ABSTRACT

Title of Dissertation: QUANTUM ADVANTAGE VIA PHYSICS-INSPIRED METHODS IN OPTIMZATION: EMPIRICAL AND THEORETICAL ANALYSIS

Yufan Zheng
Doctor of Philosophy, 2026

Dissertation Directed by: Professor Xiaodi Wu
Department of Computer Science

Continuous and combinatorial optimization are foundational to science and engineering. While convex optimization can be usually tackled efficiently, large-scale nonconvex optimization and combinatorial optimization remain a significant challenge for classical computation. Physics-inspired quantum algorithms, such as Quantum Adiabatic Algorithm (QAA) and Quantum Hamiltonian Descent (QHD), offer a promising paradigm for solving these problems by utilizing quantum dynamics to accelerate navigating complex landscapes of objective functions. However, the precise conditions under which these methods provide a provable advantage over state-of-the-art classical algorithms are still largely unexplored. In this thesis, we make progress toward answering this question.

First, we establish the global convergence of QHD for nonsmooth continuous objective functions. We also conduct complementary numerical evaluations to demonstrate that QHD consistently outperforms classical subgradient methods. Second, we present evidence of the scalable quantum advantage by QHD. We construct a family of objective functions with exponentially many local minima

and prove that QHD minimizes them in polynomial time. In contrary, a comprehensive empirical study reveals that several off-the-shelf state-of-the-art classical solvers require superpolynomial time to find global minima. Third, we investigate the complexity of Schrödinger operators, since QHD is formulated as real-space Schrödinger dynamics. We prove that simulating time-independent Schrödinger dynamics is BQP-hard and estimating their ground energy is StoqMA-complete under mild technical assumptions. Finally, we present the first provable exponential quantum speedups for QHD and QAA by constructing specific objective functions that require exponentially many classical queries but only polynomial quantum resources. To our best knowledge, these results represent the first rigorous superquadratic speedups for QHD and QAA in optimization. Together, these contributions and technical methods developed advance the understanding of the power of quantum optimization and shed light on how practical quantum advantage might be achieved.

QUANTUM ADVANTAGE VIA PHYSICS-INSPIRED METHODS IN OPTIMIZATION: EMPIRICAL AND THEORETICAL ANALYSIS

by

Yufan Zheng

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 Xiaodi Wu, Chair/Advisor
Professor Alexander Barg
Professor Andrew M. Childs
Professor Alexey V. Gorshkov
Professor Mohammad T. Hajiaghayi
Professor Haizhao Yang, Dean's representative

© Copyright by
Yufan Zheng
2026

Dedication

To my parents

Table of Contents

Dedication ii

Table of Contents iii

List of Tables vii

List of Figures viii

Chapter 1: Introduction 1
- 1.1 Quantum Speedups: Promises and Limitations . . . 1
- 1.2 Quantum Algorithms for Optimization . . . 2
- 1.3 Physics-Inspired Quantum Optimization Methods . . . 6
  - 1.3.1 Quantum Adiabatic Algorithm for Combinatorial Optimization . . . 6
  - 1.3.2 Quantum Hamiltonian Descent for Continuous Optimization . . . 7
- 1.4 Overview . . . 8

Chapter 2: Preliminaries 13
- 2.1 Notation . . . 13
- 2.2 Quantum Hamiltonian Descent . . . 14
  - 2.2.1 Hamiltonian Formulation of Accelerated Gradient Descent . . . 14
  - 2.2.2 Quantum Dynamics for Optimization . . . 18

Chapter 3: QHD for Nonsmooth Optimization 21
- 3.1 Introduction . . . 21
- 3.2 Continuous-Time QHD for Nonsmooth Optimization . . . 27
  - 3.2.1 Quantum Algorithms and Complexity Analysis . . . 27
  - 3.2.2 Convergence Analysis . . . 33
- 3.3 Numerical Results for Nonsmooth Nonconvex Optimization . . . 50
  - 3.3.1 Methodology . . . 51
  - 3.3.2 Benchmarking QHD against Classical Algorithms . . . 53
  - 3.3.3 QHD Dynamics: A Comparative Case Study . . . 55

Chapter 4: Superpolynomial Speedup of QHD for Nonconvex Optimization 58
- 4.1 Introduction . . . 58
- 4.2 Preliminaries . . . 65
  - 4.2.1 Convexity and Spectral Gap . . . 65
  - 4.2.2 Ground States of the Schrödinger Operator . . . 68
  - 4.2.3 Quantum Adiabatic Theorem for Unbounded Hamiltonian . . . 70
  - 4.2.4 Quantum Simulation of Schrödinger Equations . . . 72
- 4.3 Construction of Nonconvex Optimization Instances . . . 74
  - 4.3.1 The Spectral Gap of Asymmetric Double Well . . . 74
  - 4.3.2 1D Model Problem . . . 77

4.3.3 Toward High-Dimensional Nonconvexity . . . 81
4.4 Quantum Algorithms and Complexity Analysis . . . 84
4.4.1 Main Results . . . 84
4.4.2 Initial State Preparation . . . 86
4.4.3 Proof of Theorem 4.4.1 . . . 87
4.4.4 Numerical Simulation for 1D QHD . . . 92
4.5 Empirical Study of Classical Algorithms . . . 94
4.5.1 Methodology . . . 95
4.5.2 Annealing-Based and Monte-Carlo Algorithms . . . 97
4.5.3 Newton and Quasi-Newton Methods . . . 98
4.5.4 Stochastic Gradient Descent . . . 99
4.5.5 Gurobi . . . 103

Chapter 5: Computational Complexity of Schrödinger Operators 106
5.1 Introduction . . . 106
5.1.1 Overview of Results . . . 109
5.1.2 Technical Contributions . . . 111
5.1.3 Organization . . . 114
5.2 Complexity of the Transverse Field Ising Model . . . 114
5.3 Algorithms for Schrödinger Operators . . . 118
5.3.1 Estimating the Ground Energy is in StoqMA . . . 118
5.3.2 Time-Independent Simulation is in BQP . . . 121
5.4 Hardness of Schrödinger operators . . . 125
5.4.1 Proof Idea . . . 128
5.4.2 One-Dimensional Double-Well Function . . . 131
5.4.3 Perturbation Theory . . . 141
5.4.4 Proof of the Main Lemma . . . 145

Chapter 6: Provable (Sub)exponential Speedup of QAA and QHD 156
6.1 Introduction . . . 156
6.2 Technical Overview . . . 161
6.2.1 GHV Adiabatic Path . . . 162
6.2.2 Perturbative Reductions to TFD Hamiltonians (Lemma 6.4.3 and 6.4.4) . . . 163
6.2.3 Linearization of Adiabatic Path (Lemma 6.4.5, 6.4.6, and 6.4.7) . . . 165
6.2.4 Path Tilt and Extension (Lemma 6.4.8) . . . 167
6.2.5 Putting Everything Together (Theorem 6.1.1) . . . 168
6.2.6 Generalization to Continuous Space (Theorem 6.1.2) . . . 169
6.3 Preliminaries . . . 173
6.3.1 Hamiltonian Path . . . 174
6.3.2 Hamiltonian Simulation . . . 177
6.3.3 Perturbation Theory . . . 178
6.4 Proof Outline for Theorem 6.1.1 . . . 180
6.5 Reduction I: Proof of Lemma 6.4.3 . . . 190
6.5.1 First Step – Sparse Hamiltonians to Hypercube Hamiltonians . . . 192
6.5.2 Second Step – Cleaner Start and End . . . 196

6.5.3 Putting Everything Together . . . 199
6.5.4 Deferred Proofs . . . 200
6.6 Reduction II: Proof of Lemma 6.4.4 . . . 202
6.6.1 First Step – Hypercube Hamiltonians to TFD Hamiltonians . . . 204
6.6.2 Second Step – Cleaner Start and End . . . 207
6.6.3 Putting Everything Together . . . 209
6.6.4 Deferred Proofs . . . 210
6.7 Toy TFI Hamiltonian: Proof of Lemma 6.4.6 . . . 213
6.7.1 Unbounded Time Regime . . . 216
6.7.2 Bounded Time Regime . . . 218
6.7.3 Properties of Certain Tridiagonal Matrices . . . 221
6.7.4 Deferred Proofs . . . 226
6.8 Reduction III: Proof of Lemma 6.4.7 . . . 227
6.9 Reduction IV: Proof of Lemma 6.4.8 . . . 230
6.10 Extension to Continuous Optimization: Theorem 6.1.2 . . . 235
6.10.1 Hamiltonian Formulation of Continuous Optimization . . . 236
6.10.2 Proof of the Theorem . . . 241
6.10.3 Simulating TFD Hamiltonians with Schrödinger Operators . . . 248
6.10.4 Deferred Proofs . . . 254

Chapter 7: Conclusion 262
7.1 Summary of Contributions . . . 262
7.2 Future Directions . . . 263

Appendix A: Appendices for Chapter 3 265
A.1 Subgradient Methods . . . 265
A.2 Analysis of Schrödinger Equations . . . 269
A.2.1 Spectral Theory of Schrödinger Operators . . . 269
A.2.2 Regularity Theory of Schrödinger Equations . . . 270
A.2.3 Quantum Simulation of Schrödinger Equations with Nonsmooth Potentials . . . 273
A.3 Experiment Details . . . 275
A.3.1 Test Functions . . . 275
A.3.2 Effect of Changing L in QHD . . . 279
A.3.3 Learning-Rate Free Momentum Stochastic Gradient Descent (LFMSGD) . . . 280
A.3.4 Resolution Bottleneck of QHD Simulation . . . 281

Appendix B: Appendices for Chapter 4 283
B.1 Quantum Simulation with Analytic Initial Data . . . 283
B.2 Sub-Gaussian Ground States . . . 285
B.2.1 Probability Toolbox . . . 285
B.2.2 Proof of Lemma 4.4.2 . . . 287
B.3 Auxiliary Lemmas . . . 289

Appendix C: Appendices for Chapter 6 291
C.1 Adiabatic Quantum Computing . . . 291

C.1.1 Quantum Adiabatic Theorem 292
C.2 Explicit Technical Building Blocks 293

Bibliography 298

List of Tables

3.1 Comparison of (worst-case) convergence rate bounds and convergence guarantees for continuous-time QHD and the subgradient method across various nonsmooth function classes. Here, t denotes time or the iteration number. 25
3.2 Time-dependent functions and corresponding end evolution time in the rescaled continuous-time QHD (3.17). 32
3.3 Best-of-k optimality gap summary for QHD, LFMSGD, and Subgrad on 12 test functions (cf. Figure 3.1), including one 1-dimensional, nine 2-dimensional, and two 3-dimensional cases. The smallest gaps among three algorithms are marked blue. 53

List of Figures

1.1 Examples of quantum algorithms organized by speedup and practical relevance. Only a few are known to simultaneously achieve both practical relevance and superpolynomial speedups (top-right quadrant). Other quadrants contain many additional examples that are omitted. *Physics-inspired algorithms are not included* because determining whether they belong in the top-right quadrant is the central question of this thesis. . . . 3

3.1 Best-of-k optimality gap for 12 test functions (cf. Table 3.3). QHD outperforms Subgrad and LFMSGD for all choices of k on every test functions except `SCHWEFEL`, `ACKLEY`, `RANA` and `DAMAVANDI`. . . . 54

3.2 Landscape of test function `XINSHEYANG04` . . . 56

3.3 Distributions at intermediate iterations for QHD (1st row), Subgrad (2nd row), and LFMSGD (3rd row) on `XINSHEYANG04` where m denotes the iteration number. Note that in QHD, $m = t/h = 1000t$ due to time discretization, where t denotes the evolution time. . . . 56

4.1 **Construction of optimization instances. A.** A well-formed double well function. **B.** A separable function built from two identical well-formed double well functions. **C.** Contour plot of the 2-dimensional separable function $F(\boldsymbol{x})$. **D.** Contour plot of the rotated 2-dimensional function $F_U(\boldsymbol{x})$. The local geometry of $F(\boldsymbol{x})$ is preserved, while the new function $F_U(\boldsymbol{x})$ is not separable. . . . 59

4.2 **The spectral gap and ground state of asymmetric double well. A.** Plot of the double well function $w(x)$. **B.** The spectral gap $\Delta(\lambda)$ of the Hamiltonian operator $\hat{H}(\lambda)$ as a function of λ. **C, D.** The blue curve (or the red dashed curve, resp.) represents the ground state (or the harmonic approximation of the ground state, resp.) of the Hamiltonian operator $\hat{H}(\lambda)$ for $\lambda = 1, 40$. The x axis represents the one-dimensional space and the y axis shows the amplitude of the quantum states. When λ becomes larger, the harmonic approximation becomes more accurate. . . . 75

4.3 Numerical experiment for 1D QHD. . . . 93

4.4 TTS scaling of dual annealing and basin-hopping. Confidence intervals are calculated at the 95% level. Results suggest that TTS for both algorithms (lower envelopes of curves) scale exponentially in dimensionality d. . . . 98

4.5 TTS scaling of Ipopt and SQP. Confidence intervals are calculated at the 95% level. Both scale exponentially in dimensionality d. . . . 99

4.6 TTS scaling of SGD under two types of learning rate schedules respectively. Confidence intervals are calculated at the 95% level. It is clear that TTS in both settings (lower envelopes of curves) scale exponentially in dimensionality d. In some curves, the data points are missing because we do not report the TTS if less than 5 success runs are detected. . . . 102

4.7 TTS scaling of Gurobi which is roughly exponential in dimensionality d. Confidence intervals are calculated at the 95% level. 1 work unit is proportional to 1 second of CPU time. In some curves, the data points are missing because we do not report the TTS if less than 5 success runs are detected. . . . 104

5.1 First two eigenfunctions $\psi_0(x), \psi_1(x)$ and their linear combination $\widehat{s_0}(x) = (\psi_0(x) + \psi_1(x))/\sqrt{2}$ for the operator $\widehat{X} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + f_{\mathrm{dw}}$ with parameter $w = 0.05$ and $\ell = 0.5$, where f_{dw} is defined in Section 5.4.1. 113

6.1 Overview diagram of the proof structure illustrating the logical dependencies among the main lemmas and theorems. The dashed lines separate major steps, which are covered in the corresponding subsections of Section 6.2. 162

6.2 An illustration of how we reduce an edge of weight $-J$ ($0 < J \ll 1$) in a stoquastic sparse Hamiltonian to a gadget in a stoquastic hypercube Hamiltonian, where Δ is a parameter being sufficiently large. Nodes represent computational basis states. Edges with different colors correspond to different components from the second-order perturbative reduction; see Lemma 6.4.3 for details. Note that k can be *polynomially* large since it only introduces a $\sqrt{k+1}$ overhead on weights. 164

6.3 An illustration of how we reduce an edge of weight $-J$ ($0 < J \ll 1$) in a stoquastic hypercube Hamiltonian to a gadget in a TFD Hamiltonian, where Δ is a parameter that is sufficiently large. Nodes represent computational basis states. Edges with different colors correspond to different components from the second-order perturbative reduction; see Lemma 6.4.4 for details. 164

6.4 Spectrum of $A(t)$ for different λ values. A small but nonzero λ perturbs the spectrum, ensuring an $\Theta(\lambda)$ spectral gap throughout and making the instantaneous ground state transition of $A(t)$ smooth. 166

6.5 Visualizations of $|\chi_0\rangle, |\chi_1\rangle$ regarding $\widehat{X}$ and their linear combination $|\widehat{0}\rangle, |\widehat{1}\rangle$ defined in Equation (6.12). The parameter for $\widehat{X}$ is chosen to be $\lambda = 30$. The potential field f_{dw} (amplified 10 times) is plotted in a dotted line for reference. 171

6.6 The Construction of $\widehat{D}$ from a diagonal Hamiltonian D with parameter $w = 1/4$ in Definition 6.10.4. It is efficient in the sense that a query to $\widehat{D}(\cdot)$ can be implemented by one query to D. 171

A.1 Best-of-k optimality gap for 12 test functions. Minimal optimality gaps that QHD can "see" after discretization are shown as red dotted lines. 282

Chapter 1: Introduction

1.1 Quantum Speedups: Promises and Limitations

Algorithms are central to both scientific discovery and everyday technology [1]. A *quantum algorithm* is an algorithm that runs on a computational device capable of exploiting quantum-mechanical phenomena such as superposition and interference; such devices are referred to as *quantum computers*. We refer to [2] for an introduction to the fundamentals of quantum computing.

Over the past decades, numerous quantum algorithms have been developed for a wide range of computational tasks, including those in linear algebra [3], differential equations [4], optimization [5], quantum simulation [6], and cryptanalysis [7]. A quantum algorithm for a computational task makes practical sense only if it improves upon the best available classical algorithm with respect to a well-defined resource, since quantum hardware is presumably more costly to build and operate [8]. In most settings the primary resource of interest is *time* (or an equivalent proxy such as gate complexity, circuit depth, or query complexity), and we therefore seek *speedups* over state-of-the-art classical algorithms. For example, suppose that the best classical algorithm for a task runs in time T, where T is parameterized by attributes of the task such as its size. A *quadratic speedup* means that a quantum algorithm solves the same task in time $O(\sqrt{T})$.

While asymptotic speedups provide a useful guidance, turning them into practical advantage can be subtle. First, the *end-to-end* cost of a quantum solution often includes nontrivial overheads beyond the core algorithmic primitive, such as state preparation, data loading, and measurement overheads. Second, many analyses idealize the hardware model; for instance, fault-tolerant error correction can introduce huge overheads hidden by asymptotic notation. As a consequence, even modest polynomial speedups may fail to translate into advantage at realistic problem sizes. Indeed, achieving practi-

cal quantum advantage is often challenging for algorithms whose asymptotic improvements are only quadratic or slightly better, once such overheads are accounted for; instead, quantum algorithms with roughly *quartic speedups* (e.g., quantum $O(T^{\frac{1}{4}})$ versus classcal T) are far more promising for practical quantum advantage [9].

In light of the considerations above, an important observation emerges: Overall, only a small subset of known quantum algorithms simultaneously (i) offer superquadratic asymptotic speedups under realistic modeling assumptions and (ii) target tasks of clear practical relevance [10]; see Figure 1.1. This observation suggests that substantial work remains to identify quantum algorithms that can deliver *practical* quantum advantage.

1.2 Quantum Algorithms for Optimization

Optimization provides a particularly compelling testing ground for practical quantum advantage. A wide range of tasks in science, engineering, machine learning, and operations research can be formulated as optimization problems, and improvements in optimization algorithms often translate directly into real-world impact. Consequently, quantum algorithms that achieve *substantial* speedups for certain optimization problems would naturally occupy the top-right quadrant of Figure 1.1, providing clear practical advantage. This prospect has motivated significant effort toward developing quantum algorithms for optimization [5].

Quantum algorithms for combinatorial optimization. The first applications of quantum algorithms have predominantly focused on combinatorial optimization, as most of these algorithms can be implemented on currently available quantum architectures. Notable examples include Grover's Search [13, 22–24], Variational Quantum Eigensolver (VQE) [25–28], Quantum Adiabatic Algorithm (QAA) [29–

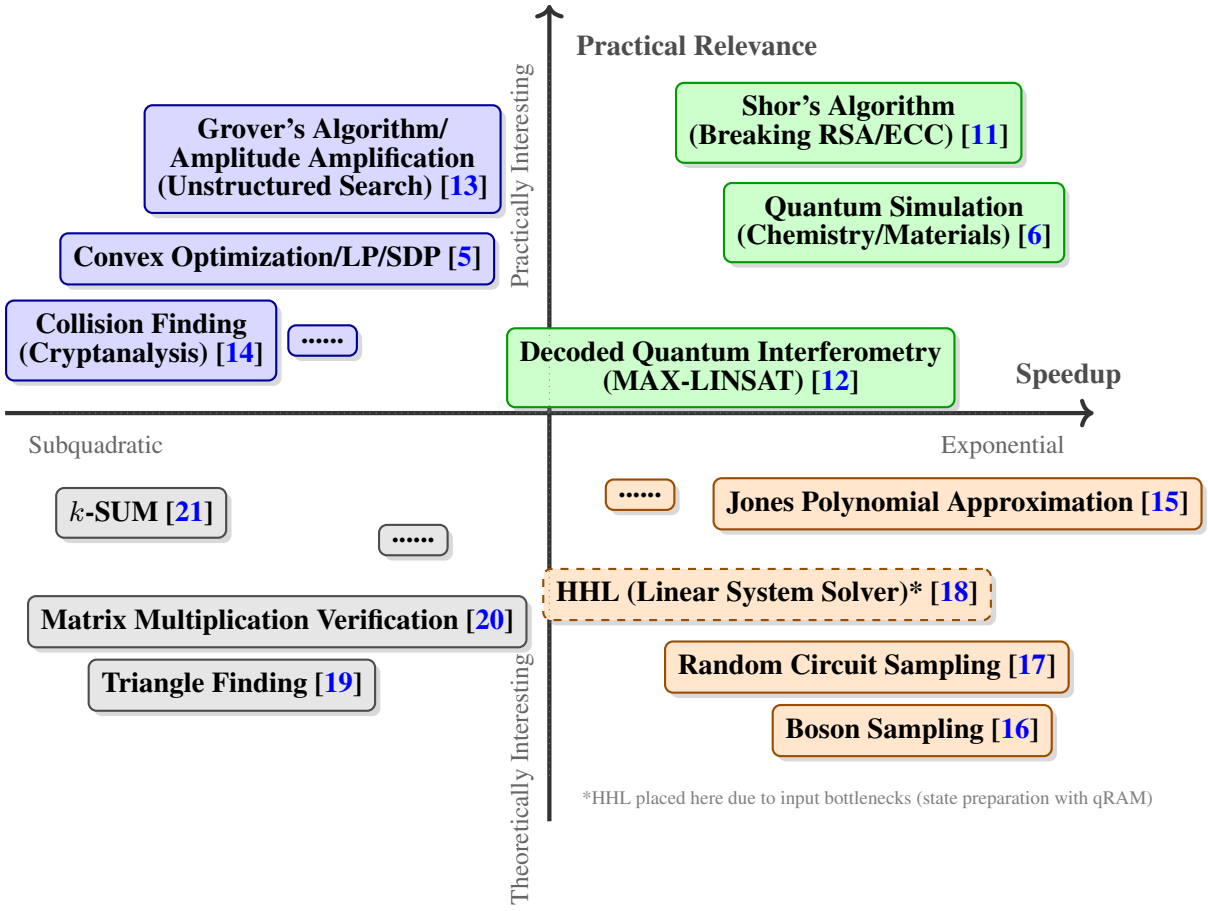


Figure 1.1: Examples of quantum algorithms organized by speedup and practical relevance. Only a few are known to simultaneously achieve both practical relevance and superpolynomial speedups (top-right quadrant). Other quadrants contain many additional examples that are omitted. *Physics-inspired algorithms are not included* because determining whether they belong in the top-right quadrant is the central question of this thesis.

