully transitioned to the ground state of $\hat{H}(\lambda_{i+1})$. No further action is required in these instances.

If instead the measurement yields $|\downarrow\rangle_a$, this indicates that the system is in an excited state of $\hat{H}(\lambda_{i+1})$. Importantly, this excited state lies within the two-dimensional subspace spanned by $|g(\lambda_i)\rangle$ and $|g(\lambda_{i+1})\rangle$, and takes the form $\sqrt{2}|g(\lambda_i)\rangle - |g(\lambda_{i+1})\rangle$, as illustrated in Figure 3.8. To correct this, the algorithm applies the operator $\hat{U}(\lambda_i)_b \sigma_a^x$, which reflects the state across $|g(\lambda_i)\rangle$ and yields $|g(\lambda_{i+1})\rangle$.

Thus, regardless of the measurement outcome, the system ends each cycle in the ground state of $\hat{H}(\lambda_{i+1})$, requiring at most two oracle calls per step. For a long path of total length L, the number of oracle calls required is approximately $3L/\sqrt{4\log 2}$ (up to small statistical corrections), and since each call takes a fixed time, the overall runtime scales linearly with L. Given that the length scales as $\sqrt{V}$, this implies that the cost of state preparation grows as $\sqrt{V}$.

Of course, this idealized scenario is not achievable in practice. Real implementations inevitably

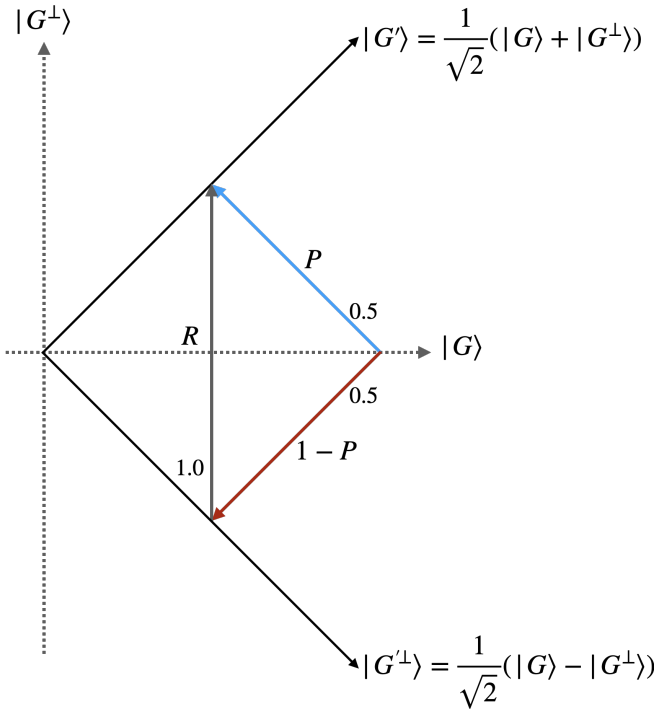

Figure 3.8: Graphical description of a single operation of the idealized version of the algorithm. Here, $|G\rangle \equiv |g(\lambda_i)\rangle$ and $|G'\rangle \equiv |g(\lambda_{i+1})\rangle$. The operator P denotes projection onto the new ground state, while R represents reflection about the current ground state.

encounter errors due to imperfect knowledge, non-ideal hardware, or the intrinsic limitations of constructing exact oracles within finite time. Prior approaches to managing such imperfections [201–203] have focused on suppressing errors at each step through increased resource use, ensuring the cumulative error remains below a target threshold. However, such strategies introduce additional overhead and generally scale no better than $L \log L$.

The present work offers a different insight: it suffices to prevent secular growth in error with increasing L. As long as the admixture of excited states remains bounded—or follows a distribution whose width does not grow with L—one can apply a final "afterburner" stage to refine the state to the desired accuracy. This final projection can be done with cost scaling as $\sqrt{V}$, and crucially, without logarithmic corrections.

The right panel of Figure 3.7 shows a modified flowchart adapted to realistic settings. This

version differs from the ideal case in just one way: after step 3, rather than terminating the cycle, the algorithm loops back to step 1. In a perfect setting, this would be redundant, as the system would already be in the correct ground state. However, in the presence of imperfections, this feedback loop mitigates the accumulation of errors with only a modest increase in cost.

Several types of errors can affect the process. One is imperfect projection: since projection cannot be done exactly in finite time, there is a possibility of "false positives," where excited components are misidentified as ground states. Another source of error is the approximate determination of points satisfying Eq. (3.64). "False negatives" can also occur when uncertainty in the ground state energy leads to rejection of a state that is actually correct, which are discussed for the rodeo algorithm in Appendix D.

One might worry that such errors could accumulate as the system traverses along the path, reducing the ground state fidelity. However, a simple heuristic shows that this is not the case. Suppose the maximum error per step is ϵ, representing the amplitude of non-ground-state contamination introduced in a single cycle. Because each cycle concludes with an application of $\hat{U}(\lambda)_a$, errors introduced in one step are counteracted in the next. Each new step may introduce its own error (bounded by ϵ), but this process effectively damps the accumulation. Thus, the total error converges geometrically and remains bounded—ensuring that errors do not grow secularly with L.

To evaluate the heuristic argument, we carried out a numerical simulation of the algorithm using a simplified toy model. Rather than working directly with Hamiltonians, the model focuses on the structure of the ground and excited states, since the algorithm depends only on these states, not on how they are generated. States are constructed essentially at random by forming linear combinations of an orthonormal basis of 1024 vectors—chosen to be large enough to mimic a generic quantum system while still allowing for manageable computations.

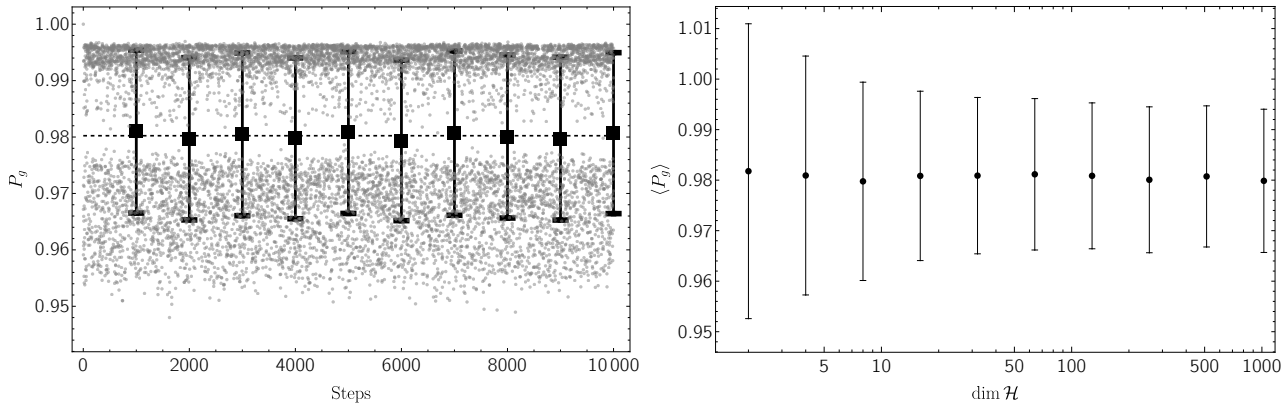

Figure 3.9: A numerical simulation of the non-ideal algorithm shown in Figure 3.7 using the toy model described in the text. The left plot is with the dimension of Hilbert space as 1024. The ground state probability, P_g, is defined as the squared overlap between the true intermediate ground state and the state produced by the algorithm. Dots indicate P_g at each step of the path, while rectangles show averages over every 1000 steps, with error bars representing the standard deviation. The dashed line marks the overall average across all steps. Note that the vertical axis is compressed, with P_g ranging from 0.94 to 1.00. The right plot depicts the behavior of the algorithm by changing the dimension of the Hilbert space. $\langle P_g \rangle$ denotes the average of the squared overlap between the true intermediate ground state and the state produced by the algorithm.

Each pair of successive ground states in the sequence was chosen to have an overlap near $1/\sqrt{2}$. To model imperfections realistically, we incorporated false positives by allowing excited states to be mistakenly identified as ground states with a probability drawn uniformly between 0 and 0.1—an intentionally conservative choice (in practice, for example, a single Rodeo "super iteration" has a false positive rate below 0.05 shown in Figure 3.4). We also included a 5% chance of false negatives, representing cases where the projection incorrectly rejects a true ground state—again, a generous estimate assuming decent knowledge of the spectrum. Finally, we introduced variation in step sizes by allowing the angle between successive ground states to differ from the ideal $\pi/4$ by up to 10%, simulating imperfections in selecting path points.

The left panel in figure 3.9 shows the results from this toy model. The ground state probability, P_g—the squared overlap between the true and computed intermediate ground states—remains very close to unity across all steps, with a minimum exceeding 0.947. While fluctuations occur, there is

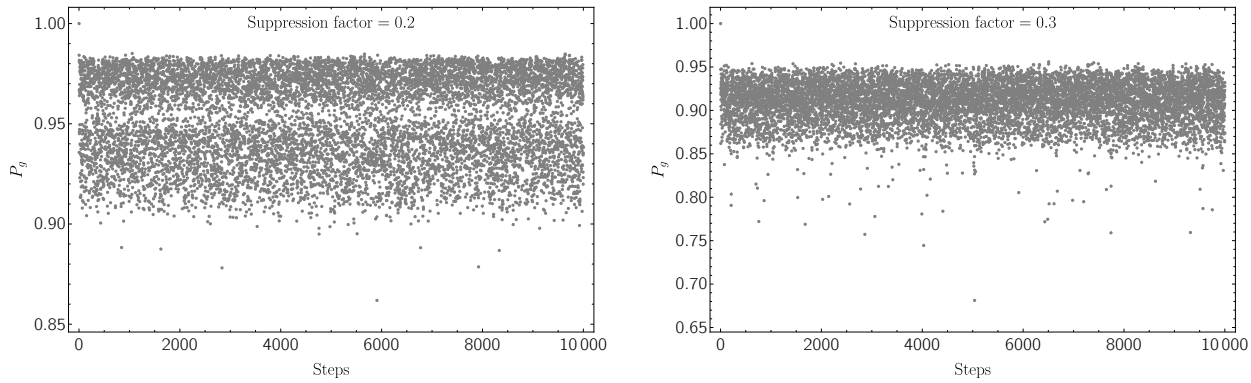

Figure 3.10: Numerical simulations of the non-ideal algorithm utilizing larger false positive error rates of 0.2 and 0.3 (left and right panels, respectively). All other simulation parameters are identical to those detailed in Figure 3.9.

no evidence of secular growth in error: the statistical distribution of P_g in steps 9000–10000 closely mirrors that of steps 1–1000, both in average value and spread. These results provide strong numerical support for the claim that the error remains bounded throughout the path.

The right panel of Figure 3.9 demonstrates that the algorithm is independent of the Hilbert space dimension. Specifically, the ensemble average of the squared overlap between the true ground state and the intermediate ground state, $\langle P_g \rangle$, unchanges as the dimension of the Hilbert space increases. This is mathematically expected; projections to the ground state only depend on the amplitude of the ground state component.

Figure 3.10 demonstrates that the false positive error rates (suppression factors) need not be small during the intermediate steps. The left and right panels illustrate simulations with suppression factors of 0.2 and 0.3, respectively. Although the overlap between the true ground state and the intermediate states is reduced compared to scenarios with stronger projections, the ultimate objective is solely the fidelity of the final state. Consequently, one can significantly optimize the computational cost of the algorithm by employing weaker projections during the intermediate stages.

Assuming this behavior persists in general, the final state after traversing a long path will have

a large overlap with the true ground state. If extremely high accuracy is required, a final refinement step—an "afterburner"—can be applied. Although the last point on the path may not satisfy the optimal overlap condition from Eq. (3.64), repeated projection and reflection operations using a much more stringent projector (with an exponentially suppressed false positive rate) will eventually yield the true ground state with exponentially small error. As shown in Sec. 3.2, the computational cost of achieving such precision grows more slowly than logarithmically with the inverse of the false positive rate when using the rodeo super iterations.

For long paths, the total cost is dominated by the traversal, not the final refinement. Each projection generally requires time proportional to the inverse of the spectral gap, Δ, so the overall cost scales as

$$C \sim V^{\frac{1}{2}} \left\langle \frac{1}{\Delta} \right\rangle, \tag{3.68}$$

where the angle brackets denote the average inverse spectral gap along the entire path.

While the numerical evidence demonstrates empirical stability, the algorithm's robustness can be enhanced by introducing a hierarchical projection scheme. Specifically, one can apply systematically stronger projections at fixed steps—for instance, increasing the projection strength by 10 times every 10 steps, 100 times every 100 steps, and so forth. Because the suppression factor decays exponentially with projection strength, these periodic, high-fidelity projections effectively purge accumulated intermediate errors at the cost of a strictly bounded constant overhead in evolution time. Unlike the strategy of deferring a single, perfect projection to the final step, this periodic regularization continuously stabilizes the intermediate states while preserving the linear scaling of the overall computational cost with respect to the path length.

To make subsequent traversals efficient, certain properties of the path must first be determined

during initial passes. At each point, one needs a reasonable estimate of the appropriate step size $\Delta\lambda$ that approximately satisfies Eq. (3.64), the ground state energy (to reduce the false negative rate), and the spectral gap (for effective projection). While the algorithm described above can be applied to any problem with a long path, it may become impractical if determining these system properties during the initial traversal is computationally expensive.

However, for local field theories in large volumes, the situation is favorable: key quantities vary slowly between adjacent points along the path. This means that one can determine the required values at a relatively sparse set of points and interpolate between them. To see why, recall from Eq. (3.64) that the step size scales as $\Delta\lambda_i = \lambda_{i+1} - \lambda_i \propto V^{-1/2}$. At large volumes, the spectral gap $\Delta(\lambda)$ depends only on λ and not on the volume. Consequently, $\Delta\lambda_{i+N}$ can be well-approximated by $\Delta\lambda_i$ for as many as $N \propto V^{1/2}$ steps. Likewise, from Eqs. (3.47)-(3.49), the spectral gap remains nearly constant over intervals of similar length.

In contrast, the ground state energy $E_0(\lambda)$ grows with volume, so it requires more careful treatment. Still, a Taylor expansion around λ_i provides a systematic approximation: an nth-order expansion remains accurate out to $N \propto V^{1/2-1/n}$ steps. Thus, even for extensive quantities like $E_0(\lambda)$, efficient estimation across the path is possible with relatively modest initial effort.

3.5 Adiabatic state preparation

The adiabatic theorem, first discovered by Born and Fock [214], stands as one of the foundational results of quantum mechanics. It asserts that a quantum system initialized in an eigenstate will remain in its instantaneous eigenstate if the system's Hamiltonian changes sufficiently slowly and the spectral gap remains open. This principle has found wide-ranging applications, from the Born-Oppenheimer

approximation in molecular dynamics [215] to the understanding of topological phases of matter [216–218].

In recent decades, the adiabatic theorem has also become central to quantum computing, where it provides a foundation for adiabatic state preparation and adiabatic quantum algorithms [188–192, 219–223]. These approaches typically involve initializing a quantum system in the ground state of a simple Hamiltonian and slowly evolving it along a path through Hamiltonian space, such that the system remains in its ground state and ultimately encodes the solution to a computational problem. The viability of this process depends critically on the evolution rate and the spectral properties of the Hamiltonian, making the adiabatic theorem both a conceptual and practical cornerstone of quantum state preparation.

While the adiabatic theorem is highly useful, its implementation in practice is complicated by the trade-off between speed and precision. Since physical systems evolve over finite times, perfect adiabatic behavior cannot be maintained, resulting in errors. Assessing the magnitude of these errors and how they scale with the evolution time T is essential for evaluating whether adiabatic methods are viable in realistic settings.

In realistic quantum simulations, the Hamiltonian never evolves infinitely slowly, and as a result, diabatic transitions—transitions to excited states of the instantaneous Hamiltonian—are unavoidable. These transitions introduce errors that limit the fidelity of adiabatic state preparation. Understanding and controlling such diabatic errors is especially important in the context of quantum computing, where precise control over quantum states is essential. However, the behavior of these errors is nuanced and depends sensitively on properties of the Hamiltonian and its evolution. Over the past two decades, a number of rigorous approaches have been developed to characterize and bound these errors. These methods typically provide upper bounds on the total evolution time T required to keep the error below

a desired threshold ϵ. The bounds often depend on the minimum spectral gap Δ along the evolution path, as well as the norms of the Hamiltonian and its time derivatives [3, 224–229].

However, the usefulness of these bounds is limited, as they depend on the norm of the Hamiltonian, which in turn is influenced by the highest energy eigenvalues. Although quantum computers simulate systems with a finite Hilbert space and thus a maximum energy level, they often aim to approximate systems that have much larger—or even infinite—Hilbert spaces and unbounded energy spectra. As the simulation becomes more faithful to the underlying physical system, these theoretical bounds lose their tightness. This issue becomes especially significant when dealing with systems of large or infinite extent, such as those found in quantum field theories [163, 170, 198, 230–239].

In this section, we examine the "switching theorem" as a method for estimating errors, with a focus on its effectiveness in minimizing errors while conserving computational resources. Although the switching theorem provides a practical framework for relating errors to the evolution timescale, it does not fully capture error behavior in the long-path limit—an important regime for large systems such as quantum field theories. To address this, we introduce a new proxy that more accurately reflects the computational complexity and analyze how it scales with the path length.

3.5.1 Adiabatic theorem

This subsection provides a review of the adiabatic theorem and introduces the key terminology used throughout the discussion.

The evolution of a quantum system is governed by the time-dependent Schrödinger equation:

$$i\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \tag{3.69}$$

where $H(t)$ is the time-dependent Hamiltonian and $\hbar = 1$ is used for convenience. According to the adiabatic theorem, if the Hamiltonian evolves infinitely slowly, a system that begins in the nth eigenstate of the initial Hamiltonian will remain in the corresponding nth eigenstate of the instantaneous Hamiltonian throughout the evolution, arriving at the nth eigenstate of the final Hamiltonian. In this work, the focus is on the ground state, though the same framework applies to other eigenstates *mutatis mutandis*.

In realistic implementations, however, the evolution occurs over a finite time, leading to non-adiabatic transitions, or diabatic errors, where the system partially transitions out of the instantaneous eigenstate. The error accumulated during such an evolution is quantified by

$$\epsilon \equiv \left\| \left(1 - |g_f\rangle\langle g_f|\right) U(t_f, t_i) |g_i\rangle \right\|, \tag{3.70}$$

where $U(t_f, t_i)$ is the time evolution operator from the initial time t_i to the final time t_f and $|g_i\rangle$, $|g_f\rangle$ are the ground states of the initial and final Hamiltonians, respectively. This expression measures the amplitude of the evolved state that is orthogonal to the final ground state.

To facilitate analysis, the time evolution is commonly reparameterized by introducing a dimensionless parameter $s \in [0, 1]$, related to physical time by $s = (t - t_i)/T$, where $T = t_f - t_i$ is the total evolution time. Under this reparameterization, the Schrödinger equation becomes

$$i\frac{d}{ds}|\psi(s)\rangle = TH(s)|\psi(s)\rangle, \tag{3.71}$$

which makes explicit the role of the timescale T. This formulation highlights the trade-off between evolution time and accuracy: longer timescales suppress diabatic transitions but come at the cost of

increased computational resources. Understanding how the error ϵ scales with T is central to evaluating the practicality of adiabatic state preparation in both quantum computing and quantum simulation contexts.

3.5.2 Switching theorem

One method for estimating diabatic errors involves the use of the switching theorem, which provides a direct characterization of error behavior in the limit of asymptotically long evolution times. Although the adiabatic theorem was introduced much earlier, the switching theorem was developed decades later [240] and has since been explored primarily in the context of mathematical physics [241–253]. This theorem captures how the diabatic error scales with the evolution time T and the spectral gap structure of the Hamiltonian. The resulting error expression takes the form of a *quasi*-asymptotic series, so called because its coefficients depend on T only through relative phase differences originating at the initial and final points. Notably, the structure of the path taken by the Hamiltonian influences errors only through phase factors, allowing the error bound given by the switching theorem to be effectively independent of the specific trajectory taken.

Mathematically, the operator difference between the time-evolved projector $U|g_i\rangle\langle g_i|U^\dagger$ and the desired projector $|g_f\rangle\langle g_f|$ takes the form

$$U|g_i\rangle\langle g_i|U^\dagger - |g_f\rangle\langle g_f| = \sum_{n=1}^{n_{\max}(T)} \frac{B_n}{T^n} + R, \tag{3.72}$$

where the operator B_n depend on T only through phase factors and the remainder R decays more faster than any polynomial in $1/T$. The number of terms $n_{\max}(T)$ indicates the optimal truncation order at which the approximation remains valid. This expansion implies a corresponding asymptotic form for

the error magnitude itself:

$$\epsilon = \sum_{n=1}^{n_{\max}(T)} \frac{b_n}{T^n} + r, \tag{3.73}$$

where the coefficients b_n and r mirror the properties of B_n and R, respectively. Crucially, both B_n and b_n depend only on the initial and final Hamiltonians, specifically through the matrix elements of the nth derivative of the Hamiltonian with respect to the rescaled time variable. The intermediate path between endpoints affects the error only through phase factors, underscoring the endpoints nature of the leading error contributions in the switching theorem framework.

This means that if the Hamiltonian has only a finite number of nonzero derivatives at the endpoints, then the series terminates after a finite number of terms, as all higher-order coefficients vanish. In the limit of large T, the switching theorem further indicates that, aside from oscillatory behavior from phase factors, the leading error scales as $1/T^k$, where k is the smallest integer such that b_k is nonzero.

This implies that if only a finite number of derivatives are nonzero, then there will only be a finite number of nonzero coefficients and the series will truncate. For asymptotically large values of T, the switching theorem implies that up to oscillations due to the phase factors, the errors behave as $(1/T)^k$ where k is the smallest integer for which b_k is nonzero.

The coefficients b_n can be determined using perturbation theory [184, 254–256]. In typical scenarios where the first time derivative of the Hamiltonian at either endpoint is nonzero, the leading contribution to the *quasi*-asymptotic expansion at large T comes from the $1/T$ term. The explicit expression for b_1 is

$$b_1 = \sqrt{\sum_{j \neq g} \left| e^{i w_{j,g} T} \frac{\langle j(1)|H'(1)|g(1)\rangle}{\Delta_{j,g}^2(1)} - \frac{\langle j(0)|H'(0)|g(0)\rangle}{\Delta_{j,g}^2(0)} \right|^2}, \tag{3.74}$$

where $\Delta_{j,g}(t) = E_j(t) - E_g(t)$ is the energy gap between the jth excited state the ground state, and $w_{j,g}$ is the integrated gap along the path, defined as $w_{j,g} = \int_0^1 ds\, \Delta_{j,g}(s)$. The summation runs over all excited states. From Eq. (3.74), it follows that if $H'(s)$ vanishes at both endpoints, then $b_1 = 0$, and the leading error term appears at order $1/T^2$ in Eq. (3.73). More generally, if all derivatives $H^{(m)}$ up to order n vanish at the endpoints, but $H^{(n)}$ is nonzero at least one endpoint, the dominant contribution to the error is controlled by b_n, given by

$$b_n = \sqrt{\sum_{j \neq g} \left| \mathrm{e}^{i w_{j,g} T} \frac{\langle j(1)|H^{(n)}(1)|g(1)\rangle}{\Delta_{j,g}^{n+1}(1)} - \frac{\langle j(0)|H^{(n)}(0)|g(0)\rangle}{\Delta_{j,g}^{n+1}(0)} \right|^2}. \tag{3.75}$$

3.5.3 Hyperadiabatic regime and typical error

Within its range of applicability, the switching theorem yields increasingly accurate results as more terms in the series are included—up to an optimal number that typically depends on the evolution time T. When the Hamiltonian has only a finite number of nonzero derivatives at the endpoints, the expansion terminates at a finite order. We define the *hyperadiabatic regime* as the range of T values for which the asymptotic expansion remains valid and is well-approximated by its leading term. This regime is a subset of the broader adiabatic regime, where the system remains close to the instantaneous ground state and the adiabatic theorem is approximately satisfied.

Note that according to Eq. (3.75), the coefficient b_n carries an explicit dependence on T through the oscillatory phase factors $\mathrm{e}^{i w_{j,g} T}$. As a result, although the error exhibits a leading scaling of $1/T^n$, its actual dependence on T is more intricate due to these phases. Furthermore, the presence of these phase factors implies that the error is not determined exclusively by the endpoint—it also reflects the properties of the Hamiltonian along the entire path of the evolution.

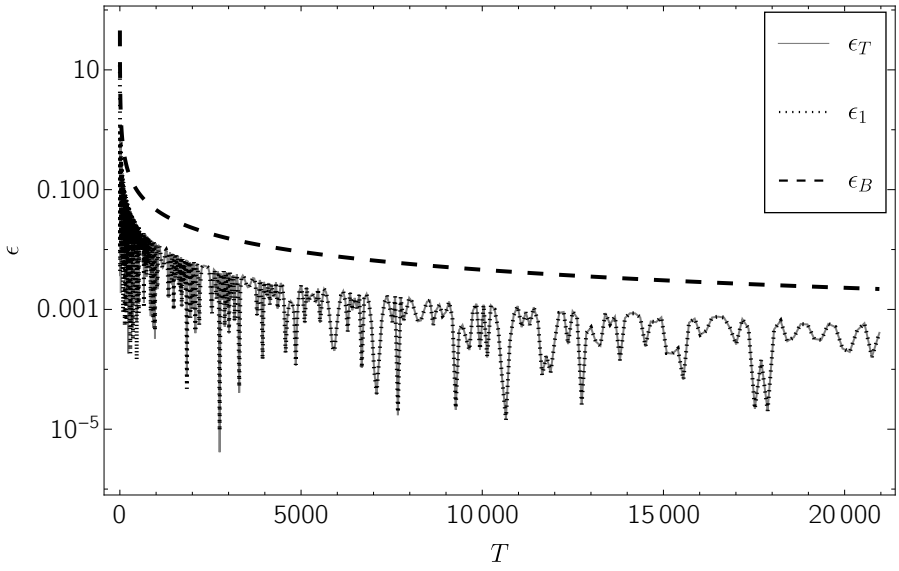

Figure 3.11: Comparison of the true error (ϵ_T, solid), the switching theorem estimate (ϵ_1, dotted), and the rigorous upper bound based on the norms of the Hamiltonian's derivatives and the minimum spectral gap (ϵ_B, dashed), following the well-known result from [3], for the Hamiltonian defined in Eq. (3.76).

Figure 3.11 illustrates the various difficulties in estimating adiabatic errors using both the upper bound based on the norms of the Hamiltonian's derivatives and minimal spectral gap, and the switching theorem. The figure compares these estimates against the exact error, ϵ_T, computed by directly solving the Schrödinger equation, using the Hamiltonian

$$H(s) = \begin{bmatrix} s(1-s) & 0.2 \\ 0.2 & -s(1-s) \end{bmatrix}, \tag{3.76}$$

with $s \in [0,1]$. Since $H'(s)$ is nonzero at the endpoints, the switching estimate $\epsilon_1 = b_1/T$ (where b_1 is given in Eq. (3.74)) applies. The upper bound ϵ_B, derived in Ref. [3], relies on the minimum spectral gap and the norms of the Hamiltonian's derivatives:

$$\epsilon_B = \frac{1}{T}\left.\frac{m(s)\|\dot{H}(s)\|}{\Delta_{1,g}(s)}\right|_{\text{u.b.}} + \frac{1}{T}\int_0^1 ds\left(\frac{m(s)\|\ddot{H}(s)\|}{\Delta_{1,g}^2(s)} + 7m(s)\sqrt{m(s)}\frac{\|\dot{H}(s)\|^2}{\Delta_{1,g}^3(s)}\right), \tag{3.77}$$

where $f|_{\text{u.b.}}$ represents $f(0) + f(1)$, and $m(s)$ denotes the number of eigenvalues of the spectrum of $H(s)$ restricted to the projection operator $P(s) = |g(s)\rangle\langle g(s)|$. As shown in Figure 3.11, the upper bound hugely overestimates the true error. In contrast, the switching estimate closely tracks the actual error at large T, though both exhibit oscillations—an intrinsic feature of adiabatic dynamics. These fluctuations highlight the difficulty of reliably estimating errors in practice, where the exact timescale T is often uncertain.

This subsection explores how these issues can be avoided with a slight change of focus on a quantity that differs from the asymptotic expression for the error itself. The goal is to identify a quantity associated with the typical value of the error that has the properties of exhibiting a simple power-law scaling with T in the hyperadiabatic regime, and whose value depends solely on the endpoints. This can be achieved by considering the typical value of the error for timescales in the vicinity of T by averaging the error over a range of T values. The range should be much larger than the inverse of $w_{1,g}$ (the average spectral gap between the ground state and the lowest excited state) but much smaller than T.