32] and Quantum Approximate Optimization Algorithm (QAOA) [33–35]. More recent breakthroughs, such as the Short-Path Algorithm [36], which outperforms Grover's Search with a theoretical guarantee, and Decoded Quantum Interferometry [12], which reduces optimization problems to *classical* decoding tasks, have opened new avenues for exploring quantum optimization.

Quantum algorithms for continuous optimization. In recent decades, significant progress has been made in developing quantum algorithms for continuous optimization. Quantum Gradient Descent methods implement gradient descent in n-dimensional space using only $\mathrm{polylog}(n)$ qubits and they are efficient if the number of iterations is small [37] or the gradient is an affine function [38]. A quadratic speedup (in the instance dimension n) has been achieved for general convex optimization [39, 40]. For more structured problems, such as linear programming and semidefinite programming, faster quantum algorithms have been devised using Multiplicative Weights Update (MWU) framework [41–44] or interior-point methods [45–49]. In contrast, much less is known about *provable* quantum speedups for nonconvex optimization. A mild quantum speedup has been demonstrated for finding critical points in stochastic optimization [50], while a cubic speedup (in the instance dimension n) has been achieved for escaping saddle points [51].

Most of the algorithms for continuous optimization discussed above share a common feature: they leverage quantum acceleration in intermediate steps of classical optimization methods. Recently, a new research direction has emerged, focusing on directly harnessing quantum dynamics in carefully designed systems for optimization. Notable examples include Quantum Hamiltonian Descent (QHD) [52] and Quantum Langevin Dynamics (QLD) [53]. These methods provably converge to global minima on smooth objectives, albeit not always efficiently, and have demonstrated strong empirical performance in tackling nonconvex optimization problems.

Among the algorithms mentioned above, a particular class of *physics-inspired* quantum algorithms for general combinatorial and/or nonconvex optimization—including QAA, QAOA, VQE, QHD, and QLD—deserves special attention. These algorithms have the *potential* to achieve superquadratic or even exponential speedups, unlike most of the other optimization algorithms discussed above, which typically provide at most subquadratic speedups. Moreover, many of them are compatible with near-term quantum hardware, as they rely on relatively shallow circuits or analog quantum dynamics, in contrast to other quantum algorithms that typically require large-scale fault-tolerant quantum computers. Indeed, some of them have already been demonstrated experimentally on existing quantum devices [26, 52, 54, 55].

At the same time, these physics-inspired algorithms are known to be notoriously difficult to analyze rigorously. For broad classes of optimization problems, even their precise time complexity and achievable quantum speedups remain poorly understood. Although worst-case complexity considerations suggest that generic exponential quantum speedups for NP-hard optimization problems are unlikely, meaningful advantages may still arise in structured or average-case settings for certain optimization tasks, and understanding when such speedups occur remains an important open question [5].

This thesis hence investigates this broader question through the study of QAA and QHD. These algorithms are more amenable to theoretical analysis because the physical principles underlying their formulation are relatively well understood, allowing the use of mathematical tools from the theory of Schrödinger operators [56], adiabatic perturbation theory [57], and Hamiltonian complexity [58]. More specifically, we investigate the following question:

To what extent can QAA and QHD provide quantum speedups for optimization?

1.3 Physics-Inspired Quantum Optimization Methods

Quantum Adiabatic Algorithm (QAA) and Quantum Hamiltonian Descent (QHD) are two physics-inspired approaches to optimization that use quantum dynamics to search for high-quality solutions. Roughly speaking, QAA is motivated by adiabatic evolution while staying at the instantaneous equilibrium state (which resembles the idea behind simulated annealing), whereas QHD is inspired by *damped* (i.e., energy reduces over time) Hamiltonian dynamics that resemble gradient descent with friction in optimization. QAA is most commonly applied to combinatorial optimization, whereas QHD is designed for continuous optimization.

1.3.1 Quantum Adiabatic Algorithm for Combinatorial Optimization

Quantum Adiabatic Algorithm (QAA), introduced by Farhi et al. [31] in 2000, has emerged as a distinct paradigm for solving classical optimization problems using quantum mechanics. In the combinatorial setting, one may think of QAA as addressing problems of the form

$$\min_{x\in\{0,1\}^n} f(x),$$

where $f\colon \{0,1\}^n \to \mathbb{R}$ is the objective function.[1]

The basic idea of QAA is simple: one starts from a quantum system whose lowest-energy state is easy to prepare, and then gradually deforms it into another system whose lowest-energy state encodes the solution to the optimization problem. From a physical perspective, this process resembles quantum annealing, where the system is slowly guided toward an equilibrium configuration corresponding to

[1] The domain need not be $\{0,1\}^n$, although most studies of QAA focus on functions defined over this domain.

the ground state of the final Hamiltonian. If this deformation is carried out sufficiently slowly, the quantum adiabatic theorem [57, 59] (from which QAA derives its name) implies that the system remains close to its instantaneous ground state during the evolution and therefore ends near a global minimizer of f. This natural physics-inspired quantum algorithm has led to extensive empirical and experimental studies, as well as several important theoretical insights (notably [60] and [61]); see [62] for a comprehensive review.

1.3.2 Quantum Hamiltonian Descent for Continuous Optimization

Quantum Hamiltonian Descent (QHD), introduced by Leng et al. [52] in 2023, is a more recent quantum algorithm for continuous optimization. QHD focuses on problems of the form

$$\min_{x \in X} f(x),$$

where $X \subseteq \mathbb{R}^n$ is a bounded domain and $f \colon X \to \mathbb{R}$ is a real-valued objective function.

Similar to QAA, QHD treats optimization as a quantum dynamical process by deforming a quantum system. The method is inspired by damped Hamiltonian systems in classical mechanics, where the combination of conservative dynamics and damping drives the system toward low-energy configurations. In particular, it evolves a quantum state over the search space so that the wavefunction initially spreads broadly and then gradually concentrates near regions where the objective value is small, ideally close to a global minimizer of f. Using the so-called Hamiltonian embedding technique [63], initial numerical experiments [52] have been carried out on D-Wave quantum machines [64], and theoretical analysis has begun to emerge recently [65].

Besides targeting different types of optimization problems, a key difference between QAA and

QHD is that QHD does *not* require the deformation to be carried out slowly, especially for convex optimization tasks. Indeed, Leng et al. derived QHD from a Hamiltonian formulation related to accelerated gradient methods, which are inherently diabatic.

1.4 Overview

The rest of this thesis is structured as follows. In Chapter 2, we present the essential definitions and notations used throughout, as well as related fundamental concepts of quantum computing. The remaining chapters present the main results of the thesis, as summarized in the following.

Theoretical analysis of QHD convergence for nonsmooth optimization. While convergence of QHD is known [52], the existing convergence analysis breaks down when the objective function is nonsmooth. In Chapter 3, we establish convergence of QHD for general continuous objective functions that are not necessarily smooth. In particular, we establish the global convergence of QHD using energy reduction arguments tailored to each problem class. The convergence rates are proven for strongly convex (Theorem 3.2.12) and generally convex (Theorem 3.2.14) optimization problems. We also prove that QHD converges to the global minimum of any (possibly nonsmooth and nonconvex) locally Lipschitz-continuous objective function in finite time (Theorem 3.2.16), which, to the best of our knowledge, is the *first rigorous* result in the literature on quantum optimization achieving such a guarantee. In contrast, classical methods like the subgradient method can fail to converge under the locally Lipschitz condition [66, 67], and no convergence result is known for the stochastic subgradient method assuming the locally Lipschitz condition only—let alone convergence to the global minimum.[2]

[2]Exceptions exist when there are more assumptions on objective functions (e.g., being semialgebraic or definable in an o-minimal structure), under which the stochastic subgradient method converges to first-order stationary points [68].

Empirical performance of QHD for nonsmooth nonconvex optimization. The convergence result for nonconvex objective functions established in Section 3.2 is qualitative rather than quantitative, since it does not provide a bound on the convergence time. To complement this theoretical result, Section 3.3 presents an empirical evaluation of QHD on nonsmooth nonconvex objective functions. Using classical simulation of QHD, we benchmark it against representative classical methods, including the vanilla subgradient method [69] and learning-rate-free momentum stochastic gradient descent [70], across 12 test functions with varying dimension and landscape. The numerical results (Table 3.3) show that QHD often achieves substantially smaller optimality gaps and outperforms classical algorithms within the same number of queries to the objective function. We then conduct a comparative case study of the optimization dynamics of QHD against those of classical algorithms, providing further intuition for its observed performance.

Superpolynomial speedup of QHD for nonconvex optimization. The empirical results presented in Chapter 3 do not address how the observed quantum advantage scale with the dimension of the objective function. In Chapter 4, we investigate the extent to which QHD can outperform classical algorithms when the dimension of the objective function scales. We construct a family of objective functions with exponentially many local minima. For this family, we prove that QHD finds the global minimizer with polynomial query and time complexity (Theorem 4.1.2). On the classical side, we carry out a comprehensive empirical study of representative off-the-shelf methods, including dual annealing [71], basin-hopping [72], stochastic gradient descent, Ipopt [73], sequential quadratic programming [74], and Gurobi [75]. The numerical results (Section 4.5) indicate that all tested classical methods need numbers of queries that scale superpolynomially. Together, these results provide evidence of a quantum-classical performance separation for a carefully constructed family of nonconvex

optimization problems.

Computational complexity of Schrödinger operators. While the result in Chapter 4 provides evidence of a quantum advantage for QHD, it does not establish a *provable* advantage over classical computation. To obtain rigorous theoretical results, in Chapter 5, we study the computational complexity of Schrödinger operators of the form $H := -\Delta + V$, where Δ is the Laplace operator. Schrödinger operators are central to QHD, since QHD simulates quantum dynamics generated by them. We consider two fundamental computational tasks: simulating the dynamics generated by time-independent H and estimating its ground energy. On the simulation side, we prove that Schrödinger dynamics is BQP-hard (Theorem 5.4.2), meaning that simulating such systems is at least as hard as universal quantum computation. On the ground-state side, we prove that estimating the ground energy of a Schrödinger operator is StoqMA-complete (Theorem 5.4.1 and 5.3.3). These results are obtained by establishing a connection between Schrödinger operators and transverse field Ising Hamiltonians. A direct consequence is that quantum algorithms involving Schrödinger dynamics [51–53,76–81] cannot be efficiently dequantized by a general-purpose classical simulation algorithm, unless BPP = BQP which is widely believed to be false.

Provable exponential quantum speedup for QAA. Although Chapter 5 establishes the computational power of Schrödinger dynamics in a complexity-theoretic sense, it does not provide an explicit optimization problem with a provable quantum-classical separation. To address this, we need to temporarily shift our focus from QHD to QAA. In Chapter 6, we study combinatorial optimization and prove an exponential quantum speedup for QAA. Specifically, we construct a family of objective functions $f : \{0,1\}^n \to [0,1]$ for which any classical algorithm requires exponentially many queries to find the

global minimizer, while QAA succeeds using only polynomially many queries and polynomial time (Theorem 6.1.1). To our best knowledge, this is the *first provable* superquadratic quantum speedup for QAA in the context of optimization since it was proposed in 2000. The proof proceeds through a sequence of efficient perturbative reductions from the Gilyén–Hastings–Vazirani oracle separation for adiabatic quantum computing [82].

Provable exponential quantum speedup for QHD. Finally, in Section 6.10, we lift the above quantum-classical separation from combinatorial optimization to continuous optimization. Specifically, we construct a family of objective functions $f : [0,1]^n \to [0,1]$ for which any classical algorithm requires exponentially many queries to find the global minimizer, whereas a slightly generalized version of QHD succeeds using only polynomially many queries and polynomial time (Theorem 6.1.2). The proof builds on the connection between Schrödinger operators and transverse field Ising Hamiltonians developed in Chapter 5, and extends it so as to transfer the quantum advantage of QAA to QHD: note that QAA (resp., QHD) is typically formulated in terms of transverse field Ising Hamiltonians (resp., Schrödinger operators). This is the first provable quantum speedup for QHD.

List of papers. This thesis is based on the following papers.

[79] *Quantum Hamiltonian Descent for Non-Smooth Optimization*

Jiaqi Leng*, Yufan Zheng*, Zhiyuan Jia, Lei Fan, Chaoyue Zhao, Yuxiang Peng, and Xiaodi Wu.

Chapter 3 is based on this paper.

[77] *A Quantum-Classical Performance Separation in Nonconvex Optimization*

Jiaqi Leng, Yufan Zheng, and Xiaodi Wu.

Chapter 4 is based on this paper.

[83] *On the Computational Complexity of Schrödinger Operators*

Yufan Zheng, Jiaqi Leng, Yizhou Liu, and Xiaodi Wu.

Chapter 5 is based on this paper.

[84] *(Sub)exponential Quantum Speedup for Optimization*

Jiaqi Leng*, Kewen Wu*, Xiaodi Wu*, and Yufan Zheng*.

Chapter 6 is based on this paper.

Acknowledgment of AI Assistance. The author acknowledges the use of generative artificial intelligence tools to assist with polishing the English writing in this thesis. All research ideas, analyses, results, and interpretations presented in this thesis are the author's own work.

Chapter 2: Preliminaries

We assume that the reader is familiar with basic concepts in quantum computing, which can be found in the textbook [2]. Below we only introduce necessarily preliminaries used *throughout* the thesis.

2.1 Notation

We use $\mathbb{R}, \mathbb{C}$ to denote real number and complex numbers respectively; and use $\mathbb{N}$ to denote all non-negative integers and use $\mathbb{N}_+$ for all positive integers. For a positive integer n, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. For real numbers $\ell \leq r$, we use $[\ell, r]$ to denote the set $\{t \in \mathbb{R} \colon \ell \leq t \leq r\}$.

Asymptotics. Throughout this thesis, we use the standard asymptotic notations $O(\cdot)$, $\Theta(\cdot)$, and $\Omega(\cdot)$ to represent *predetermined* expressions that do not depend on any other parameters.[1] Additionally, we use $\mathrm{poly}(\cdot)$ to denote a function that is bounded below by 1 and bounded above by some implicit polynomial in terms of the parameters within. We further define $\mathrm{polylog}(f) \coloneqq \mathrm{poly}(\log(f))$, $\widetilde{O}(f) \coloneqq O(f \,\mathrm{polylog}\, f)$, and $\widetilde{\Omega}(f) \coloneqq \Omega(f/\,\mathrm{polylog}\, f)$. We emphasize that *all* implicit functions represented in these asymptotic forms can be computed explicitly and efficiently.

Norms. For a matrix (Hamiltonian) A, we use $\|A\|$ to denote its operator norm and use $\|A\|_{\mathrm{F}}$ to denote its Frobenius norm. For a vector (state) v, we use $\|v\|$ to denote its Euclidean (i.e., ℓ_2) norm. The distance between two vectors (states) u, v is measured in ℓ_2 norm; and we say they are ϵ-close if $\|u - v\| \leq \epsilon$.

[1] For example, the statement "if $f \geq O(g)$ then $\epsilon = \Theta(\eta)$" means that there exist functions $g_0 = O(g)$ and $\eta_0 = \Theta(\eta)$ such that if $f \geq g_0$, then $\epsilon = \eta_0$.

2.2 Quantum Hamiltonian Descent

Quantum Hamiltonian Descent (QHD) [52] is a quantum optimization algorithm motivated by the interplay between classical gradient-based optimization algorithms and differential equations. The formulation of QHD is closely related to the continuous-time limit of Nesterov's accelerated gradient descent (NAG) method [85, 86]. In this section, we provide an overview of the QHD algorithm, including its origin from classical accelerated gradient descent methods, continuous-time formulation, and discrete-time implementation.

2.2.1 Hamiltonian Formulation of Accelerated Gradient Descent

Consider the optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{2.1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth convex function.

NAG and differential equations. There is a well-established historical connection between ordinary differential equations (ODEs) and optimization. For example, the standard gradient descent (GD) can be regarded as the forward Euler discretization of the gradient flow equation:

$$\dot{X}_t = -\nabla f(X_t), \quad X_t = x_0. \tag{2.2}$$

where x_0 is the initial guess. Accelerated gradient descent algorithms are first-order methods that achieve a convergence rate faster than the standard GD. Nesterov [87] proposed the first accelerated

gradient descent algorithm (known as Nesterov's accelerated gradient descent, or NAG). The update rules of NAG are given as follows:

$$x_k = y_{k-1} - s\nabla f(y_{k-1}), \tag{2.3a}$$

$$y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}), \tag{2.3b}$$

where x_0 is an initial guess, $y_0 = 0$, and $s > 0$ is a fixed step size. For decades, the mechanism of acceleration in NAG appeared mysterious, and an extensive body of research works has been dedicated to a more intuitive understanding of accelerated gradient descent algorithms. In a seminal work [85], the authors show that the continuous-time limit of NAG can be described as a simple second-order ordinary differential equation (ODE):

$$\ddot{X}_t + \frac{3}{t}\dot{X}_t + \nabla f(X_t) = 0, \quad X_t = x_0. \tag{2.4}$$

This continuous-time dynamical system perspective of NAG has revolutionized the understanding and design of accelerated gradient descent through systematically introducing of techniques from continuous mathematics, in particular, the theory of differential equations.

The ODE perspective of NAG was later extended to a broader framework that characterizes accelerated gradient descent by Hamiltonian dynamics [86]. Specifically, the Hamiltonian function that models NAG is the following,

$$H(X, P, t) = \frac{1}{2t^3}\|P\|^2 + t^3 f(X), \tag{2.5}$$

where X and P represent the position and momentum variables of the system, respectively. The Hamiltonian flow generated by the function $H(X_t, P_t, t)$ is described by the following system of ODEs:

$$\dot{X}_t = \frac{\partial H}{\partial P_t} = \frac{1}{t^3} P_t, \tag{2.6a}$$

$$\dot{P}_t = -\frac{\partial H}{\partial X_t} = -t^3 \nabla f(X_t), \tag{2.6b}$$

with initial data $X_0 = x_0$ and $P_0 = 0$. It can be readily verified that the Hamiltonian flow is equivalent to the ODE model (2.4) by plugging (2.6b) into (2.6a).

The Bregman-Lagrangian framework. Motivated by the connection between NAG and Hamiltonian flows, [86] proposed a more general framework to model accelerated gradient descent using Hamiltonian/Lagrangian mechanics. This framework is based on the so-called Bregman Lagrangian:

$$\mathcal{L}(X, \dot{X}, t) = e^{\alpha_t + \gamma_t} \left(\frac{1}{2} |e^{-\alpha_t} \dot{X}|^2 - e^{\beta_t} f(X) \right), \tag{2.7}$$

where $f(x)\colon \mathbb{R}^d \to \mathbb{R}$ is the objective function, $t \geq 0$ represents the time variable, $X \in \mathbb{R}^d$ is the position, $\dot{X} \in \mathbb{R}^d$ is the velocity, and $\alpha_t, \beta_t, \gamma_t$ are arbitrary smooth functions that control the damping of energy in the system. The Euler-Lagrange equation associated with the Bregman Lagrangian reads:

$$\ddot{X}_t + (\dot{\gamma}_t - \dot{\alpha}_t)\dot{X}_t + e^{2\alpha_t + \beta_t} \nabla f(X_t) = 0, \tag{2.8}$$

which is a second-order differential equation. In particular, the ODE model of NAG (2.4) is a special case of the Euler-Lagrange equation (2.8) by choosing $\alpha_t = -\log(t)$ and $\beta_t = \gamma_t = 2\log(t)$. Equivalently, we can write down the Hamiltonian function corresponding to the Bregman Lagrangian via the

Legendre transformation [86, Section 3],

$$\mathcal{H}(X, P, t) = e^{\alpha_t + \gamma_t} \left(\frac{1}{2} |e^{-\gamma_t} P|^2 + e^{\beta_t} f(X) \right), \tag{2.9}$$

where P and X represent the momentum and position variables, respectively. The Euler-Lagrange equation (2.8) can be directly recovered from the Hamilton's equations associated with the Hamiltonian (2.9).

The convergence of the Euler-Lagrange equation can be established via an energy reduction argument. First, we consider the following energy function

$$\mathcal{E}(t) = \|e^{-\alpha_t} \dot{X}_t + X_t - x^*\|^2 + e^{\beta_t} \left(f(X_t) - f(x^*) \right). \tag{2.10}$$

Under the so-called *ideal scaling conditions*, i.e.,

$$\dot{\beta}_t \leq e^{\alpha_t}, \quad \dot{\gamma}_t = e^{\alpha_t}, \tag{2.11}$$

we can prove that the energy function $\mathcal{E}(t)$ is nonincreasing in time:

$$\mathcal{E}(t) \leq \mathcal{E}(0) \quad \forall t \geq 0. \tag{2.12}$$

In other words, $\mathcal{E}(t)$ is a *Lyapunov function* associated with the dynamical system. As a result, we obtain the convergence rate of the Euler-Lagrange equation [86, Theorem 1.1]:

$$f(X_t) - f(x^*) \leq \mathcal{O}(e^{-\beta_t}). \tag{2.13}$$

2.2.2 Quantum Dynamics for Optimization

Quantization of accelerated gradient descent. Hamiltonian evolution serves as a bridge between two fundamental realms of physics: classical and quantum mechanics. Through this connection, the classical Hamiltonian formulation of accelerated gradient descent has inspired **Quantum Hamiltonian Descent (QHD)** [52], which is expressed as a quantum Hamiltonian evolution governed by the following partial differential equation (PDE),

$$i\partial_t \Psi(t,x) = \hat{H}(t)\Psi(t,x), \quad \hat{H}(t) = e^{\alpha_t - \gamma_t}\left(-\frac{1}{2}\Delta\right) + e^{\alpha_t+\beta_t+\gamma_t} f(x), \tag{2.14}$$

Here, $i = \sqrt{-1}$ is the imaginary unit, and $\Delta = \sum_{j=1}^{d} \partial_{x_j}^2$ is the Laplace operator. The operator $\hat{H}(t)$ is called the quantum Hamiltonian, which represents the total energy of the system. The quantum Hamiltonian $\hat{H}(t)$ in QHD can be regarded as the direct quantization of the Bregman Hamiltonian (2.9), where we invoke the following change of variables (a procedure known as *canonical quantization* [88]):

$$P_j \mapsto \hat{p}_j = -i\partial_{x_j}, \quad X_j \mapsto \hat{x}_j, \quad \forall j \in [d]. \tag{2.15}$$

Note that $\hat{p}_j$ and $\hat{x}_j$ should be interpreted as operators rather than scaler functions. In particular, given a smooth test function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$, we have

$$(\hat{p}_j\varphi)(x) \coloneqq -i\partial_{x_j}\varphi(x), \quad (\hat{x}_j\varphi)(x) \coloneqq x_j\varphi(x). \tag{2.16}$$

The function $\Psi(t,\boldsymbol{x})\colon [0,\infty)\times\mathbb{R}^d \to \mathbb{C}$ is the quantum wave function that describes the state of

the quantum system at time $t \geq 0$. It is known that the evolutionary PDE (2.14) preserves the L^2-norm of the wave function. In other words, given an initial state $\Psi(0, \boldsymbol{x}) = \Psi_0(\boldsymbol{x})$ with a unit L^2-norm, i.e., $\|\Psi_0\| = 1$, the wave function always has a unit L^2-norm during the time evolution:

$$\int_{\mathbb{R}^d} |\Psi(t, \boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} |\Psi_0(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x} = 1, \quad \forall t \geq 0. \tag{2.17}$$

Therefore, the modulus square of the wave function $|\Psi(t, \boldsymbol{x})|^2$ can be interpreted as the probability density of the quantum particle in the real space $\mathbb{R}^d$ at a certain time t.

Similar to the Bregman-Lagrangian framework, the asymptotic convergence of the quantum PDE can be shown under the ideal scaling conditions (2.11). We consider the following quantum energy function:

$$\mathcal{W}(t) \coloneqq \int_{\mathbb{R}^d} \overline{\Psi(t, \boldsymbol{x})} \hat{O}(t) \Psi(t, \boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \quad \hat{O}(t) \coloneqq \frac{1}{2} \sum_{j=1}^{d} \|e^{-\gamma_t} \hat{p}_j + \hat{x}_j\|^2 + e^{\beta_t} f(x). \tag{2.18}$$

For a smooth convex objective function f such that $f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{x}^* = 0$, we can prove that $\mathcal{W}(t)$ is nonincreasing in time. As a consequence, we obtain the convergence rate of QHD [52, Theorem 1]:

$$\mathbb{E}[f(X_t)] - f(\boldsymbol{x}^*) \leq \mathcal{O}(e^{-\beta_t}), \tag{2.19}$$

where X_t is a random variable distributed following the probability density function $|\Psi(t, \boldsymbol{x})|^2$. Note that the QHD dynamics (2.14) itself does not require f to be smooth or convex.

QHD as quantum optimization algorithms. The QHD dynamics (2.14) can be exploited to solve optimization problems by simulating the time evolution using a quantum computer. Simulating quantum

Hamiltonian evolution is a fundamental task in quantum computing. It can be shown that, under reasonable assumptions, the QHD dynamics can be efficiently simulated on gate-based quantum computers with $\tilde{\mathcal{O}}(T)$ quantum queries to the function f and $\tilde{\mathcal{O}}(dT)$ total gate complexity [77, 89], where d is the problem dimension and T is the total evolution time. Alternatively, it is also possible to simulate the QHD dynamics using an analog quantum Ising Hamiltonian simulator via the *Hamiltonian embedding* technique [63]. This approach leads to resource-efficient implementations of QHD to solve large-scale nonlinear and nonconvex optimization problems on near-term quantum devices [52, 90].

Chapter 3: QHD for Nonsmooth Optimization

3.1 Introduction

Nonsmooth optimization involves the study of functions that lack differentiability at certain points or across entire regions of their domains. This lack of smoothness often arises in practical problems where robustness, sparsity, or discrete decision-making is required. Nonsmooth optimization plays a vital role in various fields such as machine learning, signal processing, and finance. For example, in classification problems, techniques like support vector machines use hinge loss functions that are inherently nonsmooth [91]. Similarly, robust regression techniques leverage L_1 norm regularization to enforce sparsity and manage outliers, which are particularly relevant in fields like finance and bioinformatics [92]. Engineering and physics frequently involve variational problems where nonsmooth constraints model real-world constraints effectively [93]. In addition, imaging problems such as deblurring and denoising often rely on nonsmooth objectives to enhance computational efficiency and accuracy [94].

We will consider the following unconstrained minimization problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x}), \tag{3.1}$$

where the real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$ is assumed to be continuous but not necessarily differentiable (i.e., nonsmooth). For ease of theoretical treatment, we also assume that the function f is finite over $\mathbb{R}^d$ and has a unique global minimum f^* attained at the point $\boldsymbol{x}^*$.

Classical algorithms, such as subgradient and proximal gradient methods, have been widely adopted to address this problem. Subgradient methods [95–97] generalize gradients for nondifferen-

tiable functions and are applicable across diverse optimization scenarios. However, their convergence rates can be slow, particularly in high-dimensional or nonconvex problems [92]. Despite being foundational, their scalability remains limited in practice, as confirmed by broader research on efficient subgradient applications [98]. Proximal gradient methods [99, 100] are particularly suited for composite optimization problems, where they combine gradient descent with proximal operators to handle nonsmooth terms. These methods have proven effective in specific contexts, such as sparse modeling and total variation minimization [101]. However, similar to subgradient methods, they often struggle with parameter tuning and efficiency in nonconvex scenarios [102].

Beyond these, techniques like smoothing and bundle methods attempt to bridge gaps in classical approaches. Smoothing methods enhance computational tractability and convergence, particularly for problems with inherent nonsmoothness [103]. Bundle methods focus on large-scale optimization challenges by leveraging memory-efficient strategies [104]. Despite these advancements, inefficiencies in handling nonconvexity and high-dimensionality remain prominent issues, underscoring the importance of exploring newer paradigms.

In recent years, quantum computing has emerged as a transformative technology that reshapes the landscape of computational methods. Quantum computers have demonstrated exponential speedups for fundamental problems in scientific computing, such as solving eigenvalue problems [105] and linear systems [18], as well as simulating differential equations [106, 107]. Building on quantum subroutines for numerical computation, numerous quantum algorithms for optimization have been developed and shown to be particularly promising in high-dimensional regimes, where classical approaches often face challenges related to robustness and efficiency (e.g., [37–40, 46, 47]).

However, despite the extensive literature on quantum algorithms for differentiable and convex optimization problems, quantum algorithms for nonsmooth problems remain relatively under-

explored. [108] show that there is no quantum speedup over classical methods for nonsmooth convex optimization in the worst-case scenario. In contrast, a recent work by [109] presents a quantum algorithm for finding a Goldstein stationary point of Lipschitz continuous functions that outperforms the best-known classical algorithms by a polynomial factor, pioneering the investigation of quantum advantage in nonsmooth, nonconvex optimization. Overall, the development of efficient quantum algorithms for nonsmooth optimization—ones that achieve both accelerated convergence and improved solution quality—remains an open avenue for pursuing quantum advantage. In light of these challenges, it is natural to ask whether QHD can be adapted to the nonsmooth setting, although it is originally proposed for optimizing smooth objective functions.

Indeed, an interesting feature of QHD is that it solely relies on function values without requiring gradient information. Intuitively, this feature makes QHD particularly suitable for nonsmooth optimization, where the gradients may not exist. However, the convergence analysis of QHD [52] heavily relies on the differentiability of the objective function and cannot be trivially generalized to nonsmooth optimization. Moreover, the original QHD and its convergence analysis involve abstract time-dependent parameters, making it unclear how to choose these parameters to effectively address various nonsmooth optimization problems in practice. It remains an open question whether QHD is effective for nonsmooth optimization and, if so, what potential quantum advantages it may offer over classical approaches such as subgradient methods.

Our results. In this chapter, we study the performance of QHD for nonsmooth optimization problems. Our main contributions are summarized as follows:

- We propose three variants of QHD, each formulated as a differential equation to address specific classes of nonsmooth optimization problems with distinct convexity structures: *(i)* **QHD-SC** for

nonsmooth strongly convex problems, *(ii)* **QHD-C** for nonsmooth convex problems, and *(iii)* **QHD-NC** for nonsmooth nonconvex problems. These variants, detailed in Section 3.2.1, are collectively referred to as **continuous-time QHD**. The corresponding continuous-time dynamics can be directly employed as quantum optimization algorithms by simulating their time evolution on a quantum computer (see Algorithm 3.1). We analyze the query and gate complexity of these algorithms in Corollary 3.2.5.

- We establish the global convergence of continuous-time QHD using energy reduction arguments tailored to each problem class. The convergence rates are proven for strongly convex (Theorem 3.2.12) and generally convex (Theorem 3.2.14) optimization problems. We also prove that **QHD-NC** converges to the global minimum of any (possibly nonsmooth and nonconvex) locally Lipschitz objective function in finite time (Theorem 3.2.16), which, to the best of our knowledge, is the *first rigorous* result in the literature on quantum optimization achieving such a guarantee. In contrast, classical methods like the subgradient method can fail to converge under the locally Lipschitz condition [66, 67], and no convergence result is known for the stochastic subgradient method assuming the locally Lipschitz condition only—let alone convergence to the global minimum.[1]

- Additionally, we perform numerical experiments to evaluate the performance of QHD in nonsmooth nonconvex optimization. We observe that QHD outperforms the subgradient method and its stochastic variants in 10 out of 12 test problems. In some cases, QHD achieves an optimality gap that is *orders of magnitude* lower than those of the classical methods. Our findings make QHD a promising approach to addressing nonsmooth optimization instances in both theoretical

[1] Exceptions exist when there are more assumptions on objective functions (e.g., being semialgebraic or definable in an o-minimal structure), under which the stochastic subgradient method converges to first-order stationary points [68].

and practical applications.

	Continuous-Time QHD	(Stochastic) Subgradient Method
μ-Strongly Convex	$O(e^{-\sqrt{\mu}t})$ (Theorem 3.2.12)	$\Theta(t^{-1}\log t)$ [69, 110, 111]
Convex	$O(t^{-2})$ (Theorem 3.2.14)	$\Theta(t^{-1/2})$ [112–114]
Locally Lipschitz	Global Minimum (Theorem 3.2.16)	Convergence Not Guaranteed [66, 67]

Table 3.1: Comparison of (worst-case) convergence rate bounds and convergence guarantees for continuous-time QHD and the subgradient method across various nonsmooth function classes. Here, t denotes time or the iteration number.

In Table 3.1, we present the convergence results for both continuous-time QHD across various classes of nonsmooth functions. For comparison, we also include the convergence rates of the subgradient method in both deterministic and stochastic settings. Continuous-time QHD achieves a faster convergence rate (or guarantees global convergence) compared to the subgradient method in all three categories.

Technical summary. In this chapter, we analyze the query and time complexity of continuous-time QHD, assuming a standard gate-model quantum computer. The complexity analysis of continuous-time QHD relies on a strengthened version (Proposition 3.2.4) of the time-dependent Hamiltonian simulation algorithm in [89]. We also provide a convergence analysis of continuous-time QHD for various classes of nonsmooth optimization problems. A brief summary follows below.

- Nonsmooth convex optimization: In the seminal work [52], the convergence results are obtained by constructing a specific Lyapunov function. Mathematically, a Lyapunov function is a time-dependent function associated with a dynamical system that is nonincreasing in time. Lyapunov analysis is a standard tool in the study of long-term behaviors of differential equations. By showing the Lyapunov function is nonincreasing in time, [52] proves the expectation

value of the objective converges to the global minimum $f(x^*)$ at a rate $\mathcal{O}(e^{-\beta_t})$, where β_t is a time-dependent function in the quantum Hamiltonian. A detailed overview of QHD is provided in Section 2.2. However, the choice of β_t (and other time-dependent parameters) is not clear in our setting. Meanwhile, to show the monotonicity of the Lyapunov function, [52] requires a well-defined derivative of the objective f, which does not directly apply to a nonsmooth objective f. To circumvent the technical issue, we *(i)* introduce two Lyapunov functions, each tailored to **QHD-C** and **QHD-SC**, respectively, and then *(ii)* employ a weaker notion of derivatives to rewrite the operator commutator in the Lyapunov function via the integral by parts formula (see Lemma 3.2.9). It is worth noting that our new argument requires a C^∞ regularity estimate of the solution to the Schrödinger equation, as discussed in Section A.2.2.

- Nonsmooth nonconvex optimization: [52] provides a heuristic argument to show the global convergence of QHD based on a quantum adiabatic approximation for smooth nonconvex optimization problems. This heuristic argument heavily relies on the fact that the function f is twice differentiable at the global minimizer x^*, which may not hold in nonsmooth optimization. We rigorously establish the asymptotic convergence of **QHD-NC** to the global minimum of $f(x)$ by exploiting a correspondence (often called an *energy argument* in physics) between the expected function value $\mathbb{E}[f(X_t)]$ and the energy spectrum of the quantum Hamiltonian $H_{\text{NC}}(t)$. We prove the global convergence of **QHD-NC** by analyzing the spectral properties of the quantum Hamiltonian via Weyl's law [115,116], which enables us to get rid of the smoothness assumption at the global minimizer.

Code Availability. The source code of the empirical study are available at `https://github.com/lwins-lights/QHD-experiment`.

Organization. This rest of this chapter is organized as follows. Section 3.2 presents the three variants of continuous-time QHD for various classes of optimization problems and provides the corresponding convergence analysis. Section 3.3 presents a numerical evaluation of QHD for nonsmooth nonconvex optimization problems and demonstrates its advantage over classical algorithms.

3.2 Continuous-Time QHD for Nonsmooth Optimization

3.2.1 Quantum Algorithms and Complexity Analysis

Motivated by the classical study of accelerated gradient descent [117], we propose three variants of QHD, each corresponding to a specific class of optimization problems. In each class, the optimization problems are characterized by the convexity of the objective function.

Definition 3.2.1 (Convexity)**.** A function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *convex* if for all $0 \le t \le 1$ and any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we have

$$f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \le tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y}). \tag{3.2}$$

Definition 3.2.2 (Subgradient)**.** Let $f\colon \mathbb{R}^d \to \mathbb{R}$. We say a vector $\boldsymbol{g} \in \mathbb{R}^d$ is a *subgradient* of f at $\boldsymbol{x}$ if,

$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \boldsymbol{g}^\top(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^d. \tag{3.3}$$

The subgradient of a function f at a fixed point is not unique. The *subdifferential* of f at $\boldsymbol{x} \in \mathbb{R}^d$ is defined as the set of all subgradients of f at $\boldsymbol{x}$, denoted by

$$\partial f(\boldsymbol{x}) \coloneqq \{\boldsymbol{g} \in \mathbb{R}^d \colon f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \boldsymbol{g}^\top(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^d\}. \tag{3.4}$$

Definition 3.2.3 (Strong convexity)**.** Let $\mu > 0$. A function f is a *μ-strongly convex* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ and any $\boldsymbol{g} \in \partial f(\boldsymbol{x})$, we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{g}^\top(\boldsymbol{y} - \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2. \tag{3.5}$$

Now, we list the variants of QHD for nonsmooth optimization problems with different convexity structures:

(a) QHD for strongly convex optimization problems (**QHD-SC**):

$$i\partial_t \Psi(t) = H_{\mathrm{SC}}(t)\Psi(t), \quad H_{\mathrm{SC}}(t) = e^{-2\sqrt{\mu}t}\left(-\frac{1}{2}\Delta\right) + e^{2\sqrt{\mu}t}f(x), \quad t \in [0, \infty), \tag{3.6}$$

where f is a μ-strongly convex function.

(b) QHD for convex optimization problems (**QHD-C**):

$$i\partial_t \Psi(t) = H_{\mathrm{C}}(t)\Psi(t), \quad H_{\mathrm{C}}(t) = \frac{1}{t^3}\left(-\frac{1}{2}\Delta\right) + t^3 f(x), \quad t \in [T_0, \infty), \tag{3.7}$$

where f is a convex function, $T_0 > 0$ is the starting time.[2]

(c) QHD for nonconvex optimization problems (**QHD-NC**):

$$i\partial_t \Psi(t) = H_{\mathrm{NC}}(t)\Psi(t), \quad H_{\mathrm{NC}}(t) = \frac{1}{\alpha t^{1/3}}\left(-\frac{1}{2}\Delta\right) + \alpha t^{1/3} f(x), \quad t \in [T_0, \infty), \tag{3.8}$$

where $\alpha > 0$ is a parameter that controls the rate of the evolution. The choice of α, which

[2]We require $T_0 > 0$ to ensure the quantum Hamiltonian $H_{\mathrm{C}}(t)$ is well-defined for all t.

depends on f and the target optimality gap, will be detailed in the proof of Theorem 3.2.16.

Note that in all three cases, the QHD Hamiltonian takes the following form:

$$H(t) = \frac{1}{\lambda(t)}\left(-\frac{1}{2}\Delta\right) + \lambda(t)f(x), \tag{3.9}$$

where $\lambda(t)$ is a monotonically increasing function in time t. In Algorithm 3.1, we give the quantum algorithm that implements continuous-time QHD to solve the optimization problem (3.1).

We assume our quantum algorithm only has access to the *quantum evaluation oracle* (i.e., zeroth-order oracle), which is defined as a unitary map O_f such that for any $|\boldsymbol{x}\rangle$,

$$O_f\left(|\boldsymbol{x}\rangle \otimes |0\rangle\right) = |\boldsymbol{x}\rangle \otimes |f(\boldsymbol{x})\rangle\,. \tag{3.10}$$

Note that the quantum evaluation oracle can be coherently accessed.

Algorithm 3.1: Continuous-time QHD

Classical inputs: time-dependent parameter $\lambda(t)$ (**QHD-SC**: $\lambda(t) = e^{2\sqrt{\mu}t}$; **QHD-C**: $\lambda(t) = t^3$; **QHD-NC**: $\lambda(t) = \alpha t^{1/3}$), starting time T_0, stopping time T_f

Quantum inputs: zeroth-order oracle O_f, an initial guess state $|\Psi_0\rangle$

Output: an approximate solution to the minimization problem (3.1)

1. Initialize the quantum computer to the quantum state $|\Psi_0(\boldsymbol{x})\rangle$
2. Simulate the quantum dynamics described by (3.9) for time $[T_0, T_f]$ using Corollary 3.2.5.
3. Measure the final quantum state $|\Psi(T)\rangle$ using the computational basis.

In the last step, the computational basis measurement corresponds to measuring the wave function $\Psi(T, \boldsymbol{x})$ using the position quadrature observable $\hat{\boldsymbol{x}}$. The outcome is a random variable X_T distributed as the probability density $|\Psi(T, \boldsymbol{x})|^2$.

Complexity analysis. The complexity of Algorithm 3.1 is mainly determined by the cost of simulating the Schrödinger equation governed by the time-dependent Hamiltonian operator (3.9). Previous results (e.g., [77]) focus on the quantum simulation with a smooth potential function, which do not apply to our nonsmooth setting. To address this issue, we prove the following theorem for simulating Schrödinger equation with nonsmooth (but bounded) potential functions, as detailed in Proposition 3.2.4. The proof of the result can be found in Section A.2.3, which relies on a new Sobolev estimate that characterizes the growth of derivatives of the wave function $\Psi(t,x)$ in continuous-time QHD.

Proposition 3.2.4. *Consider the time-dependent Schrödinger equation:*

$$i\frac{\partial}{\partial s}\left|\Psi(s,x)\right\rangle = \left[-\frac{1}{2}\Delta + \varphi(s)f(x)\right]\left|\Psi(s,x)\right\rangle, \tag{3.11}$$

where the function $\Psi(s,x)\colon [0,s_f]\times\Omega\to\mathbb{C}$ is subject to a period and analytic initial condition $\Psi_0(x)$ over the box $\Omega=[-M,M]^d$ for some $M>0$ with periodic boundary condition. Suppose the potential function $f(x)$ is bounded in Ω. We define

$$\xi = \int_0^{s_f}\varphi(s)\mathrm{d}s, \quad S = \sup_{x\in\Omega}|f(x)|. \tag{3.12}$$

Then, the Schrödinger equation (3.11) can be simulated for time $s\in[0,s_f]$ up to accuracy η with $\widetilde{\mathcal{O}}\left(\xi S\right)$ queries to the quantum evaluation oracle O_V, and an additional gate complexity

$$\widetilde{\mathcal{O}}\left(\xi S(d+\mathrm{poly}(\xi S))\right), \tag{3.13}$$

where the $\widetilde{\mathcal{O}}(\cdot)$ notation suppresses poly-logarithmic factors in η.

To leverage the simulation algorithm as described in Proposition 3.2.4, we need to map the QHD equation to the standard form as in (3.11). We consider a change of variable $s \mapsto t(s)$, where $t\colon [0, s_f] \to [T_0, T_f]$ represents a time scaling (the function $t(s)$ will be determined later), and introduce a new function $\widetilde{\Psi}(s) \coloneqq \Psi(t(s))$, where $\Psi(t)$ solves the QHD equation:

$$i\partial_t \Psi(t) = H(t)\Psi(t), \quad H(t) = \frac{1}{\lambda(t)}\left(-\frac{1}{2}\Delta\right) + \lambda(t) f(x). \tag{3.14}$$

By the chain rule, we have

$$i\partial_s \widetilde{\Psi}(s) = \frac{\mathrm{d}t}{\mathrm{d}s}\left[\frac{1}{\lambda(t(s))}\left(-\frac{1}{2}\Delta\right) + \lambda(t(s)) f(x)\right]\widetilde{\Psi}(s). \tag{3.15}$$

To obtain the standard form with a Hamiltonian $-\Delta + \varphi(s) f$, we require

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \lambda(t(s)), \tag{3.16}$$

which is a differential equation and can be solved for each choice of $\lambda(t)$:

1. **QHD-SC**: $\lambda(t) = e^{2\sqrt{\mu}t}$ for $t \in [0, T_f]$, $t(s) = \frac{1}{2\sqrt{\mu}} \log\left(\frac{1}{1-2\sqrt{\mu}s}\right)$ for $s \in \left[0, \frac{1}{2\sqrt{\mu}}\left(1 - e^{-2\sqrt{\mu}T_f}\right)\right]$.
2. **QHD-C**: $\lambda(t) = t^3$ for $t \in [T_0, T_f]$, $t(s) = \frac{1}{\sqrt{T_0^{-2} - 2s}}$ for $s \in [0, (T_0^{-2} - T_f^{-2})/2]$.
3. **QHD-NC**: $\lambda(t) = \alpha t^{1/3}$ for $t \in [T_0, T_f]$, $t(s) = (T_0^{2/3} + \frac{2\alpha}{3}s)^{3/2}$ for $s \in \left[0, \frac{3}{2\alpha}\left(T_f^{2/3} - T_0^{2/3}\right)\right]$.