Formally, the "typical error" at a large timescale T, denoted $\bar{\epsilon}(T)$, can be defined as

$$\bar{\epsilon}(T) \equiv \frac{1}{2\sqrt{T\tau_0}} \int_{T-\sqrt{T\tau_0}}^{T+\sqrt{T\tau_0}} dT' \, \epsilon(T'), \tag{3.78}$$

where τ_0 is some arbitrary positive constant. As T grows, this averaging window spans a timescale that is wide enough to smooth over the rapid oscillations in $\epsilon(T)$ yet remains narrow compared to T itself. As a result, the average becomes effectively independent of the precise value of τ_0. When the first nonzero derivative of the Hamiltonian at the endpoints is of order n, the switching theorem (specifically Eq. (3.75)) implies that $\bar{\epsilon}(T)$ scales as $\bar{b}_n/T^n$ in the hyperadiabatic regime. The prefactor

$\bar{b}_n$ is determined by the endpoint properties of the Hamiltonian and is given by

$$\bar{b}_n = \sqrt{\sum_{j\neq g}\left|\frac{\langle j(0)|H^{(n)}(0)|g(0)\rangle}{\Delta_{j,g}^{n+1}(0)}\right|^2 + \sum_{j\neq g}\left|\frac{\langle j(1)|H^{(n)}(1)|g(1)\rangle}{\Delta_{j,g}^{n+1}(1)}\right|^2} \equiv \sqrt{(\bar{b}_n^0)^2 + (\bar{b}_n^1)^2}. \tag{3.79}$$

Although the typical error is formally introduced in Eq. (3.78), its structure can be anticipated by analyzing the square of the switching coefficient in Eq. (3.75):

$$\begin{aligned} b_n^2 = \sum_{j\neq g}\Bigg(&\left|\frac{\langle j(1)|H^{(n)}(1)|g(1)\rangle}{\Delta_{j,g}^{n+1}(1)}\right|^2 + \left|\frac{\langle j(0)|H^{(n)}(0)|g(0)\rangle}{\Delta_{j,g}^{n+1}(0)}\right|^2 \\ &+ \mathrm{e}^{iw_{j,g}T}\frac{\langle j(1)|H^{(n)}(1)|g(1)\rangle}{\Delta_{j,g}^{n+1}(1)}\frac{\langle g(0)|H^{(n)}(0)|j(0)\rangle}{\Delta_{j,g}^{n+1}(0)} \\ &+ \mathrm{e}^{-iw_{j,g}T}\frac{\langle g(1)|H^{(n)}(1)|j(1)\rangle}{\Delta_{j,g}^{n+1}(1)}\frac{\langle j(0)|H^{(n)}(0)|g(0)\rangle}{\Delta_{j,g}^{n+1}(0)}\Bigg). \end{aligned} \tag{3.80}$$

The last two terms oscillate with T due to the phase factors $\mathrm{e}^{\pm iw_{j,g}T}$ and, when averaged over a suitable timescale, effectively cancel out. This leaves only the first two, time-independent terms, resulting in the expression for $\bar{b}_n$ in Eq. (3.79). While other formulations of typical error—such as using a root-mean-square over time—are possible, they also eliminate the oscillatory contributions upon averaging and yield the same behavior as Eq. (3.78).

There are multiple advantages to using the typical error, $\bar{\epsilon}(T)$, in place of the true (or instantaneous) error $\epsilon(T)$. Two primary motivations for introducing $\bar{\epsilon}(T)$ are first, that in the asymptotic limit it exhibits a simple power-law dependence on T, and second, that its value depends only on the endpoint behavior of the Hamiltonian. This makes the asymptotic scaling entirely insensitive to the details of how the evolution path is traversed. Indeed, the average of the squared error, $\bar{\epsilon^2}(T)$, is given by $\left((\bar{b}_n^0)^2 + (\bar{b}_n^1)^2\right)/T^{2n}$, a sum of contributions from the initial and final points alone.

Beyond these core advantages, $\bar{\epsilon}(T)$ has several practical merits. In realistic scenarios, it is often

not feasible to exert fine-grained control over the interpolation path—or even to know its detailed structure well enough to compute the relevant phase factors in the oscillatory terms. In such cases, the typical error plays a role analogous to the expected error value, providing a reliable estimate. Moreover, there are situations in which $\epsilon(T)$ closely tracks $\bar{\epsilon}(T)$ for nearly all large T, particularly when the sum in Eq. (3.75) receives comparable contributions from many excited states. In these cases, the oscillatory phase terms tend to destructively interfere, suppressing fluctuations.

Another practical benefit is that $\bar{\epsilon}(T)$ can provide an effective upper bound on the true error in the hyperadiabatic regime. From the form of Eq. (3.75) and Eq. (3.79), one finds that

$$\epsilon(T) \leq \sqrt{2}\bar{\epsilon}(T). \tag{3.81}$$

Such a bound may prove particularly useful in the context of quantum computation, where guarantees on error magnitude are often essential. However, it is important to note that this is not a strict upper bound valid for all T; it applies only once the system has reached the hyperadiabatic regime. At earlier times, before this regime is entered, the error behavior may still be influenced by the total path length or other non-asymptotic features.

Figure 3.12 presents a comparison between the true error, ϵ_T, the typical error, $\bar{\epsilon}$, and its corresponding upper bound, $\sqrt{2}\bar{\epsilon}$. The plot demonstrates that $\bar{\epsilon}$ effectively smooths out the fluctuations present in the true error. Moreover, in the hyperadiabatic regime, the bound $\sqrt{2}\bar{\epsilon}$ closely tracks the envelope of the true error's fluctuations, confirming its accuracy as an asymptotic upper limit.

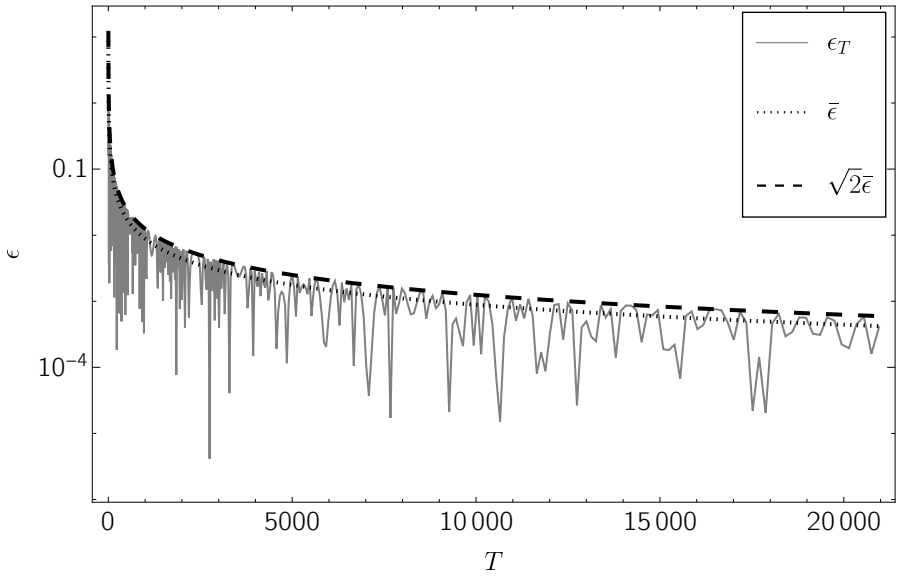

Figure 3.12: Comparison of the true error (ϵ_T, solid), the typical error ($\bar{\epsilon}$, dotted), the asymptotic upper bound in the hyperadiabatic regime ($\sqrt{2}\bar{\epsilon}$, dashed), for the Hamiltonian in given Eq. (3.76).

3.5.4 Utility of the switching theorem

The fact that Eq. (3.79) depends only on the endpoints of the evolution—and not on the details of the intermediate Hamiltonian—presents a potential advantage. In the hyperadiabatic regime, it becomes possible to alter the instantaneous properties of the Hamiltonian at the initial and final times, while leaving the bulk of the intermediate evolution unchanged[2], to reduce the error. The central question then becomes: what determines whether the system is actually in the hyperadiabatic regime?

The switching theorem becomes especially valuable for quantum computing applications when the error is dominated by the leading nonzero term in the $1/T$ expansion from Eq. (3.73). When $b_1 \neq 0$, this dominance is achieved when

$$b_1 \geq b_2 T^{-1}, \cdots, b_n T^{-n+1}, \cdots \tag{3.82}$$

[2]Note that changing the intermediate path can be dangerous since one might encounter phase transitions as discussed in this section earlier.

For a given set of higher-order coefficients $\{b_2\ ,\ b_3\ ,\ \cdots\}$, this means that a larger b_1 causes the leading-order behavior to emerge at shorter timescales T. While this might seem to suggest that increasing b_1 helps reveal the scaling behavior more quickly, doing so actually increases the error—defeating the purpose.

A more effective strategy for error reduction in the hyperadiabatic regime is to suppress the lower-order derivatives of the Hamiltonian at the endpoints while preserving the overall path. According to the switching theorem, if $H^{(k)}(t_i) = H^{(k)}(t_f) = 0$ for all $k \leq n$, the leading term in the error scales as $1/T^{n+1}$. Thus, increasing n should, in principle, reduce the error—assuming the system remains in the hyperadiabatic regime after the modification. However, there is a risk: such endpoint smoothing may push the required timescale T for the hyperadiabatic approximation to hold to impractically large values. It is therefore essential to assess whether this tradeoff occurs.

3.5.4.1 Changing scaling behavior and the onset of hyperadiabaticity

To better understand how the hyperadiabatic regime emerges, this subsection explores a set of simple, numerically tractable model Hamiltonians. These models confirm the expectations set by the switching theorem: at sufficiently long times, the system's behavior does indeed approach the asymptotic form predicted by the theorem. However, the timescale required for this asymptotic behavior to become accurate depends sensitively on details of the interpolation between endpoints. In some cases, this timescale can become quite large.

Interestingly, it is possible for the system to evolve slowly enough to satisfy standard adiabatic conditions without yet entering the hyperadiabatic regime. In such situations, the error can exhibit an approximate power-law scaling in $1/T^n$, where n is *smaller* than the leading value expected from the

switching theorem. Only at longer times does the error's scaling converge to the predicted asymptotic form.

To illustrate how the timescale for hyperadiabatic behavior can depend on fine details at the endpoints, consider the following example. Start with a generic Hamiltonian $H_0(s)$ that vanishes at both endpoints but has nonzero first derivatives: $H_0(0) = H_0(1) = 0$ and $H_0'(0)$, $H_0'(1) \neq 0$. In this case, the typical error in the hyperadiabatic regime scales as $\bar{\epsilon} = \bar{b}_1/T$.

Now, define a family of functions

$$g(s;k) = \frac{s}{s+k}, \tag{3.83}$$

where k is a small positive constant. This function is designed to be zero at $s = 0$ and approaches 1 for $s \gg k$. Using this, construct a modified Hamiltonian:

$$H_{k,n}(s) \equiv g(s;k)^n H_0(s) g(1-s;k)^n, \tag{3.84}$$

which keeps the intermediate path nearly identical to $H_0(s)$ but modifies the endpoints such that the first n derivatives vanish. The first nonzero derivative of this modified Hamiltonian at the endpoints is:

$$H_{k,n}^{(n+1)}(0) = \frac{n! H_0'(0)}{k^n}, \; H_{k,n}^{(n+1)}(1) = \frac{n! H_0'(1)}{k^n}. \tag{3.85}$$

Substituting this into the expression for the typical error yields:

$$\bar{\epsilon}_n(T) = \frac{1}{T^n}\sqrt{\frac{n!}{k^n}}\sqrt{\sum_{j\neq g}\left(\frac{\langle j(0)|H'(0)|g(0)\rangle}{\Delta_{j,g}^{n+1}(0)} + \frac{\langle j(1)|H'(1)|g(1)\rangle}{\Delta_{j,g}^{n+1}(1)}\right)}. \tag{3.86}$$

To ensure this is a valid estimate of the actual error, one needs $\bar{\epsilon}_n(T) \ll 1$, implying a condition on the timescale:

$$T \gg \frac{(n!)^{1/2n}}{\sqrt{k}}. \tag{3.87}$$

Thus, as $k \to 0$ (i.e., as the modified Hamiltonian more closely resembles the original), the timescale required to reach the hyperadiabatic regime grows. More significantly, increasing n to achieve faster error decay (e.g., $1/T^{n+1}$) also pushes the necessary timescale up rapidly—approximately factorially.

The key insight is this: even if one successfully engineers the endpoints of the Hamiltonian to suppress low-order derivatives, large derivatives at higher order can still arise. These large derivatives delay the onset of the hyperadiabatic scaling behavior predicted by the switching theorem. Therefore, while such modifications can in principle reduce long-time errors, they may only do so at impractically long times.

3.5.4.2 Examples: two-level and three-level systems

We analyze specific examples in two-and three-level systems to investigate how the error approaches the asymptotic scaling behavior predicted by the switching theorem. These examples are intentionally designed to reflect the practical challenges discussed earlier. We begin with a baseline Hamiltonian $H_0(t)$, which we then perturb near the endpoints so that its first derivatives vanish, while keeping the intermediate evolution nearly unchanged. This allows us to study how such endpoint modifications impact error evolution when the actual dynamics deviates slightly from the intended $H_0(t)$.

Consider a two-level system governed by the Hamiltonian:

$$H_0(t) = \begin{bmatrix} 0 & E_1 f_0(t) \\ E_1 f_0(t) & E_0 \end{bmatrix}, \tag{3.88}$$

with a time-dependent function

$$f_0(t) \equiv t(1-t). \tag{3.89}$$

Since $H_0'(t)$ is nonzero at the endpoints, the switching theorem predicts that in the large-T limit, the typical error scales as $\bar{b}_1/T$ where $\bar{b}_1$ is defined in Eq. (3.79).

Although this model might seem trivial because the initial and final states are the same, this does not affect the key insights. The conclusions remain valid for nontrivial protocols where the initial and final states differ, as is typical in practical adiabatic state preparation. For the numerical results shown, we choose $(E_0, E_1) = (1, 1)$, which gives an energy gap of one at the endpoints (in arbitrary units). We are not concerned with the detailed time dependence of the gap here. The actual error is computed by numerically solving the Schrödinger equation and averaging the result using Eq. (3.78); we denote this as $\bar{\epsilon}(T)$, and compare it with the switching theorem estimate $\bar{\epsilon}_n(T)$ from Eq. (3.79).

According to the expansion in Eq. (3.73), if the first derivative of the Hamiltonian vanishes at both endpoints but the second derivative does not, the leading error in the large-T limit scales as $\bar{\epsilon} \to \bar{b}_n/T^2$, with $\bar{b}_2$ defined as in Eq. (3.79). Motivated by this, we ask whether we can suppress the error by modifying the Hamiltonian only near the endpoints to eliminate the first derivative, while keeping the rest of the evolution intact. In the case of the Hamiltonian from Eq. (3.88), this involves modifying the off-diagonal time-dependent term so that its first derivative vanishes at the boundaries. However, any such change at the endpoints will inevitably influence the interior behavior as well,

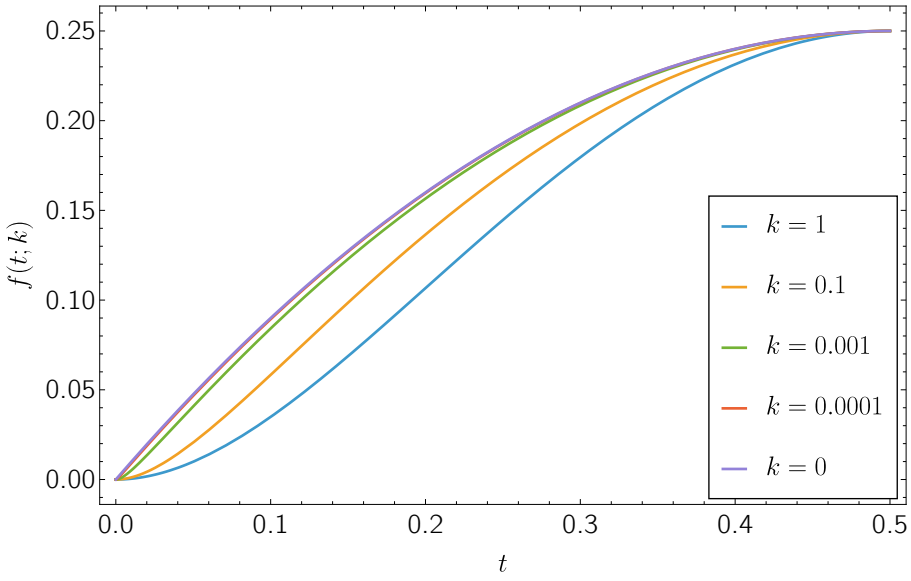

Figure 3.13: The function f(t;k) plotted for various values of k. When $k = 0$, the function reduces to $f_0(t) = t(1-t)$, which has nonzero first derivatives at the endpoints. For $k > 0$, the function is smoothly modified so that the first derivatives vanish at $t = 0$ and $t = 1$.

though potentially only slightly.

To implement this, we introduce a modified Hamiltonian:

$$H(t;k) = \begin{bmatrix} 0 & E_1 f(t;k) \\ E_1 f(t;k) & E_0 \end{bmatrix}, \tag{3.90}$$

where

$$f(t;k) \equiv \left(\frac{1}{1+2k}\right)^2 \frac{t}{t+k} t(1-t) \frac{1-t}{1-t+k}. \tag{3.91}$$

The additional factor $\frac{t}{t+k}$ and $\frac{1-t}{1-t+k}$ smooth the function near $t = 0$ and $t = 1$, ensuring that the first derivatives vanish when $k > 0$. When $k = 0$, the function reduces to $f_0(t)$, and we recover the original Hamiltonian $H_0(t)$, which exhibits error scaling as $\bar{b}_1/T$. For small but nonzero k, the goal is to reduce the error at moderately large T without significantly altering the evolution much. The normalization factor ensures that the midpoint value $f(1/2;k)$ matches $f_0(1/2)$, keeping the intermediate behavior

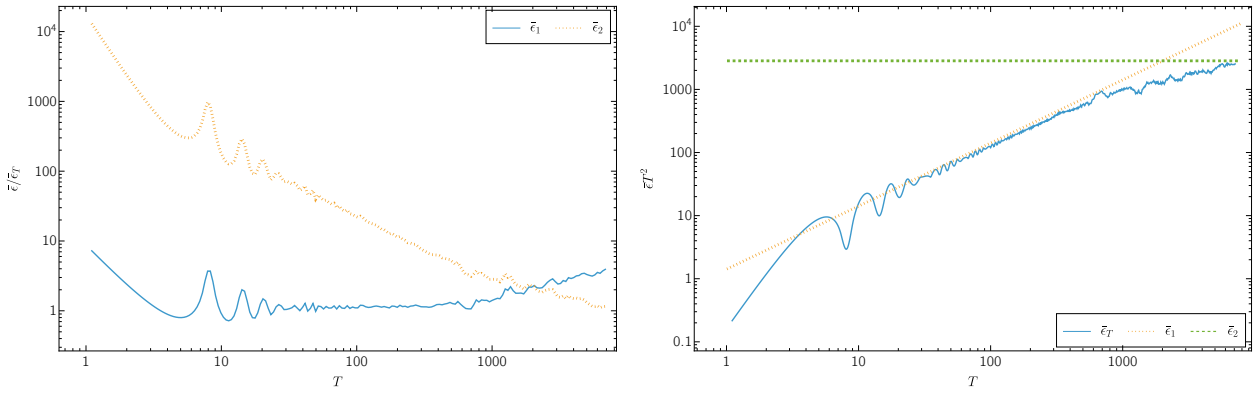

Figure 3.14: Comparison of average errors $\bar{\epsilon}_T$, $\bar{\epsilon}_1$, and $\bar{\epsilon}_2$, with $k = 10^{-3}$ and $(E_0, E_1) = (1, 1)$. ***Left panel***: The ratio of the switching estimates to the true average error, with the solid line showing $\bar{\epsilon}_1/\bar{\epsilon}_T$ and the dotted line showing $\bar{\epsilon}_2/\bar{\epsilon}_T$. ***Right panel***: The values of $\bar{\epsilon}_T$, $\bar{\epsilon}_1$, and $\bar{\epsilon}_2$ plotted as a function of the evolution time T, with the y-axis showing $\bar{\epsilon} \times T^2$ to highlight the scaling behavior.

as close as possible to the original.

Figure 3.13 illustrates $f(t; k)$ for various values of k. Since the function is symmetric around $t = 1/2$, only the interval $0 \leq t \leq 1/2$ is shown. When k is small, $f(t; k)$ closely resembles $f_0(t)$, but with the key difference that its first derivatives at the endpoints vanish. To ensure the Hamiltonian path in the adiabatic limit remains close to that of $H_0(t)$, we restrict k to small values that minimally perturb the trajectory.

Figure 3.14 compares the actual average error $\bar{\epsilon}_T$ with the first-order and second-order switching theorem estimates $\bar{\epsilon}_1$, and $\bar{\epsilon}_2$, using $k = 10^{-3}$. The estimate $\bar{\epsilon}_1$ is calculated based on the original Hamiltonian $H_0(t)$, while $\bar{\epsilon}_2$ is based on the modified Hamiltonian $H(t)$. The left panel plots the ratios $\bar{\epsilon}_1/\bar{\epsilon}_T$ and $\bar{\epsilon}_2/\bar{\epsilon}_T$, and the right panel shows the rescaled errors $\bar{\epsilon} \cdot T^2$, making the long-time scaling behavior more apparent.

These plots highlight a transition in the accuracy of the switching theorem estimates: for small T, $\bar{\epsilon}_1$ more accurately tracks the true error, while in the large-T limit, the true error aligns with $\bar{\epsilon}_2$. In the left panel, we see that $\bar{\epsilon}_1/\bar{\epsilon}_T \approx 1$ at short timescales, signaling better agreement in this regime.

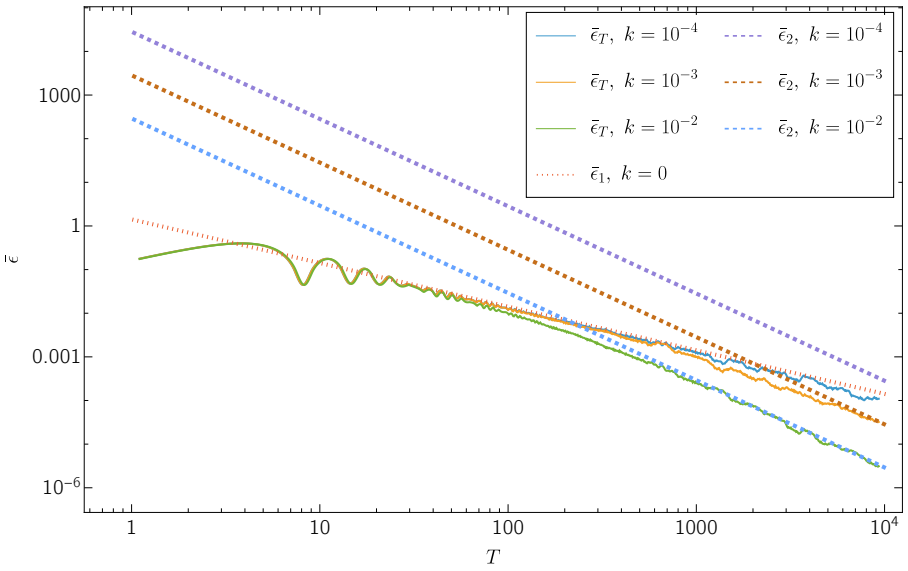

Figure 3.15: Comparison of average errors $\bar{\epsilon}_T$, $\bar{\epsilon}_1$, and $\bar{\epsilon}_2$ for two values of k, 10^{-4} and 10^{-3}, with $(E_0, E_1) = (1, 1)$. Note that the switching estimates $\bar{\epsilon}_1$ and $\bar{\epsilon}_1$ can exceed 1, as they are estimates, whereas the true average error $\bar{\epsilon}_T$, computed from Eq. (3.78), always remains below 1.

As T increases, this ratio drifts away from unity while $\bar{\epsilon}_2/\bar{\epsilon}_T$ approaches one, confirming that the second-order switching estimate captures the asymptotic scaling behavior more reliably. The right panel conveys the same transition from $1/T$ to $1/T^2$ scaling.

Figure 3.15 presents a broader comparison across different k values. Solid curves represent the numerically computed average errors $\bar{\epsilon}_T$, dotted lines show $\bar{\epsilon}_1$, and dashed lines denote $\bar{\epsilon}_2$. Across all values of k, the true error initially follows the first-order estimate $\bar{\epsilon}_1$, then transitions to the second-order scaling $\bar{\epsilon}_2$ in the hyperadiabatic regime. However, the onset of this asymptotic scaling depends sensitively on k. For each k, the true error exhibits $1/T$ scaling for moderate T, which matches the behavior predicted by $H_0(t)$. Only at significantly larger times $T \gg O(k^{-1}\Delta_{1,0}^{-1})$ does the error begin to reflect the improved $1/T^2$ scaling associated with the smoothed Hamiltonian $H(t)$.

Notably, for smaller k, the crossover from $\bar{\epsilon}_1$ to $\bar{\epsilon}_2$ occurs at large T, and the magnitude of the true error becomes smaller. This reflects the fact that as $k \to 0$, the modified Hamiltonian becomes more similar to $H_0(t)$, while introducing more abrupt behavior near the endpoints. Consequently, the

system requires longer times to enter the hyperadiabatic regime.

In a few words, these results confirm that scaling of errors in powers of $1/T$ by switching theorem is achieved at long times, but the timescale at which this regime sets in is controlled by the size of the endpoint derivatives. Therefore, achieving substantial error reductions at moderate times may require more than minimal endpoint modifications.

We have examined how a two-level system transitions to the asymptotic regime described by the switching theorem. We expect this qualitative behavior to extend to systems with multiple accessible excited states. To explore this, we conduct a similar analysis for a three-level system. The reference Hamiltonian $H_0(t)$, analogous to the two-level case, is defined as:

$$H_0(t) = \begin{bmatrix} 1 & E_1 f_0(t) & E_2 f_0(t) \\ E_1 f_0(t) & 2 & E_3 f_0(t) \\ E_2 f_0(t) & E_3 f_0(t) & 3 \end{bmatrix}, \tag{3.92}$$

where $f_0(t)$ is defined in Eq. (3.89).

Following the same strategy as before, we define a smoothed Hamiltonian with independent smoothing parameters:

$$H(t;\mathbf{k}) = \begin{bmatrix} 1 & E_1 f(t;k_1) & E_2 f(t;k_2) \\ E_1 f(t;k_1) & 2 & E_3 f(t;k_3) \\ E_2 f(t;k_2) & E_3 f(t;k_3) & 3 \end{bmatrix}, \tag{3.93}$$

where $f(t;k)$ is given in Eq. (3.91). As expected, $H(t,\mathbf{k}) \to H_0(t)$ as $\mathbf{k} = (k_1,k_2,k_3) \to \mathbf{0}$, and the time derivative of H at the endpoints vanishes when $\mathbf{k} \neq 0$, i.e., $\partial_t H(t_i,\mathbf{k}) = \partial_t H(t_f,\mathbf{k}) = 0$.

We consider two cases for $\mathbf{k}$ and $\mathbf{E} = (E_1, E_2, E_3)$:

- Case 1: $E_1 = E_2 = E_3 = 1$ and $\mathbf{k} = (k, k, k)$:

$$H(t; \mathbf{k}) = \begin{bmatrix} 1 & f(t;k) & f(t;k) \\ f(t;k) & 2 & f(t;k) \\ f(t;k) & f(t;k) & 3 \end{bmatrix}. \tag{3.94}$$

- Case 2 : $(E_1, E_2, E_3) = (1, 0, 1)$ and $\mathbf{k} = (k, 0, 0)$:

$$H(t; \mathbf{k}) = \begin{bmatrix} 1 & E_1 f(t;k_1) & 0 \\ E_1 f(t;k_1) & 2 & E_3 f_0(t) \\ 0 & E_3 f_0(t) & 3 \end{bmatrix}. \tag{3.95}$$

Case 1 can be seen as a natural extension of the earlier two-level example, where first-order diabatic transitions between any pair of states are suppressed by construction. Case 2, while reminiscent of the two-level setup, introduces new behavior: for $k \neq 0$, first-order transitions from the ground state to either excited state are eliminated, but transitions between excited states remain allowed at first order in $1/T$.