Therefore, to simulate continuous-time QHD, it is equivalent to simulate the following Schrödinger equation

$$i\partial_s \Psi(s) = \mathcal{H}(s)\Phi(s), \quad \mathcal{H}(s) = -\frac{1}{2}\Delta + \varphi(s) f(x), \quad s \in [0, s_f], \tag{3.17}$$

where the rescaled time-dependent function $\varphi(s)$ and the end evolution time s_f is given in Table 3.2. Therefore, we can invoke Proposition 3.2.4 to simulate continuous-time QHD. The overall query complexity is summarized in Corollary 3.2.5.

Variants of QHD	(Rescaled) time-dependent function $\varphi(s)$	End Evolution Time s_f
QHD-SC	$\varphi(s) = (1 - 2\sqrt{\mu}s)^{-2}$	$s_f = \frac{1}{2\sqrt{\mu}}\left(1 - e^{-2\sqrt{\mu}T_f}\right)$
QHD-C	$\varphi(s) = (T_0^{-2} - 2s)^{-3}$	$s_f = (T_0^{-2} - T_f^{-2})/2$
QHD-NC	$\varphi(s) = \alpha^2\left(T_0^{2/3} + \frac{2\alpha}{3}s\right)$	$s_f = \frac{3}{2\alpha}\left(T_f^{2/3} - T_0^{2/3}\right)$

Table 3.2: Time-dependent functions and corresponding end evolution time in the rescaled continuous-time QHD (3.17).

Corollary 3.2.5. *Given an analytic initial data, the query complexity of simulating QHD is: (i)* $\widetilde{\mathcal{O}}\left(\frac{1}{\sqrt{\mu}}e^{2\sqrt{\mu}T_f}\right)$ *for* ***QHD-SC****, (ii)* $\widetilde{\mathcal{O}}\left(T_f^4 - T_0^4\right)$ *for* ***QHD-C****, and (iii)* $\widetilde{\mathcal{O}}\left(\alpha(T_f^{4/3} - T_0^{4/3})\right)$ *for* ***QHD-NC****.*

Proof. By the change of variable, we have

$$\int_0^{s_f} \varphi(s)\mathrm{d}s = \int_0^{s_f} \lambda(t(s))\frac{\mathrm{d}t}{\mathrm{d}s}\mathrm{d}s = \int_{T_0}^{T_f} \lambda(t)\mathrm{d}t. \tag{3.18}$$

Then, we can compute the query complexity based on Proposition 3.2.4. □

Remark 3.2.6. We note that the query complexity of **QHD-SC** scales exponentially in the evolution time T_f. In the following section, we will show that **QHD-SC** converges to the global minimum at a linear rate, meaning that the evolution time can be $\mathcal{O}(\log(1/\epsilon))$, where ϵ is a target optimality gap. As a result, the overall query complexity of **QHD-SC** still scales polynomially in $1/\epsilon$.

3.2.2 Convergence Analysis

3.2.2.1 Weak Derivative and Subgradient

In this section, we introduce the notion of weak derivatives for nondifferentiable functions.

Definition 3.2.7. A function $f\colon \mathbb{R}^d \to \mathbb{R}$ is a *locally integrable* function, denoted by $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, if

$$\int_{\mathbb{R}^d} |f(\boldsymbol{x})\varphi(\boldsymbol{x})| \, \mathrm{d}\boldsymbol{x} < +\infty \tag{3.19}$$

for any test function $\varphi \in C_c^\infty(\mathbb{R}^d)$, i.e., $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ is a compactly supported, smooth function.

Definition 3.2.8. Let $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ and $\boldsymbol{g} = (g_1, \ldots, g_d)^\top$ with $g_j \in L^1_{\text{loc}}(\mathbb{R}^d)$ for all $j = 1, \ldots, d$. We say that g_j is a *weak partial derivative* of f if

$$\int f \partial_j \varphi \, \mathrm{d}x = -\int g_j \varphi \, \mathrm{d}x \tag{3.20}$$

for all test functions $\varphi \in C_c^\infty(\mathbb{R}^d)$. If $\boldsymbol{g}$ exists, we say f is *weakly differentiable*, and $\boldsymbol{g}$ in the *weak gradient* of f.

A weak derivative of f, if exists, is unique up to a set of measure zero [118]. Similar to strong derivatives, we have the integral by parts formula for weak derivatives, as summarized in the following lemma.

Lemma 3.2.9 (Integral by parts)**.** *Suppose that $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ admits a weak gradient $\boldsymbol{g}$. Let $\Phi(\boldsymbol{x})\colon \mathbb{R}^d \to$*

$\mathbb{R}^d$ *such that* $\Phi_j(\boldsymbol{x}) \in W^{1,1}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$ *for each* $j = 1, \ldots, d$. *Then, we have*

$$\int (\nabla \cdot \Phi) f \, \mathrm{d}\boldsymbol{x} = -\int \Phi \cdot \boldsymbol{g} \, \mathrm{d}\boldsymbol{x}. \tag{3.21}$$

Proof. It suffices to prove the lemma for $d = 1$. Since $\Phi(x) \in W^{1,1}(\mathbb{R}) \cap C^\infty(\mathbb{R})$, we can construct a sequence of test functions $\{\varphi_k\}_{k=1}^\infty \subset C_c^\infty(\mathbb{R})$ such that *(i)* $\varphi_k(x) \leq \Phi(x)$ for all x, and *(ii)* $\varphi_k(x) \to \Phi(x)$ for every x. Then, by the dominated convergence theorem, we have

$$\int (\partial_x \Phi)(x) f(x) \mathrm{d}x = \lim_{k\to\infty} \int (\partial_x \varphi_k)(x) f(x) \mathrm{d}x = \lim_{k\to\infty} -\int \varphi_k(x) g(x) \mathrm{d}x = -\int \Phi(x) g(x) \mathrm{d}x \tag{3.22}$$

□

In general, even if a function f is differentiable almost everywhere with a classical derivative f' defined over $\mathbb{R}^d \setminus \Omega_f$, the classical derivative f' does not necessarily coincide with the weak derivative of f. For example, we consider the indicator function $\chi_{(0,1)}$ defined over the interval $(0, 1)$. It is straightforward to check that the weak derivative of $\chi_{(0,1)}$ is $\delta_0 - \delta_1$, i.e., the difference between two Dirac measures. However, the classical derivative of $\chi_{(0,1)}$ is 0 except for $x \in \{0, 1\}$. In this chapter, we assume f is Lipschitz continuous, which implies that f is differentiable almost everywhere and f' automatically corresponds to a weak derivative of f.

Theorem 3.2.10. *Assume* $f \colon \mathbb{R}^d \to \mathbb{R}$ *is locally Lipschitz. Then,* f *is differentiable almost everywhere, and its gradient equals its weak gradient almost everywhere.*

Proof. This is a direct consequence of Theorem 4 and Theorem 5 in Section 5 of [118]. □

Corollary 3.2.11. *Assume $f : \mathbb{R}^d \to \mathbb{R}$ is locally Lipschitz. If $\boldsymbol{g}$ is a subgradient of f, then $\boldsymbol{g}$ equals the weak derivative of f almost everywhere.*

Proof. This follows from the fact that the subgradient $\boldsymbol{g}$ must coincide with the strong gradient ∇f wherever it exists. □

3.2.2.2 Case I: Nonsmooth Strongly Convex Optimization

The main convergence result for **QHD-SC** is summarized as follows.

Theorem 3.2.12. *Suppose that $f(x)$ is locally integrable, Lipschitz continuous, μ-strongly convex, and potentially nonsmooth. Let $\Psi(t, \boldsymbol{x})$ be the solution to QHD-SC (3.6) subject to a smooth initial state $\Psi_0(\boldsymbol{x})$, and X_t be a random variable following the distribution $|\Psi(t, \boldsymbol{x})|^2$. Then, we have*

$$\mathbb{E}[f(X_t)] - f(\boldsymbol{x}^*) \leq O(e^{-\sqrt{\mu}t}). \tag{3.23}$$

Proof. We prove Theorem 3.2.12 using a Lyapunov function argument. For $t \geq 0$, let $\Psi(t, \boldsymbol{x})$ be the solution to **QHD-SC**, i.e.,

$$i\partial_t \Psi(t) = H_{\mathrm{SC}}(t)\Psi(t), \quad H_{\mathrm{SC}}(t) = e^{-2\sqrt{\mu}t}\left(-\frac{1}{2}\Delta\right) + e^{2\sqrt{\mu}t} f(x), \tag{3.24}$$

subject to an initial state $\Psi(0, \boldsymbol{x}) = \Psi_0(\boldsymbol{x})$. Without loss of generality, we assume f has a unique

global minimum $f(\boldsymbol{x}^*) = 0$ with $\boldsymbol{x}^* = 0$. We consider the following Lyapunov function:

$$\mathcal{E}(t) = \left\langle \hat{O}_{\mathrm{SC}}(t) \right\rangle_t, \quad \hat{O}_{\mathrm{SC}}(t) \coloneqq f + \frac{1}{4} e^{-4\sqrt{\mu}t} \sum_{j=1}^{d} p_j^2 + \frac{1}{4} \sum_{j=1}^{d} J_j^2, \quad J_j = e^{-2\sqrt{\mu}t} p_j + 2\sqrt{\mu} x_j. \tag{3.25}$$

Here, the operators p_j and x_j are the momentum and position operators for the j-th coordinate, as described in (3.48). As shown in Lemma 3.2.13, the Lyapunov function $\mathcal{E}(t)$ is strictly nonincreasing in time:

$$\dot{\mathcal{E}}(t) \leq -\frac{\sqrt{\mu}}{4} \mathcal{E}(t) < 0. \tag{3.26}$$

Moreover, by Grönwall's inequality, for any $t \geq 0$, we have

$$\mathcal{E}(t) \leq e^{-\frac{\sqrt{\mu}}{4}t} \mathcal{E}(0) \tag{3.27}$$

for any $t \geq 0$. Meanwhile, note that the operators p_j^2 and J_j^2 are positive for each $j = 1, \ldots, d$. By the definition of the Lyapunov function, we have

$$\mathcal{E}(t) = \langle f \rangle_t + \frac{1}{4} e^{-4\sqrt{\mu}t} \sum_{j=1}^{d} \langle p_j^2 \rangle_t + \frac{1}{4} \sum_{j=1}^{d} \langle J_j^2 \rangle_t \geq \langle f \rangle_t. \tag{3.28}$$

Combining (3.27) and (3.28), we show the exponential convergence of **QHD-SC**:

$$\langle f \rangle_t \leq e^{-\frac{\sqrt{\mu}}{4}t} \mathcal{E}(0), \tag{3.29}$$

which concludes the proof of Theorem 3.2.12. □

Lemma 3.2.13. *Let f be a locally integrable and Lipschitz continuous function. For $t \geq 0$, we denote $\Psi(t, \boldsymbol{x})$ as the solution to **QHD-SC** with an initial state $\Psi_0(\boldsymbol{x}) \in C^\infty(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Let the function $\mathcal{E}(t)$ be the same as in (3.25). Then, for any $t \geq 0$, $\Psi(t, \boldsymbol{x})$ is a smooth function in $\boldsymbol{x}$ and*

$$\dot{\mathcal{E}}(t) = \sqrt{\mu}\left[-e^{-4\sqrt{\mu}t}\sum_{j=1}^{d}\langle p_j^2\rangle_t + \int \nabla\cdot\left(|\Psi(t,\boldsymbol{x})|^2\boldsymbol{x}\right) f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right]. \tag{3.30}$$

Moreover, if f is μ-strongly convex with $\boldsymbol{x}^* = 0$ and $f(\boldsymbol{x}^*) = 0$, the following holds for all $t \geq 0$:

$$\dot{\mathcal{E}}(t) \leq -\frac{\sqrt{\mu}}{4}\mathcal{E}(t). \tag{3.31}$$

Proof. The smoothness of the solution at any t follows from the standard regularity theory of Schrödinger equations, see Theorem A.2.8. Similar to in the proof of Lemma 3.2.15, the time derivative of the Lyapunov function can be expressed as follows:

$$\partial_t\mathcal{E}(t) = \langle\partial_t\hat{O}_{\mathrm{SC}}\rangle_t + \langle i[H_{\mathrm{SC}}(t), \hat{O}_{\mathrm{SC}}]\rangle_t, \tag{3.32}$$

with

$$\langle\partial_t\hat{O}_{\mathrm{SC}}(t)\rangle_t = \left\langle -2\sqrt{\mu}e^{-4\sqrt{\mu}t}\sum_{j=1}^{d}p_j^2 - \mu e^{-2\sqrt{\mu}t}\sum_{j=1}^{d}\{x_j, p_j\}\right\rangle_t, \tag{3.33}$$

and

$$\langle i[H_{\mathrm{SC}}(t), \hat{O}_{\mathrm{SC}}(t)]\rangle_t = \frac{\mu}{2} e^{-2\sqrt{\mu}t} \sum_{j=1}^{d} \langle i[p_j^2, x_j^2]\rangle_t + \frac{\sqrt{\mu}}{4} e^{-4\sqrt{\mu}t} \sum_{j=1}^{d} \langle i[p_j^2, \{x_j, p_j\}]\rangle_t \tag{3.34}$$

$$+ \sqrt{\mu} \int \nabla \cdot \left(|\Psi|^2 \boldsymbol{x}\right) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{3.35}$$

For any $j = 1, \ldots, d$, we have the following commutation relations [52, Appendix B.1, Lemma 2]:

$$i[p_j^2, x_j^2] = 2\{x_j, p_j\}, \quad i[p_j^2, \{x_j, p_j\}] = 4p_j^2. \tag{3.36}$$

By adding the two parts of $\partial_t \mathcal{E}(t)$ together and using the above commutation relations, we obtain

$$\partial_t \mathcal{E}(t) = \sqrt{\mu} \left(-e^{-4\sqrt{\mu}t} \sum_{j=1}^{d} \langle p_j^2 \rangle_t + \int \nabla \cdot \left(|\Psi(t, \boldsymbol{x})|^2 \boldsymbol{x}\right) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right), \tag{3.37}$$

which proves the first part of the lemma. When f is Lipschitz continuous and μ-strongly convex, there is a subgradient $\boldsymbol{g}$ that coincides with its weak derivative almost everywhere. By Lemma 3.2.9, we have

$$\int \nabla \cdot \left(|\Psi|^2 \boldsymbol{x}\right) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = -\langle \boldsymbol{x} \cdot \boldsymbol{g}(\boldsymbol{x})\rangle_t \leq -\left\langle f + \frac{\mu}{2}\|\boldsymbol{x}\|^2 \right\rangle_t. \tag{3.38}$$

Therefore, by plugging (3.38) into (3.37), we have

$$\partial_t \mathcal{E}(t) \leq -\sqrt{\mu} \left\langle e^{-4\sqrt{\mu}t} \sum_{j=1}^{d} p_j^2 + f + \frac{\mu}{2} \sum_{j=1}^{d} x_j^2 \right\rangle_t = \mathcal{J}_1(t) + \mathcal{J}_2(t), \tag{3.39}$$

where

$$\mathcal{J}_1(t) \coloneqq -\sqrt{\mu}\left\langle \frac{1}{8}e^{-4\sqrt{\mu}t}\sum_{j=1}^{d} p_j^2 + \frac{\mu}{2}\sum_{j=1}^{d} x_j^2 \right\rangle_t, \quad \mathcal{J}_2(t) \coloneqq -\sqrt{\mu}\left\langle f + \frac{7}{8}e^{-4\sqrt{\mu}t}\sum_{j=1}^{d} p_j^2 \right\rangle_t. \tag{3.40}$$

Furthermore, by invoking the operator inequality $\frac{1}{2}(A+B)^2 \leq A^2 + B^2$ (see (3.41)) and the positivity of f and p_j^2 (see (3.43)), we obtain that

$$\mathcal{J}_1(t) = -\frac{\sqrt{\mu}}{8}\sum_{j=1}^{d}\left\langle e^{-4\sqrt{\mu}t}p_j^2 + 4\mu x_j^2\right\rangle_t \leq -\frac{\sqrt{\mu}}{16}\sum_{j=1}^{d} J_j^2, \tag{3.41}$$

$$\mathcal{J}_2(t) = -\frac{\sqrt{\mu}}{4}\mathcal{E}(t) + \frac{\sqrt{\mu}}{16}\left(\sum_{j=1}^{d} J_j^2\right) - \left(\frac{3\sqrt{\mu}}{4}\langle f\rangle_t + \frac{13\sqrt{\mu}}{16}e^{-4\sqrt{\mu}t}\sum_{j=1}^{d}\langle p_j^2\rangle_t\right) \tag{3.42}$$

$$\leq -\frac{\sqrt{\mu}}{4}\mathcal{E}(t) + \frac{\sqrt{\mu}}{16}\left(\sum_{j=1}^{d} J_j^2\right). \tag{3.43}$$

As a consequence, we have

$$\dot{\mathcal{E}}(t) \leq \mathcal{J}_1(t) + \mathcal{J}_2(t) \leq -\frac{\sqrt{\mu}}{4}\mathcal{E}(t), \tag{3.44}$$

which proves the second part of the lemma. □

3.2.2.3 Case II: Nonsmooth Convex Optimization

The main convergence result for **QHD-C** is summarized as follows.

Theorem 3.2.14. *Suppose that $f(x)$ is locally integrable, Lipschitz continuous, convex, and potentially nonsmooth. Let $\Psi(t,\boldsymbol{x})$ be the solution to **QHD-C** (3.7) subject to a smooth initial state $\Psi_0(\boldsymbol{x})$, and*

X_t be a random variable following the distribution $|\Psi(t)|^2$. Then,

$$\mathbb{E}[f(X_t)] - f(\boldsymbol{x}^*) \leq O(t^{-2}). \tag{3.45}$$

Proof. Let $\Psi(t, \boldsymbol{x})$ be the solution to the differential equation of **QHD-C** for $t \geq T_0 > 0$ subject to an initial state $\Psi(0, \boldsymbol{x}) = \Psi_0(\boldsymbol{x})$. For a self-adjoint operator $\hat{O}$, we define the expected value of $\hat{O}$ as a function of time t:

$$\langle \hat{O} \rangle_t \coloneqq \int \overline{\Psi(t, \boldsymbol{x})} \hat{O} \Psi(t, x) \mathrm{d}\boldsymbol{x} = \left\langle \Psi(t) \middle| \hat{O} \middle| \Psi(t) \right\rangle. \tag{3.46}$$

With this notation, we have $\mathbb{E}[f(X_t)] = \langle f \rangle_t$.

Let f be a convex objective function. Without loss of generality, we assume $f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{x}^* = 0$. We consider the following Lyapunov function,

$$\mathcal{E}(t) = \left\langle \hat{O}_{\mathrm{C}}(t) \right\rangle_t, \quad \hat{O}_{\mathrm{C}}(t) \coloneqq t^2 f + \frac{1}{2} \sum_{j=1}^{d} \left(\frac{1}{t^4} p_j^2 + \frac{2}{t^2} \{p_j, x_j\} + 4 x_j^2 \right). \tag{3.47}$$

Here, the operators p_j and x_j denote the momentum and position operators for the j-th coordinate. They should be interpreted as operators acting on a smooth test function $\varphi(\boldsymbol{x}) \colon \mathbb{R}^d \to \mathbb{R}$ as follows:

$$(p_j \varphi)(\boldsymbol{x}) \coloneqq -i \partial_{x_j} \varphi(\boldsymbol{x}), \quad (x_j \varphi)(\boldsymbol{x}) \coloneqq x_j \varphi(\boldsymbol{x}). \tag{3.48}$$

$\{A, B\} = AB + BA$ represents the anti-commutator between operators A and B.

For each $j = 1, \ldots, d$, we observe that

$$t^{-4}p_j^2 + 2t^{-2}\{x_j, p_j\} + 4x_j^2 = \left(t^{-2}p_j + 2x_j\right)^2 \geq 0, \tag{3.49}$$

which immediately implies that $\hat{O}_{\mathrm{C}}(t) - t^2 f$ is a positive operator, i.e.,

$$t^2\langle f\rangle_t \leq \mathcal{E}(t). \tag{3.50}$$

In Lemma 3.2.15, we show that $\mathcal{E}(t)$ is a nonincreasing function in time for all $t \geq T_0$. As a result,

$$\langle f\rangle_t \leq \frac{\mathcal{E}(t)}{t^2} \leq \frac{\mathcal{E}(T_0)}{t^2}, \tag{3.51}$$

which proves Theorem 3.2.14. □

Lemma 3.2.15. *Let f be a locally integrable and Lipschitz continuous function. For $t \geq T_0$, we denote $\Psi(t, \boldsymbol{x})$ as the solution to **QHD-C** with an initial state $\Psi_0(\boldsymbol{x}) \in C^\infty(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Let the function $\mathcal{E}(t)$ be the same as in (3.47). Then, for any $t \geq T_0$, $\Psi(t, \boldsymbol{x})$ is a smooth function in $\boldsymbol{x}$ and*

$$\dot{\mathcal{E}}(t) = 2t\left[\langle f\rangle_t + \int \nabla \cdot \left(|\Psi(t, \boldsymbol{x})|^2 \boldsymbol{x}\right) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\right]. \tag{3.52}$$

Moreover, if f is a convex function with $\boldsymbol{x}^* = 0$ and $f(\boldsymbol{x}^*) = 0$, we have $\dot{\mathcal{E}}(t) \leq 0$ for all $t \geq T_0$.

Proof. The smoothness of the solution at any t follows from the standard regularity theory of Schrödinger equations, see Theorem A.2.8. Recall that we have $\mathcal{E}(t) = \left\langle \Psi(t) \middle| \hat{O}_{\mathrm{C}} \middle| \Psi(t) \right\rangle$, with $\hat{O}_{\mathrm{C}} = t^2 f +$

$\frac{1}{2}\sum_{j=1}^{d}\left(\frac{1}{t^4}p_j^2+\frac{2}{t^2}\{p_j,x_j\}+4x_j^2\right)$. By the distribution law of derivative, we have

$$\partial_t\mathcal{E}(t)=\left\langle\partial_t\Psi(t)\middle|\hat{O}_{\mathrm{C}}\middle|\Psi(t)\right\rangle+\left\langle\Psi(t)\middle|\partial_t\hat{O}_{\mathrm{C}}\middle|\Psi(t)\right\rangle+\left\langle\Psi(t)\middle|\hat{O}_{\mathrm{C}}\middle|\partial_t\Psi(t)\right\rangle \tag{3.53}$$

$$=\langle\partial_t\hat{O}_{\mathrm{C}}(t)\rangle_t+\langle i[H_{\mathrm{C}}(t),\hat{O}_{\mathrm{C}}(t)]\rangle_t, \tag{3.54}$$

where the last step follow from the equation $\partial_t\Psi=-iH_{\mathrm{C}}(t)\Psi$. Here, $[\cdot,\cdot]$ is the commutator of two operators: $[A,B]=AB-BA$. Further calculation shows

$$\langle\partial_t\hat{O}_{\mathrm{C}}(t)\rangle_t=\left\langle 2tf-2t^{-3}\sum_{j=1}^{d}\{x_j,p_j\}-2t^{-5}\sum_{j=1}^{d}p_j^2\right\rangle_t, \tag{3.55}$$

and

$$\langle i[H_{\mathrm{C}}(t),\hat{O}_{\mathrm{C}}(t)]\rangle_t=\left\langle t^{-3}\sum_{j=1}^{d}i[p_j^2,x_j^2]+t^{-5}\sum_{j=1}^{d}i[p_j^2,\{x_j,p_j\}/2]\right\rangle_t+2t\int\nabla\cdot\left(|\Psi(t)|^2\boldsymbol{x}\right)f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{3.56}$$

Note that the last term in (3.56) is equivalent to $-2t\langle\nabla f(\boldsymbol{x})\cdot\boldsymbol{x}\rangle_t$ if f is continuously differentiable. However, since f is nonsmooth, we need to move the derivatives to the "smooth part" of the integral in order to make the expression well-defined. For any $j=1,\ldots,d$, we have the following commutation relations [52, Appendix B.1, Lemma 2]:

$$i[p_j^2,x_j^2]=2\{x_j,p_j\},\quad i[p_j^2,\{x_j,p_j\}]=4p_j^2. \tag{3.57}$$

Therefore, by adding up (3.55) and (3.56), we end up with

$$\partial_t \mathcal{E}(t) = 2t\langle f \rangle_t + 2t \int \nabla \cdot \left(|\Psi(t)|^2 \boldsymbol{x} \right) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{3.58}$$

which proves the first part of the lemma. When f is Lipschitz continuous, Corollary 3.2.11 implies that there exists a subgradient fo f (denoted by $\boldsymbol{g}$) that equals the weak derivative of f almost everywhere. By Lemma 3.2.9, we have

$$\int \nabla \cdot \left(|\Psi(t,\boldsymbol{x})|^2 \boldsymbol{x} \right) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = -\int |\Psi(t,\boldsymbol{x})|^2 \boldsymbol{x} \cdot \boldsymbol{g}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = -\langle \boldsymbol{x} \cdot \boldsymbol{g} \rangle_t. \tag{3.59}$$

Plugging (3.59) into (3.58), we have

$$\dot{\mathcal{E}}(t) = 2t\langle f - \boldsymbol{x} \cdot g \rangle_t \leq 0, \tag{3.60}$$

where the last inequality follows from the fact that $\boldsymbol{g}$ is a subgradient so $f + \boldsymbol{g} \cdot (\boldsymbol{x}^* - \boldsymbol{x}) \leq f(\boldsymbol{x}^*) = 0$. □

3.2.2.4 Case III: Nonsmooth Nonconvex Optimization

In this section, we prove the global convergence of QHD for nonconvex problems. For simplicity, we impose box constraints on the optimization problem

$$\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}), \tag{3.61}$$

where $\Omega \coloneqq [-1,1]^d$. We note that this assumption is made for technical reasons and does not materially affect the behavior of the quantum algorithm. Moreover, the bounded search space can be further mapped into Ω through appropriate variable transformations. Since the domain is no longer $\mathbb{R}^d$, we make the QHD wave function $\Psi(t)$ in (3.8) satisfy a Dirichlet boundary condition, namely

$$\Psi(t, \boldsymbol{x}) = 0, \quad \text{for } \boldsymbol{x} \in \partial\Omega. \tag{3.62}$$

Define $|\phi\rangle$ as the ground state of $-\Delta$, where Δ is the Laplace operator. Its wave function can be written analytically:

$$\phi(\boldsymbol{x}) = \phi(x_1, \ldots, x_d) = \prod_{1 \le j \le d} \cos\Big(\frac{\pi}{2} x_j\Big). \tag{3.63}$$

Arguably, $|\phi\rangle$ is easy to prepare since it is a product state with simple components. Now we are ready to state the main convergence result for **QHD-NC**.

Theorem 3.2.16. *For any $\delta > 0$, there exist $\alpha = \alpha(f, \delta) > 0$ and $T_0 = T_0(f, \delta) > 0$ for **QHD-NC** (3.8) with the initial state being $|\phi\rangle$, such that the final solution $|\psi\rangle \coloneqq \Psi(T)$ satisfies $\langle\psi|f|\psi\rangle \le \min f(\boldsymbol{x}) + \delta$, where $T = T(f, \delta)$. In other words, let X_T be a random variable following the distribution $|\Psi(T)|^2$, we have*

$$\mathbb{E}[f(X_T)] \le \min f(\boldsymbol{x}) + \delta. \tag{3.64}$$

The proof of Theorem 3.2.16 heavily relies on the analysis of spectra of Schrödinger operators $H = -\Delta + V$ where $V \colon \Omega \to \mathbb{R}$ is a continuous potential function. Refer to Section A.2.1 for important properties of Schrödinger operators. Let us introduce necessary tools before proceeding to the formal proof.

Adiabatic evolution of unbounded quantum Hamiltonian. Consider the dynamics described by the Schrödinger equation

$$i\epsilon\frac{\mathrm{d}}{\mathrm{d}t}\left|\psi^{\epsilon}(t)\right\rangle = H(t)\left|\psi^{\epsilon}(t)\right\rangle, \tag{3.65}$$

subject to an initial state $|\psi^{\epsilon}(0)\rangle$. Denote by $U^{\epsilon}(t)$ the propagator of the dynamics described by (3.65), so the solution at time t is given by

$$\left|\psi^{\epsilon}(t)\right\rangle = U^{\epsilon}(t)\left|\psi^{\epsilon}(0)\right\rangle. \tag{3.66}$$

The parameter $\epsilon > 0$ controls the time scale on which the quantum Hamiltonian $H(t)$ varies. For a small ϵ, the system evolves slowly and the dynamics are relatively simple: if the system begins with an eigenstate of $H(0)$, it remains close to an eigenstate of $H(t)$. This process is called *adiabatic quantum evolution*.

Formally, let $P(t)$ be the rank-1 projector onto the ground state of $H(t)$. The quantum adiabatic theorem states that the propagator $U^{\epsilon}(t)$ approximately maps the ground state of $H(0)$ to that of $H(t)$ for sufficiently small ϵ. Various formulations of the quantum adiabatic theorem exist in the literature, while most of them assume the Hamiltonian is a bounded linear operator. In QHD, however, the Hamiltonian is an unbounded operator defined on Ω. Here, we introduce a quantum adiabatic theorem for unbounded Hamiltonian, which allows us to have a neat analysis without discretizing the unbounded operator in QHD.

Theorem 3.2.17 ([57, Sec. 1.1]). *Let $\dot{H}(t)$ and $\ddot{H}(t)$ be bounded.*[3] *Assume that the spectral gap of*

[3] i.e., they have finite *graph norm*, which can be bounded by their usual 2-norm; See [57].

$H(t)$ is always bounded above some $c > 0$. Then, there exists a constant C such that

$$\|(I - P(t))U^{\epsilon}(t)P(0)\| \leq C\epsilon(1+t). \tag{3.67}$$

Characterization of the low-energy spectrum. Our proof idea is basically relating the dynamics in (3.65) to that of **QHD-NC** (3.8). Since the quantum adiabatic theorem (Theorem 3.2.17) ensures that the system in (3.65) will approximately stay in the instantaneous ground state of $H(t)$ throughout, it is thus important to characterize the ground state. This ultimately reduces to analyzing the spectral properties of $H = -h^2\Delta + V$ in the limit $h \to 0$, which is called *semiclassical analysis* in the mathematics and physics literature. A seminal result from semiclassical analysis is the *Weyl's law*, which asymptotically characterizes the low-energy spectrum of the Schrödinger operator when $h \to 0$.

Definition 3.2.18. Define $N(E, H)$ as the number of eigenvalues no greater than E, counting multiplicities, of a self-adjoint operator H.

Lemma 3.2.19 (Weyl's law [115, 116])**.** *Let $\Omega \subseteq \mathbb{R}^d$ be open and bounded. Fix a continuous[4] function $V \colon \Omega \to \mathbb{R}$. Define $H(h) \coloneqq -h^2\Delta + V$ which acts on L^2-functions that vanish on $\partial\Omega$. We have*

$$N(E, H(h)) = \left(\frac{1}{(4\pi)^{d/2}\Gamma(d/2+1)} \cdot h^{-d} \int_{\Omega} \max\{E - V(\boldsymbol{x}), 0\}^{d/2} \mathrm{d}\boldsymbol{x}\right)(1 \pm o(1)),^{5} \tag{3.68}$$

when $E \to +\infty$ or $h \to +0$, where $\Gamma(\cdot)$ is the gamma function.

Below we prove the main convergence result for **QHD-NC** (3.8).

[4] The continuity assumption can be weakened to $V \in L^{d/2}(\Omega)$; See [116].

[5] $A = B(1 \pm o(1))$ is equivalent to $A/B \to 1$.

Proof of Theorem 3.2.16. Let $\Omega \coloneqq [-1,1]^d$ be the domain of f. Without loss of generality assume $\inf_{x\in\Omega} f(\boldsymbol{x}) = 0$. Set $H(t) \coloneqq -\Delta + tf(\boldsymbol{x})$ for $t \in [0, t_{\max}]$ with $t_{\max}$ to be determined later. Let $\|f\|$ be the operator norm of f, or equivalently $\|f\| \coloneqq \sup_{\boldsymbol{x}\in\Omega}\{f(\boldsymbol{x})\}$. Note that $\|f\| < \infty$ because f is continuous on its compact domain Ω. $\dot{H}(t)$ and $\ddot{H}(t)$ are bounded since $f(\boldsymbol{x})$ is bounded. Applying Lemma A.2.1 with Remark A.2.2 we know that for any $t \in [0, t_{\max}]$, $H(t)$ has a positive spectral gap $g(t)$. Moreover, since $g\colon [0, t_{\max}] \to \mathbb{R}_+$ is obviously continuous and $[0, t_{\max}]$ is compact, $g(\cdot)$ has a *minimum* over $[0, t_{\max}]$ and hence $H(t)$ satisfies the spectral gap condition in Theorem 3.2.17. Now, by Theorem 3.2.17, there exists a constant C such that

$$\|(1 - P(t))U^\epsilon(t)P(0)\| \leq C\epsilon(1 + t), \tag{3.69}$$

where we recall that $U^\epsilon(t)$ is the propagator of the dynamics in

$$i\epsilon\frac{\mathrm{d}}{\mathrm{d}t}\left|\psi^\epsilon(t)\right\rangle = H(t)\left|\psi^\epsilon(t)\right\rangle \tag{3.70}$$

such that $\left|\psi^\epsilon(t)\right\rangle = U^\epsilon(t)\left|\psi^\epsilon(0)\right\rangle$ for any initial state $\left|\psi^\epsilon(0)\right\rangle$, and $P(t)$ is the rank-1 projector onto the ground state of $H(t)$, denoted by $\left|\phi_0(t)\right\rangle$. Note that $\left|\phi\right\rangle = P(0)\left|\phi\right\rangle$ as $\left|\phi\right\rangle = \left|\phi_0(0)\right\rangle$ is the ground state of $H(0) = -\Delta$. Therefore we have

$$\|(I - P(t))U^\epsilon(t)\left|\phi\right\rangle\| = \|(I - P(t))U^\epsilon(t)P(0)\left|\phi\right\rangle\| \leq C\epsilon(1 + t), \tag{3.71}$$

where the inequality follows from (3.69). Now, we set

$$\left|\psi^\epsilon(0)\right\rangle = \left|\phi\right\rangle, \tag{3.72}$$

and decompose $|\psi^\epsilon(t)\rangle$ into two orthogonal components being its projection onto $P(t)$ and the remaining part:

$$|\psi^\epsilon(t)\rangle = U^\epsilon(t)\,|\phi\rangle = P(t)U^\epsilon(t)\,|\phi\rangle + (I - P(t))U^\epsilon(t)\,|\phi\rangle\,. \tag{3.73}$$

Since $U^\epsilon(t)$ is unitary and $P(t)$ projects onto the instantaneous ground state $|\phi_0(t)\rangle$, we can rewrite (3.73) as

$$|\psi^\epsilon(t)\rangle = \alpha(t)\,|\phi_0(t)\rangle + \beta(t)\,|r^\epsilon(t)\rangle\,, \tag{3.74}$$

where $\beta(t)\,|r^\epsilon(t)\rangle \coloneqq (I-P(t))U^\epsilon(t)\,|\phi\rangle$ with $|r^\epsilon(t)\rangle$ normalized, $|\beta(t)| \le C\epsilon(1+t)$ and $|\alpha(t)| = 1 - |\beta(t)|^2$. We will use this decomposition of $|\psi^\epsilon(t)\rangle$ to bound its expectation on $f(\cdot)$, i.e., $\langle\psi^\epsilon(t)|f|\psi^\epsilon(t)\rangle$.

Define $S(t) \coloneqq \frac{1}{t}H(t) = -\left(\frac{1}{\sqrt{t}}\right)^2\Delta + f(\boldsymbol{x})$. Obviously $|\phi_0(t)\rangle$ is also the ground state of $S(t)$. Recall that we assumed $\inf_{x\in\Omega} f(\boldsymbol{x}) = 0$ in the beginning of the proof. Thus the continuity of $f(\cdot)$ implies that for any $E > 0$,

$$\int_\Omega \max\{E - f(\boldsymbol{x}), 0\}^{d/2}\mathrm{d}\boldsymbol{x} > 0. \tag{3.75}$$

Therefore we can estimate $N(\delta/10, S(t))$ by plugging into Weyl's law $h \coloneqq 1/\sqrt{t}$ and $E \coloneqq \delta/10$:

$$N(\delta/10, S(t)) = \left(\frac{1}{(4\pi)^{d/2}\Gamma(d/2+1)}\cdot t^{d/2}\int_\Omega \max\{\delta/10 - f(\boldsymbol{x}), 0\}^{d/2}\mathrm{d}\boldsymbol{x}\right)(1 \pm o(1)), \tag{3.76}$$

which implies $\lim_{t\to+\infty} N(\delta/10, S(t)) \to \infty$. It follows that the ground energy of $S(t)$ will be eventually smaller than $\delta/10$ as t increases. Therefore, there exists a sufficiently large $t_{\max}$ such that

$$\langle\phi_0(t_{\max})|f|\phi_0(t_{\max})\rangle \le \langle\phi_0(t_{\max})|S(t_{\max})|\phi_0(t_{\max})\rangle \le \delta/10, \tag{3.77}$$

where the first inequality follows from $-\Delta$ being positive semidefinite. Write $\alpha \coloneqq \alpha(t_{\max})$ and

$\beta \coloneqq \beta(t_{\max})$. We have

$$\langle\psi^\epsilon(t_{\max})|f|\psi^\epsilon(t_{\max})\rangle \tag{3.78a}$$

$$= |\alpha|^2 \langle\phi_0(t_{\max})|f|\phi_0(t_{\max})\rangle + |\beta|^2 \langle r^\epsilon(t_{\max})|f|r^\epsilon(t_{\max})\rangle$$

$$+ 2\,\mathrm{Re}\{\alpha\beta \langle\phi_0(t_{\max})|f|r^\epsilon(t_{\max})\rangle\} \qquad \text{(by (3.74))} \tag{3.78b}$$

$$\le |\alpha|^2\delta/10 + |\beta|^2\|f\| + 2|\alpha\beta|\|f\| \qquad \text{(by (3.77))} \tag{3.78c}$$

$$\le |\alpha|^2\delta/10 + C\epsilon(1+t_{\max})\|f\| + 2|\alpha|C\epsilon(1+t_{\max})\|f\| \qquad (|\beta|^2 \le |\beta| \le C\epsilon(1+t_{\max})) \tag{3.78d}$$

$$\le \delta/10 + C'\epsilon(1+t_{\max})\|f\|, \qquad (|\alpha| \le 1) \tag{3.78e}$$

where C' is some constant that depends on C and we assume $C\epsilon(1+t_{\max}) \ll 0$. Let $\epsilon \coloneqq \delta/(10C'(1+t_{\max})\|f\|)$ we obtain

$$\langle\psi^\epsilon(t_{\max})|f|\psi^\epsilon(t_{\max})\rangle \le 2\delta/10. \tag{3.79}$$

Now we relate the dynamics of (3.70) to QHD via a time dilation trick. Let

$$H_{\mathrm{NC}}(\tau) \coloneqq \frac{1}{\sqrt{t/2}}\left(-\frac{1}{2}\Delta\right) + \sqrt{t/2}f(\boldsymbol{x}), \quad t \coloneqq \frac{1}{2}(3\epsilon\tau)^{2/3}. \tag{3.80}$$

It is straightforward that $\frac{\mathrm{d}}{\mathrm{d}\tau} = \epsilon\frac{1}{\sqrt{2t}}\frac{\mathrm{d}}{\mathrm{d}t}$. Hence (3.70) gives

$$i\frac{\mathrm{d}}{\mathrm{d}\tau}|\psi^\epsilon(\tau)\rangle = i\epsilon\frac{1}{\sqrt{2t}}\frac{\mathrm{d}}{\mathrm{d}t}|\psi^\epsilon(t)\rangle = \frac{1}{\sqrt{2t}}H(t)|\psi^\epsilon(t)\rangle = H_{\mathrm{NC}}(\tau)|\psi^\epsilon(\tau)\rangle, \tag{3.81}$$

where $|\psi^\epsilon(\tau)\rangle \coloneqq |\psi^\epsilon(t)\rangle$ with $t = \frac{1}{2}(3\epsilon\tau)^{2/3}$. As the result, for any $t_0 > 0$, running QHD according to the dynamics (3.81) with the initial state $|\psi^\epsilon(t_0)\rangle$ for $\tau \in [T_0, T]$ will yield the final state $|\psi^\epsilon(t_{\max})\rangle$, where $T \coloneqq \frac{2\sqrt{2}}{3}\frac{t_{\max}^{3/2}}{\epsilon}$ and $T_0 \coloneqq \frac{2\sqrt{2}}{3}\frac{t_0^{3/2}}{\epsilon}$. Note that for sufficiently small $t_0 = t_0(f,\delta)$ we can always

have that $|\psi^\epsilon(t_0)\rangle$ is $(\delta/5)$-close to $|\psi^\epsilon(0)\rangle = |\phi\rangle$ in ℓ_2 distance due to the continuity of $|\psi^\epsilon(\cdot)\rangle$ from (3.70) and (3.72). Consequently, replacing the initial condition $|\psi^\epsilon(0)\rangle = |\phi\rangle$ by

$$|\psi^\epsilon(t_0)\rangle = |\phi\rangle\,, \tag{3.82}$$

we have

$$\langle\psi^\epsilon(t_{\max})|f|\psi^\epsilon(t_{\max})\rangle \leq 2\delta/10 + \delta/5 + \delta/5 < \delta, \tag{3.83}$$

by using triangle inequalities with (3.79), as desired in the theorem statement. □

3.3 Numerical Results for Nonsmooth Nonconvex Optimization

In this section, we focus on the empirical performance of QHD for nonsmooth nonconvex optimization problems since we have already established the convergence of both continuous- and discrete-time QHD for nonsmooth convex optimization in earlier sections. We evaluate the performance of QHD and two variants of subgradient methods, namely, Subgrad (vanilla subgradient algorithm)[6] and LFMSGD (learning rate-free momentum stochastic gradient descent) [70]. We test them on 12 nonsmooth, nonconvex, and box-constrained functions, including one 1-dimensional, nine 2-dimensional, and two 3-dimensional cases, as detailed in Section A.3.1. For details on QHD, Subgrad, and LFMSGD, refer to Section 2.2, Section A.1, and Section A.3.3, respectively. All source code is publicly available [120].

[6] In our implementation of Subgrad, the algorithm returns the final iterate instead of the best among all iterates, requiring only *one* function value query to $f(\cdot)$ throughout and allowing us to allocate other queries to $\partial_C f$ for a better performance with the same query number limit. This is a common practice in subgradient methods [119].

3.3.1 Methodology

We use classical simulation to evaluate the performance of QHD. For fairness among QHD, Subgrad, and LFMSGD, each algorithm is allowed a total of 10,000 queries to the function value (resp., subgradient) for QHD (resp. Subgrad and LFMSGD) in a single run. Subgrad and LFMSGD will start with 10,000 initial points (in 10,000 independent runs) chosen from the domain uniformly at random. The experiment is conducted on a consumer laptop (Intel Core i7-8750H CPU 2.20GHz).