Importantly, because the switching theorem expressions, Eq. (3.79), only account for transitions from the ground state, and such transitions are absent at first order in this case, the leading-order nontrivial contribution to the switching error for the ground state is $\bar{\epsilon}_2$ when $k \neq 0$.

We start with Case 1, investigating how varying k influences the behavior of the true error relative to the switching error estimates, $\bar{\epsilon}_1$ and $\bar{\epsilon}_2$. These are computed as before using the reference Hamiltonian $H_0(t)$ and the modified Hamiltonian $H(t)$, respectively.

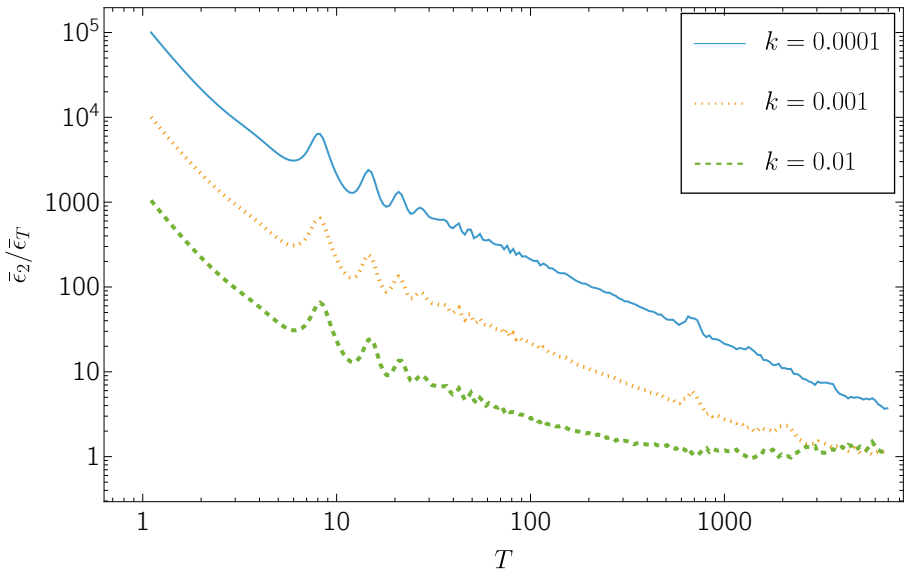

Figure 3.16: Comparison of the error ratio $\bar{\epsilon}_2/\bar{\epsilon}_T$for the three-level system with identical off-diagonal couplings, as defined in Eq. (3.94), for various values of k.

Figure 3.16 presents the ratio $\bar{\epsilon}_2/\bar{\epsilon}_T$ for several values of k. The results show that for larger k, the true error closely follows $\bar{\epsilon}_2$ even at smaller timescales. In contrast, for smaller k, the true error gradually converges to $\bar{\epsilon}_2$ as T increases, mirroring the behavior observed in Figure 3.15.

Turning to Case 2, we apply the same methodology as before. As shown in Figure 3.17, the true error aligns with $\bar{\epsilon}_1$ at small T, transitioning to follow $\bar{\epsilon}_2$ at larger T, consistent with the trends seen in earlier cases.

Finally, we compare the true errors of Case 1 and Case 2, fixing $k = k_1$, to understand how changes in the excited-state structure impact ground state preparation. Figure 3.18 demonstrates that in the hyperadiabatic regime, both cases exhibit nearly identical behavior. This suggests that modifying transitions between excited states has minimal influence on the asymptotic error scaling. Instead, the dominant contribution comes from transitions between the ground and first excited states, as predicted by Eq. (3.79). Nevertheless, the figure also indicates that at small T, other transitions can still play a significant role in determining the total error.

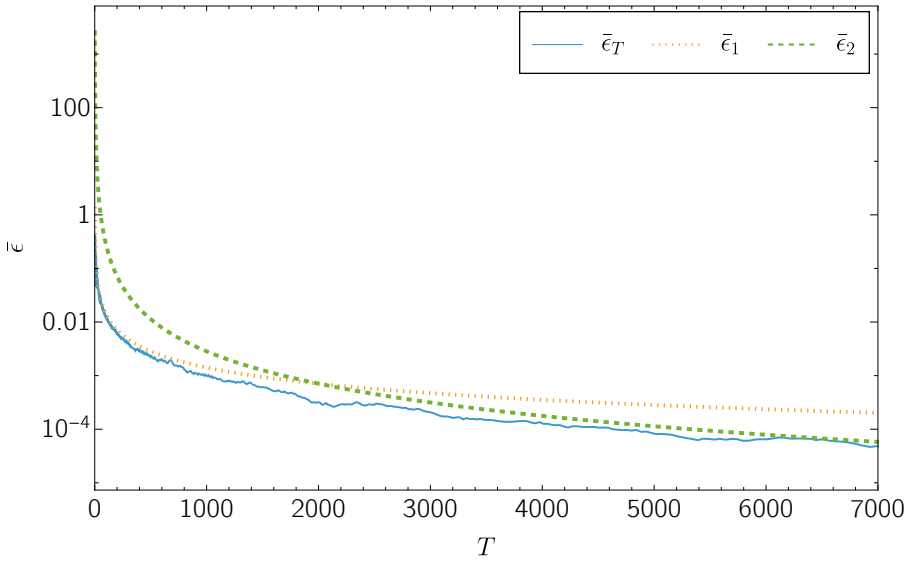

Figure 3.17: Average errors $\bar{\epsilon}_T$, $\bar{\epsilon}_1$, and $\bar{\epsilon}_2$ for Case 2 with $k = 10^{-3}$. For the given Hamiltonian parameters, the first-order switching error $\bar{\epsilon}_1$ vanishes; thus the curve labeled $\bar{\epsilon}_1$ is computed using the unmodified Hamiltonian $H_0(t)$ as defined in Eq. (3.92)

3.5.4.3 Systems with essential singularities at the endpoints

In earlier examples, we modified the Hamiltonians so that their first derivatives vanished at the endpoints, while leaving the intermediate evolution largely unchanged. In the hyperadiabatic regime, this led to the expected T^{-2} scaling of the error. For small values of T, agreement with this asymptotic scaling improved as differences between the modified and original Hamiltonians became negligible during the bulk of the evolution. This approach can be extended further by ensuring that higher-order derivatives also vanish at the endpoints. The asymptotic scaling behavior suggests that the error decays more rapidly—with a more negative power of T—as the order of the lowest non-vanishing derivative at the endpoints increases.

A particularly extreme endpoint modification involves setting all time derivatives to zero, corresponding mathematically to the introduction of an essential singularity. One such construction for the

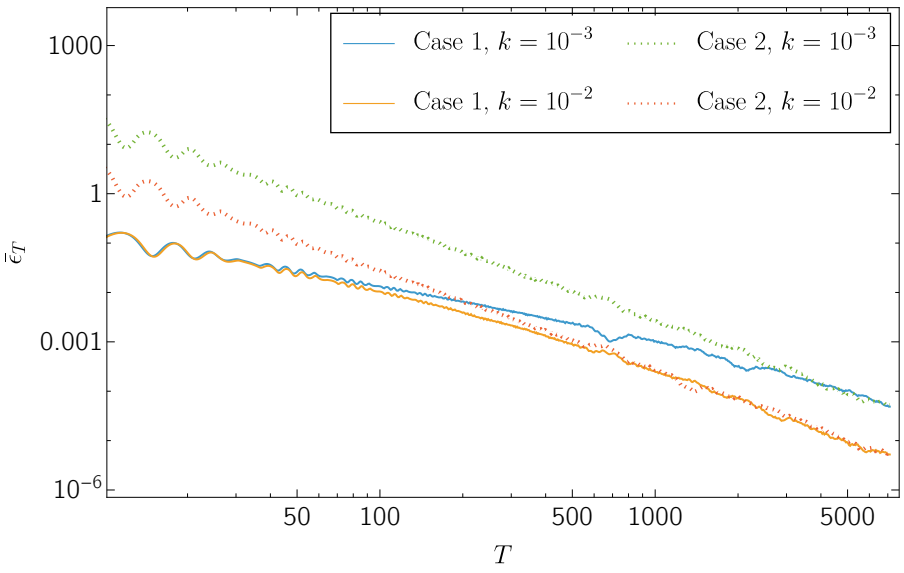

Figure 3.18: Comparison of Case 1 and Case 2, as defined in Eqs. (3.94) and(3.95). Solid lines correspond to Case 1, and dotted lines correspond to Case 2.

two-level Hamiltonian is given by:

$$H(t;k) = \begin{bmatrix} 0 & E_1 f_{\text{exp}}(t;k) \\ E_1 f_{\text{exp}}(t;k) & E_0 \end{bmatrix},$$
$$\text{with } f_{\text{exp}}(t;k) = \begin{cases} t(1-t)\exp\left(-\frac{k}{t(1-t)}\right), & \text{if } 0 \le t \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.96}$$

As $k \to 0$, the difference between $H(t;k)$ in Eq. (3.96) and $H_0(t)$ in Eq. (3.88) becomes negligible at intermediate times. However, since all derivatives of $H(t;k)$ vanish at the endpoints for $k \neq 0$, the switching theorem is no longer applicable for estimating the error. One can verify directly that all time derivatives vanish at the endpoints. Nevertheless, the structure of the switching theorem motivates this kind of endpoint modification, especially when $f(t;k)$ is viewed as the limit of analytic functions with well-defined endpoint derivatives.

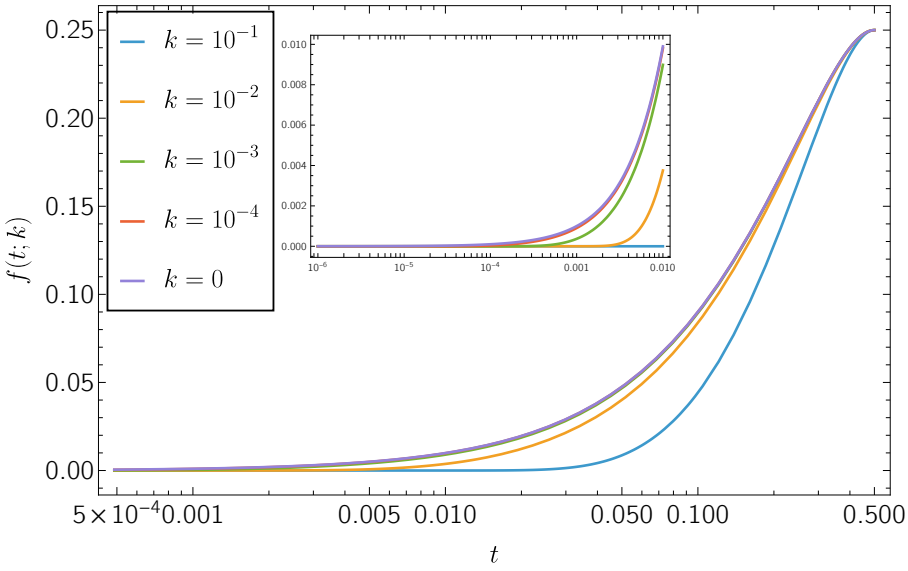

Figure 3.19: Plots of $f_{\text{exp}}(t;k)$ for different values of k. When $k = 0$, the function reduces to $f_0(t)$, which has nonzero first-order derivatives at the endpoints. The inset highlights the behavior of $f_{\text{exp}}(t;k)$ near the endpoint.

In Figure 3.19, the function $f_{\text{exp}}(t;k)$ is compared with $f_0(t)$ from Eq. (3.89). We then study the errors produced by evolving with the Hamiltonian in Eq. (3.96) for $k = 10^{-4}$, 10^{-3}, 10^{-2}, and compare them with the errors from evolving with $H_0(t)$ in Eq. (3.88).

As in earlier cases, when k is small, the Hamiltonian closely resembles Eq. (3.88), so we compare its error to $\bar{\epsilon}_1$ computed from the $k = 0$ Hamiltonian. These results are shown in Figure 3.20[3]. Here, we set $E_0 = E_1 = 1$ as in previous two-level examples.

Figure 3.20 shows that although all derivatives of the Hamiltonian vanish, for small T, the true error still follows the estimate from the $k = 0$ case. As T increases, the error eventually transitions to an asymptotic regime, decaying faster than any power law. Similar to previous observations, decreasing k delays the onset of this exponential decay. Notably, for small k and moderate T, $\bar{\epsilon}_T$ does not yet exhibit clear signs of exponential behavior.

[3]The Schrödinger equation is solved from $t = 10^{-6}$ to $t = 1 - 10^{-6}$ rather than from 0 to 1, to avoid numerical issues due to the singularity at the endpoints.

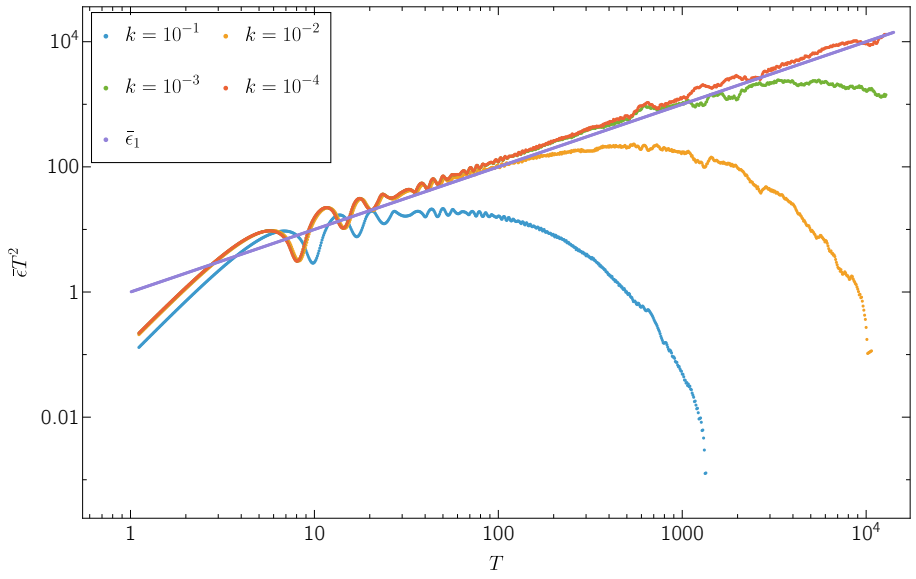

Figure 3.20: Average true errors are shown for different values of k as defined in Eq. (3.96), with $\bar{\epsilon}_1$ computed using the Hamiltonian corresponding to $k = 0$.

3.5.4.4 Discussion

The practical value of the asymptotic error scaling predicted by the switching theorem hinges on the timescale at which it becomes a good approximation of the true error. We explored this by comparing switching estimates with numerically computed errors for several model Hamiltonians. We observed that the agreement between the true error and the switching prediction improves with increasing evolution time—i.e., deep in the hyperadiabatic regime.

The switching theorem asserts that for sufficiently large evolution times T, the diabatic error scales with a characteristic power of $1/T$, determined entirely by the behavior of the Hamiltonian at the endpoints. This naturally leads to the question: can one reduce errors by only modifying the Hamiltonian's time derivatives at the endpoints, while leaving the intermediate evolution largely unchanged? Mathematically, this means enforcing the vanishing of a finite number of time derivatives at the endpoints. We investigated this idea using simple two- and three-level systems, where these effects can be

studied in detail. We confirmed that the switching theorem accurately predicts the asymptotic scaling, but the onset of that regime—i.e., when the approximation becomes accurate—strongly depends on how smooth the endpoints are.

Our results show that the usefulness of the switching theorem for practical adiabatic state preparation is not guaranteed. It is most helpful when the leading nonzero term in the asymptotic expansion provides a good estimate of the error even at relatively short times. However, if that regime is only reached at very large T, or if the corresponding error is already negligible, then the switching theorem offers limited practical guidance. In other words, for the switching theorem to be a useful predictive tool, the onset of asymptotic scaling must occur early enough—and with sufficiently large errors—that scaling behavior remains observable.

Consider a scenario where one employs a fault-tolerant analog quantum simulator, and Trotter errors are not a concern. Even then, applying the switching theorem requires precise knowledge and control of the Hamiltonian near its endpoints. In practice, many adiabatic protocols involve abrupt switching at the start and end of the evolution, which can be difficult to control and may introduce large diabatic errors not captured by the switching estimate. Our results show that deliberately modifying the endpoint behavior—while keeping the bulk of the evolution intact—can lead to error suppression, thus making strategic use of the switching theorem.

In contrast, digital quantum simulations must also contend with Trotterization errors, which complicate matters further. In this context, the clean asymptotic scaling behavior described by the switching theorem may no longer be directly applicable. Understanding how Trotter errors interact with the switching behavior remains an open and important problem.

Throughout this work, we focused on small modifications to the Hamiltonian, parameterized by a small auxiliary parameter k, that smooth out the time evolution near the endpoints without altering

the global path. This reflects a common scenario in which the Hamiltonian path is largely fixed, either by physical constraints or prior design.

Of course, many different paths could in principle interpolate between the same initial and final Hamiltonians. But not all of them are suitable for adiabatic preparation—for instance, paths that pass through quantum phase transitions are likely to fail [257–260]. This motivates our focus on modifying only the endpoints of an otherwise fixed, physically reasonable path. Throughout this thesis, our focus is not finding the best path, but how to travel the given path with less computational cost.

Given the cost of reaching very small diabatic errors via pure adiabatic evolution, a more efficient strategy in practice may be to prepare an approximate ground state adiabatically, and then refine it using post-processing techniques. Projection-based algorithms—such as those discussed in Refs. [150, 209–213] and the rodeo algorithm mentioned earlier—can then be used to project onto the true ground state. In such hybrid approaches, the asymptotic scaling predicted by the switching theorem may be of limited relevance, since the diabatic errors are already negligible by the time this regime is reached.

3.5.5 Adiabatic state preparation for long path

This subsection focuses on a different direction of errors: how they relate to the overall computational cost and the length of the path in parameter space, where the parameter determines the Hamiltonian. The objective here is to find a dimensionless measure that serves as a proxy for computational cost—something more insightful than simply relying on the total evolution time—particularly in the regime of long path lengths.

Although adiabatic evolution can be implemented on digital quantum computers [195, 196, 199, 253, 261–263], analog quantum simulators offer a more natural setting to study such dynamics. These

platforms avoid complications arising from Trotterization errors [193, 194], which simplifies analysis. Nevertheless, the ideas discussed here can be extended to digital setups by incorporating the cost associated with Trotter errors, but the central cost metrics considered in this subsection are more general.

Understanding the typical behavior in the long path length limit is especially relevant, since many important physical problems—including gapped lattice quantum field theories—fall into this category. One of the overarching goals of this work is to characterize how the computational effort scales with path length when the error is kept small. While it might seem natural to take the total runtime as a proxy for computational complexity, we argue that this is misleading. A crucial insight is a kind of no-go theorem for time-based scaling: for any proposed scaling relation between time and path length at fixed error, there exists a Hamiltonian that violates it. In other words, one can always find a different Hamiltonian evolution that achieves the same final error but with a runtime that scales with path length in any arbitrary way. As a result, time alone is not a robust indicator of computational cost in adiabatic evolution.

To address this, we propose a dimensionless measure that captures how "adiabatic" the traversal of the Hamiltonian path is, and suggest using this as a more reliable indicator of cost. There is also an apparent paradox: shorter paths seem to produce smaller errors for a given runtime, since the state changes less and encounters fewer opportunities for transitions. However, this intuition can conflict with implications of the switching theorem. In fact, path length does matter, especially in scenarios where all terms in the switching expansion vanish and the remaining error is captured entirely by the remainder term. If we are interested in how long it takes to achieve a small fixed error, and that time grows as the path length L increases, then it becomes crucial to understand how the expansion coefficients b_n scale with L.

To formalize the discussion, we define a "velocity" of the path through parameter space. The instantaneous velocity at a given parameter value, $v(\lambda)$, and the average velocity over the entire path, $\overline{v}$, are given by:

$$v(\lambda) \equiv \frac{d\lambda}{dt} \quad \text{and} \quad \overline{v} \equiv \frac{L}{T}, \tag{3.97}$$

where $T = t_f - t_i$ is the total evolution time, and L is the path length. Since L is dimensionless, $v(\lambda)$ has units of inverse time. We can factor out the average velocity $\overline{v}$ as a scaling parameter and consider the limit $\overline{v} \to 0$, while keeping the shape of the trajectory fixed—this means the local velocity profile $v(\lambda)$ scales with $\overline{v}$ but retains its relative variation with λ.

3.5.5.1 A counterexample for scaling of time in path length

For sufficiently long paths, the Zeno-inspired method discussed in Section 3.4 offers highly efficient scaling, with the required computational resources increasing only linearly with the path length. This raises the question of whether adiabatic evolution can achieve comparable scaling in the long-path regime.

The importance of path length in the state preparation along a path in the Hamiltonian space has been recognized for over a decade, with several insightful papers emphasizing its role [201–203, 208, 264, 265]. The most assertive claim appears in [208], which presents an argument rooted in computational complexity, leveraging Grover's algorithm [179]. The abstract of that work asserts that, for processes with a constant spectral gap, the minimum time required to approximately prepare the final eigenstate scales proportionally to the ratio of the path length to the gap. This suggests that no adiabatic protocol can beat this lower bound. Similar assertions appear throughout the paper. However, the soundness of this claim has been questioned [266], and its interpretation requires care.

At first glance, the claim might be taken to imply that any Hamiltonian path with a fixed gap must require time scaling linearly with path length in order to maintain a given level of accuracy. However, this interpretation does not hold up when one examines the methods used in Ref. [208]. In fact, one can easily construct explicit counterexamples that defy this supposed lower bound.

Take, for example, a simple three-level system with a time-dependent Hamiltonian:

$$H(t) = \Delta \left(1 + \frac{t}{\tau}\right) X^\dagger H_D X, \tag{3.98}$$

where

$$H_D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{1+\frac{t}{\tau}} & 0 \\ 0 & 0 & 2 \end{bmatrix} \text{ and } X = \begin{bmatrix} \cos\left(\frac{t}{\tau} + \frac{t^2}{2\tau^2}\right) & 0 & -\sin\left(\frac{t}{\tau} + \frac{t^2}{2\tau^2}\right) \\ 0 & 1 & 0 \\ \sin\left(\frac{t}{\tau} + \frac{t^2}{2\tau^2}\right) & 0 & \cos\left(\frac{t}{\tau} + \frac{t^2}{2\tau^2}\right) \end{bmatrix}. \tag{3.99}$$

Here, Δ and τ are constants representing energy and time scales, respectively. The energy gap remains constant at Δ throughout the evolution. If the system starts in the ground state at $t = 0$, the path length up to time t is:

$$L(t) = \frac{t}{\tau} + \frac{t^2}{2\tau^2}. \tag{3.100}$$

Meanwhile, the error ϵ (as defined in Eq. (3.70)) reaches a maximum value of

$$\epsilon_{\max} = \frac{1}{\sqrt{1 + \Delta^2 \tau^2}}, \tag{3.101}$$

and repeatedly returns to this value as $L \to \infty$. That is, we can analyze the error at points where it has reached or nearly reached this maximum. Solving for the time as a function of path length, while

keeping the gap and error fixed, we find:

$$t(L) = \sqrt{2L\tau^2}\left(\sqrt{1+(2L)^{-1}} - \sqrt{(2L)^{-1}}\right) \xrightarrow{L\to\infty} \sqrt{2L\tau^2}. \tag{3.102}$$

This implies that the evolution time grows with $\sqrt{L}$ rather than linearly, directly contradicting the apparent claim from Ref. [208] that the time must scale at least linearly with L at fixed error and gap.

This example highlights a broader point: the error accumulated during the adiabatic traversal of a path depends both on the local "velocity" of the Hamiltonian change and on the relevant energy scales driving transitions. Since time carries units, a dimensional quantity must set the scale, and these energy differences do just that. The analysis in Ref. [208] implicitly assumes that the only relevant energy scale is the fixed gap Δ, but this is not generally valid. As shown above, one can construct Hamiltonians where the characteristic energy differences evolve along with the velocity in a controlled way, yielding arbitrary scaling of runtime with L while maintaining fixed error.

In fact, it is possible to engineer systems where different Hamiltonians produce mathematically equivalent dynamics, simply by adjusting the local velocity and energy scales together. We will further demonstrate how such constructions allow for arbitrary time scalings at fixed error, independent of gap. Importantly, in many such systems, the gap remains constant while the other energy scales vary, so the gap alone is insufficient to set the scale of the evolution.

The takeaway is that the time-to-error scaling with path length, by itself, is not a reliable indicator of computational difficulty. Instead, it is necessary to develop more meaningful, dimensionless metrics to assess the cost of state preparation via adiabatic evolution.

3.5.5.2 Insight from time-periodic Hamiltonians

We consider the regime where the adiabatic error ϵ is small. Since the time-evolution operator is unitary, it can be expressed as $U(t_f, t_i) = \exp(iA_{f,i})$, where $A_{f,i}$ is a Hermitian operator. When ϵ is small, it is sufficient to expand the exponential to first order in A to estimate the error via Eq. (3.70), leading to the approximation:

$$\epsilon = \left\| (1 - |g_f\rangle\langle g_f|)\, A_{f,i} |g_i\rangle \right\|. \tag{3.103}$$

Suppose the full evolution along a long path is decomposed into N shorter segments, each with a small error. If

$$U(t_f, t_i) = U(t_f, t_{N-1}) U(t_{N-1}, t_{N-2}) \cdots U(t_2, t_1) U(t_1, t_i), \tag{3.104}$$

then the total error becomes

$$\epsilon = \left\| (1 - |g_f\rangle\langle g_f|) \sum_{j=0}^{N-1} A_{j+1,j} |g_i\rangle \right\|. \tag{3.105}$$

Now consider the special case of a time-periodic Hamiltonian, $H(t+\tau) = H(t)$, and an evolution that spans exactly N periods, starting in the ground state. Let L_{sc} denote the path length for a single cycle, so that the total length is $L = NL_{\text{sc}}$. Assume the trajectory is nontrivial—that is, it does not simply reverse itself halfway through a cycle.

Divide the evolution into N segments, each corresponding to one cycle: $U(t_f, t_i) = U_0^N$ with $U_0 = \exp(iA_0)$ being the single-cycle evolution operator. Then, from Eq. (3.105), and assuming small error, we find

$$\epsilon = N\epsilon_{\text{sc}} \text{ where } \epsilon_{\text{sc}} = \left\| (1 - |g_0\rangle\langle g_0|)\, A_0 |g_0\rangle \right\|, \tag{3.106}$$

where $|g_0\rangle$ is the ground state at both the beginning and end of a cycle.

Because each cycle starts and ends at the same point, all coefficients b_n in the asymptotic series of Eq. (3.73) vanish, and the single-cycle error is determined entirely by the remainder term: $\epsilon_{\mathrm{sc}} = r_{\mathrm{sc}}$, which decays faster than any power law in $1/\overline{v}_0 = T_0/L_0$. Thus, we obtain:

$$\epsilon = N \exp\left(-f(1/\overline{v}_{\mathrm{sc}})\right) = \frac{L}{L_{\mathrm{sc}}} \exp\left(-f(1/\overline{v})\right), \tag{3.107}$$

where f is a positive monotonic function satisfying $\log(x) = o(f(x))$—that is, it grows faster than logarithmically.

Solving for the time T, we find:

$$T = L\, f^{-1}\left(\log\left(\frac{L/L_{\mathrm{sc}}}{\epsilon}\right)\right). \tag{3.108}$$

While we do not know the precise form of f, if $f(x) = x$ then T scales as $L \log(\frac{L/L_0}{\epsilon})$, which grows faster than linearly with L. More generally, this result implies that, to keep the error fixed, the total time must grow faster than linearly (superlinearly) with the path length.

The core intuition behind this result is simple: as the system evolves over many cycles at a fixed average speed, small errors from each cycle accumulate. Maintaining a fixed total error thus requires reducing the error per cycle, which entails slowing down the evolution—hence increasing the total time.

While this argument is grounded in the periodic case, the underlying principle—that errors accumulate along the path—should apply more broadly. Therefore, one generally expects that maintaining a fixed error leads to computational cost growing faster than linearly in L. However, the three-level

model in Eq. (3.98) demonstrates that in some cases, the required time can grow more slowly than L. This tension points to the need for a more refined understanding that generalizes the periodic result to broader classes of systems, while incorporating the insight from such models that time can sometimes scale sublinearly with L for fixed error.

3.5.5.3 A no-go theorem for the relation between velocities and errors

The behavior seen in the model of Eq. (3.98) illustrates a general no-go theorem: for any proposed bound on how the total time T must scale with path length L at fixed small error, one can construct a time-dependent Hamiltonian that violates it. This means that a universal bound on $T(L)$ at fixed ϵ does not exist. However, this does not invalidate the intuition that computational costs tend to grow faster than L—it simply implies that time alone is not a reliable proxy for computational cost.

To build intuition for this result, consider first a trivial observation: if one rescales a Hamiltonian by a constant factor along a path, the evolution can be sped up by the same factor without changing the error. Consequently, reducing the total runtime proportionally preserves the final state. But this rescaling does not change how T scales with L, which is our primary interest. A more insightful case is when the Hamiltonian is rescaled by a function that varies along the path—this changes the relationship between L and T while keeping the error fixed, and leads directly to the no-go result.

The technical proof begins with the time-dependent Schrödinger equation, $H|\psi\rangle = i\frac{d}{dt}|\psi\rangle$, and introduces a unitary transformation X that locally diagonalizes the Hamiltonian:

$$H = X^\dagger H_d X, \tag{3.109}$$

where H_d is the instantaneous diagonal Hamiltonian, and X depends on λ (implicitly time-dependent).

Redefining the state as $|\phi\rangle = X|\psi\rangle$, the Schrödinger equation becomes

$$\left(H_d + i\dot{X}X^\dagger\right)|\phi\rangle = i\frac{d}{dt}|\phi\rangle. \tag{3.110}$$

Expressed in terms of λ (with $d/dt = vd/d\lambda$), this becomes

$$\left(\frac{H_d}{v} + iX'X^\dagger\right)|\phi\rangle = i\frac{d}{d\lambda}|\phi\rangle. \tag{3.111}$$

Now, consider two Hamiltonians, $H_A(\lambda)$ and $H_B(\lambda)$, corresponding to two different traversal schedules $l_A(t)$ and $l_B(t)$ of the same path. Define their velocities $v_{A,B}(\lambda) = \left.\frac{dl_{A,B}(t)}{dt}\right|_{t=l^{-1}_{A,B}(\lambda)}$, and suppose they are related by

$$H_B(\lambda) = \frac{H_A(\lambda)v_B(\lambda)}{v_A(\lambda)}, \tag{3.112}$$

Then, the transformed states $|\phi_A(\lambda)\rangle$ and $|\phi_B(\lambda)\rangle$ evolve identically for all λ if they start from the same initial condition, implying the physical states $|\psi_A(t)\rangle$ and $|\psi_B(t)\rangle$ coincide at matching points along the path. Thus, the errors are identical, even though the total evolution times differ:

$$T_{A,B} = \int_{\lambda_i}^{\lambda_f} d\lambda \frac{1}{v_{A,B}}. \tag{3.113}$$

This gives

$$T_B = T_A\frac{\overline{v}_A}{\overline{v}_B} \quad \text{with} \quad \overline{v}_{A,B} = \frac{L}{\int_{\lambda_i}^{\lambda_f}\frac{d\lambda}{v_{A,B}}}, \tag{3.114}$$

where $\overline{v}_{A,B}$ are the average velocities.

It follows that by choosing $v_B(\lambda)$ to grow sufficiently quickly with λ, one can make T_B scale with L arbitrarily more slowly than T_A while keeping the error fixed. For example, setting $v_B(\lambda) =$

$v_A(\lambda)(1+\lambda^a)$ and $H_B(\lambda) = H_A(\lambda)(1+\lambda^a)$ with $a > 1$ yields

$$\frac{T_B}{T_A} \sim \lambda^{-a}, \tag{3.115}$$

so for large enough a, even if T_A scales as a power of λ, T_B can scale more slowly than any fixed power, including linearly. This shows that no general lower bound on the scaling of T with λ can exist at fixed error.

The construction can also be adapted to preserve a constant or bounded spectral gap, as considered in Ref. [208]. This is done by extending system B's Hilbert space to include two components: $H = H_{B,1} \oplus H_{B,2}$, where $H_{B,1}$ behaves as described above and $H_{B,2}$ is time-independent and has a fixed energy gap above the ground state. As long as $H_{B,1}$ starts with a larger gap than $H_{B,2}$ and its velocity grows monotonically, this guarantees a controlled gap throughout the trajectory.

3.5.5.4 General case

The no-go theorem should not be interpreted as implying that the computational cost of adiabatic state preparation fails to grow with the path length λ. Rather, it points to the fact that evolution time is simply not a reliable proxy for computational cost when analyzing how cost scales with λ at fixed error. This issue has been understood for some time—see, for example, Ref. [190]—where it is emphasized that time, being a dimensionful quantity, cannot on its own capture meaningful scaling behavior in terms of dimensionless parameters like the error ϵ or the path length L. To relate time to computational difficulty, one needs an additional scale-setting quantity.

One candidate proposed in Ref. [190] is the product $T \cdot \max\|H\|$, where T is the total evolution time and $\max\|H\|$ refers to the maximum operator norm of the Hamiltonian along the path. This

quantity has the correct units to be compared to error scaling, and intuitively it incorporates both time and energy scale. However, it has a significant shortcoming: $\max\|H\|$ reflects the largest eigenvalues of the Hamiltonian, which may be irrelevant to the evolution of the ground state and thus not an accurate reflection of actual computational cost in adiabatic state preparation.

This leads naturally to the question: is there a quantity that better captures the genuine computational difficulty of adiabatic state preparation? Let us denote such a hypothetical quantity by Q_D. To serve as a meaningful proxy for computational cost, and to potentially admit general bounds on its scaling with L, Q_D should satisfy several reasonable criteria:

1. **Superlinear scaling in time-periodic systems**:

   For systems where the Hamiltonian is periodic in time, the intuition is that errors accumulate over many cycles. Holding the total error fixed therefore requires reducing the error per cycle as the number of cycles grows, which in turn necessitates slower evolution—i.e., longer total time. A good proxy for cost should reflect this by growing superlinearly with L in such settings.

2. **Dimensionlessness**:

   Q_D should be a dimensionless quantity. This ensures that it characterizes computational difficulty independently of physical units, and hence independently of specific hardware or implementation details. It would also put it on the same footing as the path length L and the target error ϵ.

3. **Invariance under reparameterization**:

   As shown earlier, two different Hamiltonians related by a time-rescaling transformation (i.e., Eq. (3.112)) generate the same state trajectory at different physical speeds. Thus, they should be regarded as equally difficult in any abstract, model-independent sense. A proper measure of

computational cost should assign the same value of Q_D to both systems.

4. **Sensitivity to adiabaticity**:

   Q_D should track how close the evolution is to the adiabatic limit. This is essential, since staying near the adiabatic limit not only increases computational cost (e.g., by requiring slower evolution) but is also the central criterion for maintaining low error in adiabatic algorithms.

This framework makes clear why traversal time itself fails as a universal cost proxy: the no-go theorem shows that time can be rescaled arbitrarily without affecting the physical error, simply by appropriately modifying the Hamiltonian's amplitude as a function of position along the path. This destroys any general relationship between time and computational difficulty. However, that does not invalidate the original intuition that computational resources should grow with path length—it simply means that time is the wrong variable to express this intuition.

The situation is different for a quantity like Q_D, which, by construction, might retain this expected relationship even in cases where time does not. Specifically, one can ask whether for a wide class of systems—perhaps excluding only a narrow, well-characterized subclass— Q_D must scale faster than linearly with L at fixed error. At present, no general proof exists for such a statement, but this motivates formulating and testing various conjectures.

Before doing so, however, it is essential to clarify the structure of such conjectures. The general setup involves considering how the asymptotic scaling of Q_D with path length L behaves when the error ϵ is held fixed and small. This setup allows the velocity along the path to be adjusted to maintain the error at the desired value while letting L vary, isolating the scaling behavior of the cost measure from other complicating factors. The key idea is that all other relevant properties (e.g., the Hamiltonian structure, the gap profile, etc.) are held fixed while L increases.

The aim is to rescue and reformulate the original intuition about adiabatic cost scaling—namely, that errors accumulate with distance and therefore resource demands should increase accordingly—in a framework where it can be meaningfully analyzed. The no-go theorem rules this out for time, but the possibility remains open for better-behaved cost measures like Q_D, which may preserve the spirit of this intuition under broad conditions.

Consider a quantum system with a fixed Hilbert space and a Hamiltonian that depends smoothly on a single real parameter, which we denote by λ. Without loss of generality, let the evolution start at $\lambda = 0$ when $t = 0$, and define the system's path through parameter space via a differentiable and non-decreasing function $\lambda(t)$, i.e., with $\lambda'(t) \geq 0$. This monotonicity guarantees that $\lambda(t)$ can be inverted to yield $t(\lambda)$, allowing one to express the instantaneous velocity along the path as a function of λ:

$$v(\lambda) = \frac{1}{t'(\lambda)}. \tag{3.116}$$

By construction, $v(\lambda)$ is continuous and non-negative. Assume also that $H(\lambda)$ is defined for arbitrarily large values of λ, ensuring that one can consistently take the limit of large path length L.

To fully specify a trajectory, one must give not only the path $H(\lambda)$ but also the velocity profile $v(\lambda)$ and the final point $\lambda_f = L$, with the convention that the path starts at $\lambda_i = 0$, so $L = \lambda_f - \lambda_i$. Suppose we are interested in a family of trajectories that all follow the same path and have the same relative velocity along the path, but differ in their overall speed. Such a family can be characterized by a reference velocity $v_{\mathrm{ref}}(\lambda)$ and a scaling parameter s_c, so that:

$$v(\lambda) = \frac{v_{\mathrm{ref}}(\lambda)}{s_c}. \tag{3.117}$$

Any property of the evolution—such as the final error—can then be viewed as depending on this set of inputs: the Hamiltonian path $H(\lambda)$, the reference velocity $v_{\text{ref}}(\lambda)$, the scaling factor s_c, and the path length L. More generally, a quantity associated with the evolution can be denoted as:

$$Q(s_c, L; v_{\text{ref}}, H), \tag{3.118}$$

indicating that it is a function of s_c and L, and a functional of $v_{\text{ref}}(\lambda)$ and $H(\lambda)$. In particular, the error $\epsilon(s_c, L; v_{\text{ref}}, H)$ can be viewed this way. Holding L, v_{ref}, and H fixed, the error becomes a function of the scaling parameter s_c. Although this dependence is not necessarily invertible—multiple values of s_c might yield the same error—one can define $s_c(\epsilon, L; v_{\text{ref}}, H)$ as the *minimum* value of s_c needed to achieve a given error ϵ, for fixed L, v_{ref}, and H. in the regime of small error, this minimal s_c is typically unique. With this, any trajectory-based quantity Q can be re-expressed as a function of ϵ and L, and as a functional of v_{ref} and H.

The conjectures of interest concern a quantity Q_D, representing some measure of computational difficulty associated with the full trajectory. The basic conjecture is that, for a given Hamiltonian path $H(\lambda)$, and ixed error and velocity profile, this quantity grows faster than linearly with the path length:

Conjecture: Q_D grows superlinearly in L at fixed small ϵ.

To formulate such a conjecture precisely, one must first specify the choice of Q_D and then identify the domain over which the conjecture is expected to hold. Depending on these assumptions, one can define various versions of the basic conjecture, with different degrees of generality:

1. **Typical conjecture**:

The conjecture holds for a broad class of physically relevant systems with high probability, though isolated counterexamples may exist.

2. **Zero-measure exception conjecture**:

   The conjecture holds for all systems and trajectories except for a set of measure zero. This interpretation presumes that one can meaningfully sample the space of Hamiltonians and reference trajectories and that counterexamples are vanishingly rare in that measure.

3. **Criterion-based conjecture**:

   There exists a clearly defined, computable condition such that if the condition holds, the conjecture is guaranteed to be true. Variants of this scenario include:

   (a) The conjecture fails if the criterion fails.

   (b) The conjecture may still hold even if the criterion fails.

   (c) The conjecture holds with high probability for physically relevant systems, even when the criterion fails.

   (d) The conjecture fails only on a set of measure zero when the criterion fails, under the assumption that the space of systems is properly enumerable.

One could also imagine an even stronger version of the conjecture in which every sensible definition of Q_D satisfies the conjecture for all systems and trajectories. However, this possibility can likely be excluded. For instance, in certain cases, such as when a time-dependent system reduces to one with a time-independent effective Hamiltonian $H_d + i\dot{X}X^\dagger$, well-behaved definitions of Q_D may scale only linear with L at fixed error, and not superlinearly. Thus, the most credible versions of the conjecture are likely those aligned with Scenario 3, involving criteria that imply (but do not fully characterize)

the conjecture's validity.

Still, a distinct and potentially more universal conjecture might hold: that Q_D grows at least linearly (rather than superlinearly) with L for all systems and all trajectories when the error is small and held fixed. While this form is weaker in the sense that it allows linear growth, it is stronger in that it is asserted to hold universally, without exception.

As will be discussed next, it is possible to identify quantities satisfying the four conditions outlined earlier—conditions that make a cost measure both physically meaningful and technically robust—and there is good reason to believe that at least some formulation of the conjectures above should hold for these choices of Q_D.

3.5.5.5 A new measure for quantifying adiabaticity

Although total evolution time has been shown to be an inadequate measure of computational difficulty, it may have initially appeared reasonable because of its connection to adiabaticity: all else being equal, longer evolution times typically correspond to more adiabatic behavior. However, as the no-go theorem makes clear, all else is rarely equal—so relying on total time alone is insufficient. The challenge, then, is to identify a quantity that more faithfully captures adiabaticity and does so in a way that satisfies Conditions 1–4.

To this end, it helps to revisit how adiabatic evolution actually works. Consider the Scrödinger equation parameterized by λ, Eq. (3.111). In the regime of small error, the solution leads to the following expression for the transition error:

$$\epsilon = \left\| \sum_{k \neq g} \int_{\lambda_i}^{\lambda_f} d\lambda \, \exp\left(i \int_{\lambda_i}^{\lambda} d\tilde{\lambda} \frac{E_k(\tilde{\lambda}) - E_g(\tilde{\lambda})}{v(\tilde{\lambda})} \right) \langle k(\lambda) | X' X^\dagger | g(\lambda) \rangle \right\|. \tag{3.119}$$

Here, the summation over $k \neq g$ accounts for contributions from all excited states. The key point is that adiabatic behavior—manifesting as small error—arises from rapid oscillations in the phase due to energy gaps, which lead to cancellations over the timescale of the Hamiltonian time evolution.

This naturally suggests a link between adiabaticity and computational cost. For instance, on a digital quantum computer, simulating nearly adiabatic dynamics with discrete time steps requires resolving these fast phase oscillations. Thus, higher adiabaticity implies a need for finer resolution—and therefore more computational resources.

To quantify how adiabatic the evolution is at a given point λ, consider the ratio:

$$a_k(\lambda) = \frac{E_k(\lambda) - E_g(\lambda)}{v(\lambda)}, \tag{3.120}$$

which governs the suppression of transitions from the ground state to the kth excited state near λ. To measure the overall suppression of such transitions at a particular λ, one can take a weighted average of the $a_k(\lambda)$, with weights $w_k(\lambda)$ that reflect how strongly the transition-driving operator $iX'X^\dagger$ couples the ground state to the kth excited state. Given the definition of λ in Eq. (3.45), the completeness relation ensures that

$$\sum_k |\langle k(\lambda)|g'(\lambda)\rangle|^2 = 1, \tag{3.121}$$

so that the natural choice for $w_k(\lambda)$ is

$$w_k(\lambda) = |\langle k(\lambda)|g'(\lambda)\rangle|^2. \tag{3.122}$$

This leads to the following average quantity:

$$a(\lambda) = \langle g'| \frac{H(\lambda) - E_g(\lambda)}{v(\lambda)} |g'\rangle, \tag{3.123}$$

which serves as a reasonable local measure of adiabaticity at each point along the trajectory.

In this context, it is crucial to define a global measure of adiabaticity associated with traversing a path. Denote this quantity as Q_D. A natural definition is to integrate the local adiabaticity $a(\lambda)$ over the entire path:

$$Q_D \equiv \int_{\lambda_i}^{\lambda_f} d\lambda \, \langle g'| \frac{H(\lambda) - E_g(\lambda)}{v(\lambda)} |g'\rangle. \tag{3.124}$$

This definition aligns with the intuition: it captures the cumulative suppression of non-adiabatic transitions over the course of the evolution.

Given that for any excited state k, the energy difference satisfies $E_k(\lambda) - E_g(\lambda) \geq \Delta(\lambda)$, where $\Delta(\lambda)$ is the instantaneous spectral gap, and that the gap obeys $\Delta(\lambda) \geq \Delta$, the minimum gap over the entire trajectory, it follows that

$$Q_D \geq \int_{\lambda_i}^{\lambda_f} d\lambda \, \langle g'| \frac{\Delta(\lambda)}{v(\lambda)} |g'\rangle \geq \Delta \cdot T, \tag{3.125}$$

where $T = t_f - t_i$ is the total evolution time.

Although the no-go theorem constrains how T scales with the path length L, this inequality does not directly restrict the scaling of Q_D. Thus, it remains an open question that Q_D admits a scaling bound consistent with one of the superlinear scenarios previously discussed.

Importantly, Q_D satisfies Condition 1. For time-periodic Hamiltonian, Q_D scales superlinearly with T. If the path covers an integer number of periods, say N_p, then $T = N_p\tau$ where τ is the period,

and $Q_D = N_p Q_{D_{\text{sc}}}$, with $Q_{D_{\text{sc}}}$ the value of Q_D over a single cycle. Hence $Q_D = (Q_{D_{\text{sc}}}/\tau)T$, making it directly proportional to T. Even when the path length is not an exact multiple of period, the ratio Q_D/T approaches $Q_{D_{\text{sc}}}/\tau$ asymptotically as the path length increases, preserving the proportionality in the large-L limit. Since T grows superlinearly with L at fixed error, so must Q_D.

Furthermore, Q_D is dimensionless, satisfying Condition 2. By construction, it meets Condition 3, and since it was motivated by a desire to quantify adiabaticity, it also satisfies Condition 4. Thus, this definition of Q_D satisfies all four criteria, making it a plausible candidate for quantifying computational difficulty.

It also seems highly likely that Q_D satisfies at least one of the superlinear scenarios. The reasoning is simple: longer paths (in terms of L) provide more opportunities for errors to build up.

Nevertheless, the definition of Q_D in Eq. (3.124) is not unique. Alternative forms could also serve this role. For instance,

$$Q_{D_2} \equiv \int_{\lambda_i}^{\lambda_f} d\lambda \, \frac{\sqrt{\langle g' | \left(H(\lambda) - E_g(\lambda)\right)^2 | g' \rangle}}{v(\lambda)}, \tag{3.126}$$

or

$$Q_{D_{1/2}} \equiv \int_{\lambda_i}^{\lambda_f} d\lambda \, \frac{\left(\langle g' | \sqrt{H(\lambda) - E_g(\lambda)} | g' \rangle\right)^2}{v(\lambda)} \tag{3.127}$$

could also be used. The first emphasizes higher energy excitations, while the second gives more weight to lower-lying states. Both satisfy Conditions 1-4.

More generally, for any positive, monotonically increasing, and invertible function $f(x)$, one could define:

$$Q_{D_f} \equiv \int_{\lambda_i}^{\lambda_f} d\lambda \, \frac{f^{-1}\left(\langle g' | f\left(H(\lambda) - E_g(\lambda)\right) | g' \rangle\right)}{v(\lambda)}, \tag{3.128}$$

and still satisfy the same set of conditions. Thus, there is a broad class of valid choices for Q_D.

It remains unclear whether all these versions of Q_D lead to the same qualitative behavior. While it is plausible that each yields superlinear scaling with L at fixed small error, they may differ in how restrictive they are in terms of applicability or in what they emphasize about the dynamics of the system.

It seems quite plausible that Scenario 1—the weakest version of the conjecture—holds for all of these formulations of Q_D. If so, and if Q_D truly reflects computational difficulty, this has important implications: it suggests that adiabatic state preparation may be fundamentally ill-suited for problems involving very long paths.

3.5.5.6 Setup for numerical test

We outline a numerical approach to test the conjecture for a range of time-dependent Hamiltonians. Importantly, these methods are designed specifically to examine how the proposed proxies for computational cost scale with path length under adiabatic evolution—they are not intended as practical prescriptions for adiabatic algorithms in real-world applications.

While the conjecture is most naturally formulated using a path-length parameter λ, directly simulating dynamics in terms of λ is technically challenging. Therefore, we instead parameterize evolution by physical time, which is more convenient for solving the Schrödinger equation numerically.

To ensure the system remains within a fixed error tolerance, the velocity $v(\lambda)$ along the path must decrease as the path length increases. We consider a family of velocity profiles that preserve relative velocity along the path but differ by an overall scaling factor, as Eq. (3.117). While increasing s_c reduces errors, it cannot be made arbitrarily large without increasing the computational cost. Thus,

we define s_c as the value that yields an error equal to a specified threshold ϵ_{th}.

A subtlety arises: the relationship between s_c and error is not necessarily monotonic—it can oscillate about an increasing trend. As a result, multiple values of s_c may satisfy the threshold condition. While the smallest such s_c minimizes cost, it is also more sensitive to realistic errors where s_c cannot be perfectly controlled, making it less reliable in practice. To ensure robustness, we adopt the conservative approach of selecting the largest s_c that meets the error criterion. In general, s_c depends on the Hamiltonian, the path length L, and the error threshold: $s_c = s_c(\epsilon_{\text{th}}, L; H)$.

With s_c fixed, we can express the proxies for computational cost of maintaining adiabaticity in terms of time. The time-variable versions of the conjecture quantities—originally defined in terms of path length—are given by:

$$Q_{D_1}(\epsilon_{\text{th}}, T_{\text{end}}) = s_c(\epsilon_{\text{th}}, T_{\text{end}}) \int_0^{T_{\text{end}}} dt \; \frac{\langle \dot{g}| \left(H(t) - E_g(t)\right) |\dot{g}\rangle}{\langle \dot{g}|\dot{g}\rangle}, \tag{3.129}$$

$$Q_{D_2}(\epsilon_{\text{th}}, T_{\text{end}}) = s_c(\epsilon_{\text{th}}, T_{\text{end}}) \int_0^{T_{\text{end}}} dt \; \frac{\sqrt{\langle \dot{g}| \left(H(t) - E_g(t)\right)^2 |\dot{g}\rangle}}{\sqrt{\langle \dot{g}|\dot{g}\rangle}}, \tag{3.130}$$

$$Q_{D_{1/2}}(\epsilon_{\text{th}}, T_{\text{end}}) = s_c(\epsilon_{\text{th}}, T_{\text{end}}) \int_0^{T_{\text{end}}} dt \; \frac{\left(\langle \dot{g}| \sqrt{H(t) - E_g(t)} |\dot{g}\rangle\right)^2}{|\langle \dot{g}|\dot{g}\rangle|^2}. \tag{3.131}$$

Here, T_{end} marks the end of the time evolution and corresponds to a particular path length.

To simulate Q_D as a function of path length L, we follow this procedure:

1. For a chosen Hamiltonian and desired path length (encoded via T_{end}) determine the appropriate scale factor s_c by starting with a large initial guess s' and decreasing it until the evolution error matches ϵ_{th}. This gives the appropriate s_c as defined above.

2. Increase the path length by scaling T_{end} by a factor $k > 1$: $T'_{\text{end}} = kT_{\text{end}}$. Set the initial guess

for the new s_c to be $s' = \kappa s_c$, with $\kappa > 1$.

3. Repeat step 1 for the new path length.

By iterating this process, one can study how each Q_D scales with L, maintaining a consistent relative velocity profile throughout.

Although this method is framed in terms of time—a dimensionful quantity—rather than the dimensionless path length used in the theoretical formulation, it is fully equivalent. The time-based parameterization is simply a reparameterization adopted for practical convenience in numerical simulations.

3.5.5.7 Random Hamiltonians

To test the conjectures, it is necessary to study systems simple enough to permit accurate numerical simulations over long evolution times. For this purpose, we restrict our analysis to low-dimensional Hamiltonians, specifically working within a four-dimensional Hilbert space so that the Hamiltonians are represented by 4×4 matrices. Four-level systems offer more complexity than two-level ones by allowing transitions to multiple excited states, yet remain computationally manageable. Moreover, since quantities such as the diabatic error and Q_D are primarily governed by properties of the ground state and low-lying excitations, the general features of the behavior are unlikely to depend strongly on the Hilbert space dimension. Thus, small-dimensional systems provide a meaningful and practical context for exploring generic behavior.

Figure 3.21 displays how the path length scales with Hilbert space dimension, along with a logarithmic fit of the form $a \log L + b$. The figure demonstrates that the rate of increase in path length slows as the dimension increases. If we interpret the system as representing a spin$-\frac{1}{2}$ system, then the

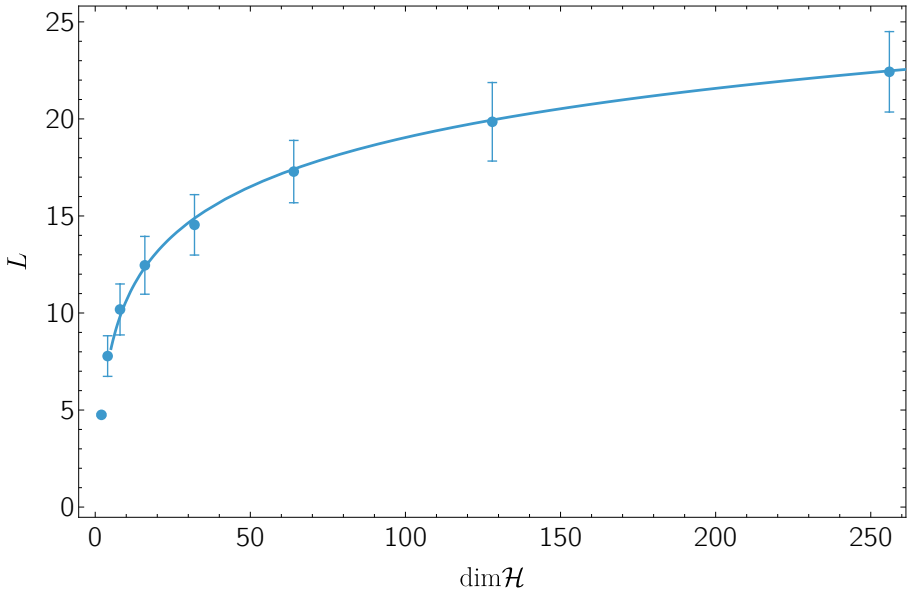

Figure 3.21: Comparison of path lengths for Hamiltonians of varying dimensions, evaluated at a fixed end time $T_{\text{end}} = 10$. For each dimension, 100 random Hamiltonians—sampled according to Eq. (3.132)—are used to compute the mean and standard deviation.

path length would scale with system volume, which itself grows faster than the behavior seen in local field theories (see Section 3.3.2). This observation suggests that using low-dimensional systems offers an efficient approach for investigating the scaling of diabatic errors with path length, particularly when working with classical computers.

Although numerical errors may arise during time evolution from the ground state of the initial Hamiltonian to approximate the final ground state, we find that for the systems studied, the dominant source of deviation comes from diabatic corrections rather than numerical inaccuracies, even for long simulation times.

Since our objective is to understand generic behavior, we focus on Hamiltonians that are randomly generated within a constrained structure. The use of random Hamiltonians helps to avoid system-specific biases and provides a way to probe typical behavior across a representative ensemble. Some non-random examples to explore the behavior of Q_D on the spectral gap are discussed in

Appendix E.

The chosen form of the Hamiltonians is designed to meet three criteria:

1. The time dependence of the Hamiltonian is non-periodic.

2. The energy gaps evolve with time, and all gaps between the ground state and excited states are on a similar scale.

3. The Hamiltonian is real and symmetric, representing a time-reversal invariant system.

Condition 1 avoids special features of periodic systems, which can exhibit non-generic scaling. For periodic Hamiltonians, we checked that keeping the diabatic error fixed requires the evolution time to grow faster than linearly with the path length—a feature that will naturally be reflected in the proxies discussed here. Condition 2 prevents any single excited state from artificially dominating the error due to fine-tuning. Condition 3 reflects our broader interest in time-reversal symmetric systems, though similar studies could be conducted using general Hermitian Hamiltonians, albeit with separate considerations for real and imaginary components.

The specific form of random Hamiltonians that was chosen for these studies was:

$$H(t) = \left(2.1\sin(t) + \sin(\sqrt{2}t)\right) \times H_1 + \left(2.7\cos(t) + \cos(\sqrt{2}t)\right) \times H_2, \tag{3.132}$$

where H_1 and H_2 are independently sampled 4×4 real symmetric matrices with elements drawn from a standard normal distribution (mean zero, unit variance). This form satisfies the above criteria. The incommensurate frequencies ensure non-periodicity, and random matrices are known to exhibit level repulsion, making it unlikely for the gap between the ground and first excited states to be anomalously small.