Parameter Setup. For QHD, we introduce a parameter L to rescale a test function into hypercube $[-L, L]^d$. Specifically, a test function $g : \bigtimes_{1\leq i\leq d}[l_i, u_i] \to \mathbb{R}$ is transformed into the function $f : [-L, L]^d \to \mathbb{R}$ as follows:

$$f(x_1, \ldots, x_d) \coloneqq g(x'_1, \ldots, x'_d), \quad x'_i \coloneqq l_i + \frac{u_i - l_i}{2L}(x_i + L). \tag{3.84}$$

We parameterize QHD with L and set $\lambda(t) = t^3$ in (3.9), with total evolution time $T = 10$, time step size $h = 1/1000$, and spatial discretization number $N = 512$.[7] We choose the same $\lambda(t)$ as in **QHD-C** (3.7) rather than **QHD-NC** (3.8) because it yields significantly stronger empirical performance. This is analogous to the case of stochastic gradient descent (SGD), where learning rate schedules used in practice bear little resemblance to those recommended by theory [121]. For Subgrad parameterized by η, the learning rate follows $s_j = \eta/\sqrt{j}$ in light of Theorem A.1.1. LFMSGD, parameterized by σ, is configured with the default setting $\beta = 0.9$ as recommended in [70], and the Gaussian noise added to each evaluation of the subgradient follows $\mathcal{N}(\mathbf{0}, \sigma^2)$.

[7] Adjusting L is equivalent to modifying $\lambda(t)$ and T in QHD; refer to Section A.3.2.

Performance Metric. All three algorithms—QHD, Subgrad, and LFMSGD—exhibit randomness: QHD and LFMSGD have internal sources of randomness, while LFMSGD and Subgrad depend on the random initial point. To evaluate their performance, we use the *best-of-k optimality gap* as a metric.[8] This metric is inspired by a common practice of running a randomized algorithm multiple times and selecting the best result. Formally, for an integer k, the best-of-k optimality gap is defined as:

$$\mathbb{E}\left[\min_{1\le i\le k} f(\tilde{\boldsymbol{x}}_i) - f_{\min}\right], \tag{3.85}$$

where $f(\cdot)$ is the test function, $f_{\min} \coloneqq \min_{\boldsymbol{x}} f(\boldsymbol{x})$ is the global minimum, $\tilde{\boldsymbol{x}}_i$ is the solution returned by the algorithm in the ith run, and the expectation is taken with respect to the algorithm's internal randomness and/or random initialization. The metric reduces to the expected optimality gap when $k = 1$. We test all three algorithms on 12 test functions with $k \in \{1, 3, 10, 30, 100\}$. The best-of-$k$ optimality gap is estimated using the Monte Carlo method for Subgrad and LFMSGD. In contrast, for QHD, it is computed exactly since the full distribution of the final output of the algorithm is known in our classical simulation.

Parameter Optimization. QHD, Subgrad, and LFMSGD are parameterized and their performance depend heavily on the choice of corresponding parameters L, η, and σ, respectively. Therefore, we employ Bayesian optimization [122] to select the optimal parameter that minimizes the best-of-k optimality gap. Bayesian optimization is particularly suited for hyperparameter tuning in settings where function evaluations are costly. For each combination of $k \in \{1, 3, 10, 30, 100\}$ and test function, we limit the number of evaluations of the best-of-k optimality gap to 100 when optimizing parameters.

[8] Here, k denotes the number of runs and should not be confused with its usage as the iteration number in previous sections.

3.3.2 Benchmarking QHD against Classical Algorithms

Test Function	k	QHD Gap	LFMSGD Gap	Subgrad Gap	Test Function	k	QHD Gap	LFMSGD Gap	Subgrad Gap
SCHWEFEL (1D)	1	2.72e+1	1.37e+2	1.90e+2	KEANE (2D)	1	1.70e-1	2.17e-1	5.31e-1
	3	3.29e-1	3.18e+1	7.91e+1		3	2.58e-2	5.81e-2	4.45e-1
	10	1.49e-3	5.51e-1	1.17e+1		10	3.81e-3	1.20e-2	2.95e-1
	30	1.79e-4	1.15e-5	1.05e-1		30	1.17e-3	3.22e-3	1.24e-1
	100	2.37e-6	4.20e-13	7.98e-9		100	3.37e-4	7.67e-4	1.23e-2
ACKLEY (2D)	1	3.30e-1	1.45e+1	1.09e+0	RANA (2D)	1	1.33e+2	1.14e+2	4.96e+1
	3	2.98e-2	8.93e+0	3.41e-1		3	5.35e+1	5.21e+1	5.05e+1
	10	2.67e-2	3.62e+0	9.45e-2		10	2.05e+1	2.10e+1	2.17e+1
	30	2.67e-2	2.14e+0	4.48e-2		30	6.86e+0	5.78e+0	5.62e+0
	100	2.67e-2	3.22e-1	2.38e-2		100	1.21e+0	7.28e-1	3.08e-1
BUKIN06 (2D)	1	8.19e+0	5.48e+1	1.71e+1	WF (2D)	1	4.25e+0	6.01e+0	4.66e+0
	3	2.29e+0	2.65e+1	5.90e+0		3	1.12e+0	5.41e+0	2.78e+0
	10	5.92e-1	1.40e+1	2.81e+0		10	4.73e-2	3.54e+0	6.48e-1
	30	3.67e-1	7.95e+0	1.53e+0		30	2.48e-5	1.41e+0	6.17e-2
	100	3.54e-1	4.10e+0	8.26e-1		100	3.28e-13	2.57e-1	1.86e-2
CARROMTABLE (2D)	1	1.73e+1	1.95e+1	2.10e+1	XINSHEYANG04 (2D)	1	2.11e-1	9.34e-1	9.63e-1
	3	8.45e+0	1.29e+1	1.69e+1		3	3.63e-2	7.06e-1	8.92e-1
	10	1.03e+0	3.72e+0	9.81e+0		10	1.47e-3	3.38e-1	6.84e-1
	30	2.00e-2	3.24e-1	2.74e+0		30	1.63e-6	1.02e-1	3.23e-1
	100	1.41e-2	1.77e-2	3.65e-2		100	1.27e-16	2.67e-2	3.14e-2
CROWNEDCROSS (2D)	1	5.88e-1	1.24e+0	1.10e+0	DROPWAVE (3D)	1	1.06e-2	6.90e-1	8.04e-1
	3	4.78e-1	1.06e+0	9.25e-1		3	8.93e-6	5.02e-1	6.93e-1
	10	4.09e-1	9.43e-1	7.26e-1		10	1.81e-16	2.99e-1	5.22e-1
	30	3.54e-1	8.24e-1	6.65e-1		30	2.64e-43	1.69e-1	3.57e-1
	100	3.15e-1	7.25e-1	6.57e-1		100	4.75e-140	8.94e-2	1.71e-1
DAMAVANDI (2D)	1	2.00e+0	1.99e+0	1.97e+0	LAYEB04 (3D)	1	9.65e-1	2.08e+1	1.81e+1
	3	1.95e+0	1.97e+0	1.88e+0		3	2.92e-2	1.85e+1	1.47e+1
	10	1.73e+0	1.88e+0	1.60e+0		10	1.71e-2	1.60e+1	1.09e+1
	30	1.26e+0	1.62e+0	9.96e-1		30	1.35e-2	1.36e+1	8.09e+0
	100	4.65e-1	9.04e-1	1.88e-1		100	1.31e-2	1.11e+1	5.16e+0

Table 3.3: Best-of-k optimality gap summary for QHD, LFMSGD, and Subgrad on 12 test functions (cf. Figure 3.1), including one 1-dimensional, nine 2-dimensional, and two 3-dimensional cases. The smallest gaps among three algorithms are marked blue.

As shown in Figure 3.1, QHD consistently outperforms Subgrad and LFMSGD across 8 out of 12 test functions, achieving significantly smaller optimality gaps for all tested values of k. Notably, for `WF`, `XINSHEYANG04` and `DROPWAVE`, QHD demonstrates an overwhelming advantage, reducing the gap by multiple orders of magnitude compared to the classical methods for large k. For `KEANE`, `CROWNEDCROSS`, `BUKIN06`, `CARROMTABLE` and `LAYEBO4`, QHD also achieves superior performance, maintaining a decisive lead across increasing values of k. These results highlight the effectiveness of QHD in finding high-quality solutions where classical algorithms struggle to close

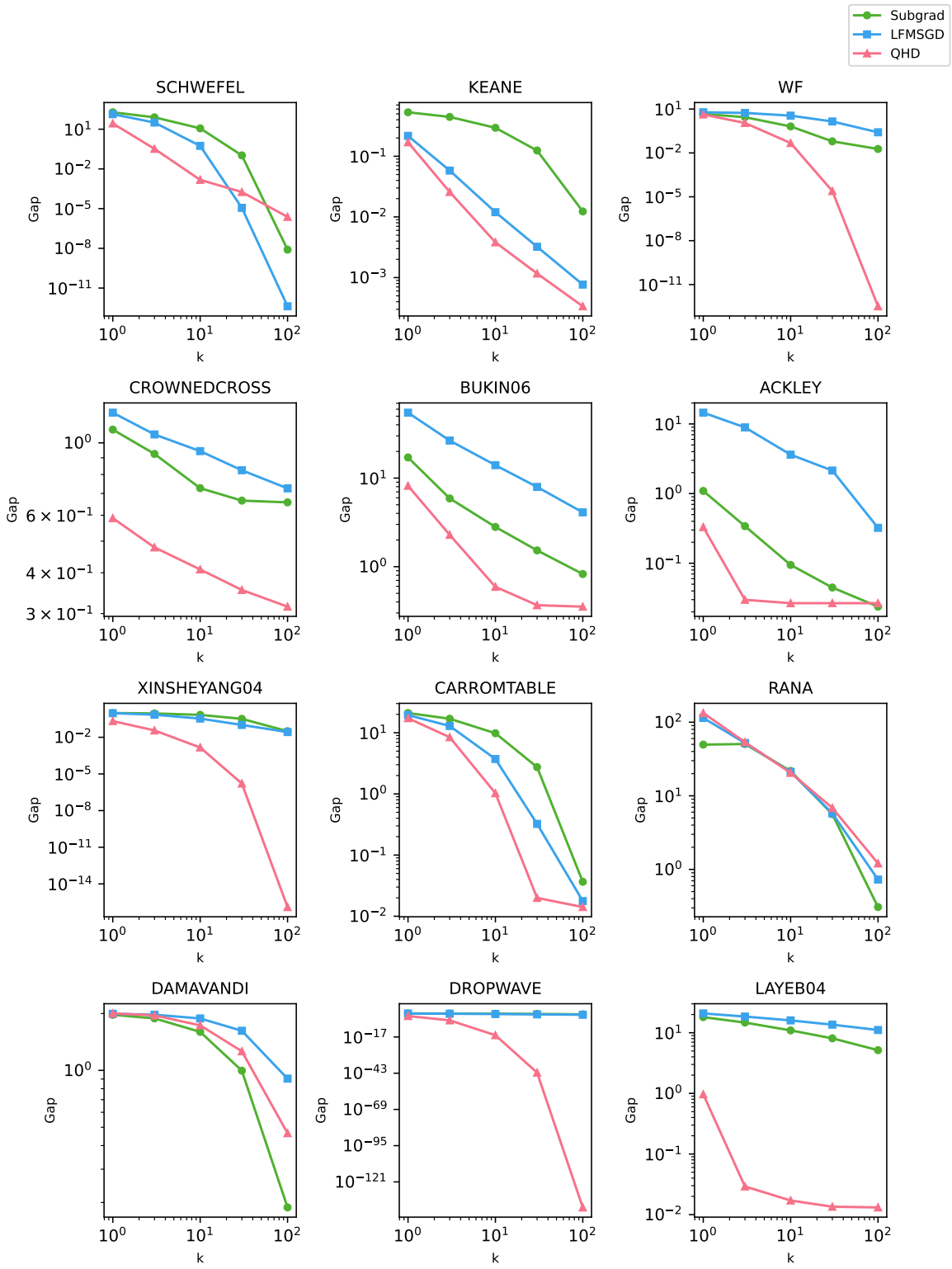


Figure 3.1: Best-of-k optimality gap for 12 test functions (cf. Table 3.3). QHD outperforms Subgrad and LFMSGD for all choices of k on every test functions except SCHWEFEL, ACKLEY, RANA and DAMAVANDI.

the optimality gap, emphasizing its robustness across a diverse set of nonconvex test functions. The numerical data corresponding to Figure 3.1 is presented in Table 3.3.

For SCHWEFEL and ACKLEY, QHD initially maintains an advantage over the classical methods but fails to sustain it as k increases. Specifically, for SCHWEFEL, QHD outperforms both Subgrad and LFMSGD when $k \leq 10$ but loses its lead for $k \geq 30$. This can be attributed to the structure of SCHWEFEL, where the global minimum has a large basin of attraction, making it easier for any local search algorithm to succeed with randomized initialization and enough repetitions. In contrast, for ACKLEY, QHD outperforms the classical methods up to $k = 30$ but stagnates at $k = 100$. This is primarily due to the spatial discretization in our QHD simulation, which limits its ability to "see" solutions closer to the global minimum; see Section A.3.4 for a detailed discussion.

For the remaining two test functions, RANA and DAMAVANDI, QHD yields an optimality gap comparable with classical methods, and no significant advantage is observed for QHD. This suggests that these functions may not exhibit structures where QHD can succesfully utilize. Nonetheless, QHD's competitive performance on them, combined with its superiority on the majority of test functions, reinforces its effectiveness as a powerful optimization algorithm.

3.3.3 QHD Dynamics: A Comparative Case Study

To illustrate how QHD outperforms Subgrad and LFMSGD in most cases, we need to show the intermediate iterations for these three algorithms. We use test function XINSHEYANG04 as an example. The landscape of the 2-dimensional multimodal function XINSHEYANG04 has its global minimum located at the center as shown in Figure 3.2. We also illustrate the intermediate iteration process of three algorithms on test function XINSHEYANG04 in Figure 3.3. We summarize our ob-

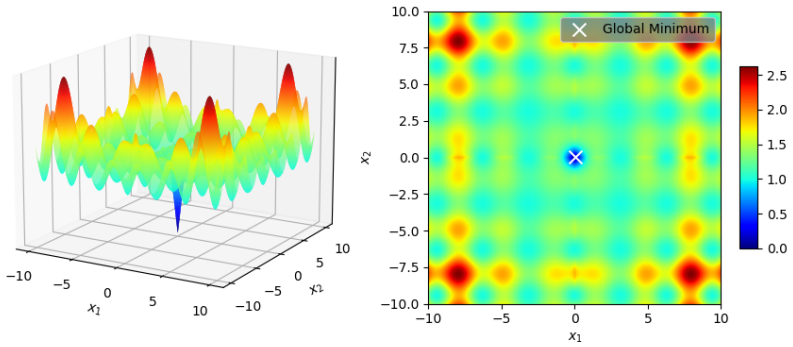


Figure 3.2: Landscape of test function XINSHEYANG04

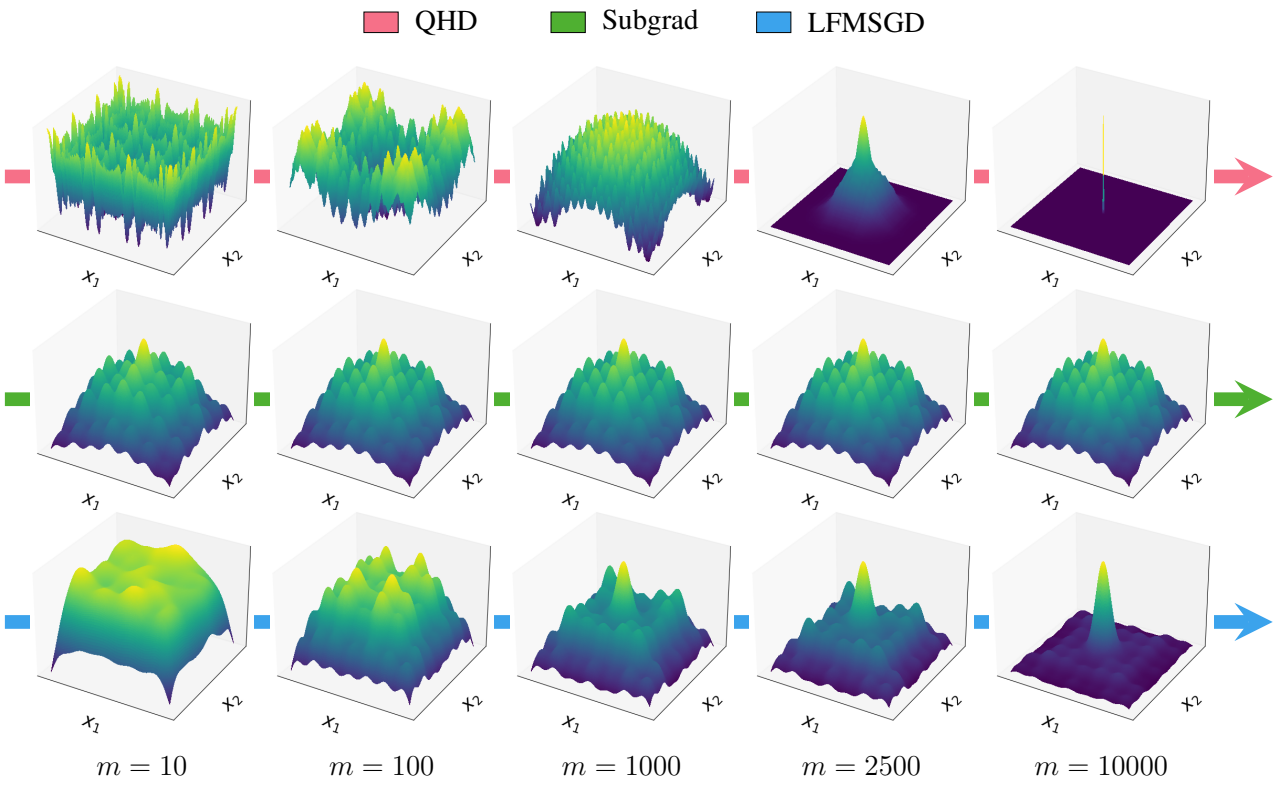


Figure 3.3: Distributions at intermediate iterations for QHD (1st row), Subgrad (2nd row), and LFMSGD (3rd row) on XINSHEYANG04 where m denotes the iteration number. Note that in QHD, $m = t/h = 1000t$ due to time discretization, where t denotes the evolution time.

servations below.

QHD demonstrates a strong ability to explore the landscape and escape local minima. Initially, the distribution appears highly fluctuating, reflecting a *kinetic phase*, during which the algorithm aver-

ages the initial wave function over the whole search space to reduce the risk of poor initialization [52]. As the evolution time t (or iteration number m) increases, QHD dynamically adapts, gradually concentrating on more promising regions while retaining some level of exploration (*global search phase*). By the end of the evolution, the distribution becomes more localized (*descent phase*), indicating convergence toward the global minimum while avoiding premature trapping in suboptimal regions. Eventually, the distribution narrows into a sharp peak at the global minimum.

LFMSGD also exhibits the ability to escape local minima, albeit being less efficient. In the early iterations, the distributions are relatively broad due to the random initialization. However, unlike QHD, the refinement process is less efficient, as evidenced by the gradual and somewhat diffusive concentration of probability mass over subsequent iterations. As the evolution progresses, the distribution becomes more peaked, though some residual spread persists. Eventually, LFMSGD converges, but the slower contraction of the distribution suggests a less directed search compared to QHD, highlighting its inefficiency despite its stochastic capability to escape local minima.

Subgrad, in contrast, exhibits a starkly different behavior. Once the algorithm enters a basin of attraction, determined by the random initialization, it remains trapped in the corresponding local minimum encountered early in the process, and this entrapment persists across all iterations. This aligns with known limitations of subgradient methods, which lack intrinsic mechanisms for escaping local optima, making them ineffective at navigating nonconvex landscapes.

In summary, QHD demonstrates the most effective escape dynamics and convergence efficiency, leveraging its quantum-inspired evolution to navigate complex landscapes. LFMSGD benefits from stochasticity to avoid local traps but at the cost of slower refinement. Subgradient descent, being fully deterministic, remains consistently trapped, reinforcing its limitations in highly nonconvex settings.

Chapter 4: Superpolynomial Speedup of QHD for Nonconvex Optimization

4.1 Introduction

Nonconvex optimization is a central object in machine learning and operations research. In practice, iterative optimization algorithms using local gradient information (e.g., stochastic gradient descent) have been proven successful for large-scale nonconvex optimization, especially when the optimization landscape is well-conditioned. However, lots of nonconvex problems naturally arising from application domains possess sophisticated optimization landscape with a huge number of saddle points and spurious local minima, posing great challenges for gradient-based algorithms. Heuristic optimization algorithms (e.g., simulated annealing, particle swarm, etc.), on the other hand, could explore the nonconvex landscape more efficiently, while few theoretical guarantees are established and the performance of heuristic algorithms varies drastically between instances.

In this chapter, we further investigate of behavior of QHD on nonconvex continuous optimization problems. We construct a specific family of optimization instances such that: (1) the number of local minima on these problems scales exponentially with the problem dimension; (2) no explicit sparsity or separability pattern can be employed to enable an efficient classical solution. Due to these features, the constructed problem instances are intractable for most classical optimization algorithms. Empirically, we find that several state-of-the-art classical optimization algorithms and solvers would require a *super-polynomial* time to solve these hard instances (see Section 4.5). Meanwhile, we prove that QHD solves these hard optimization problems in *polynomial* time. Our findings constitute new evidence of a significant performance separation between quantum and classical algorithms in nonconvex optimization.

Problem formulation. We construct a family of optimization instances representing nonconvex optimization problems with sophisticated landscapes. These hard problem instances are constructed as follows. First, we consider certain well-formed one-dimensional "double well" functions with two local minima (see Figure 4.1A). These well-formed double well functions can be expressed as degree-4 polynomials (see Section 4.3 for more details). Suppose that $w(x)$ is a well-formed 1D double well function with global minimizer at $x = x^*$, we then define a d-dimensional separable function $F(\boldsymbol{x}) := \sum_{k=1}^{d} w(x_k)$. We note that the function $F(\boldsymbol{x})$ has 2^d local minima and a unique global minimizer at $\boldsymbol{x}^* = (x^*, \ldots, x^*)$. In Figure 4.1B, we show a two-dimensional separable function with 4 local minima.