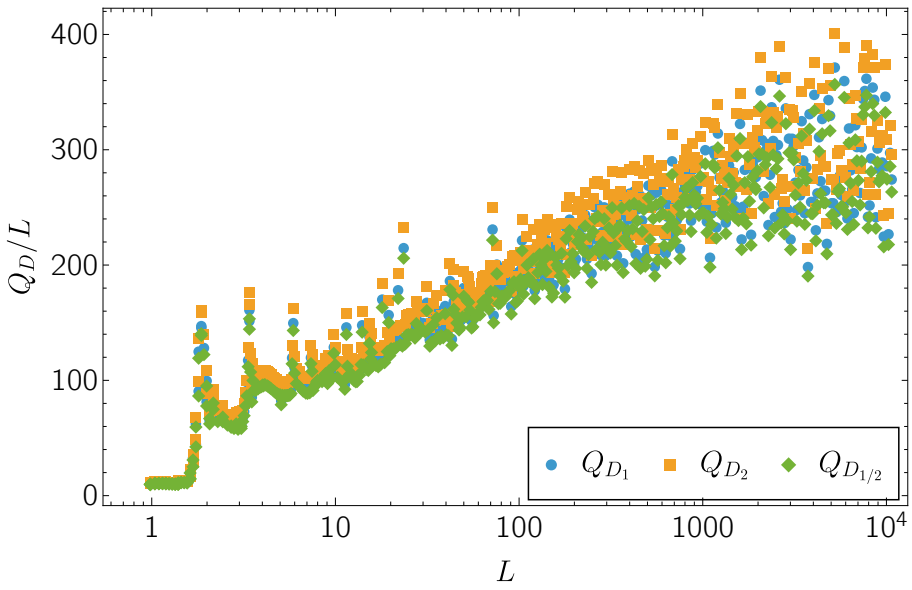

Figure 3.22: Comparison of different Q_D proxies defined in Eq. (3.131) for a representative Hamiltonian constructed using Eq. (3.132).

Figure 3.22 compares the various proxies Q_D defined in Eq. (3.131) for a single random Hamiltonian. The results show nearly identical behavior among the proxies, suggesting they are all capturing the same scaling features. This pattern persists across other randomly chosen Hamiltonians, as verified numerically; we omit those additional figures for clarity. For the remainder of this work, we use Q_{D_1} as the representative form of Q_D.

There are two main strategies to achieve large path lengths: one can simulate small systems for long times, or consider larger systems. Since solving time-dependent Schrödinger equations scales as $O(N^2)$ for explicit solvers (and worse for implicit methods), and because path length generally grows linearly with end time T_{end}, it is computationally advantageous to use low-dimensional models and extend the evolution time rather than increasing Hilbert space dimension.

In Figure 3.23, we show Q_D as a function of the path length L for four different randomly generated Hamiltonians of the above form. The vertical axis shows the quantity Q_D/L, allowing us to assess whether Q_D grows superlinearly with L. The results confirm the conjectured superlinear scal-

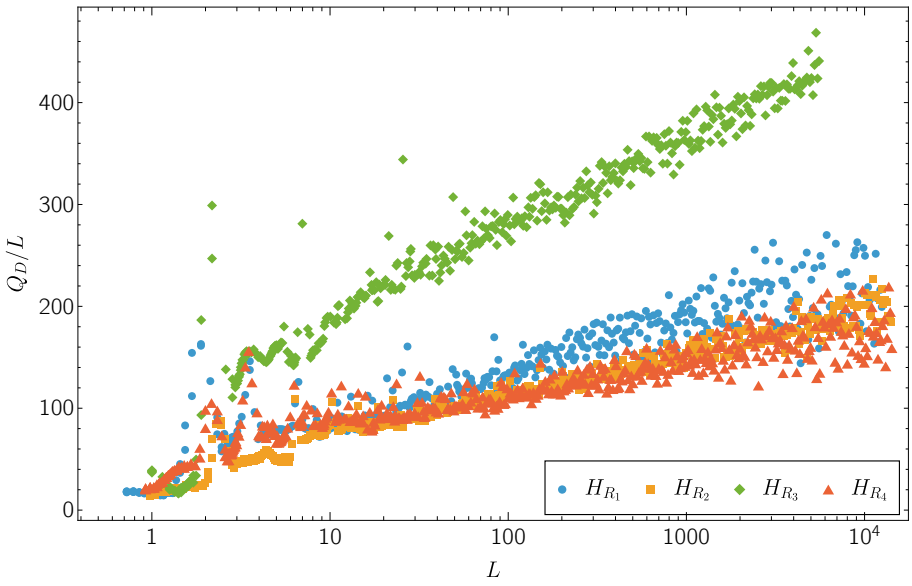

Figure 3.23: Q_D as a function of path length, computed using four different random Hamiltonians of the form given in Eq. (3.132). The threshold error is set to $\epsilon_{\text{th}} = 0.1$. Each curve labeled R_i corresponds to a different random choice of matrices H_1 and H_2 in Eq. (3.132).

ing: although fluctuations in Q_D/L are observed across different instances and values of L, all cases exhibit a clear secular upward trend. Due to these fluctuations, isolating the exact functional form of the growth is nontrivial, but the trend is consistent with an approximate $Q_D \propto L \log L$ scaling. Regardless of the detailed form, the central result is unambiguous: Q_D grows more rapidly than linearly with path length, in agreement with the conjecture.

3.5.5.8 Discussion

In quantum computing, understanding how computational cost scales with path length during ground state preparation is essential, especially since some methods—such as the Zeno-inspired method—are known to scale linearly with path length. Therefore, assessing whether adiabatic state preparation remains competitive for systems with long path length is an important question.

There are special cases where the computational proxy Q_D grows linearly with path length L

while maintaining a fixed error. One such case involves Hamiltonians of the form:

$$H(t) = \exp\left(-i\frac{vA\,t}{\sqrt{\langle g_0|A^2|g_0\rangle}}\right) H_0 \exp\left(i\frac{vA\,t}{\sqrt{\langle g_0|A^2|g_0\rangle}}\right) \tag{3.133}$$

where $|g_0\rangle$ is the ground state of H_0, A is a Hermitian operator with zero expectation value in the ground state, and v is a velocity parameter. This construction ensures that the evolution corresponds to a unitary transformation of the original Hamiltonian, preserving its spectrum. An explicit example is a particle in one dimension with $H_0 = \frac{p^2}{2m} + V(x)$ and $A = p$, where the evolution corresponds to a spatial translation. For such paths, the diabatic error averaged over time saturates at a value proportional to v in the limit of small v and long path length, implying that the average error is independent of L. Meanwhile, Q_D increases linearly with L, as expected from Eq. (3.124).

However, these cases are highly engineered and do not represent generic behavior. The key question is how Q_D scales in typical scenarios. The numerical examples studied, though based on small systems, show compelling evidence for superlinear scaling when errors are held fixed. This suggests that superlinear behavior is not only possible but likely generic, and expected to persist even for systems with much larger Hilbert spaces. In realistic applications of adiabatic evolution, it is unlikely that one would encounter the kind of fine-tuned structure present in Eq. (3.133); therefore, proxies like Q_D are expected to scale superlinearly in typical cases.

This implies that for very long paths, methods with provably linear scaling may be preferable. Nonetheless, our findings also indicate that the superlinear growth in Q_D is mild, consistent with a scaling of $Q_D \sim L \log L$. Thus, if adiabatic state preparation offers other computational advantages—such as implementation simplicity or robustness—it could still be more practical than alternative methods, except in regimes where path lengths become exceptionally large. At present, there is no compelling

reason to assume such advantages exist for the adiabatic approach. In fact, alternative methods might provide benefits beyond path length scaling. A natural direction for future work is to identify the regimes in which adiabatic state preparation remains competitive or preferable in the long-path limit.

Chapter 4: Conclusion

This thesis explores methodologies for studying non-perturbative physics using both classical and quantum computational frameworks. Chapter 2 addresses techniques for reducing the variance in Monte Carlo calculations, while Chapter 3 focuses on efficient strategies for preparing eigenstates on quantum computers—an essential step in many quantum algorithms.

In Chapter 2, two primary approaches to variance reduction are examined: modifying the sampling distribution via reweighting, and identifying alternative observables that yield the same expectation value but with reduced variance.

Section 2.3 investigates the infinite variance problem encountered in fermionic observables. This issue originates from zeros of the fermion determinant and is mitigated through reweighting with an extra time-slice probability distribution. A prior method introduces bias during reweighting, which can be corrected using a secondary Monte Carlo estimation. However, in scenarios where the sign problem is present, the secondary estimation also inherits the sign problem, highlighting the need for further research in this direction.

Section 2.4 examines the use of contour deformation and complex normalizing flows as tools for alleviating the sign problem. These methods integrate naturally with machine learning frameworks and are applied to scalar field theory at complex coupling. Additionally, the zeros of the partition function are probed using both normalizing flows and the behavior of the average sign phase.

In Section 2.5, we study the control variates method for reducing the signal-to-noise problem in lattice field theory. Through the construction of Stein control variates, this approach can also be enhanced by employing machine learning techniques. Signal-to-noise ratios for toy models can be largely improved by incorporating the symmetry of theories into the construction of control variates. Additionally, its universality is discussed in Appendix C.

To apply contour deformation or control variates methods on realistically large lattices, transfer learning becomes essential. In this thesis, we primarily employ fully connected feed-forward networks, which are difficult to generalize to larger lattice sizes. For scalability, alternative architectures such as convolutional neural networks [267] or transformers [268] should be explored, particularly in ways that incorporate symmetries for manifolds or control variates.

Chapter 3 is devoted to the problem of eigenstate preparation on quantum devices. The discussion centers on algorithms that follow a continuous path in Hamiltonian space, as introduced in Section 3.3.

Section 3.2 analyzes the Rodeo algorithm proposed in [2], identifying that fluctuations in excited-state suppression stem from intrinsic randomness. Enhancements are proposed to remove these fluctuations and improve suppression factors for a fixed computational cost.

Section 3.4 develops a projection-based scheme for traversing a Hamiltonian path. By incorporating reflections following failed projections, a linear-cost algorithm in the path length L is constructed. This method demonstrates quantum advantage by enabling eigenstate preparation in local field theories with a computational cost scaling as $O(\sqrt{V})$, where V is the volume of the system.

Section 3.5 examines various facets of the adiabatic theorem. Section 3.5.3 introduces the notion of typical errors, while Section 3.5.4 explores the practical effectiveness of the switching theorem in mitigating these errors. As the system evolves along a longer path, the accumulation of errors becomes more likely, suggesting that the cost of adiabatic state preparation should increase with path length. This consideration is particularly relevant for quantum field theories, which are expected to involve long Hamiltonian paths. This hypothesis is investigated in Section 3.5.5, where the scaling behavior of the adiabatic theorem is studied using several simplified models. Further analysis with more physically realistic systems is warranted to fully characterize the cost scaling in practical applications.

Appendix A: Path integral formulation

The path integral formulation, developed by Feynman [269], provides an alternative to canonical quantization for describing quantum field theories. Rather than promoting fields and their conjugate momenta to operators and imposing commutation relations, this approach describes quantum dynamics as a sum over all possible histories the field can take, with each history contributing an amplitude to the process.

For a given Lagrangian density $\mathcal{L}(\phi)$, the central object is the transition amplitude between an initial state $|\phi_i, t_i\rangle$ and a final state $|\phi_f, t_f\rangle$, which is given by

$$\langle \phi_f, t_f | \phi_i, t_i \rangle = \int_{\phi(t_i)=\phi_i}^{\phi(t_f)=\phi_f} \mathcal{D}\phi \, \mathrm{e}^{iS[\phi]}, \tag{A.1}$$

where

- $\mathcal{D}\phi$ denotes integration over all field configurations consistent with the boundary conditions,
- $S[\phi] = \int d^4x \mathcal{L}(\phi)$ is the classical action of the field.

This formulation extends naturally to interacting quantum field theories, where the generating functional (with source $J(\phi)$) is defined as

$$Z[J] = \int \mathcal{D}\phi \, \exp\left(i \int d^4x \left[\mathcal{L}(\phi(x)) + J(x)\phi(x) \right] \right). \tag{A.2}$$

From this generating functional, time-ordered correlation functions (Green's functions) are obtained

via functional differentiation:

$$\langle 0|T\phi(x_1)\phi(x_2)\cdots\phi(x_n)|0\rangle = \left.\frac{\delta^n Z[J]}{\delta J(x_1)\delta J(x_2)\cdots\delta J(x_n)}\right|_{J=0}. \tag{A.3}$$

The path integral formulation is manifestly Lorentz invariant and naturally incorporates symmetries of the classical action. Moreover, gauge invariance can be implemented directly at the level of the path integral through the use of gauge fixing and Faddeev-Popov ghosts.

A.1 Euclidean path integral

The Minkowski path integral Eq. (A.2) is not mathematically well-defined in general due to the rapid oscillations of the integrand, especially in systems with many degrees of freedom. This makes numerical evaluation—such as Monte Carlo methods—impractical.

To address this, a Wick rotation is employed:

$$t \to -i\tau, \tag{A.4}$$

which transforms the theory to Euclidean time τ. Under this transformation, the Minkowski action S becomes the Euclidean action S_E, and the path integral becomes

$$Z_E = \int \mathcal{D}\phi \mathrm{e}^{-S_E[\phi]}. \tag{A.5}$$

The exponential is now real and damped, except for some systems including those with finite baryon chemical potential, making the Euclidean path integral well-suited for both rigorous analysis

and numerical simulation.

The Euclidean path integral closely resembles the partition function in statistical mechanics. For a system at inverse temperature $\beta = 1/T$, the partition function is given by

$$Z = \mathrm{Tr}(\mathrm{e}^{-\beta H}), \tag{A.6}$$

which, when expressed in the path integral language, becomes equivalent to the Euclidean path integral with periodic (or antiperiodic) boundary conditions in Euclidean time:

- Bosons: periodic boundary conditions, $\phi(\tau = 0) = \phi(\tau = \beta)$,
- Fermions: anti-periodic boundary conditions, $\psi(\tau = 0) = -\psi(\tau = \beta)$.

Thus, quantum field theory at finite temperature corresponds to Euclidean field theory on a compactified time direction of length β. In the zero-temperature limit $\beta \to \infty$, one recovers the vacuum (ground state) physics.

This correspondence allows for the direct application of statistical methods to the study of quantum systems, particularly in lattice field theory where spacetime is discretized, and the path integral becomes a finite-dimensional integral suitable for Monte Carlo calculations.

A.2 Fermionic path integral

In contrast to bosonic fields, fermionic fields obey Fermi-Dirac statistics and anticommute. To formulate a path integral over fermionic degrees of freedom, we must introduce Grassmann variables—mathematical objects that anticommute with each other and with fermionic operators.

Grassmann numbers θ, η satisfy the anticommutation relation:

$$\theta\eta = -\eta\theta. \tag{A.7}$$

In particular, this implies that any Grassmann variable squares to zero:

$$\theta^2 = 0. \tag{A.8}$$

As a result, any function of Grassmann variables is necessarily a polynomial of finite degree. For example, a general function $f(\theta)$ of a single Grassmann variable has the form:

$$f(\theta) = A + B\theta, \tag{A.9}$$

where A and B are complex numbers.

Integration over Grassmann variables is defined axiomatically by demanding linearity and invariance under shifts. The defining rules are

$$\int d\theta = 0, \quad \int d\theta\, \theta = 1. \tag{A.10}$$

For multiple Grassmann variables, the integration rule generalizes:

$$\int d\theta_1 \cdots d\theta_n\, \theta_n \cdots \theta_1 = 1, \tag{A.11}$$

where the order of the differentials and the Grassmann variables is essential due to anticommutativity.

The fermionic path integral is built from Gaussian integrals over Grassmann variables. Consider the simplest Gaussian integral:

$$\int D\bar{\psi} D\psi \, \mathrm{e}^{-\bar{\psi} M \psi},$$

where $\bar{\psi}$, ψ are Grassmann-valued fields and M is a matrix (or a differential operator in field theory). The result is the determinant:

$$\int D\bar{\psi} D\psi \, \mathrm{e}^{-\bar{\psi} M \psi} = \det M. \tag{A.12}$$

This is in contrast to bosonic Gaussian integrals, which yield $(\det M)^{-1/2}$ for real fields. The determinant structure from fermionic integration is crucial in quantum field theory, as it naturally encodes the effects of virtual fermion loops and obeys the correct antisymmetric statistics.

A.2.1 Coherent states

In systems with fermions, each site of the lattice (or each mode in quantum mechanics) corresponds to a two-dimensional Hilbert space, due to the Pauli exclusion principle. The occupation number basis consists of the vacuum state $|0\rangle$ and the single-particle state $|1\rangle = \hat{\psi}^\dagger |0\rangle$, where $\hat{\psi}^\dagger$ is the fermion creation operator and $\hat{\psi}$ is the fermion annihilation operator satisfying

$$\{\hat{\psi}, \hat{\psi}^\dagger\} = 1, \quad \hat{\psi}^2 = (\hat{\psi}^\dagger)^2 = 0. \tag{A.13}$$

To construct a path integral for fermions in the operator formalism, we introduce fermionic coherent states, which are eigenstates of the annihilation operator with Grassmann-valued eigenvalues:

$$|\eta\rangle \equiv \mathrm{e}^{\hat{\psi}^\dagger \eta} |0\rangle, \quad \hat{\psi}|\eta\rangle = \eta|\eta\rangle, \tag{A.14}$$

where η is a Grassmann number.

These coherent states form an overcomplete basis with the following resolution of identity:

$$I = \int d\bar{\eta} d\eta \, \mathrm{e}^{-\bar{\eta}\eta} |\eta\rangle\langle\eta|. \tag{A.15}$$

Also, the coherent states allow us to express the trace of any operator A:

$$\mathrm{Tr}(A) = \int d\bar{\eta} d\eta \, \mathrm{e}^{-\bar{\eta}\eta} \langle -\eta | A | \eta \rangle. \tag{A.16}$$

To see this, we can expand:

$$\begin{aligned} \int d\bar{\eta} d\eta \, \mathrm{e}^{-\bar{\eta}\eta} \langle -\eta | A | \eta \rangle &= \int d\bar{\eta} d\eta \, (1 - \bar{\eta}\eta)(\langle 0| + \bar{\eta}\langle 1|) A (|0\rangle - \eta |1\rangle) \\ &= \langle 0|A|0\rangle + \langle 1|A|1\rangle = \mathrm{Tr}(A), \end{aligned} \tag{A.17}$$

which encodes the nature of anti-periodicity in the time direction discussed in Section A.1.

The fermionic partition function is rewritten in the coherent-state formalism as

$$Z = \mathrm{Tr}(\mathrm{e}^{-\beta H}) = \int d\bar{\eta}_0 d\eta_0 \, \mathrm{e}^{-\bar{\eta}_0 \eta_0} \langle -\eta_0 | \mathrm{e}^{-\beta H} | \eta_0 \rangle. \tag{A.18}$$

To express this as a path integral, we discretize the time interval $[0, \beta]$ into N steps of size $\delta = \beta/N$, and insert the resolution of identity between each short-time evolution operator $\mathrm{e}^{-\delta H}$. This yields

$$Z = \int \prod_{t=0}^{N-1} d\bar{\eta}_t d\eta_t \, \mathrm{e}^{\sum_t (\bar{\eta}_{t+1} - \bar{\eta}_t)\eta_t - \delta \sum_t H(\bar{\eta}_{t+1}, \eta_t)}, \tag{A.19}$$

where $\bar{\eta}_N \equiv -\bar{\eta}_0$ due to the anti-periodic boundary condition for fermions in the imaginary time

formalism. In the continuum limit $\delta \to 0$ with $N\delta = \beta$, this becomes

$$Z = \int D\bar{\eta} D\eta \, \exp\left(- \int_0^\beta dt \left[-\frac{d\bar{\eta}}{dt}(t)\eta(t) + H(\bar{\eta}(t), \eta(t)) \right] \right). \tag{A.20}$$

The term $\frac{d\bar{\eta}}{dt}\eta$ plays the role of the kinetic term for fermionic degrees of freedom, and the full Euclidean action is

$$S_F = \int_0^\beta dt \left[\bar{\eta}(t)\frac{d\eta}{dt}(t) + H(\bar{\eta}(t), \eta(t)) \right]. \tag{A.21}$$

A.2.2 Hubbard-Stratonovich transformation

The Hubbard-Stratonovich (HS) transformation [270, 271] is a powerful technique used to decouple quartic (four-fermion) interactions by introducing auxiliary bosonic fields, thereby reducing the fermionic action to an exactly integrable bilinear form. This method is especially useful in fermionic systems where interactions take the form $(\bar{\psi}\psi)^2$, and it plays a key role in both analytical treatments and numerical simulations of strongly correlated systems.

A central difficulty in simulating fermionic theories on classical computers arises from the Grassmann nature of fermionic fields, which cannot be directly represented in terms of ordinary numerical variables. Consequently, one must analytically integrate out all Grassmann variables before performing numerical computations. However, four-fermion interactions prevent a straightforward application of Gaussian integration. The HS transformation provides a way to bypass this obstacle by replacing quartic fermion terms with bilinear fermion couplings to an auxiliary bosonic field.

The basic identity behind the transformation is a Gaussian integral:

$$\mathrm{e}^{-\frac{g^2}{2}(\bar{\psi}\psi)^2} \propto \int d\phi \, \mathrm{e}^{-\frac{1}{2}\phi^2 + g\bar{\psi}\psi\phi}. \tag{A.22}$$

where ϕ is a real auxiliary scalar field.

In some contexts—such as contour deformation techniques for evaluating highly oscillatory integrals—it is advantageous to use compact auxiliary variables instead of real ones. Compact variables allow the deformed contour to remain within the same homology class, which is important for preserving the correctness of the integral under deformation.

An example of a compact version of the HS transformation is given by the identity:

$$\frac{1}{2\pi I_0(\beta)} \int_0^{2\pi} d\theta \, \mathrm{e}^{\beta \cos\theta} \mathrm{e}^{iX \sin\theta} = \mathrm{e}^{-\frac{g}{2}X^2}, \tag{A.23}$$

where $I_n(z)$ are modified Bessel functions and β is determined from the condition $g = I_1(\beta)/\beta I_0(\beta)$.

A.3 Path integral representation for the Hubbard model

The representation in Eq. (2.20) is derived by a Trotterization which separates quadratic and quartic terms in the Hamiltonian:

$$Z = \mathrm{Tr}\left[\left(\mathrm{e}^{-\epsilon H_2}\mathrm{e}^{-\epsilon H_4}\right)^N\right] + O(\epsilon^2), \tag{A.24}$$

with

$$\begin{aligned} H_2 &= -\kappa \sum_{\langle x,y\rangle} \sum_a \hat{\psi}^\dagger_{a,x}\hat{\psi}_{a,y} - \mu \sum_x (\hat{n}_{1,x} - \hat{n}_{2,x}), \\ H_4 &= \frac{U}{2} \sum_x (\hat{n}_{1,x} - \hat{n}_{2,x})^2. \end{aligned} \tag{A.25}$$

Above, $\hat{n}_{1,x}$ represents the number operator for up-electrons at site x, while $\hat{n}_{2,x}$ corresponds to the number operator for down-holes.

We apply a HS transformation with a compact auxiliary field by observing that Eq. (A.23) holds for any operator $\hat{X}$ with eigenvalues $\{0, \pm 1\}$:

$$\frac{1}{2\pi I_0(\beta)} \int_0^{2\pi} d\theta \; \mathrm{e}^{\beta \cos\theta} \mathrm{e}^{i\hat{X}\sin\theta} = \mathrm{e}^{-\frac{\epsilon U}{2}\hat{X}^2}, \tag{A.26}$$

where β instead satisfies $\mathrm{e}^{-\epsilon U/2} = I_0(\sqrt{\beta^2 - 1})/I_0(\beta)$. If we exploit the coherent state property for Grassmann integrals [272],

$$\langle \eta | \mathrm{e}^{\sum_{x,y} \hat{\psi}_x^\dagger \hat{X}_{xy} \hat{\psi}_y} | \eta' \rangle = \mathrm{e}^{\sum_{x,y} \bar{\eta}_x (\mathrm{e}^X)_{xy} \eta'_y}, \tag{A.27}$$

and let $\hat{X} = \hat{n}_{1,x} - \hat{n}_{2,x}$, the partition function can be written as

$$Z = \int [d\phi]_N \; \mathrm{e}^{-S_0(\phi)} \det\left(M_1(\phi) M_2(\phi)\right) + O(\epsilon^2). \tag{A.28}$$

Here, the fermion matrix M_a is

$$M_a(\phi)_{(t,x),(t',y)} = \delta_{x,y}\delta_{t,t'} + A^a(\phi_t)_{x,y} b_t \delta_{t,t'+1}\,, \tag{A.29}$$

where $b_t = -1$ for $t = 2N - 1$ and 1 otherwise, and

$$A^a(\phi_t)_{x,y} = \begin{cases} \mathrm{e}^{-\tilde{H}_2^a} & \text{if } t \text{ is even,} \\ \mathrm{e}^{-\tilde{H}_4^a(\phi_{(t-1)/2})} & \text{if } t \text{ is odd,} \end{cases} \tag{A.30}$$

with

$$(H_2^a)_{x,y} = -\epsilon\kappa\delta_{<x,y>} + \varepsilon_a\epsilon\mu\,,$$
$$\tilde{H}_4^a(\phi_t)_{x,y} = -i\varepsilon_a\delta_{x,y}\sin\phi_{t,x}\,. \tag{A.31}$$

Note that we have N auxiliary fields, $\phi_0, \cdots, \phi_{N-1}$, from HS transformations for H_4, while we need $2N$ coherent state insertions for H_2 and H_4.

With the following identity (called the BSS formula) for any set of square matrices $\{A_t\}$ [40],

$$\det\begin{bmatrix} \mathbb{I} & 0 & \cdots & A_n \\ A_1 & \mathbb{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & A_{n-1} & \mathbb{I} \end{bmatrix} = \det\left(\mathbb{I} - (-1)^n A_n\cdots A_1\right), \tag{A.32}$$

one can derive that

$$Z = \int [d\phi]_N \mathrm{e}^{-S_0(\phi)} \det M_1(\phi) \det M_2(\phi) + O(\epsilon^2), \tag{A.33}$$

where the fermion matrix becomes

$$M_a(\phi) = \mathbb{I} + B^a(\phi_N)\cdots B^a(\phi_1), \tag{A.34}$$

where $B^a(\phi_t) \equiv A^a(\phi_{2t-1})A^a(\phi_{2t}) = \mathrm{e}^{-\tilde{H}_2^a}\mathrm{e}^{-\tilde{H}_4^a(\phi_{(t-1)/2})}$.

The "expansion method" in Fig. 2.4 refers to an alternative approach for evaluating $F(\phi)$ in Eq. (2.33), which expresses it in terms of fermion contractions with respect to an action defined over N time-slices, as discussed in [39]. First, begin with the Grassmann integral identity

$$\det M_{N+1}(\phi,\phi^*) = \mathrm{Tr}\left[\hat{B}(\phi^*)\hat{B}(\phi_N)\cdots\hat{B}(\phi_1)\right], \tag{A.35}$$

where $\hat{B}(\phi) = \mathrm{e}^{-\epsilon H_2}\mathrm{e}^{i\hat{X}\sin(\phi)}$ with $\hat{X} = \hat{n}_1 - \hat{n}_2$. Note that the hat is used to emphasize that B is an operator. By reversing the HS transformation over ϕ^* with $\mathrm{e}^{-\mathrm{S}_0(\phi^*)}$, one obtains

$$\begin{aligned} F(\phi) &= \mathrm{Tr}\left[\mathrm{e}^{-\epsilon H_2}\mathrm{e}^{-\epsilon H_4}\hat{B}(\phi_N)\cdots\hat{B}(\phi_1)\right] \\ &= \mathrm{Tr}\left[(1-\epsilon H)\hat{B}(\phi_N)\cdots\hat{B}(\phi_1)\right] + O(\epsilon^2). \end{aligned} \tag{A.36}$$

Using the properties of the Grassmann integration, the first term can be written as

$$F_1(\phi) = (1 - \epsilon h_1(\phi))\det M_N(\phi), \tag{A.37}$$

where

$$\begin{aligned} h_1(\phi) = \kappa \sum_{\langle x,y\rangle}\left((M_1^{-1})_{y,x} + (M_2^{-1})_{y,x}\right) + \frac{U}{2}\sum_x\Big((M_1^{-1})_{x,x} + (M_2^{-1})_{x,x} - 2(M_1^{-1})_{x,x}(M_2^{-1})_{x,x}\Big) \\ + \mu\sum_x\left((M_1^{-1})_{x,x} - (M_2^{-1})_{x,x}\right), \end{aligned} \tag{A.38}$$

and $M_a^{-1}(\phi)$ is the inverse of Eq. (A.34). Thus, $F(\phi) = F_1(\phi) + O(\epsilon^2)$. Higher-order terms can be computed, but the number of fermion contractions increases exponentially, and one cannot avoid systematic errors.

Appendix B: Contour deformation in terms of homology

While contour deformation can be viewed as a higher-dimensional generalization of Cauchy's theorem from complex analysis, a deeper understanding requires the language of homology theory from algebraic topology.

We begin with the notion of chain groups C_n. If X is a simplicial complex (a generalization of triangles in higher dimensions), then C_n is the free abelian group generated by the set of all oriented n-simplices in X. Concretely, an element of C_n is a finite formal sum

$$c = \sum_i c_i \sigma_i, a_i \in \mathbb{Z},\ \sigma_i \text{ an } n\text{-simplex}. \tag{B.1}$$

For each n, there is a boundary operator $\partial_n : C_n \to C_{n-1}$, which encodes how an n-simplex is built from its $(n-1)$-dimensional faces. If $\sigma = [v_0, v_1, \ldots, v_n]$ is an oriented n-simplex, then

$$\partial_n \sigma = \sum_{j=0}^{n} (-1)^j [v_0, \ldots, \widehat{v_j}, \ldots, v_n], \tag{B.2}$$

where $\widehat{v_j}$ means the vertex v_j is omitted.

These boundary maps assemble into a chain complex:

$$\cdots \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0, \tag{B.3}$$

with the key property that the composition of successive boundary maps vanishes:

$$\partial_n \circ \partial_{n+1} = 0 \text{ for all } n. \tag{B.4}$$

This expresses the geometric fact that the boundary of a boundary is empty.

Using this structure, we define

- The group of n-cycles:

$$\ker(\partial_n) = \{c \in C_n | \partial_n c = 0\}.$$

- The group of n-boundaries:

$$B_n = \text{im}(\partial_{n+1}) = \{b \in C_n | b = \partial_{n+1} d \text{ for some } d \in C_{n+1}\}.$$

Cycles are chains whose boundaries vanish (such as closed loops or surfaces), while boundaries are those that arise as the boundary of a higher-dimensional object. From Eq. (B.4), every boundary is a cycle, so $B_n \subset Z_n$.

To measure nontrivial cycles—those that are not themselves boundaries—we take the quotient:

$$H_n(X) = Z_n / B_n = \ker \partial_n / \text{im}\, \partial_{n+1}. \tag{B.5}$$

The group $H_n(X)$, called the nth homology group of X, classifies genuine n-dimensional holes in X. An element $[c] \in H_n(X)$ is an equivalence class of cycles:

$$[c] = \{c + \partial_{n+1} d | d \in C_{n+1}\}. \tag{B.6}$$

If $[c] \neq 0$, then c represents a nontrivial topological feature of the space.

In the context of path integrals, we can interpret the integration contour as an n-chain and the integrand as an n-form in the space of differential forms $\Omega^n(X)$. The foundation of contour deformation lies in the following generalization of the fundamental theorem of calculus:

Theorem 1 (Stokes' theorem) *Let c be a smooth p-chain in a smooth manifold M and w be a smooth $(p-1)$-form on M. Then*

$$\int_{\partial c} w = \int_c dw.$$

Now suppose γ_1 and γ_2 are two n-cycles in the same homology class in an n-dimensional real manifold M, embedding it into a complex manifold of dimension $2n$. Then, there exists an $(n+1)$-chain $c \in C_{n+1}(M)$ such that

$$\gamma_1 = \gamma_2 + \partial c. \tag{B.7}$$

For any closed n-form $w \in \Omega^n(M)$[1] (so $dw = 0$), Stokes' theorem gives

$$\int_{\gamma_1} w = \int_{\gamma_2} w + \int_c dw = \int_{\gamma_2} w. \tag{B.8}$$

This result lies at the heart of the contour deformation method: as long as the integrand is a closed n-form and the deformed contour lies in the same homology class, the integral remains unchanged.

It is important to note that Stokes' theorem implicitly assumes that the integrand—the differential form—is well-defined and finite on the entire integration chain. For example, consider the Gaussian integral

$$Z = \int_{[-\infty,\infty]} dx \ \exp(-x^2). \tag{B.9}$$

[1] Note that every n-form in an n-dimensional manifold is closed.

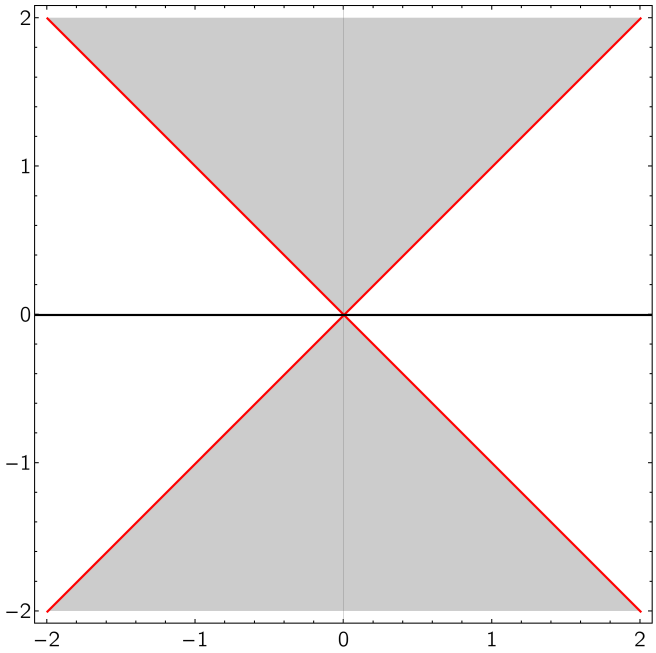

Figure B.1: Regions of a Gaussian integral divided by their homology classes. Grey regions lie outside the homology class of the real axis $\mathbb{R}$, and deformation into them would change the value of the integral.

This integral is divergent along certain complex contours, such as the imaginary axis $[-i\infty, i\infty]$, because $\exp(-x^2)$ diverges as $x \to \pm i\infty$. This is illustrated in Figure B.1. Hence, although two contours may be formally in the same homology class, the integral may not be well-defined on both if the integrand becomes singular or non-integrable on one of them. Care must be taken to avoid such divergences when applying contour deformation techniques.

Appendix C: Proofs of the universality of the Stein control variates

In this appendix, three ways of proving the universality of the Stein control variates, Eq. (2.76), are shown.

C.1 A proof using functional analysis

We begin by introducing the notation $\langle O \rangle_\Pi \equiv \int O \, d\Pi$, where Π is a probability distribution with density π. In our context, this corresponds to the path integral measure $d\Pi = [D\phi] \, e^{-S(\phi)}$. Reference [117] presents the following two results:

Proposition 1 *Let $\mathcal{G}$ be a normed space and $\mathcal{L} : \mathcal{G} \to L^2(\Pi)$ be a bounded linear operator. Consider a sequence of nested sets $\mathcal{V}_1 \subseteq \mathcal{V}_2 \subseteq \dots$ such that $\underset{p \in \mathbb{N}}{\cup} \mathcal{V}_p$ is dense in $\mathcal{U}$. If there is a $g \in \mathcal{G}$ that solves the Stein equation $\mathcal{L}g = O - \langle O \rangle_\Pi$, then $\lim\limits_{p \to \infty} \inf\limits_{v \in \mathcal{V}_p} \mathrm{Var}_\Pi[O - \mathcal{L}g] = 0$.*

Proposition 2 *Consider the vector space $\mathcal{G} = W^{2,2}(\Pi) \cap W^{1,4}(\Pi)$ equipped with norm $||g||_\mathcal{G} \equiv \max(||g||_{W^{1,4}(\Pi)}, ||g||_{W^{2,2}(\Pi)})$. Then $\mathcal{L}_{SL} : \mathcal{G} \to L^2(\Pi)$ is a bounded linear operator with $||\mathcal{L}_{SL}||_{\mathcal{G} \to L^2(\Pi)} \leq 2(||\nabla \log \pi||^2_{L^4(\Pi)} + 1)^{\frac{1}{2}}$.*

Furthermore, suppose that

(i) $\int ||x||_2^K d\Pi(x) < \infty$ *for some* $K > 8$,

(ii) $(\nabla \log \pi)(x) \cdot (x/||x||_2) \leq -r||x||_2^\alpha$ *for some* $\alpha < -1$, $r > 0$, *and all* $||x||_2 > M$ *for some* $M > 0$,

(iii) $|f(x)| \leq C_1 + C_2||x||_2^\beta$ *for some* $C_1, C_2 \geq 1$ *and* $\beta < K/4 - 2$.

Then there exists a $g \in \mathcal{G}$ that solves the Stein equation $\mathcal{L}_{SL} g = O - \langle O \rangle$.

Here, the space $W^{k,p}(\Pi)$ refers to the Sobolev space of functions whose weak derivatives up to order k lie in $L^p(\Pi)$. For a vector-valued function $h : \mathbb{R}^n \to \mathbb{R}^m$, the norm is defined as

$$|h|_{L^p(\Pi)} = \left(\sum_{i=1}^{m} |h_i|^2_{L^p(\Pi)} \right)^{1/2}. \tag{C.1}$$

The space $L^p(\Pi)$ consists of all measurable functions $f : \Pi \to \mathbb{R}$ such that

$$\int |f(x)|^p d\Pi < \infty. \tag{C.2}$$

For complete proofs and further details, see Ref. [117].

The core message of Proposition 1 is that even if we cannot find an exact solution to the Stein equation, we can approximate it arbitrarily well using expressive models (such as neural networks), provided that they lie in a dense subset of the function space.

Proposition 2 ensures that a solution to the Stein equation exists under mild regularity and growth conditions on both the observable O and the probability density π. This establishes a solid foundation for constructing neural control variates in practice.

C.2 A proof using differential topology

The direction of the proof using differential topology was first outlined in [114]. In this section, we elaborate on this approach in detail for theories of physical interest.

To begin, we reinterpret control variates using the language of differential forms. A function $f : X \to \mathbb{R}$ can be naturally associated with an n-form $f(x) dx_1 \wedge \cdots \wedge dx_n$. The expectation

value of f over X, in this context, corresponds to the integral $\int_X f$. Here, we temporarily omit the Boltzmann weight for clarity; it can be reintroduced straightforwardly, or equivalently, one may view X as equipped with a volume form that encodes the probability measure.

Before proceeding, let us recall two important subspaces of the space of r-form $\Omega^r(X)$:

$$\begin{aligned} Z^r(X) &= \{w \in \Omega^r(X) | dw = 0\} \quad \text{(closed } r\text{-form)}, \\ Z^r(X) &= \{w \in \Omega^r(X) | w = d\eta \text{ for some } \eta \in \Omega^{r-1}(X)\} \quad \text{(exact } r\text{-form)}. \end{aligned} \tag{C.3}$$

The rth de Rham cohomology group is defined as the quotient $H^r(X;\mathbb{R}) = Z^r(X)/B^r(X)$. In particular, the top cohomology group $H^n(X,\mathbb{R})$ plays a central role in our discussion. Since the exterior derivative d of an n-form is always zero (as there are no $(n+1)$-forms on an n-dimensional manifold), any n-form is automatically closed. Thus, the question of whether a control variate f can be written as a Stein control variate reduces to whether a closed n-form with vanishing integral is also exact—that is, whether it lies in $B^n(X)$.

This leads to the key topological result:

Theorem 2 *If X is a compact, connected, n-dimensional manifold, the top de Rham cohomology $H^n(X)$ is*

$$H^n(X) = \begin{cases} \mathbb{R} & \text{if } X \text{ is orientable,} \\ 0 & \text{if } X \text{ is not orientable.} \end{cases}$$

(See [273, 274] for a proof.) It follows that if X is non-orientable, all n-forms are exact, and thus any control variate with zero expectation value must be a Stein control variate. If X is orientable, then the space of closed n-forms modulo exact forms is isomorphic to $\mathbb{R}$. In this case, a linear functional

$f \mapsto \int_X f$ induces an isomorphism $H^n(X) \cong \mathbb{R}$, so any n-form with vanishing integral must lie in the trivial cohomology class—i.e., it must be exact. Hence, on any compact, connected, orientable manifold, a control variate with zero expectation value can be written as a Stein control variate $f = dg$, for some $g \in \Omega^{r-1}(X)$.

Let us now consider the case of gauge theories in physics. Many physically relevant target spaces X are products of Lie groups, such as $U(n)$ or $SU(n)$. Every Lie group is parallelizable and therefore orientable. Furthermore, compactness follows from the Heine–Borel theorem:

Theorem 3 *A matrix Lie group $G \subset GL(n;\mathbb{C})$ is compact if and only if it is compact as a subset of $M_n(\mathbb{C}) \cong \mathbb{R}^{2n}$.*

Since the Cartesian product of compact manifolds is compact, it follows that the configuration space for lattice gauge theories with compact gauge groups is also compact and orientable. Therefore, on such spaces, any control variate with zero expectation can be expressed as a Stein control variate—establishing the universality of the Stein construction in these settings.

Finally, we briefly address the case $X = \mathbb{R}^n$, which is non-compact and thus falls outside the scope of Theorem 2. However, the Poincaré lemma provides a useful substitute:

Theorem 4 (Poincaré lemma) *Every closed p-form on an open ball in $\mathbb{R}^n$ is exact for p with $1 \leq p \leq n$.*

Therefore, locally (or globally under suitable decay conditions), any closed form can be written as an exact form. In particular, on $\mathbb{R}^n$, if we restrict attention to compactly supported or sufficiently well-behaved n-forms, control variates with zero expectation can still be written in the Stein form $f = dg$. This provides further support for the universality of Eq. (2.76) even in non-compact settings.

C.3 A proof using a differential equation

This section starts with a proof suggested by Scott Lawrence [275] using the heat equation. After this, we consider its application to gauge theories.

Let us consider an arbitrary control variate $f : M^V \to \mathbb{R}$ with a uniform probability distribution on a connected space M and V number of degrees of freedom. It can be easily generalized to a general probability distribution. The condition $\langle f \rangle = 0$ implies

$$\int d^V x \, f(x) = 0. \tag{C.4}$$

To construct a Stein control variate representation of f, we employ the heat equation as a smoothing flow that naturally leads to the divergence form. Specifically, consider the heat equation in the extended space:

$$\frac{d}{dt} f(x,t) = \nabla_x^2 f(x,t), \tag{C.5}$$

with initial condition $f(x,0) = f(x)$, where $t \in \mathbb{R}^{\geq 0}$ is an auxiliary time variable independent of the configuration space M^V. By the properties of the heat equation on connected domains, we have

$$\lim_{t\to\infty} f(x,t) = \frac{1}{\mathrm{Vol}(M^V)} \int d^V x \, f(x) = 0, \tag{C.6}$$

where $\mathrm{Vol}(V) = \int d^V x$. Physically, this reflects that heat (or information) diffuses uniformly across the domain as $t \to \infty$.

Define an auxiliary function by integrating the gradient flow of f:

$$g(x) = -\int_0^\infty dt\, \nabla_x f(x,t). \tag{C.7}$$

Then we find

$$\begin{aligned} \nabla \cdot g &= -\nabla_x \cdot \int_0^\infty dt\, \nabla_x f(x,t) = -\int_0^\infty dt\, \nabla_x^2 f(x,t) \\ &= -\int_0^\infty dt\, \frac{d}{dt} f(x,t) = f(x,0) - f(x,\infty) = f(x). \end{aligned} \tag{C.8}$$

Thus, we obtain a solution to the Stein identity: $f = \nabla_x \cdot g$. The construction is not unique: any constant vector can be added to g without affecting $\nabla_x \cdot g$.

To generalize this to non-uniform distributions with a Boltzmann weight $\mathrm{e}^{-S(x)}$, consider the modified heat equation:

$$\frac{d}{dt}\left(f(x,t)\mathrm{e}^{-S(x)}\right) = \nabla_x^2 (f(x,t)\mathrm{e}^{-S(x)}), \tag{C.9}$$

with initial condition $f(x,0) = f(x)$. The solution satisfies

$$\lim_{t\to\infty} f(x,t) = \mathrm{e}^{S(x)} \frac{1}{\mathrm{Vol}(V^M)} \int d^V x\; f(x)\mathrm{e}^{-S(x)} = 0, \tag{C.10}$$

due to $\langle f \rangle = 0$. Then defining

$$g(x) = -\int_0^\infty dt\, \nabla_x (f(x,t)\mathrm{e}^{-S(x)}), \tag{C.11}$$

we compute

$$\begin{aligned}\nabla_x \cdot (g(x)\mathrm{e}^{-S(x)}) &= -\int_0^\infty dt\, \nabla_x^2 (f(x,t)\mathrm{e}^{-S(x)}) = \mathrm{e}^{-S(x)} \int_0^\infty dt\, \frac{d}{dt} f(x,t) \\ &= \mathrm{e}^{-S(x)}(f(x,0) - f(x,\infty)) = f(x)\mathrm{e}^{-S(x)}.\end{aligned} \tag{C.12}$$

Hence, g satisfies the Stein identity, Eq. (2.76):

$$f(x) = \nabla_x \cdot g(x) - g(x) \cdot \nabla_x S(x). \tag{C.13}$$

Among all approaches, this construction is particularly well-suited for gauge theories because it naturally preserves gauge invariance and clarifies the structure of the auxiliary function g. In gauge theories, control variates must be gauge-invariant due to Elitzur's theorem [276], which implies that any gauge-variant observable averages to zero. For both physical relevance and computational efficiency, one should restrict the search space for g to gauge-invariant functions.

Indeed, the differential equation preserves gauge invariance. From Eq. (C.5), a small-time evolution gives

$$f(x, \Delta t) \approx f(x) + \Delta t \times \nabla_{x,\mu}^2 f(x), \tag{C.14}$$

where μ denotes the direction of the link. This equation implies that $f(x, \Delta t)$ is gauge-invariant. Iteratively, this ensures $f(x,t)$ remains gauge-invariant for all t, implying that the integrated object $g(x)$ is also gauge-invariant.

A potential concern is that the gradient operator $\nabla_{x,\mu}$ might break gauge invariance. However, in lattice gauge theory, the fundamental degrees of freedom are link variables $U_\mu(x)$, and gauge trans-

formations act as

$$U_\mu(x) \mapsto X(x)U_\mu(x)Y(x+\hat{\mu})^\dagger. \tag{C.15}$$

While the gauge transformation depends on the site x, the gradient operator only acts on the link:

$$\nabla_{x,\mu}(X(x)U_\mu(x)Y(x+\hat{\mu})^\dagger) = X(x)\left(\nabla_{x,\mu}U_\mu(x)\right)Y(x+\hat{\mu})^\dagger. \tag{C.16}$$

Therefore, it ensures that gauge-invariant inputs yield gauge-invariant outputs in $g(x)$.

The remaining question is whether restricting to gauge-invariant g limits the expressiveness of Stein control variates. Fortunately, this is not the case. It was shown in [277] that any gauge-invariant observable can be approximated arbitrarily well by combinations of Wilson loops:

$$\sum_{n\geq 0}\sum_{W_1,\ldots,W_n} a(W_1,\ldots,W_n)\operatorname{Tr} W_1\cdots\operatorname{Tr} W_n, \tag{C.17}$$

where W_i denote Wilson loops.

Therefore, one can construct g as a function of Wilson loops:

$$g = g(\operatorname{Tr} W_1,\cdots,\operatorname{Tr} W_n). \tag{C.18}$$

ensuring gauge invariance. In the abelian $U(1)$ case, this reduces to products of plaquettes:

$$g = g(P_1,\cdots,P_n), \tag{C.19}$$

where $P = \operatorname{Tr} U_1U_2U_3^\dagger U_4^\dagger$, which is just a multiplication for $U(1)$ case, and such plaquette combina-

tions generate all Wilson loops.

In non-abelian gauge theories, constructing all gauge-invariant functions from Wilson loops is more intricate due to non-commutativity and the necessity of loop multiplications before taking traces. In such cases, gauge-equivariant neural networks [278–281] offer a principled and practical tool for learning such functions while respecting symmetry constraints.

Appendix D: Key factors in practical implementations of rodeo projections

D.1 Incorporating initial state information

If complete knowledge of the spectral function of the initial state were available, one could optimize the suppression strategy differently. Rather than minimizing

$$Q = \max_{E \geq \Delta} s\left(\frac{E}{\Delta}\frac{T}{T_0}\right)$$

to bound the total excited-state suppression S_E, one could directly seek a set of (super) iterations that either minimize S_E for a fixed total time or minimize the time needed to achieve a desired value of S_E. However, this kind of optimization would need to be repeated individually for each known spectral function, making it impractical in most cases.

That said, if one does know the spectral function and has access to a table like Table 3.2, it becomes possible to go beyond upper bounds. One could, for instance, choose a fixed total time and compute the exact value of S_E, or, more pragmatically, determine the minimum time needed to reach a required suppression level for a particular problem. This can yield significantly tighter and more efficient suppression than simply relying on conservative upper bounds.

Of course, having complete spectral information is rare. More realistically, one might have partial knowledge about the distribution of spectral weight. Even such partial knowledge can lead to significantly improved bounds compared to Eq. (3.41).

To leverage this, it is useful to define an upper-bounding function for the suppression factor,

denoted s^{UB}, that is monotonically decreasing and satisfies

$$s^{\mathrm{UB}}\left(\frac{E}{\Delta}\frac{T}{T_0}\right) \geq s\left(\frac{E}{\Delta}\frac{T}{T_0}\right) \tag{D.1}$$

for all relevant energies.

Suppose one has reason to believe that most of the spectral weight lies at energies well above the ground state. Define E_0 as the energy above which a fraction f of the excited-state spectral weight resides:

$$f \equiv \frac{\int_{E_0}^{\infty} dE\, \rho(E)}{\int_{\Delta-\epsilon}^{\infty} dE\, \rho(E)}. \tag{D.2}$$

Then, the suppression factor S_E can be bounded as:

$$S_E \leq (1-f)\, s^{\mathrm{UB}}\left(\frac{\Delta}{\Delta}\frac{T_{tot}}{T_0}\right) + f\, s^{\mathrm{UB}}\left(\frac{E_0}{\Delta}\frac{T_{tot}}{T_0}\right) \leq Q. \tag{D.3}$$

For example, using the 3-super-iteration Whac-a-Mole scheme from Table 3.2, if 99% of the excited-state weight lies above $E_0 = 3\Delta$ (i.e., $f = 0.99$), then the bound improves to $S_E \leq 5.591 \times 10^{-7}$—over 40 times tighter than the original conservative bound. With $f = 0.9999$ and $E_0 = 8\Delta$, the bound tightens dramatically to $S_E \leq 1.194 \times 10^{-8}$, which is a factor of 2000 smaller.

D.2 Cost of super iterations

We have proposed using computation time, measured in units of T_0, as a reasonable proxy for the computational cost. While this is generally a good approximation, there is an important caveat. In addition to the time required for the system to evolve under the Hamiltonian, each iteration introduces a fixed overhead: coupling the auxiliary qubit to the main system (via Hadamard gates and conditional

operations), followed by measurement of the auxiliary qubit. Although the overhead per iteration is small, it becomes problematic when considering super iterations, which—as defined in Eq. (3.40)—formally involve an infinite number of such steps. This leads to an unbounded computational cost.

A natural solution is to replace the idealized infinite super iteration with a finite one using a large but manageable number of iterations, N. The suppression factor for such a truncated super iteration becomes:

$$s_{\sup \mathrm{N}}(\zeta_{\sup}) = \prod_{k=1}^{N} \cos^2\left(\frac{\pi\zeta_{\sup}}{2^k}\right) = \frac{j_0^2\left(\pi\zeta_{\sup}\right)}{j_0^2\left(\frac{\pi\zeta_{\sup}}{2^N}\right)}. \tag{D.4}$$

For most practical purposes, the denominator is very close to one unless $\zeta_{\sup}$ becomes very large (on the order of 2^{N-2}), meaning that truncating the super iteration has negligible impact on the suppression factor over a wide range of energies.

One potential concern arises when $\zeta_{\sup}$ is an integer multiple of 2^N, in which case the suppression factor becomes unity and suppression is lost. However, this scenario is unlikely to be relevant in practice. It does, however, imply that theoretical upper bounds derived using such truncated super iterations are only valid up to a calculable maximum energy. If the spectral weight beyond that energy is negligible, the bound remains meaningful.

Encouragingly, even modest values of N result in very high-energy cutoffs. For example, using a single super iteration, the maximum energy (in units of Δ) for which the bound remains valid is approximately:

$$\frac{E_{\max}}{\Delta} \approx \frac{2^N - 1.430}{0.8129}. \tag{D.5}$$

So for $N = 15$, the bound holds for energies up to around $40,308\Delta$. When using multiple super iterations, the suppression remains valid over even broader energy ranges. In short, one can always choose N large enough to avoid practical limitations, ensuring that finite super iterations serve as a

good approximation to the ideal case.

D.3 False negative errors

So far, our focus has been on minimizing "false positives"—instances where components other than the ground state survive after the algorithm has run. We have not yet addressed the issue of "false negatives," in which the ground component is mistakenly lost. In the ideal scenario—where the ground state energy is known exactly and both time evolution and timing control are flawless—false negatives cannot occur, as the ground state survives each iteration with probability one. However, in realistic settings, small errors in the estimated ground state energy and imperfections in time evolution can lead to situations where the auxiliary qubit is measured in the down state even though the system remains in the ground state.

If one has a rough estimate of the uncertainty in the ground state energy, the quality of the time evolution, and the likelihood that the system is in the ground state, it is possible to estimate the probability of such a false negative during a single iteration. If this possibility is low but non-negligible, a practical strategy is to treat the outcome conservatively: when a down measurement is observed, continue the algorithm but flag the event as a potentially unsuccessful iteration. If the auxiliary qubit is subsequently measured in the up state in the following super iterations, it is reasonable to conclude that the earlier down measurement was a false negative and proceed as if the state had remained in the ground state.

However, if a second down measurement occurs during or shortly after the next super iteration, it is safer to assume that the iteration truly failed. This sort of accidental misclassification should be extremely rare, provided the uncertainty in the ground state energy is not excessively large.

D.4 Uncertainties in the energy of the first excited state

In our formulation, we assumed that the energies of both the ground state and the first excited state were known. As discussed earlier, imprecise knowledge of the ground state energy can lead to false negatives, but there are practical strategies to manage this. Uncertainty in the value of Δ—the energy gap between the ground and first excited states—is even easier to handle.

There are two types of possible error in estimating Δ: overestimation and underestimation.

If Δ is underestimated, the consequence is minimal. The unit time T_0 will be slightly longer than necessary, so the total runtime (our stand-in for computational cost) will be a bit higher than required to reach the target suppression. However, the upper bound for the suppression factor, as given in Table 3.2 or its generalization, remains valid because the total scaled energy ζ_{tot} will correspond to a higher actual energy.

In contrast, overestimating Δ is more problematic. In this case, the scaled energy ζ_{tot} corresponds to lower actual energies, meaning that suppression is being evaluated in a region where the suppression factor is typically larger than what the computed upper bound predicts. As a result, the supposed upper bound may no longer be valid.

The simple solution is to conservatively underestimate Δ—choose a value safely below the estimated one by more than the uncertainty. This approach preserves the validity of the upper bound while only slightly increasing the computational cost.

Appendix E: Scaling behaviors of Q_D for Hamiltonians whose gap changes monotonically

For random Hamiltonian cases, the gap in the spectrum neither grows nor shrinks secularly, it merely fluctuates. The no-go theorem is based on secular changes in the gaps along the path, which implied that computational time for calculation with fixed errors could not have any well-defined scaling with path length. The various proxies for computational difficulties are chosen so that they would be unaffected by secular changes in the overall scale. Thus it is important to test whether, as expected these proxies are, in fact, insensitive to secular shifts in the scale. In this subsection, simple systems whose spectral gap changes secularly will be considered.

One can start with a following simple 2×2 Hamiltonian:

$$H(t) = \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \lambda(t) \end{bmatrix} \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}, \tag{E.1}$$

where $\lambda(t)$ is an eigenvalue. This Hamiltonian is not periodic unless $\lambda(t)$ is periodic. For the given Hamiltonian, it is easy to determine Q_D:

$$Q_D(L) = s_c(\epsilon_{\text{th}}, L) \int_0^L dt\ |\lambda(t)|. \tag{E.2}$$

Note that it can be represented by L since $\frac{d\lambda}{dt} = 1$ along the path so it is actually the path length parametrization.