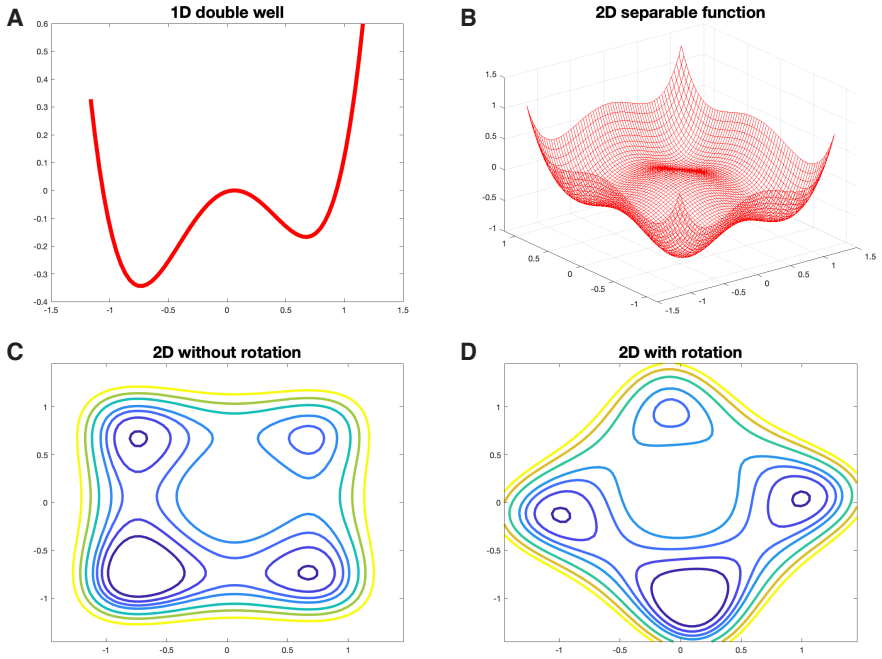


Figure 4.1: **Construction of optimization instances. A.** A well-formed double well function. **B.** A separable function built from two identical well-formed double well functions. **C.** Contour plot of the 2-dimensional separable function $F(\boldsymbol{x})$. **D.** Contour plot of the rotated 2-dimensional function $F_U(\boldsymbol{x})$. The local geometry of $F(\boldsymbol{x})$ is preserved, while the new function $F_U(\boldsymbol{x})$ is not separable.

Although the function $F(\boldsymbol{x})$ has exponentially many local minima, it is separable in each inde-

pendent variable x_k. The separability structure of $F(\boldsymbol{x})$ can be easily detected from its closed-form expression. Even if the closed-form expression is not available and the function is given via a black-box query model, a simple classical algorithm such as coordinate descent can split the problem into d independent sub-problems, each can be solved in constant time.

To create hard optimization instances for classical algorithms, we apply a random rotation to *hide* the separability structure in $F(\boldsymbol{x})$. More precisely, let U be an orthogonal matrix and we define a new function,

$$F_U(\boldsymbol{x}) \coloneqq F(U\boldsymbol{x}). \tag{4.1}$$

The newly generated function $F_U(\boldsymbol{x})$ still has 2^d local minima as the rotation preserves the geometry of the optimization landscape. However, unlike the original function $F(\boldsymbol{x})$, the rotated function $F_U(\boldsymbol{x})$ is no longer separable (see Figure 4.1C, D).

More importantly, the random rotations on nonconvex functions are difficult to revert. For example, if the double-well function $w(x)$ is a degree-4 polynomial, so is the rotated function $F_U(\boldsymbol{x})$ because the rotation by U is an affine transformation. However, to learn the rotation U from the closed-form expression of $F_U(\boldsymbol{x})$ amounts to diagonalizing a 4-tensor. Diagonalizing general 4-tensors is known to be NP-hard [123]. Therefore, it does not appear to be a trivial task for classical algorithms to revert the rotation and decompose $F_U(\boldsymbol{x})$ into d separate sub-problems.

Definition 4.1.1 (Nonconvex optimization instances)**.** Given an appropriate choice of the 1D double-well function $w(x)$ (more details available in Section 4.3), we define a class of optimization instances

$\mathscr{C}(w, d)$ for dimension $d \geq 1$:

$$\mathscr{C}(w, d) \coloneqq \{F_U(\boldsymbol{x}) : U \in \mathcal{Q}(d)\}, \tag{4.2}$$

where $F_U(\boldsymbol{x})$ is defined in (4.1) and $\mathcal{Q}(d)$ is the set of all d-by-d orthogonal matrix:

$$\mathcal{Q}(d) \coloneqq \{U \in \mathbb{R}^{d \times d} : UU^\top = I\}.$$

Main theoretical result. Due to our construction of the optimization instances, any instance $f \in \mathscr{C}(w, d)$ has 2^d local minima but a unique global minimum $f(\boldsymbol{x}^*)$. We say $\boldsymbol{x} \in \mathbb{R}^d$ is an δ-*approximate* solution to the optimization problem f if $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta$. Our main theoretical contribution is to prove that QHD can find a δ-approximate solution to any $f \in \mathscr{C}(w, d)$ in polynomial time.

We assume our quantum algorithm only has access to the *quantum evaluation oracle* (i.e., zeroth-order oracle), which is defined as a unitary map O_f such that for any $|\boldsymbol{x}\rangle$,

$$O_f \left(|\boldsymbol{x}\rangle \otimes |0\rangle\right) = |\boldsymbol{x}\rangle \otimes |f(\boldsymbol{x})\rangle \tag{4.3}$$

Note that the quantum evaluation oracle can be coherently accessed. Namely, for any $m \in \mathbb{N}$, let $|\boldsymbol{x}_1\rangle, \ldots, |\boldsymbol{x}_d\rangle \in \mathbb{R}^d$ and $\mathbf{c} \in \mathbb{C}^m$ such that $\|\mathbf{c}\| = 1$, we have

$$O_f \left(\sum_{j=1}^{m} c_j \, |\boldsymbol{x}_j\rangle \otimes |0\rangle \right) = \sum_{j=1}^{m} c_j \, |\boldsymbol{x}_j\rangle \otimes |f(\boldsymbol{x}_j)\rangle \, . \tag{4.4}$$

Our main theoretical result is summarized in the following theorem, where the $\widetilde{\mathcal{O}}$ notation suppresses poly-logarithmic factors in d and $1/\delta$.

Theorem 4.1.2 (Informal version of Theorem 4.4.1)**.** *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be an instance in the class $\mathscr{C}(w,d)$. For any $\delta > 0$, Quantum Hamiltonian Descent can produce an δ-approximate solution to f with probability at least $2/3$ using $\widetilde{\mathcal{O}}\left(\frac{d^3}{\delta^2}\right)$ quantum queries to f and additional $\widetilde{\mathcal{O}}\left(\frac{d^4}{\delta^2}\right)$ 1- and 2-qubit elementary gates.*

Empirical study of classical optimization algorithms. In our empirical study, we select 6 representatives from a comprehensive selection of state-of-the-art optimization algorithms in the following four major categories:

1. Annealing-based and Monte-Carlo algorithms: dual annealing [71], basin-hopping [72];
2. Gradient method: stochastic (or perturbed) gradient descent;
3. Newton and quasi-Newton methods: interior-point method (Ipopt [73]), sequential quadratic programming [74];
4. Branch-and-bound algorithm: Gurobi [75].

We test each optimization algorithm/solver on thousands of random instances from $\mathscr{C}(w,d)$ for a fixed degree-4 polynomial w and different values of d, and then measure the run time and success rate. Next, we compute the average *time-to-solution* (TTS) for each classical algorithm/solver as a function of the dimension d. All the empirical scaling results are computed using the same software and hardware environment to ensure comparability of the collected run time data.

Our data suggests that, for all tested classical algorithms/solvers, the measured TTS scales as a super-polynomial function in dimension d. On the contrary, our theoretical result implies that the TTS of QHD scales polynomially in dimension d. More discussions and the visualization of empirical data are available in Section 4.5.

Discussions on the implication and limitation. The original QHD paper [77] has demonstrated a large-scale empirical implementation of QHD on a class of boxed-constrained quadratic programming problems up to 75 dimensions by using D-Wave machines as a quantum Ising Hamiltonian simulator and embedding the QHD's dynamics into the evolution of quantum Ising Hamiltonian with the so-called Hamiltonian embedding technique. However, due to the limitation of the D-Wave machine (e.g., decoherence and the limited connectivity that restricts the sparsity of the realizable box-constrained quadratic programming instances), QHD's empirical performance fails to beat state-of-the-art branch-and-bound solvers (such as Gurobi). In this chapter, our newly constructed instances are empirically hard for Gurobi, which requires super-polynomial time to solve. However, our old technique fails to extend to implement QHD on the newly constructed instances on any existing quantum hardware. Instead, we theoretically prove that such instances can be solved in polynomial time on an ideal quantum computer.

Our empirical study covers a wide spectrum of classical state-of-the-art optimization algorithms, where the numerical results suggest the run time of all tested classical algorithms scales super-polynomially in dimension d. This empirical evidence implies that QHD could potentially replace the role of these classical solvers and become one competitive *off-the-shelf* optimization solution in practice. However, it does not exclude the possibility of efficient classical algorithms specially designed for our tested problem instances. We expect any such efficient classical algorithm, if exists, would, nevertheless, require a special design to leverage the hidden separable structure in the instances[1]. Moreover, our empirical study is by no means complete as it is bound by limited computing resources and time. In light of that, we make all of our empirical study publicly available and invite the whole community to

[1]Note that Theorem 4.1.2 can be established with more general instances than those used in the empirical test. Any specially designed classical algorithm should aim to work with the general condition and its hidden structure.

test it further.

We note further that our construction in the empirical test is unlikely to yield any super-polynomial oracle separation directly. This is because our optimization instances are d-dimensional degree-4 polynomials whose precise form can be fully determined with a polynomial, precisely $O(d^4)$, number of queries of the function value. Thus, any bigger than $\Omega(d^4)$ classical lower bound needs to be established in the plain model, which is likely beyond the reach of known techniques. We want to emphasize that our instances, however, could be more inspiring about concrete features of optimization problems that can be efficiently solved by QHD, compared with oracle separations, if possible, obtained via some technical routes. For example, any reduction from instances efficiently solvable by Shor's algorithm to optimization ones will likely not reveal any new quantum capability, as it is essentially running Shor's algorithm with an optimization disguise.

Finally, we want to emphasize that our specific construction of the hard optimization instances, however, should not be treated as a limitation of the applicability of the QHD algorithm, but rather a limitation of our method in analyzing QHD's behavior on general optimization instances. Indeed, as empirically evaluated in Section 3.3 and [77], QHD could perform well on optimization instances much beyond the specific construction used in this chapter. Analyzing the QHD's performance on these instances would require investigating the spectrum of Schrödinger's operator with general potential energy functions, which we believe is an interesting but challenging problem in pure math.

Code Availability. The source codes of the classical empirical study are available at `https://github.com/lwins-lights/QHD/tree/main/classical`.

4.2 Preliminaries

Notation. Throughout this chapter, we let $\mathcal{H} = L^2(\mathbb{R}^d)$ and $\mathcal{D} = C_0^\infty(\mathbb{R}^d)$. The Hilbert space $\mathcal{H}$ is equipped with the standard inner product $\langle u, v\rangle = \int_{\mathbb{R}^d} \bar{u}v \, \mathrm{d}x$, which induces the L^2-norm in $\mathcal{H}$: $\|u\|^2 = \langle u, u\rangle$.

4.2.1 Convexity and Spectral Gap

Our theoretical analysis in this chapter heavily relies on the spectrum of the Schrödinger operator,

$$\hat{H} = -\nabla^2 + f, \tag{4.5}$$

where ∇^2 is the Laplace operator[2] in the real space $\mathbb{R}^d$, and $f\colon \mathbb{R}^d \to \mathbb{R}$ is a continuous objective function that we want to minimize. The physical meaning of f is a "potential field" that confines the motion of a quantum particle.

When f is non-negative and it diverges at infinity (i.e., $\lim_{\|x\|\to\infty} f(x) = \infty$), the standard theory of elliptic operators [56, Theorem 10.7] guarantees that the spectrum of $\hat{H}$ is discrete and all eigenvalues are positive. We can arrange the eigenvalues of $\hat{H}$ in ascending order: $E_0 < E_1 \le \dots$. The difference between the first two eigenvalues (i.e., $E_1 - E_0$) is often called the *spectral gap*. It is well-known that the spectral gap of the Schrödinger operator (4.5) has a deep connection to the convexity of the potential field f. Since the eighties, several authors (van den Berg [124], Ashbaugh and Benguria [125], and Yau [126]) have independently observed that the spectral gap of a Schrödinger

[2] We do not use Δ for the Laplace operator in this chapter to save it for other purpose.

operator with a convex potential field has a lower bound $3\pi^2/D^2$, where D is the diameter of the domain on which the Schrödinger operator is defined. This observation, known as the Fundamental Gap Conjecture, has been proven by Andrews and Clutterbuck [127] in 2011 using an involved analysis of the heat equation.

We now discuss a simple example known as the quantum harmonic oscillator. A quantum harmonic oscillator is described by a Schrödinger operator with a quadratic potential field,

$$H = \hbar^2 \left(-\frac{1}{2}\nabla^2 \right) + \left(\frac{1}{2}\omega^2 x^2 \right), \tag{4.6}$$

where ω is a fixed frequency and $\hbar$ is the so-called Planck constant. The eigenvalues of the quantum harmonic oscillator are given by $E_n = (2n+1)\frac{\hbar}{2}\omega$. We refer the readers to [128] for more discussions on quantum harmonic oscillators.

If we consider the rescaled operator,

$$\frac{1}{\hbar}H = \hbar \left(-\frac{1}{2}\nabla^2 \right) + \frac{1}{\hbar} \left(\frac{1}{2}\omega^2 x^2 \right),$$

we find the eigenvalues of $H/\hbar$ are $E'_n = (n + 1/2)\omega$, which only depend on the curvature of the potential field regardless of the value of $\hbar$. The spectral gap of the rescaled quantum harmonic oscillator $H/\hbar$ seems to be an invariant quantity determined by the geometry of the quadratic potential field.

Inspired by the case of quantum harmonic oscillators, we define a one-parameter family of Schrödinger operators for a given objective function f,

$$\hat{H}(\lambda) = \frac{1}{\lambda} \left(-\frac{1}{2}\nabla^2 \right) + \lambda f, \tag{4.7}$$

where $\lambda > 0$ is an interpolating parameter and we assume f is non-negative and diverges at infinity. For any positive λ, the spectrum of $\hat{H}(\lambda)$ is discrete and we arrange the eigenvalues of $\hat{H}(\lambda)$ in ascending order,

$$E_0(\lambda) < E_1(\lambda) \leq E_2(\lambda) \leq \dots.$$

The spectral gap of $\hat{H}(\lambda)$ is denoted by

$$\Delta(\lambda) \coloneqq E_1(\lambda) - E_0(\lambda). \tag{4.8}$$

If f is a convex quadratic function, $\hat{H}(\lambda)$ is precisely a quantum harmonic oscillator. According to our discussion above, all the eigenvalues of the operator $\hat{H}(\lambda)$ only depend on f, regardless of the value of λ. Nevertheless, when f is nonconvex and non-quadratic, the eigenvalues of $\hat{H}(\lambda)$ change along with λ. The spectral gap $\Delta(\lambda)$ depends on the nonconvexity of f [127, 129] and it also changes with λ. In particular, when f is a double-well function (described by a quartic polynomial), the spectral gap can be made arbitrarily small by tuning the coefficients in the quartic polynomial [130].

Although the spectral gap $\Delta(\lambda)$ could change drastically with λ when f is nonconvex, we observe that $\min_{\lambda>0} \Delta(\lambda)$ is usually not arbitrarily small. In Section 4.3.1, we study the spectral gap of the Schrödinger operator $\hat{H}(\lambda)$ with a (nonconvex) double well potential function using numerical methods. It turns out that the spectral gap $\Delta(\lambda)$ of nonconvex f is robust under interpolation and its minimal value (i.e., $\min_{\lambda>0} \Delta(\lambda)$) appears to be determined solely by the geometry of f. Based on this observation, we prove that QHD can solve 1D nonconvex problems with two local minima (see Lemma 4.3.3). This result is later generalized to high dimensions in Proposition 4.3.4 and eventually

leads to our main theoretical result (see Theorem 4.4.1).

4.2.2 Ground States of the Schrödinger Operator

In the Schrödinger operator (4.5), there are two major components: the kinetic operator $-\nabla^2$ and the potential operator f. When the kinetic operator dominates, the low-energy subspace of the Schrödinger operator tends to behave like that of a free particle. In another case, if the potential f plays a leading role, the ground state of the Schrödinger operator exhibits properties as like a Gaussian distribution. For example, if we consider the following 1D Schrödinger operator,

$$\hat{H}(\lambda) = \frac{1}{\lambda}\left(-\frac{1}{2}\nabla^2\right) + \lambda f, \tag{4.9}$$

where $f\colon \mathbb{R} \to \mathbb{R}$ is a smooth one-dimensional potential function with a single, non-degenerate zero at $x^* = 0$: $f(0) = 0$, $\nabla f(0) = 0$, and $f''(0) > 0$. The function f admits a Taylor expansion near the global minimum:

$$f(x) = \frac{1}{2}f''(x^*)(x - x^*)^2 + O(x^3).$$

As λ becomes large, the potential landscape observed by the quantum particle is essentially the quadratic part of the potential f near x^*. Let $\gamma = \sqrt{f''(x^*)}$, we can define a comparison Schrödinger operator

$$\hat{K}(\lambda) = \frac{1}{\lambda}\left(-\frac{1}{2}\nabla^2\right) + \lambda\left(\frac{1}{2}\gamma^2(x - x^*)^2\right). \tag{4.10}$$

This new operator $\hat{K}(\lambda)$ is a quantum harmonic oscillator and it gives a nice approximation of the low-energy subspace of $\hat{H}(\lambda)$ for large λ.

Theorem 4.2.1 (Theorem 4.1, [131]). *Assume $f(x)$ is polynomially bounded (i.e., $|f(x)| \leq C(1 + |x|^m)$ for some constant C and a fixed integer m). Let n be a fixed integer such that $E_n(\lambda)$ is a simple*[3] *eigenvalue of $\hat{H}(\lambda)$. Let $\Phi_n(\lambda)$ be the corresponding eigenfunction. Denote e_n and $\phi_n(\lambda)$ to be the n-th eigenvalue and eigenfunction of $\hat{K}(\lambda)$. Then, for sufficiently large λ, we have that*

$$E_n(\lambda) - e_n = \mathcal{O}(\lambda^{-1}), \ \|\Phi_n(\lambda) - \phi_n(\lambda)\| = \mathcal{O}(\lambda^{-1/2}). \tag{4.11}$$

The ground state of the quantum harmonic oscillator $\hat{K}(\lambda)$ is a Gaussian state

$$\phi_0(\lambda) = \left(\frac{\lambda\gamma}{\pi}\right)^{1/4} e^{-\frac{\gamma\lambda(x-x^*)^2}{2}}, \tag{4.12}$$

and the eigenvalues of $\hat{K}$ are given by $e_n = \left(n + \frac{1}{2}\right)\gamma$.

As λ increases, the Gaussian state $\phi_0(\lambda)$ is localized in the vicinity of the global minimizer x^* and its tail (i.e., the probability of landing far away from x^*) is exponentially small. One would naturally guess that $\Phi_0(\lambda)$, the actual ground state of the Schrödinger operator (4.9), also has a very small tail for large λ. However, this is not immediately clear from Theorem 4.2.1. In Section 4.4, we will prove a stronger result to characterize the smallness of the tail of $\Phi_0(\lambda)$ (see Lemma 4.4.2). It turns out that, given a fast-growing potential field f, the ground state $\Phi_0(\lambda)$ corresponds to a *sub-Gaussian* distribution whose variance scales proportional to $1/\lambda$.

[3] A *simple* eigenvalue means the multiplicity is 1. In other words, the corresponding eigenspace is non-degenerate.

4.2.3 Quantum Adiabatic Theorem for Unbounded Hamiltonian

Given a quantum Hamiltonian $H(t)$, $t \in [0, t_f]$, we consider the dynamics described by the Schrödinger equation,

$$i\epsilon \frac{\mathrm{d}}{\mathrm{d}t} |\psi^\epsilon(t)\rangle = H(t) |\psi^\epsilon(t)\rangle , \tag{4.13}$$

subject to an initial state $|\psi^\epsilon(0)\rangle = |\psi_0\rangle$. We suppose that $U^\epsilon(t)$ is the propagator of the dynamics described by (4.13), so the solution at time t is given by

$$|\psi^\epsilon(t)\rangle = U^\epsilon(t) |\psi_0\rangle . \tag{4.14}$$

The parameter $\epsilon > 0$ controls the time scale on which the quantum Hamiltonian $H(t)$ varies. For a small ϵ, the system evolves slowly and the dynamics are relatively simple: if the system begins with an eigenstate of $H(0)$, it remains close to an eigenstate of $H(t)$. This process is called *adiabatic quantum evolution*.

Formally, We define a new Hamiltonian operator,

$$H_a(t) = H(t) + i\epsilon[\dot{P}, P], \tag{4.15}$$

where $P(t)$ is a rank-1 projector onto the ground state of $H(t)$, $\dot{P}$ is the time derivative of P, and $[\cdot, \cdot]$ is the commutator. Let $U_a^\epsilon(t)$ be the propagator of the quantum dynamics generated by (4.15), i.e.,

$U_a^\epsilon(t)$ is given as the solution of

$$i\epsilon\frac{\mathrm{d}}{\mathrm{d}t}U_a^\epsilon(t) = H_a(t)U_a^\epsilon(t), \tag{4.16}$$

subject to $U_a^\epsilon(0) = \mathbf{1}$. The propagator $U_a^\epsilon(t)$ is called the *adiabatic intertwiner* because it preserves the ground-energy subspace of $H(t)$ [132]. Precisely, we have that

$$U_a^\epsilon(t)P(0) = P(t)U_a^\epsilon(t). \tag{4.17}$$

The quantum adiabatic theorem states that adiabatic intertwiner (i.e., the exact adiabatic propagator) $U_a^\epsilon(t)$ is a good approximation of the actual propagator $U^\epsilon(t)$ for sufficiently small ϵ. Various formulations of the quantum adiabatic theorem exist in the literature, while most of them assume the Hamiltonian is a bounded linear operator. In Quantum Hamiltonian Descent, however, the Hamiltonian is an unbounded operator defined in the real space $\mathbb{R}^d$. Here, we introduce a quantum adiabatic theorem for unbounded Hamiltonian [57, 133], which allows us to have a neat analysis without discretizing the unbounded operator in QHD.