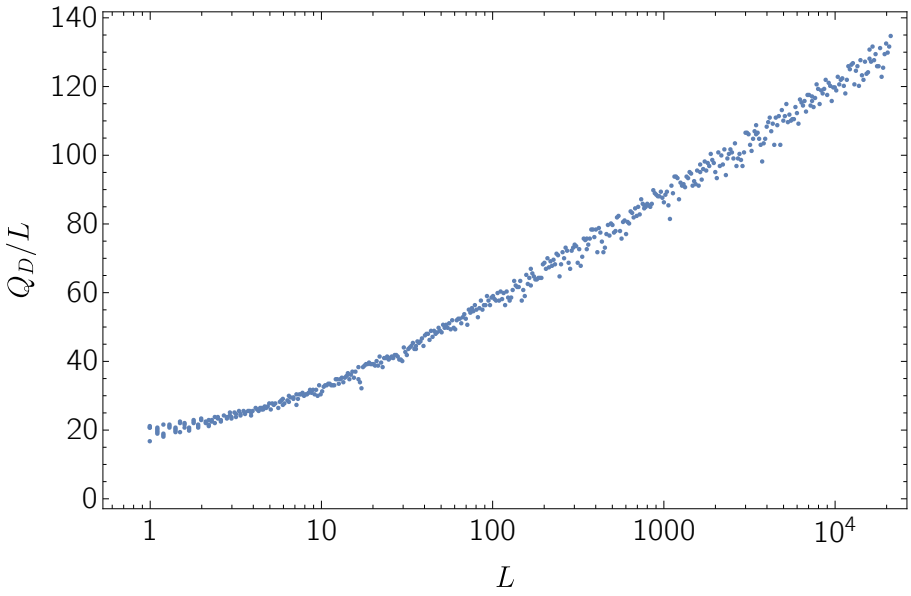

Figure E.1: Q_D with respect to path length using the log gap-opening Hamiltonian, Eq. (E.3). The threshold error ϵ_{th} is chosen as 0.1.

Therefore, one can consider two cases: the gap of the Hamiltonian increases or decreases. For gap-opening case, since $\lambda(t)$ is an increasing function, its integral is faster than linear in L. However, since the gap increases, it is easier to have the adiabatic limit with smaller scaling factors. Therefore, it is imperative to check numerically that Q_D increases faster than linearly in L, which is shown in Fig. E.1. The rate of changes for the eigenvalue is chosen as logarithmically increasing:

$$\lambda_1(t) = \log(2+t). \tag{E.3}$$

Fig. E.1 shows that similar to the random matrix case, Q_D increases with $L \log L$, which suggests that the scaling factor decreases slower than logarithmically in L for this case.

For gap-closing case, the situation is clearer. For example, let us choose the eigenvalue similar to the previous example:

$$\lambda_2(t) = 20 - \log(2+t), \tag{E.4}$$

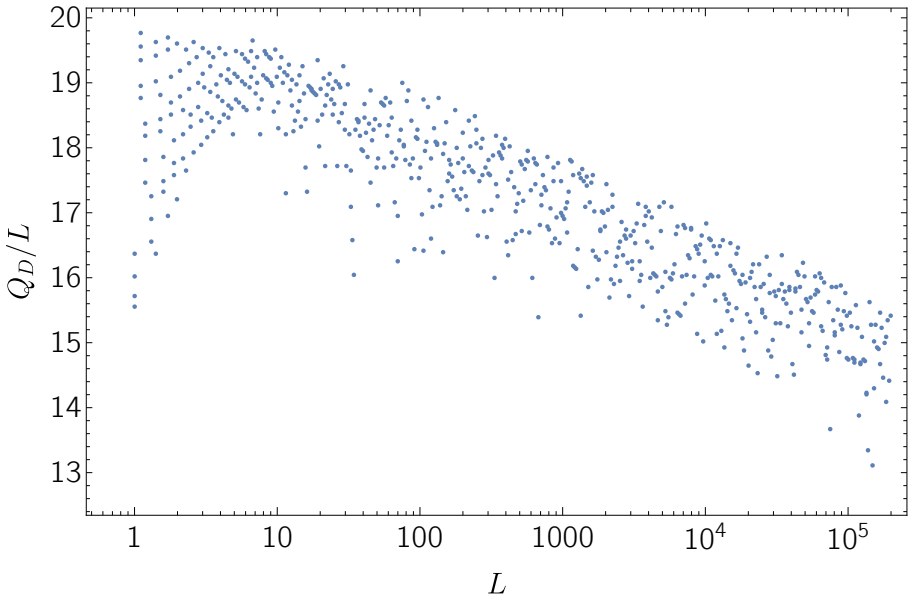

Figure E.2: Q_D with respect to path length using the log gap-closing Hamiltonian, Eq. (E.4). The threshold error ϵ_{th} is chosen as 0.1.

whose gap decreases logarithmically and closes at $t \approx e^{20} \approx 4.8 \times 10^8$. Then the integral in Q_D depends on L as $21L - L\log(2+L)$. While the scaling factor s_c increases as the path is traversed since the gap is closing, Q_D/L decreases, which is shown in Fig. E.2.

From the log-closing example, one might conclude that there is a counterexample for the conjecture. However, it does not contradict the conjecture. It is because in $L \to \infty$, the spectral gap becomes zero and increases as the gap-opening case and Q_D increases faster than linearly as Figure E.1. In the meanwhile, this example suggests that Q_D can increase slower than linearly in large, but finite L.

For previous gap-closing example, Eq. (E.4), the spectral gap ends up zero at very large path length. Therefore, one can consider a case where the gap gets smaller as the path is traversed, but it never closes, such as

$$\lambda_3(t) = \frac{1}{t+1}. \tag{E.5}$$

For this case, the integral in Q_D is $\log(L+1)$ and the full behavior of Q_D depends on how fast the

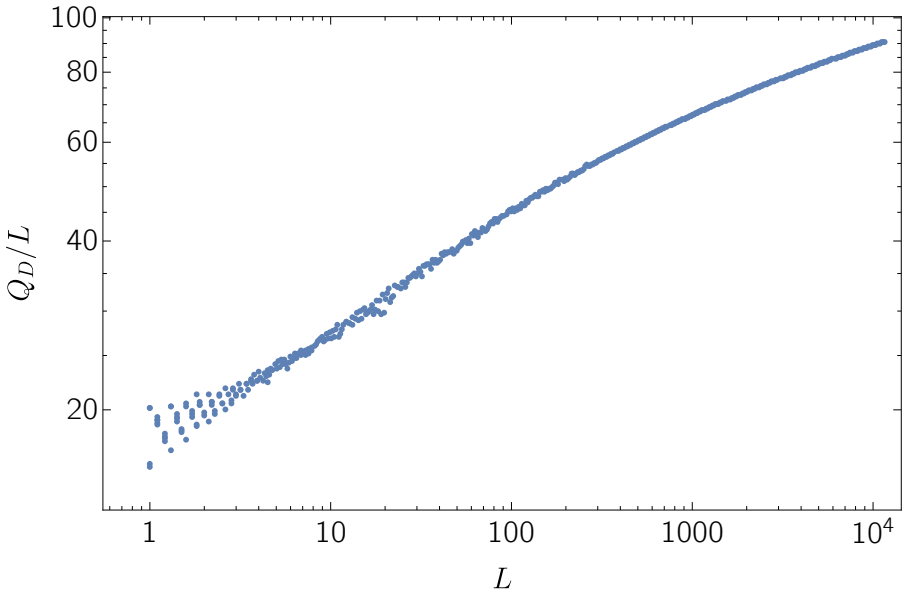

Figure E.3: Q_D with respect to path length using the polynomial gap-closing Hamiltonian, Eq. (E.5). The threshold error ϵ_{th} is chosen as 0.1.

scaling factor s_c increases. Fig. E.3 demonstrates that Q_D increases faster than linearly in L and it increases as $L \log L$, similar to the random Hamiltonians and the gap-opening case, Eq. (E.3).

Bibliography

[1] J. P. F. LeBlanc, Andrey E. Antipov, Federico Becca, Ireneusz W. Bulik, Garnet Kin-Lic Chan, Chia-Min Chung, Youjin Deng, Michel Ferrero, Thomas M. Henderson, Carlos A. Jiménez-Hoyos, E. Kozik, Xuan-Wen Liu, Andrew J. Millis, N. V. Prokof'ev, Mingpu Qin, Gustavo E. Scuseria, Hao Shi, B. V. Svistunov, Luca F. Tocchio, I. S. Tupitsyn, Steven R. White, Shiwei Zhang, Bo-Xiao Zheng, Zhenyue Zhu, and Emanuel Gull. Solutions of the two-dimensional hubbard model: Benchmarks and results from a wide range of numerical algorithms. *Phys. Rev. X*, 5:041041, Dec 2015.

[2] Kenneth Choi, Dean Lee, Joey Bonitati, Zhengrong Qian, and Jacob Watkins. Rodeo Algorithm for Quantum Computing. *Phys. Rev. Lett.*, 127(4):040505, 2021.

[3] Sabine Jansen, Mary-Beth Ruskai, and Ruedi Seiler. Bounds for the adiabatic approximation with applications to quantum computation. *J. Math. Phys.*, 48(10):102111, 2007.

[4] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. *Phys. Lett. B*, 47:365–368, 1973.

[5] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.

[6] David J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, 1973.

[7] H. David Politzer. Reliable Perturbative Results for Strong Interactions? *Phys. Rev. Lett.*, 30:1346–1349, 1973.

[8] Siegfried Bethke. Experimental tests of asymptotic freedom. *Prog. Part. Nucl. Phys.*, 58:351–386, 2007.

[9] Gerard 't Hooft and M. J. G. Veltman. Regularization and Renormalization of Gauge Fields. *Nucl. Phys. B*, 44:189–213, 1972.

[10] W. Pauli and F. Villars. On the Invariant regularization in relativistic quantum theory. *Rev. Mod. Phys.*, 21:434–444, 1949.

[11] Kenneth G. Wilson. Confinement of Quarks. *Phys. Rev. D*, 10:2445–2459, 1974.

[12] K. Symanzik. Continuum Limit and Improved Action in Lattice Theories. 1. Principles and φ^4 Theory. *Nucl. Phys. B*, 226:187–204, 1983.

[13] K. Symanzik. Continuum Limit and Improved Action in Lattice Theories. 2. O(N) Nonlinear Sigma Model in Perturbation Theory. *Nucl. Phys. B*, 226:205–227, 1983.

[14] Andrei Alexandru, Paulo F. Bedaque, Andrea Carosso, and Hyunwoo Oh. Infinite variance problem in fermion models. *Phys. Rev. D*, 107(9):094502, 2023.

[15] Hyunwoo Oh, Andrei Alexandru, Paulo F. Bedaque, and Andrea Carosso. A solution for infinite variance problem of fermionic observables. *PoS*, LATTICE2023:021, 2024.

[16] Scott Lawrence, Hyunwoo Oh, and Yukari Yamauchi. Lattice scalar field theory at complex coupling. *Phys. Rev. D*, 106(11):114503, 2022.

[17] Paulo F. Bedaque and Hyunwoo Oh. Leveraging neural control variates for enhanced precision in lattice field theory. *Phys. Rev. D*, 109(9):094519, 2024.

[18] Hyunwoo Oh. Control variates with neural networks. *PoS*, LATTICE2024:051, 2025.

[19] Hyunwoo Oh. Training neural control variates using correlated configurations. *Phys. Rev. D*, 112(7):074501, 2025.

[20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[21] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57:97–109, 1970.

[22] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195:216–222, 1987.

[23] Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6(6):721–741, 1984.

[24] M. H. Quenouille. Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):68–84, 1949.

[25] M. H. Quenouille. Notes on bias in estimation. *Biometrika*, 43(3/4):353–360, 1956.

[26] Martin Luscher, Stefan Sint, Rainer Sommer, Peter Weisz, and Ulli Wolff. Nonperturbative O(a) improvement of lattice QCD. *Nucl. Phys. B*, 491:323–343, 1997.

[27] William A. Bardeen, A. Duncan, E. Eichten, G. Hockney, and H. Thacker. Light quarks, zero modes, and exceptional configurations. *Phys. Rev. D*, 57:1633–1641, 1998.

[28] G. Schierholz, M. Gockeler, A. Hoferichter, R. Horsley, D. Pleiter, Paul E. L. Rakow, and P. Stephenson. Resolving exceptional configurations in quenched lattice QCD. *Nucl. Phys. B Proc. Suppl.*, 73:889–891, 1999.

[29] L. Del Debbio, Leonardo Giusti, M. Luscher, R. Petronzio, and N. Tantalo. Stability of lattice QCD simulations and the thermodynamic limit. *JHEP*, 02:011, 2006.

[30] Roberto Frezzotti, Pietro Antonio Grassi, Stefan Sint, and Peter Weisz. Lattice QCD with a chirally twisted mass term. *JHEP*, 08:058, 2001.

[31] Martin Luscher and Filippo Palombi. Fluctuations and reweighting of the quark determinant on large lattices. *PoS*, LATTICE2008:049, 2008.

[32] Chuan Miao, Harvey B. Meyer, and Hartmut Wittig. Twisted-mass reweighting for O(a) improved Wilson fermions. *PoS*, LATTICE2011:041, 2011.

[33] Mattia Bruno et al. Simulation of QCD with $N_f = 2 + 1$ flavors of non-perturbatively improved Wilson fermions. *JHEP*, 02:043, 2015.

[34] J. Hubbard. Electron correlations in narrow energy bands. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 276(1365):238–257, 1963.

[35] J. G. Bednorz and K. A. Muller. Possible high Tc superconductivity in the Ba-La-Cu-O system. *Z. Phys. B*, 64:189–193, 1986.

[36] B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, and J. Zaanen. From quantum matter to high-temperature superconductivity in copper oxides. *Nature*, 518(7538):179–186, 2015.

[37] Maksim Ulybyshev and Fakher Assaad. Mitigating spikes in fermion Monte Carlo methods by reshuffling measurements. *Phys. Rev. E*, 106(2):025318, 2022.

[38] Cagin Yunus and William Detmold. Infinite variance in Monte Carlo sampling of lattice field theories. *Phys. Rev. D*, 106(9):094506, 2022.

[39] Hao Shi and Shiwei Zhang. Infinite Variance in Fermion Quantum Monte Carlo Calculations. *Phys. Rev. E*, 93(3):033303, 2016.

[40] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar. Monte Carlo Calculations of Coupled Boson - Fermion Systems. 1. *Phys. Rev. D*, 24:2278, 1981.

[41] Jose H. Blanchet, Nan Chen, and Peter W. Glynn. Unbiased monte carlo computation of smooth functions of expectations via taylor expansions. In *2015 Winter Simulation Conference (WSC)*, pages 360–367, 2015.

[42] Sarat Babu Moka, Dirk P. Kroese, and Sandeep Juneja. Unbiased estimation of the reciprocal mean for non-negative random variables. In *2019 Winter Simulation Conference (WSC)*, pages 404–415, 2019.

[43] Luca Guido Molinari. Determinants of block tridiagonal matrices. *Linear Algebra and its Applications*, 429(8):2221–2226, 2008.

[44] Matthias Troyer and Uwe-Jens Wiese. Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. *Phys. Rev. Lett.*, 94:170201, 2005.

[45] Andrei Alexandru, Gokce Basar, Paulo F. Bedaque, and Neill C. Warrington. Complex paths around the sign problem. *Rev. Mod. Phys.*, 94(1):015006, 2022.

[46] Marco Cristoforetti, Francesco Di Renzo, and Luigi Scorzato. New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble. *Phys. Rev. D*, 86:074506, 2012.

[47] Abhishek Mukherjee, Marco Cristoforetti, and Luigi Scorzato. Metropolis Monte Carlo integration on the Lefschetz thimble: Application to a one-plaquette model. *Phys. Rev. D*, 88(5):051502, 2013.

[48] H. Fujii, D. Honda, M. Kato, Y. Kikukawa, S. Komatsu, and T. Sano. Hybrid Monte Carlo on Lefschetz thimbles - A study of the residual sign problem. *JHEP*, 10:147, 2013.

[49] M. Cristoforetti, F. Di Renzo, G. Eruzzi, A. Mukherjee, C. Schmidt, L. Scorzato, and C. Torrero. An efficient method to compute the residual phase on a Lefschetz thimble. *Phys. Rev. D*, 89(11):114505, 2014.

[50] Yuya Tanizaki. Lefschetz-thimble techniques for path integral of zero-dimensional $O(n)$ sigma models. *Phys. Rev. D*, 91(3):036002, 2015.

[51] Yuya Tanizaki, Yoshimasa Hidaka, and Tomoya Hayata. Lefschetz-thimble analysis of the sign problem in one-site fermion model. *New J. Phys.*, 18(3):033002, 2016.

[52] Hirotsugu Fujii, Syo Kamata, and Yoshio Kikukawa. Lefschetz thimble structure in one-dimensional lattice Thirring model at finite density. *JHEP*, 11:078, 2015. [Erratum: JHEP 02, 036 (2016)].

[53] Hirotsugu Fujii, Syo Kamata, and Yoshio Kikukawa. Monte Carlo study of Lefschetz thimble structure in one-dimensional Thirring model at finite density. *JHEP*, 12:125, 2015. [Erratum: JHEP 09, 172 (2016)].

[54] Andrei Alexandru, Gökçe Basar, and Paulo Bedaque. Monte Carlo algorithm for simulating fermions on Lefschetz thimbles. *Phys. Rev. D*, 93(1):014504, 2016.

[55] Abhishek Mukherjee and Marco Cristoforetti. Lefschetz thimble Monte Carlo for many-body theories: A Hubbard model study. *Phys. Rev. B*, 90(3):035134, 2014.

[56] Yuya Tanizaki and Takayuki Koike. Real-time Feynman path integral with Picard–Lefschetz theory and its applications to quantum tunneling. *Annals Phys.*, 351:250–274, 2014.

[57] Takuya Kanazawa and Yuya Tanizaki. Structure of Lefschetz thimbles in simple fermionic systems. *JHEP*, 03:044, 2015.

[58] Yuya Tanizaki, Hiromichi Nishimura, and Kouji Kashiwa. Evading the sign problem in the mean-field approximation through Lefschetz-thimble path integral. *Phys. Rev. D*, 91(10):101701, 2015.

[59] Francesco Di Renzo and Giovanni Eruzzi. Thimble regularization at work: from toy models to chiral random matrix theories. *Phys. Rev. D*, 92(8):085030, 2015.

[60] Andrei Alexandru, Gokce Basar, Paulo F. Bedaque, Gregory W. Ridgway, and Neill C. Warrington. Fast estimator of Jacobians in the Monte Carlo integration on Lefschetz thimbles. *Phys. Rev. D*, 93(9):094514, 2016.

[61] Andrei Alexandru, Gokce Basar, Paulo Bedaque, Gregory W. Ridgway, and Neill C. Warrington. Study of symmetry breaking in a relativistic Bose gas using the contraction algorithm. *Phys. Rev. D*, 94(4):045017, 2016.

[62] Edward Witten. Analytic Continuation Of Chern-Simons Theory. *AMS/IP Stud. Adv. Math.*, 50:347–446, 2011.

[63] Edward Witten. A New Look At The Path Integral Of Quantum Mechanics. 9 2010.

[64] Andrei Alexandru, Gokce Basar, Paulo F. Bedaque, Gregory W. Ridgway, and Neill C. Warrington. Sign problem and Monte Carlo calculations beyond Lefschetz thimbles. *JHEP*, 05:053, 2016.

[65] Andrei Alexandru, Gokce Basar, Paulo F. Bedaque, Gregory W. Ridgway, and Neill C. Warrington. Monte Carlo calculations of the finite density Thirring model. *Phys. Rev. D*, 95(1):014502, 2017.

[66] Andrei Alexandru, Gokce Basar, Paulo F. Bedaque, and Neill C. Warrington. Tempered transitions between thimbles. *Phys. Rev. D*, 96(3):034513, 2017.

[67] Masafumi Fukuma and Naoya Umeda. Parallel tempering algorithm for integration over Lefschetz thimbles. *PTEP*, 2017(7):073B01, 2017.

[68] Andrei Alexandru, Paulo F. Bedaque, Henry Lamm, and Scott Lawrence. Finite-Density Monte Carlo Calculations on Sign-Optimized Manifolds. *Phys. Rev. D*, 97(9):094510, 2018.

[69] Andrei Alexandru, Paulo F. Bedaque, Henry Lamm, Scott Lawrence, and Neill C. Warrington. Fermions at Finite Density in 2+1 Dimensions with Sign-Optimized Manifolds. *Phys. Rev. Lett.*, 121(19):191602, 2018.

[70] Yuto Mori, Kouji Kashiwa, and Akira Ohnishi. Toward solving the sign problem with path optimization method. *Phys. Rev. D*, 96(11):111501, 2017.

[71] M. S. Albergo, G. Kanwar, and P. E. Shanahan. Flow-based generative models for Markov chain Monte Carlo in lattice field theory. *Phys. Rev. D*, 100(3):034515, 2019.

[72] Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan. Equivariant flow-based sampling for lattice gauge theory. *Phys. Rev. Lett.*, 125(12):121601, 2020.

[73] Denis Boyda, Gurtej Kanwar, Sébastien Racanière, Danilo Jimenez Rezende, Michael S. Albergo, Kyle Cranmer, Daniel C. Hackett, and Phiala E. Shanahan. Sampling using $SU(N)$ gauge equivariant flows. *Phys. Rev. D*, 103(7):074504, 2021.

[74] Scott Lawrence and Yukari Yamauchi. Normalizing Flows and the Real-Time Sign Problem. *Phys. Rev. D*, 103(11):114509, 2021.

[75] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

[76] Daniel C. Hackett, Chung-Chun Hsieh, Sahil Pontula, Michael S. Albergo, Denis Boyda, Jiunn-Wei Chen, Kai-Feng Chen, Kyle Cranmer, Gurtej Kanwar, and Phiala E. Shanahan. Flow-based sampling for multimodal and extended-mode distributions in lattice field theory. 7 2021.

[77] Carl M. Bender and Stefan Boettcher. Real spectra in nonHermitian Hamiltonians having PT symmetry. *Phys. Rev. Lett.*, 80:5243–5246, 1998.

[78] Carl M. Bender, Dorje C. Brody, and Hugh F. Jones. Extension of PT symmetric quantum mechanics to quantum field theory with cubic interaction. *Phys. Rev. D*, 70:025001, 2004. [Erratum: Phys.Rev.D 71, 049901 (2005)].

[79] Jean Alexandre, John Ellis, Peter Millington, and Dries Seynaeve. Spontaneous symmetry breaking and the Goldstone theorem in non-Hermitian field theories. *Phys. Rev. D*, 98:045001, 2018.

[80] Olalla A. Castro-Alvaredo, Benjamin Doyon, and Takato Yoshimura. Emergent hydrodynamics in integrable quantum systems out of equilibrium. *Phys. Rev. X*, 6(4):041065, 2016.

[81] Peter N. Meisinger and Michael C. Ogilvie. PT Symmetry in Classical and Quantum Statistical Mechanics. *Phil. Trans. Roy. Soc. Lond. A*, 371:20120058, 2013.

[82] Qian Du, Kui Cao, and Su-Peng Kou. Physics of PT-symmetric quantum systems at finite temperatures. *Phys. Rev. A*, 106(3):032206, 2022.

[83] Michael C. Ogilvie, Peter N. Meisinger, and Timothy D. Wiser. $\mathcal{PT}$ symmetry in statistical mechanics and the sign problem. *International Journal of Theoretical Physics*, 50(4):1042–1051, Apr 2011.

[84] R. El-Ganainy, K. G. Makris, D. N. Christodoulides, and Ziad H. Musslimani. Theory of coupled optical *PT*-symmetric structures. *Opt. Lett.*, 32(17):2632–2634, Sep 2007.

[85] Liang Feng, Ramy El-Ganainy, and Li Ge. Non-Hermitian photonics based on parity–time symmetry. *Nature Photon.*, 11(12):752–762, 2017.

[86] Ş. K. Özdemir, S. Rotter, F. Nori, and L. Yang. Parity–time symmetry and exceptional points in photonics. *Nature Materials*, 18(8):783–798, 2019.

[87] Stefano Longhi. $\mathcal{PT}$-symmetric laser absorber. *Phys. Rev. A*, 82:031801, Sep 2010.

[88] Peter Rabl, Stefan Rotter, Peter Kirton, and Julian Huber. Emergence of PT-symmetry breaking in open quantum systems. *SciPost Phys.*, 9(4):052, 2020.

[89] Yuma Nakanishi and Tomohiro Sasamoto. PT phase transition in open quantum systems with Lindblad dynamics. *Phys. Rev. A*, 105(2):022219, 2022.

[90] Elias Starchl and Lukas M. Sieberer. Quantum quenches in driven-dissipative quadratic fermionic systems with parity-time symmetry. *Phys. Rev. Res.*, 6(1):013016, 2024.

[91] Jiaming Li, Andrew K. Harter, Ji Liu, Leonardo de Melo, Yogesh N. Joglekar, and Le Luo. Observation of parity-time symmetry breaking transitions in a dissipative Floquet system of ultracold atoms. *Nature Commun.*, 10(1):855, 2019.

[92] Yang Wu, Wenquan Liu, Jianpei Geng, Xingrui Song, Xiangyu Ye, Chang-Kui Duan, Xing Rong, and Jiangfeng Du. Observation of parity-time symmetry breaking in a single-spin system. *Science*, 364(6443):878–880, 2019.

[93] Scott Lawrence, Ryan Weller, Christian Peterson, and Paul Romatschke. Instantons, analytic continuation, and PT-symmetric field theory. *Phys. Rev. D*, 108(8):085013, 2023.

[94] Julian S. Schwinger. Brownian motion of a quantum oscillator. *J. Math. Phys.*, 2:407–432, 1961.

[95] L. V. Keldysh. Diagram technique for nonequilibrium processes. *Zh. Eksp. Teor. Fiz.*, 47:1515–1527, 1964.

[96] Andrei Alexandru, Gokce Basar, Paulo F. Bedaque, Sohan Vartak, and Neill C. Warrington. Monte Carlo Study of Real Time Dynamics on the Lattice. *Phys. Rev. Lett.*, 117(8):081602, 2016.

[97] Andrei Alexandru, Gokce Basar, Paulo F. Bedaque, and Gregory W. Ridgway. Schwinger-Keldysh formalism on the lattice: A faster algorithm and its application to field theory. *Phys. Rev. D*, 95(11):114501, 2017.

[98] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, San Diega, CA, USA, 2015.

[99] C. N. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions. i. theory of condensation. *Phys. Rev.*, 87:404–409, Aug 1952.

[100] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. ii. lattice gas and ising model. *Phys. Rev.*, 87:410–419, Aug 1952.

[101] Radford M. Neal. Annealed importance sampling. *Stat. Comput.*, 11(2):125–139, 2001.

[102] Barry Simon and Robert B. Griffiths. Griffiths-hurst-sherman inequalities and a lee-yang therorem for the $(\phi^4)_2$ field theory. *Phys. Rev. Lett.*, 30:931–933, May 1973.

[103] G. Parisi. The Strategy for Computing the Hadronic Mass Spectrum. *Phys. Rept.*, 103:203–211, 1984.

[104] G. Peter Lepage. The Analysis of Algorithms for Lattice Field Theory. In *Theoretical Advanced Study Institute in Elementary Particle Physics*, 6 1989.

[105] Tanmoy Bhattacharya, Scott Lawrence, and Jun-Sik Yoo. Control variates for lattice field theory. *Phys. Rev. D*, 109(3):L031505, 2024.

[106] Antonietta Mira, Reza Solgi, and Daniele Imparato. Zero variance markov chain monte carlo for bayesian estimators. *Statistics and Computing*, 23(5):653–662, 2013.

[107] Chris J. Oates, Theodore Papamarkou, and Mark Girolami and. The controlled thermodynamic integral for bayesian model evidence evaluation. *Journal of the American Statistical Association*, 111(514):634–645, 2016.

[108] Chris J. Oates, Mark Girolami, and Nicolas Chopin. Control functionals for monte carlo integration. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(3):695–718, 05 2016.

[109] Chris J. Oates, Jon Cockayne, François-Xavier Briol, and Mark Girolami. Convergence rates for a class of estimators based on stein's method. *Bernoulli*, 25(2):1141–1159, 2019.

[110] Alessandro Barp, François-Xavier Briol, Andrew B. Duncan, Mark Girolami, and Lester Mackey. Minimum stein discrepancy estimators. In *Advances in Neural Information Processing Systems (NeurIPS) 2019*, pages 8546–8557, 2019.

[111] L F South, T Karvonen, C Nemeth, M Girolami, and C J Oates. Semi-exact control functionals from sard's method. *Biometrika*, 109(2):351–367, 09 2021.

[112] Leah F. South, Chris J. Oates, Antonietta Mira, and Christopher Drovandi. Regularized zero-variance control variates. *Bayesian Analysis*, 18(3):865–888, 2023. Available under Creative Commons Attribution 4.0 International License (CC BY 4.0).

[113] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 583–602. University of California Press, 1972.

[114] Scott Lawrence. Schwinger-Dyson control variates for lattice fermions. 4 2024.

[115] Roland Assaraf and Michel Caffarel. Zero-variance principle for monte carlo algorithms. *Phys. Rev. Lett.*, 83:4682–4685, Dec 1999.

[116] Roland Assaraf and Michel Caffarel. Zero-variance zero-bias principle for observables in quantum monte carlo: Application to forces. *The Journal of Chemical Physics*, 119(20):10536–10552, 11 2003.

[117] Shijing Si, Chris. J. Oates, Andrew B. Duncan, Lawrence Carin, and François-Xavier Briol. Scalable control variates for monte carlo methods via stochastic optimization. In Alexander Keller, editor, *Monte Carlo and Quasi-Monte Carlo Methods*, pages 205–221, Cham, 2022. Springer International Publishing.

[118] William Detmold, Gurtej Kanwar, Michael L. Wagman, and Neill C. Warrington. Path integral contour deformations for noisy observables. *Phys. Rev. D*, 102(1):014514, 2020.

[119] William Detmold, Gurtej Kanwar, Henry Lamm, Michael L. Wagman, and Neill C. Warrington. Path integral contour deformations for observables in $SU(N)$ gauge theory. *Phys. Rev. D*, 103(9):094517, 2021.

[120] Yin Lin, William Detmold, Gurtej Kanwar, Phiala E. Shanahan, and Michael L. Wagman. Signal-to-noise improvement through neural network contour deformations for 3D $SU(2)$ lattice gauge theory. *PoS*, LATTICE2023:043, 2024.

[121] Scott Lawrence and Yukari Yamauchi. Convex optimization of contour deformations. *Phys. Rev. D*, 110(1):014508, 2024.

[122] Scott Lawrence. Perturbative Removal of a Sign Problem. *Phys. Rev. D*, 102(9):094504, 2020.

[123] Scott Lawrence and Yukari Yamauchi. Deep learning of fermion sign fluctuations. *Phys. Rev. D*, 107(11):114505, 2023.

[124] Christoph Gäntgen, Evan Berkowitz, Thomas Luu, Johann Ostmeyer, and Marcel Rodekamp. Fermionic sign problem minimization by constant path integral contour shifts. *Phys. Rev. B*, 109(19):195158, 2024.

[125] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[126] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.*, 2(4):303–314, 1989.

[127] Ruosi Wan, Mingjun Zhong, Haoyi Xiong, and Zhanxing Zhu. Neural control variates for monte carlo variance reduction. In Ulf Brefeld, Elisa Fromont, Andreas Hotho, Arno Knobbe, Marloes Maathuis, and Céline Robardet, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 533–547, Cham, 2020. Springer International Publishing.

[128] Zhuo Sun, Alessandro Barp, and François-Xavier Briol. Vector-valued control variates. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

[129] Zhuo Sun, Chris J. Oates, and François-Xavier Briol. Meta-learning control variates: variance reduction with limited data. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, UAI '23. JMLR.org, 2023.

[130] Katharina Ott, Michael Tiemann, Philipp Hennig, and François-Xavier Briol. Bayesian numerical integration with neural networks. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, UAI '23. JMLR.org, 2023.

[131] Ali Siahkoohi and Hyunwoo Oh. Conditional neural control variates for variance reduction in Bayesian inverse problems. 2 2026.

[132] Will Grathwohl, Dami Choi, Yuhuai Wu, Geoff Roeder, and David Duvenaud. Backpropagation through the void: Optimizing control variates for black-box gradient estimation. In *International Conference on Learning Representations*, 2018.

[133] Thomas Müller, Fabrice Rousselle, Alexander Keller, and Jan Novák. Neural control variates. *ACM Trans. Graph.*, 39(6), November 2020.

[134] L. F. South, C. J. Oates, A. Mira, and C. Drovandi. Regularized Zero-Variance Control Variates. *Bayesian Analysis*, 18(3):865 – 888, 2023.

[135] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

[136] Scott Lawrence and Yukari Yamauchi. Mitigating a discrete sign problem with extreme learning machines. 12 2023.

[137] Guilherme Catumba and Alberto Ramos. Stochastic automatic differentiation and the Signal to Noise problem. 2 2025.

[138] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, page III–1310–III–1318. JMLR.org, 2013.

[139] Jonathan T. Barron. Continuously differentiable exponential linear units, 2017.

[140] Richard P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21:467–488, 1982.

[141] Peter W. Shor. Polynomial time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Statist. Comput.*, 26:1484, 1997.

[142] H. W. Lenstra. Factoring integers with elliptic curves. *Annals of Mathematics*, 126(3):649–673, 1987.

[143] Carl Pomerance. Analysis and comparison of some integer factoring algorithms. *Computational Methods in Number Theory*, 0:89–139, 1982.

[144] Carl Pomerance. A tale of two sieves. *Notices of the American Mathematical Society*, 43(12):1473–1485, 1996.

[145] Richard P. Brent. An improved monte carlo factorization algorithm. *BIT Numerical Mathematics*, 20(2):176–184, Jun 1980.

[146] E. C. Behrman, Gary A. Jongeward, and Peter G. Wolynes. A Monte Carlo approach for the real time dynamics of tunneling systems in condensed phases. *The Journal of Chemical Physics*, 79(12):6277–6281, 12 1983.

[147] Mark R. Dowling, Matthew J. Davis, Peter D. Drummond, and Joel F. Corney. Monte carlo techniques for real-time quantum dynamics. *Journal of Computational Physics*, 220(2):549–567, 2007.

[148] Philippe de Forcrand. Simulating QCD at finite density. *PoS*, LAT2009:010, 2009.

[149] Gert Aarts. Introductory lectures on lattice QCD at nonzero baryon number. *J. Phys. Conf. Ser.*, 706(2):022004, 2016.

[150] A. Yu. Kitaev. Quantum measurements and the Abelian stabilizer problem. 11 1995.

[151] Nicholas J. Ward, Ivan Kassal, and Alán Aspuru-Guzik. Preparation of many-body states for quantum simulation. *The Journal of Chemical Physics*, 130(19), 05 2009. 194105.

[152] Daniel S. Abrams and Seth Lloyd. Simulation of many body Fermi systems on a universal quantum computer. *Phys. Rev. Lett.*, 79:2586–2589, 1997.

[153] R. Somma, G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme. Simulating physical phenomena by quantum networks. *Phys. Rev. A*, 65:042323, Apr 2002.

[154] Joseph C. Aulicino, Trevor Keen, and Bo Peng. State preparation and evolution in quantum computing: A perspective from hamiltonian moments. *International Journal of Quantum Chemistry*, 122(5):e26853, 2022.

[155] Kishor Bharti et al. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94(1):015004, 2022.

[156] Xiao-Ming Zhang, Tongyang Li, and Xiao Yuan. Quantum state preparation with optimal circuit depth: Implementations and applications. *Phys. Rev. Lett.*, 129:230504, Nov 2022.

[157] Almudena Carrera Vazquez and Stefan Woerner. Efficient State Preparation for Quantum Amplitude Estimation. *Phys. Rev. Applied*, 15(3):034027, 2021.

[158] Dean Lee, Joey Bonitati, Gabriel Given, Caleb Hicks, Ning Li, Bing-Nan Lu, Abudit Rai, Avik Sarkar, and Jacob Watkins. Projected Cooling Algorithm for Quantum Computation. *Phys. Lett. B*, 807:135536, 2020.

[159] B. Kraus, H. P. Büchler, S. Diehl, A. Kantian, A. Micheli, and P. Zoller. Preparation of entangled states by quantum Markov processes. *Phys. Rev. A*, 78(4):042307, 2008.

[160] David B. Kaplan, Natalie Klco, and Alessandro Roggero. Ground States via Spectral Combing on a Quantum Computer. 9 2017.

[161] Erik Gustafson. Projective Cooling for the transverse Ising model. *Phys. Rev. D*, 101(7):071504, 2020.

[162] Zohreh Davoudi, Chung-Chun Hsieh, and Saurabh V. Kadam. Scattering wave packets of hadrons in gauge theories: Preparation on a quantum computer. *Quantum*, 8:1520, 2024.

[163] Anthony N. Ciavarella, Stephan Caspar, Hersh Singh, and Martin J. Savage. Preparation for quantum simulation of the (1+1)-dimensional O(3) nonlinear σ model using cold atoms. *Phys. Rev. A*, 107(4):042404, 2023.

[164] Roland C. Farrell, Marc Illa, Anthony N. Ciavarella, and Martin J. Savage. Scalable Circuits for Preparing Ground States on Digital Quantum Computers: The Schwinger Model Vacuum on 100 Qubits. *PRX Quantum*, 5(2):020315, 2024.

[165] Alessandro Roggero, Chenyi Gu, Alessandro Baroni, and Thomas Papenbrock. Preparation of excited states for nuclear dynamics on a quantum computer. *Phys. Rev. C*, 102(6):064624, 2020.

[166] Anthony N. Ciavarella and Ivan A. Chernyshev. Preparation of the SU(3) lattice Yang-Mills vacuum with variational quantum methods. *Phys. Rev. D*, 105(7):074504, 2022.

[167] Ali Hamed Moosavian, James R. Garrison, and Stephen P. Jordan. Site-by-site quantum state preparation algorithm for preparing vacua of fermionic lattice field theories. 11 2019.

[168] Natalie Klco and Martin J. Savage. Minimally entangled state preparation of localized wave functions on quantum computers. *Phys. Rev. A*, 102(1):012612, 2020.

[169] Christopher F. Kane, Niladri Gomes, and Michael Kreshchuk. Nearly optimal state preparation for quantum simulations of lattice gauge theories. *Phys. Rev. A*, 110(1):012455, 2024.

[170] Alexander J. Buser, Tanmoy Bhattacharya, Lukasz Cincio, and Rajan Gupta. State preparation and measurement in a quantum simulation of the $O(3)$ sigma model. *Phys. Rev. D*, 102(11):114514, 2020.

[171] Luca Lumia, Pietro Torta, Glen B. Mbeng, Giuseppe E. Santoro, Elisa Ercolessi, Michele Burrello, and Matteo M. Wauters. Two-Dimensional Z2 Lattice Gauge Theory on a Near-Term Quantum Simulator: Variational Quantum Optimization, Confinement, and Topological Order. *PRX Quantum*, 3(2):020320, 2022.

[172] Ali Hamed Moosavian and Stephen Jordan. Faster Quantum Algorithm to simulate Fermionic Quantum Field Theory. *Phys. Rev. A*, 98(1):012332, 2018.

[173] Carter Ball and Thomas D. Cohen. Boltzmann distributions on a quantum computer via active cooling. *Nucl. Phys. A*, 1038:122708, 2023.

[174] Zohreh Davoudi, Niklas Mueller, and Connor Powers. Towards Quantum Computing Phase Diagrams of Gauge Theories with Thermal Pure Quantum States. *Phys. Rev. Lett.*, 131(8):081901, 2023.

[175] Fernando G. S. L. Brandão and Michael J. Kastoryano. Finite Correlation Length Implies Efficient Preparation of Quantum Thermal States. *Commun. Math. Phys.*, 365(1):1–16, 2019.

[176] Wibe A. de Jong, Kyle Lee, James Mulligan, Mateusz Płoskoń, Felix Ringer, and Xiaojun Yao. Quantum simulation of nonequilibrium dynamics and thermalization in the Schwinger model. *Phys. Rev. D*, 106(5):054508, 2022.

[177] Chi-Fang, Chen, Michael J. Kastoryano, Fernando G. S. L. Brandão, and András Gilyén. Quantum Thermal State Preparation. 3 2023.

[178] Zoe Holmes, Gopikrishnan Muraleedharan, Rolando D. Somma, Yigit Subasi, and Burak Şahinoğlu. Quantum algorithms from fluctuation theorems: Thermal-state preparation. *Quantum*, 6:825, October 2022.

[179] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Symposium on the Theory of Computing*, 1996.

[180] Thomas D. Cohen and Hyunwoo Oh. Optimizing the rodeo projection algorithm. *Phys. Rev. A*, 108(3):032422, 2023.

[181] Thomas D. Cohen and Hyunwoo Oh. Efficient vacuum-state preparation for quantum simulation of strongly interacting local quantum field theories. *Phys. Rev. A*, 109(2):L020402, 2024.

[182] Thomas D. Cohen and Hyunwoo Oh. Asymptotic errors in adiabatic evolution. *Phys. Rev. A*, 111(4):042612, 2025.

[183] Thomas D. Cohen, Andrew Li, Hyunwoo Oh, and Maneesha Sushama Pradeep. Practical limitations of the switching theorem for adiabatic state preparation. *Eur. Phys. J. D*, 80(4):41, 2026.

[184] Thomas D. Cohen and Hyunwoo Oh. Corrections to adiabatic behavior for long paths. *Phys. Rev. A*, 110(6):062601, 2024.

[185] Thomas D. Cohen, Hyunwoo Oh, and Veronica Wang. Numerical study of computational cost of maintaining adiabaticity for long paths. 12 2024.

[186] Zhengrong Qian, Jacob Watkins, Gabriel Given, Joey Bonitati, Kenneth Choi, and Dean Lee. Demonstration of the rodeo algorithm on a quantum computer. *Eur. Phys. J. A*, 60(7):151, 2024.

[187] Max Bee-Lindgren, Zhengrong Qian, Matthew DeCross, Natalie C. Brown, Christopher N. Gilbreth, Jacob Watkins, Xilin Zhang, and Dean Lee. Controlled Gate Networks Applied to Eigenvalue Estimation. 8 2022.

[188] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution, 2000.

[189] Dorit Aharonov and Amnon Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge, 2003.

[190] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev. Adiabatic quantum computation is equivalent to standard quantum computation. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 42–51, 2004.

[191] Dorit Aharonov and Amnon Ta-Shma. Adiabatic quantum state generation. *SIAM Journal on Computing*, 37(1):47–82, 2007.

[192] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 90(1):015002, 2018.

[193] H. F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959.

[194] M. Suzuki. Generalized Trotter's Formula and Systematic Approximants of Exponential Operators and Inner Derivations with Applications to Many Body Problems. *Commun. Math. Phys.*, 51:183–190, 1976.

[195] Kianna Wan and Isaac H. Kim. Fast digital methods for adiabatic state preparation. 4 2020.

[196] Eduardo A. Coello Pérez, Joey Bonitati, Dean Lee, Sofia Quaglioni, and Kyle A. Wendt. Quantum state preparation by adiabatic evolution with custom gates. *Phys. Rev. A*, 105(3):032403, 2022.

[197] Anthony N. Ciavarella, Stephan Caspar, Marc Illa, and Martin J. Savage. State Preparation in the Heisenberg Model through Adiabatic Spiraling. *Quantum*, 7:970, 2023.

[198] Bipasha Chakraborty, Masazumi Honda, Taku Izubuchi, Yuta Kikuchi, and Akio Tomiya. Classically emulated digital quantum simulation of the Schwinger model with a topological term via adiabatic state preparation. *Phys. Rev. D*, 105(9):094503, 2022.

[199] Lucas K. Kovalsky, Fernando A. Calderon-Vargas, Matthew D. Grace, Alicia B. Magann, James B. Larsen, Andrew D. Baczewski, and Mohan Sarovar. Self-Healing of Trotter Error in Digital Adiabatic State Preparation. *Phys. Rev. Lett.*, 131(6):060602, 2023.

[200] Matteo D'Anna, Marina Krstic Marinkovic, and Joao C. Pinto Barros. Adiabatic state preparation for digital quantum simulations of QED in 1 + 1D. 11 2024.

[201] Sergio Boixo, Emanuel Knill, and Rolando Somma. Eigenpath traversal by phase randomization. *Quantum Info. Comput.*, 9(9):833–855, sep 2009.

[202] S. Boixo, E. Knill, and R. D. Somma. Fast quantum algorithms for traversing paths of eigenstates, 2010.

[203] Hao-Tien Chiang, Guanglei Xu, and Rolando D. Somma. Improved bounds for eigenpath traversal. *Phys. Rev. A*, 89:012314, Jan 2014.

[204] Gerardo A. Paz-Silva, A. T. Rezakhani, Jason M. Dominy, and D. A. Lidar. Zeno effect for quantum computation and control. *Phys. Rev. Lett.*, 108:080501, Feb 2012.

[205] Alfredo Luis. Quantum-state preparation and control via the zeno effect. *Phys. Rev. A*, 63:052112, Apr 2001.

[206] Michael V. Berry. Quantal phase factors accompanying adiabatic changes. *Proc. Roy. Soc. Lond. A*, 392:45–57, 1984.

[207] J. E. Avron and A. Elgart. Adiabatic theorem without a gap condition. *Commun. Math. Phys.*, 203:445–463, 1999.

[208] S. Boixo and R. D. Somma. Necessary condition for the quantum adiabatic approximation. *Phys. Rev. A*, 81:032308, Mar 2010.

[209] David Poulin and Pawel Wocjan. Preparing ground states of quantum many-body systems on a quantum computer. *Phys. Rev. Lett.*, 102:130503, 2009.

[210] Yimin Ge, Jordi Tura, and J. Ignacio Cirac. Faster ground state preparation and high-precision ground energy estimation with fewer qubits. *J. Math. Phys.*, 60(2):022202, 2019.

[211] Yulong Dong, Lin Lin, and Yu Tong. Ground-State Preparation and Energy Estimation on Early Fault-Tolerant Quantum Computers via Quantum Eigenvalue Transformation of Unitary Matrices. *PRX Quantum*, 3(4):040305, 2022.

[212] Lin Lin and Yu Tong. Near-optimal ground state preparation. *Quantum*, 4:372, 2020.

[213] I. Stetcu, A. Baroni, and J. Carlson. Projection algorithm for state preparation on quantum computers. *Phys. Rev. C*, 108(3):L031306, 2023.

[214] M. Born and V. Fock. Beweis des adiabatensatzes. *Zeitschrift für Physik*, 51(3):165–180, Mar 1928.

[215] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[216] Matthias Vojta. Quantum phase transitions. *Reports on Progress in Physics*, 66(12):2069, nov 2003.

[217] M. Z. Hasan and C. L. Kane. Topological Insulators. *Rev. Mod. Phys.*, 82:3045, 2010.

[218] Subir Sachdev. *Quantum Phase Transitions*. Cambridge University Press, 4 2011.

[219] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem. *Science*, 292(5516):1057726, 2001.

[220] Andrew M. Childs, Edward Farhi, and John Preskill. Robustness of adiabatic quantum computation. *Phys. Rev. A*, 65:012322, 2002.

[221] Jérémie Roland and Nicolas J. Cerf. Quantum search by local adiabatic evolution. *Phys. Rev. A*, 65(4):042308, 2002.

[222] B. Apolloni, C. Carvalho, and D. de Falco. Quantum stochastic optimization. *Stochastic Processes and their Applications*, 33(2):233–244, 1989.

[223] W. van Dam, M. Mosca, and U. Vazirani. How powerful is adiabatic quantum computation? In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 279–287, 2001.

[224] Tosio Kato. On the adiabatic theorem of quantum mechanics. *Journal of the Physical Society of Japan*, 5(6):435–439, 1950.

[225] Albert Messiah. *Quantum Mechanics Volume II*. Elsevier Science B.V., 1961.

[226] Daniel Comparat. General conditions for quantum adiabatic evolution. *Phys. Rev. A*, 80:012106, Jul 2009.

[227] Donny Cheung, Peter Høyer, and Nathan Wiebe. Improved error bounds for the adiabatic approximation. *Journal of Physics A: Mathematical and Theoretical*, 44(41):415302, sep 2011.

[228] Evgeny Mozgunov and Daniel A. Lidar. Quantum adiabatic theorem for unbounded Hamiltonians with a cutoff and its application to superconducting circuits. *Phil. Trans. A. Math. Phys. Eng. Sci.*, 381(2241):20210407, 2022.

[229] Daniel Burgarth, Paolo Facchi, Giovanni Gramegna, and Kazuya Yuasa. One bound to rule them all: from Adiabatic to Zeno. *Quantum*, 6:737, June 2022.

[230] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum Computation of Scattering in Scalar Quantum Field Theories. *Quant. Inf. Comput.*, 14:1014–1080, 2014.

[231] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum Algorithms for Quantum Field Theories. *Science*, 336:1130–1133, 2012.

[232] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum Algorithms for Fermionic Quantum Field Theories. 4 2014.

[233] Junyu Liu and Yuan Xin. Quantum simulation of quantum field theories as quantum chemistry. *JHEP*, 12:011, 2020.

[234] Hrant Gharibyan, Masanori Hanada, Masazumi Honda, and Junyu Liu. Toward simulating superstring/M-theory on a quantum computer. *JHEP*, 07:140, 2021.

[235] Michael Kreshchuk, William M. Kirby, Gary Goldstein, Hugo Beauchemin, and Peter J. Love. Quantum simulation of quantum field theory in the light-front formulation. *Phys. Rev. A*, 105(3):032418, 2022.

[236] Masazumi Honda, Etsuko Itou, Yuta Kikuchi, Lento Nagano, and Takuya Okuda. Classically emulated digital quantum simulation for screening and confinement in the Schwinger model with a topological term. *Phys. Rev. D*, 105(1):014504, 2022.

[237] Andy C. Y. Li, Alexandru Macridin, Stephen Mrenna, and Panagiotis Spentzouris. Simulating scalar field theories on quantum computers with limited resources. *Phys. Rev. A*, 107(3):032603, 2023.

[238] Matteo Turco, Gonçalo M. Quinta, João Seixas, and Yasser Omar. Quantum Simulation of Bound State Scattering. *PRX Quantum*, 5(2):020311, 2024.

[239] Roland C. Farrell, Marc Illa, Anthony N. Ciavarella, and Martin J. Savage. Quantum Simulations of Hadron Dynamics in the Schwinger Model using 112 Qubits. 1 2024.

[240] Andrew Lenard. Adiabatic invariance to all orders. *Annals of Physics*, 6(3):261–276, 1959.

[241] L.M. Garrido and F.J. Sancho. Degree of approximate validity of the adiabatic invariance in quantum mechanics. *Physica*, 28(6):553–560, 1962.

[242] L. M. Garrido. Generalized Adiabatic Invariance. *Journal of Mathematical Physics*, 5(3):355–362, 03 1964.

[243] F J Sancho. mth-order adiabatic invariance for quantum systems. *Proceedings of the Physical Society*, 89(1):1, sep 1966.

[244] G. Nenciu. Adiabatic theorem and spectral concentration. *Communications in Mathematical Physics*, 82(1):121–135, Mar 1981.

[245] J. E. Avron, R. Seiler, and L. G. Yaffe. Adiabatic theorems and applications to the quantum hall effect. *Communications in Mathematical Physics*, 110(1):33–49, Mar 1987.

[246] Alain Joye and Charles-Edouard Pfister. Exponentially small adiabatic invariant for the schroedinger equation. *Communications in Mathematical Physics*, 140, 01 1991.

[247] G. Nenciu. *Asymptotic Invariant Subspaces, Adiabatic Theorems and Block Diagonalisation*, pages 133–149. Springer Netherlands, Dordrecht, 1991.

[248] G. Nenciu. Linear adiabatic theory. exponential estimates. *Communications in Mathematical Physics*, 152(3):479–496, Mar 1993.

[249] George A. Hagedorn and Alain Joye. Elementary exponential error estimates for the adiabatic approximation. *Journal of Mathematical Analysis and Applications*, 267(1):235–246, 2002.

[250] Daniel A. Lidar, Ali T. Rezakhani, and Alioscia Hamma. Adiabatic approximation with exponential accuracy for many-body systems and quantum computation. *Journal of Mathematical Physics*, 50(10):102106, 10 2009.

[251] A. T. Rezakhani, A. K. Pimachev, and D. A. Lidar. Accuracy versus run time in an adiabatic quantum search. *Phys. Rev. A*, 82:052305, Nov 2010.

[252] Alexander Elgart and George A. Hagedorn. A note on the switching adiabatic theorem. *Journal of Mathematical Physics*, 53(10):102202, 09 2012.

[253] Yimin Ge, András Molnár, and J. Ignacio Cirac. Rapid adiabatic preparation of injective projected entangled pair states and gibbs states. *Phys. Rev. Lett.*, 116:080503, Feb 2016.

[254] R. MacKenzie, E. Marcotte, and H. Paquette. Perturbative approach to the adiabatic approximation. *Phys. Rev. A*, 73:042104, Apr 2006.

[255] Gustavo Rigolin, Gerardo Ortiz, and Victor Hugo Ponce. Beyond the quantum adiabatic approximation: Adiabatic perturbation theory. *Phys. Rev. A*, 78:052508, Nov 2008.

[256] M.R. Passos, M.M. Taddei, and R.L. de Matos Filho. Error-run-time trade-off in the adiabatic approximation beyond scaling relations. *Annals of Physics*, 418:168172, 2020.

[257] A. P. Young, S. Knysh, and V. N. Smelyanskiy. First-order phase transition in the quantum adiabatic algorithm. *Phys. Rev. Lett.*, 104:020502, Jan 2010.

[258] Thomas Jörg, Florent Krzakala, Guilhem Semerjian, and Francesco Zamponi. First-order transitions and the performance of quantum algorithms in random optimization problems. *Phys. Rev. Lett.*, 104:207206, May 2010.

[259] Boris Altshuler, Hari Krovi, and Jeremie Roland. Adiabatic quantum optimization fails for random instances of np-complete problems, 2009.

[260] M. H. S. Amin and V. Choi. First-order quantum phase transition in adiabatic quantum computation. *Phys. Rev. A*, 80:062326, Dec 2009.

[261] R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. Las Heras, R. Babbush, A. G. Fowler, B. Campbell, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, E. Solano, H. Neven, and John M. Martinis. Digitized adiabatic quantum computing with a superconducting circuit. *Nature*, 534(7606):222–226, Jun 2016.

[262] Yifan Sun, Jun-Yi Zhang, Mark S Byrd, and Lian-Ao Wu. Trotterized adiabatic quantum simulation and its application to a simple all-optical system. *New Journal of Physics*, 22(5):053012, may 2020.

[263] Changhao Yi. Success of digital adiabatic simulation with large Trotter step. *Phys. Rev. A*, 104:052603, 2021.

[264] Marin Bukov, Dries Sels, and Anatoli Polkovnikov. Geometric Speed Limit of Accessible Many-Body State Preparation. *Phys. Rev. X*, 9(1):011034, 2019.

[265] Zhen-Yu Wang and Martin B. Plenio. Necessary and sufficient condition for quantum adiabatic evolution by unitary control fields. *Phys. Rev. A*, 93:052107, May 2016.

[266] Oleg Lychkovskiy. A necessary condition for quantum adiabaticity applied to the adiabatic grover search. *Journal of Russian Laser Research*, 39(6):552–557, Nov 2018.

[267] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[268] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[269] R. P. Feynman. Space-time approach to nonrelativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367–387, 1948.

[270] J. Hubbard. Calculation of partition functions. *Physical Review Letters*, 3:77–78, 1959.

[271] R. L. Stratonovich. On a method of calculating quantum distribution functions. *Soviet Physics Doklady*, 2:416, 1957.

[272] I. Montvay and G. Munster. *Quantum fields on a lattice*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 3 1997.

[273] Michael Spivak. *A Comprehensive Introduction to Differential Geometry*. Publish or Perish, Houston, Texas, 2nd edition, 1979. Volumes 1–5.

[274] John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, 2nd edition, 2012.

[275] Scott Lawrence. Personal communication.

[276] S. Elitzur. Impossibility of Spontaneously Breaking Local Symmetries. *Phys. Rev. D*, 12:3978–3982, 1975.

[277] B. Durhuus. On the structure of gauge invariant classical observables in lattice gauge theories. *Letters in Mathematical Physics*, 4(6):515–522, Nov 1980.

[278] Taco S Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, volume 48, pages 2990–2999. PMLR, 2016.

[279] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and 1 Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1321–1330. PMLR, 2019.

[280] Matteo Favoni, Andreas Ipp, David I. Müller, and Daniel Schuh. Lattice Gauge Equivariant Convolutional Neural Networks. *Phys. Rev. Lett.*, 128(3):032003, 2022.

[281] Yuki Nagai and Akio Tomiya. Gauge covariant neural network for quarks and gluons. *Phys. Rev. D*, 111(7):074501, 2025.