For a self-adjoint operator $H(t)$ with discrete spectrum at any t, the spectral gap of $H(t)$, denoted by $\Delta(t)$, is the difference between the first two eigenvalues of $H(t)$ at time t.

Theorem 4.2.2 ([133, Theorem 2.1]). *Assume that for all $t \in [0, t_f]$, there exist positive numbers c_0 and c_1 such that*

$$\dot{H}^2 \leq c_0 + c_1 H^2. \tag{4.18}$$

Moreover, we assume $H > 0$ *and the spectral gap* $\Delta(t)$ *has a uniform lower bound, i.e.,* $\Delta(t) \geq \Delta_{\min} > 0$. *We denote* ϕ_0 *as the ground state of* $H(0)$. *Then, we have that*

$$\| \left(U^{\epsilon}(t_f) - U_a^{\epsilon}(t_f)\right) \phi_0 \| \leq \epsilon \theta t_f, \tag{4.19}$$

where θ *is a constant that only depends on* c_0, c_1 *and* $\Delta_{\min}$.

Remark 4.2.3. In Theorem 4.2.2, we rescale the time variable compared to the original version in [133, Theorem 2.1]. After the time rescaling, the adiabatic error from the original theorem reads $b = \frac{\theta t_f}{1/\epsilon} = \epsilon \theta t_f$, which is precisely what we have in (4.19).

4.2.4 Quantum Simulation of Schrödinger Equations

In Quantum Hamiltonian Descent [77], a major subroutine is to simulate the Schrödinger equation, a task known as *quantum simulation*. Quantum simulation is a prominent application of quantum computation, and several efficient quantum algorithms for simulating real-space quantum dynamics have been proposed, including [89, 134–137]. In this chapter, we employ a quantum simulation algorithm due to Childs, Leng, Li, Liu, and Zhang [89]. The complexity of this algorithm has near-optimal dependence in the dimension d and accuracy η by leveraging the pseudo-spectral representation and interaction-picture quantum simulation. We consider the Schrödinger equation over the time interval $[t_0, t_1]$ for a given time-dependent potential $V(x, t)$,

$$i \frac{\partial}{\partial t} |\Psi(x,t)\rangle = \left[-\frac{1}{2} \nabla^2 + V(x,t) \right] |\Psi(x,t)\rangle \,, \tag{4.20}$$

where we specify $\Omega = [-M, M]^d$ for a sufficiently large M and $V(x,t)\colon \Omega \times [t_0, t_1] \to \mathbb{R}$ is a time-dependent potential function, $\Psi(x,t)\colon \Omega \times [t_0, t_1] \to \mathbb{C}$ is the wave function subject to certain initial condition and the periodic boundary condition.

In [89, Theorem 8], the complexity of simulating (4.20) involves an additional parameter g' that depends on the regularity of the wave function $\Psi(x,t)$. We slightly improve this result by assuming the initial condition is *analytic*, which is the case in QHD.

Theorem 4.2.4. *Suppose the potential field $V(x,t)$ is bounded, smooth in x and t, and periodic in x. Moreover, we assume the initial data $\Psi_0(x)$ is analytic on Ω and V is L-Lipschitz in t. We define $\|V\|_{\infty,1} \coloneqq \int_{t_0}^{t_1} \|V(\cdot,t)\|_\infty \,\mathrm{d}t$. Then, the Schrödinger equation (4.20) can be simulated for time $t \in [t_0, t_1]$ up to accuracy η with the following cost:*

1. *Queries to the quantum evaluation oracle O_V: $\mathcal{O}\left(\|V\|_{\infty,1} \frac{\log(\|V\|_{\infty,1}/\eta)}{\log\log(\|V\|_{\infty,1}/\eta)}\right)$,*
2. *1- and 2-qubit gates:*

$$\mathcal{O}\left(\|V\|_{\infty,1}\left(\mathrm{poly}(z) + \log^{2.5}\left(L\|V\|_{\infty,1}/\eta\right) + d\log\log(1/\eta)\right)\frac{\log(\|V\|_{\infty,1}/\eta)}{\log\log(\|V\|_{\infty,1}/\eta)}\right).$$

The proof of the theorem is available in Section B.1. Note that the quantum simulation algorithm in [89] also requires two quantum oracles other than O_V, namely, the *inverse change-of-variable* oracle O_{inv} and the *max-norm* oracle O_{norm} (see Lemma 5 [89]). In QHD, the potential function $V(x,t) \coloneqq \varphi(t)V(x)$, where $\varphi(t)$ is a time-dependent function described by a closed-form formula (see (4.53)) and $V(x)$ is given by O_V. In this case, the two oracles O_{inv} and O_{norm} are efficiently implemented without querying the function V.

4.3 Construction of Nonconvex Optimization Instances

4.3.1 The Spectral Gap of Asymmetric Double Well

In this section, we use a numerical example to illustrate the theory of Schrödinger operators that we discussed in Section 4.2. We consider the following 1-parameter family of Hamiltonian operators

$$\hat{H}(\lambda) = \frac{1}{\lambda}\left(-\frac{1}{2}\nabla^2\right) + \lambda w(x), \tag{4.21}$$

where $w(x)\colon \mathbb{R} \to \mathbb{R}$ is a one-dimensional double-well function given by a degree-4 polynomial:

$$w(x) = x^4 - \left(x - \frac{1}{32}\right)^2 - c, \tag{4.22}$$

where $c \approx -0.296$ such that the global minimum of $w(x)$ is zero.

In Figure 4.2A, we plot the graph of the function w between $x = -1$ and $x = 1$. This function has two local minima:

$$x_1 \approx -0.722,\ w(x_1) = 0,$$

$$x_2 \approx 0.691,\ w(x_2) = 0.088.$$

For any $\lambda > 0$, the spectral theory implies that the spectrum of $\hat{H}(\lambda)$ is discrete. We denote the spectrum of the Hamiltonian $\hat{H}(\lambda)$ as $\{E_0(\lambda) < E_1(\lambda) \leq \dots\}$. For large λ, the first few eigenvalues of $\hat{H}(\lambda)$ can be computed using the semi-classical approximation. Note that w has a positive Hessian at the global minimum: $w''(x_1) \approx 4.5092$. Let $\gamma^2 = w''(x_1)$. For $\lambda \gg 1$, by Theorem 4.2.1, we have

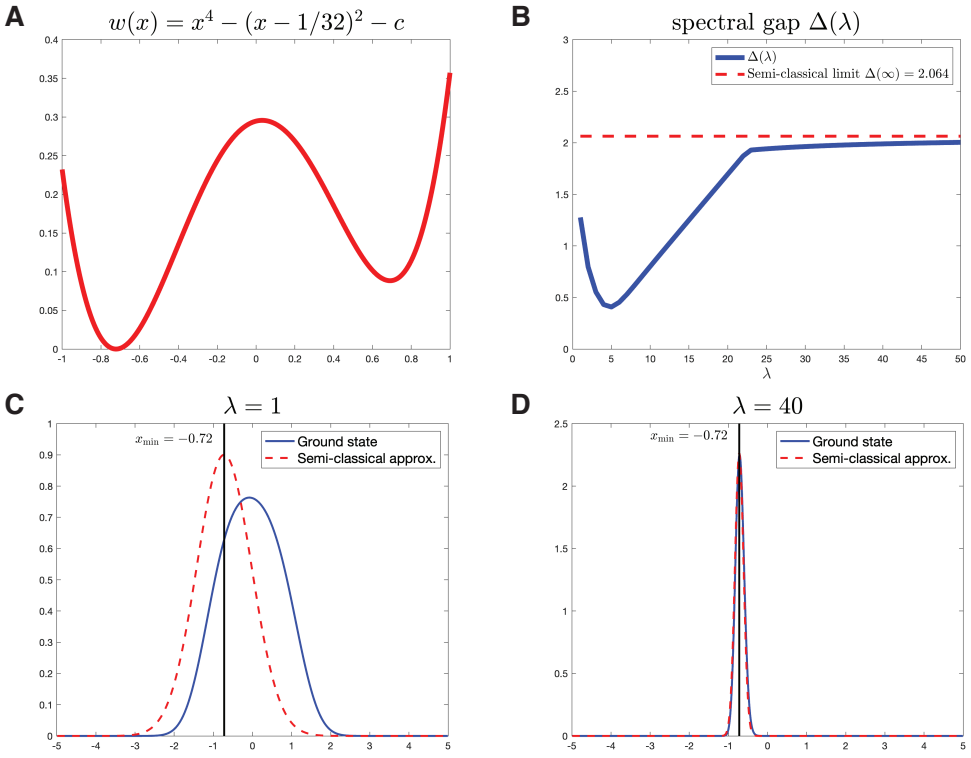


Figure 4.2: **The spectral gap and ground state of asymmetric double well. A.** Plot of the double well function $w(x)$. **B.** The spectral gap $\Delta(\lambda)$ of the Hamiltonian operator $\hat{H}(\lambda)$ as a function of λ. **C, D.** The blue curve (or the red dashed curve, resp.) represents the ground state (or the harmonic approximation of the ground state, resp.) of the Hamiltonian operator $\hat{H}(\lambda)$ for $\lambda = 1, 40$. The x axis represents the one-dimensional space and the y axis shows the amplitude of the quantum states. When λ becomes larger, the harmonic approximation becomes more accurate.

that

$$E_0(\lambda) = \frac{1}{2}\gamma + \mathcal{O}(\lambda^{-1}), \quad E_1(\lambda) = \frac{3}{2}\gamma + \mathcal{O}(\lambda^{-1}), \tag{4.23}$$

which implies that the spectral gap in the large λ limit is

$$\lim_{\lambda\to\infty} \Delta(\lambda) = \gamma \approx 2.064. \tag{4.24}$$

We employ numerical methods to compute the spectral gap of the operator $\hat{H}(\lambda)$. Our numerical results confirm our estimate of the spectral gap in the limit of large λ. In Figure 4.2B, we show the spectral gap $\Delta(\lambda) = E_1(\lambda) - E_0(\lambda)$ as a function of λ. For large λ, it is clear that the spectral gap eventually converges to the predicted value γ (indicated by the red dashed line). We also find that the spectral gap $\Delta(\lambda)$ achieves its minimum $\Delta_{\min} \approx 0.408$ at around $\lambda = 5$. For $\lambda > 5$, the spectral gap $\Delta(\lambda)$ starts to increase and gradually approaches the semi-classical limit. At first glance, it is a bit surprising to see that the spectral gap $\Delta(\lambda)$ has a lower bound for all $\lambda > 0$ even when the potential function w is nonconvex. This means that the minimal spectral gap $\Delta_{\min}$ is actually an intrinsic property that only depends on the geometry of the double-well potential w but not the interpolating parameter λ. Therefore, it is reasonable to formalize our observation by defining a class of *gapped* potential function w.

Definition 4.3.1 (Gapped 1D potential)**.** Given a 1-dimensional twice differentiable function $w\colon \mathbb{R} \to \mathbb{R}$ such that w is bounded from below and diverges to infinity as $|x| \to \infty$. For any $\lambda > 0$, let $\Delta(\lambda)$ denote the spectral gap of the Hamiltonian operator (4.21). We say this function w is **gapped** if the

spectral gap is uniformly bounded by a positive constant $\Delta_{\min} > 0$, i.e.,

$$\forall \lambda > 0 \colon \Delta(\lambda) \geq \Delta_{\min} > 0. \tag{4.25}$$

We believe this gap condition holds for many asymmetric double-well potential functions. We will use these gapped asymmetric double wells to construct optimization instances with exponentially many local minima, and we will show that there exist efficient quantum algorithms (such as QHD) that can find the global solution in polynomial time.

Another numerical observation we made is that the semi-classical approximation works for not only the spectral gap but also the ground state. In panels C and D in Figure 4.2, we plot the ground state of the Hamiltonian $\hat{H}(\lambda)$ (in blue solid line) and the semi-classical approximation given by (4.12) (in red dashed line) for $\lambda = 1, 40$. Clearly, when λ is large, the Gaussian state given by the semi-classical theory is a very accurate approximation of the ground state.

4.3.2 1D Model Problem

We now formulate our 1D model problem.

Definition 4.3.2. We call a non-negative function $w \colon \mathbb{R} \to \mathbb{R}$ a **well-formed** asymmetric double well if w satisfies the following conditions:

1. w has at least 2 local minima and a unique non-degenerate global minimum at $x = x^*$, i.e., $w'(x^*) = 0$ and $w''(x^*) \geq \gamma^2 > 0$;

2. There exist positive numbers $C > c > 0$ such that $c|x - x^*|^4 \leq w(x) \leq C(1 + |x - x^*|^4)$ for all $x \in \mathbb{R}$;

3. $w \in C^\infty$ and w is ρ-smooth, i.e., $\|w''\| \leq \rho$;[4]

4. w is a gapped function in the sense of Definition 4.3.1.

In Section 4.3.1, we use numerical methods to study the function $w(x) = x^4 - \left(x - \frac{1}{32}\right)^2 - c$, where $c \approx -0.296$ such that $\min_x w(x) = 0$. It is clear that this $w(x)$ satisfies the first two conditions in Definition 4.3.2. If we restrict $w(x)$ to a compact subset of $\mathbb{R}$ (which is the case in our quantum simulation), condition 3 is satisfied. Meanwhile, our numerical results in Figure 4.2 show that the spectral gap $\Delta(\lambda)$ achieves its minimum near $\lambda = 5$ and then $\Delta(\lambda)$ converges to the semi-classical limit, aligned with the theoretical prediction given by Theorem 4.2.1. Therefore, with the numerical evidence, we believe the function $w(x)$ is a *well-formed* asymmetric double well. This function will later be used in our empirical study (see Section 4.5).

When the function w is a well-formed asymmetric double well, we can apply the adiabatic theorem for unbounded Hamiltonian (i.e., Theorem 4.2.2) to prepare the ground state of the Schrödinger operator

$$\hat{H}(\lambda) = \frac{1}{\lambda}\left(-\frac{1}{2}\nabla^2\right) + \lambda w(x)$$

for very large λ. Meanwhile, condition 2 in Definition 4.3.2 ensures that the ground state of $\hat{H}(\lambda)$ is sub-Gaussian (for details, see Lemma 4.4.2). These nice properties will allow us to find an approximate solution to the optimization problem $\min_x w$.

Provided with a pre-fixed parameter λ_f (which we will discuss soon), we define a time-dependent

[4]Precisely speaking, a degree-4 polynomial is not ρ-smooth, as its Hessian is a quadratic function that diverges at infinity. Thanks to the localization of wave function in the quantum evolution (because w grows as $|x|^4$), we may restrict the evolution to a compact domain in quantum simulation. In this case, the double-well potential function is effectively ρ-smooth.

function

$$\lambda(t) = e^{2t - \log(\lambda_f)}, \tag{4.26}$$

Clearly, for $0 \leq t \leq \log(\lambda_f)$, we have that $1/\lambda_f \leq \lambda(t) \leq \lambda_f$.

Lemma 4.3.3. *Let $w \colon \mathbb{R} \to \mathbb{R}$ be a well-formed asymmetric double well. For $0 \leq t \leq \log(\lambda_f)$, let $H(t) \coloneqq \hat{H}(\lambda(t))$. Denote $|\phi(t)\rangle$ as the ground state of $H(t)$ for $t \in [0, \log(\lambda_f)]$. Let $|\psi^\epsilon(t)\rangle$ be the solution to the Schrödinger equation (4.13) governed by $H(t)$ with the initial condition $|\psi^\epsilon(0)\rangle = |\phi(0)\rangle$. Then, for any $0 \leq t \leq \log(\lambda_f)$, we have that*

$$\|\psi^\epsilon(t) - \tilde{\phi}(t)\| \leq \epsilon\theta t, \tag{4.27}$$

where θ is an absolute constant that only depends on f. Here, $\left|\tilde{\phi}(t)\right\rangle$ only differs from $|\phi(t)\rangle$ by a global phase, i.e.,

$$\tilde{\phi}(t) = e^{i\alpha(t)}\phi(t), \tag{4.28}$$

where $\alpha(t)$ is a real-valued function.

Proof. First, we show that $H(t)$ satisfies all conditions specified in Theorem 4.2.2. Since w is non-negative, $\hat{H}(\lambda) \geq 0$ for all positive λ. The gappedness of w guarantees a uniform lower bound on the spectral gap of $H(t)$. We claim that

$$\dot{H}^2 \leq 4\rho + 4H^2. \tag{4.29}$$

By the definition of $H(t)$, we have that

$$\dot{H} = \frac{\dot{\lambda}}{\lambda}\left(\frac{1}{2\lambda}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \lambda w\right) = 2\left(\frac{1}{2\lambda}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \lambda w\right). \tag{4.30}$$

For any test function $\psi \in C_0^\infty(\mathbb{R})$, we consider the inner product:

$$\begin{aligned}\langle \psi, [4H^2 - \dot{H}^2]\psi\rangle &= 4\langle \psi, H^2\psi\rangle - \langle \psi, (H')^2\psi\rangle \\ &= -4\left(\langle \psi, (w\psi)''\rangle + \langle \psi, w\psi''\rangle\right) \\ &= -4\left(\langle \psi, w''\psi\rangle + 2\langle \psi, w'\psi'\rangle + 2\langle \psi, w\psi''\rangle\right).\end{aligned}$$

Using integral by parts, we have that

$$\langle \psi, w\psi''\rangle = \langle w\psi, \psi''\rangle = -\langle w'\psi, \psi'\rangle - \langle w\psi', \psi'\rangle = -\langle \psi, w'\psi'\rangle - \langle \psi', w\psi'\rangle, \tag{4.31}$$

which implies that

$$\langle \psi, [4H^2 - \dot{H}^2]\psi\rangle = -4\langle \psi, w''\psi\rangle + 4\langle \psi', w\psi'\rangle. \tag{4.32}$$

By Definition 4.3.2, w is non-negative and ρ-smooth, so we have $\langle \psi', w\psi'\rangle \geq 0$ and $w'' \leq \rho$. Therefore, we have that

$$\langle \psi, [4H^2 - \dot{H}^2]\psi\rangle \geq -4\rho, \tag{4.33}$$

which implies (4.29). With all conditions satisfied, we invoke Theorem 4.2.2 and obtain that

$$\| \left(U^\epsilon(t) - U_a^\epsilon(t)\right) \phi(0) \| \leq \epsilon\theta t, \tag{4.34}$$

where θ is a constant that only depends on w. Note that

$$U^\epsilon(t) \left|\phi(0)\right\rangle = \left|\psi^\epsilon(t)\right\rangle, \tag{4.35}$$

and $U_a^\epsilon(t) \left|\phi(0)\right\rangle$ is always in the ground-energy subspace of $H(t)$, we denote

$$\tilde{\phi}(t) = U_a^\epsilon(t) \left|\phi(0)\right\rangle \tag{4.36}$$

which differs from $\left|\phi(t)\right\rangle$ by a global phase because the ground-energy subspace of $H(t)$ is 1-dimensional. Substituting (4.35) and (4.36) to (4.34), we complete the proof. □

4.3.3 Toward High-Dimensional Nonconvexity

Let $d \geq 1$ be an integer and $w(x)$ be a well-formed asymmetric double well. We define the d-dimensional objective function F:

$$F(x_1, \ldots, x_d) = \sum_{k=1}^{d} w(x_k).$$

If the 1D model function $w(x)$ has two local minima (e.g., a double-well potential), the function $F(\boldsymbol{x})$ has 2^d local minima. Let U be an arbitrary d-by-d orthogonal matrix, we define a new function

$F_U(\boldsymbol{x}) \coloneqq F(U\boldsymbol{x})$. Now, we consider the time-dependent Hamiltonian:

$$H(t) = \frac{1}{\lambda(t)}\left(-\frac{1}{2}\nabla^2\right) + \lambda(t)F_U(\boldsymbol{x}), \tag{4.37}$$

where the function $\lambda(t)$ is defined in (4.26).

Proposition 4.3.4. *For $0 \le t \le \log(\lambda_f)$, let $H(t)$ be defined as in (4.37). Denote $|\Phi(t)\rangle$ as the ground state of $H(t)$ for $t \in [0, \log(\lambda_f)]$. Let $|\Psi^\epsilon(t)\rangle$ be the solution to the Schrödinger equation (4.13) governed by $H(s)$ with the initial condition $|\Psi^\epsilon(0)\rangle = |\Phi(0)\rangle$. Then, for any $0 \le t \le \log(\lambda_f)$, we have that*

$$\|\Psi^\epsilon(t) - \tilde{\Phi}(t)\| \le \epsilon\theta dt, \tag{4.38}$$

where θ is an absolute constant that only depends on $w(x)$, and $\left|\tilde{\Phi}(t)\right\rangle$ only differs from $|\Phi(t)\rangle$ by a global phase.

Proof. We consider the following Hamiltonian operator with the separable objective function $F(\boldsymbol{x})$,

$$H_0(t) = \frac{1}{\lambda(t)}\left(-\frac{1}{2}\nabla^2\right) + \lambda(t)F(\boldsymbol{x}). \tag{4.39}$$

For $t \in [0, \log(\lambda_f)]$, let $|\Psi_0^\epsilon(t)\rangle$ be the solution to the Schrödinger equation,

$$i\epsilon\left|\Psi_0^\epsilon(t)\right\rangle = H_0(t)\left|\Psi_0^\epsilon(t)\right\rangle. \tag{4.40}$$

If the initial state $\Psi_0^\epsilon(\boldsymbol{x}, 0) = \Psi^\epsilon(U^\top\boldsymbol{x}, 0)$, since the Laplace operator is invariant under rotation, it is

clear that

$$\Psi_0^\epsilon(\boldsymbol{x}, t) = \Psi^\epsilon(U^\top \boldsymbol{x}, t) \tag{4.41}$$

for all $t \in [0, \log(\lambda_f)]$. Similarly, the ground state of $H_0(t)$ is given by

$$\Phi_0(U^\top \boldsymbol{x}, t) = \Phi(U^\top \boldsymbol{x}, t), \tag{4.42}$$

where $\Phi(\boldsymbol{x}, t)$ is the ground state of $H(t)$. Therefore, to show (4.38), it suffices to prove that there is a state $\left|\tilde{\Phi}_0(t)\right\rangle$ that differs from $|\Phi_0(t)\rangle$ by a global phase such that

$$\|\Psi_0^\epsilon(t) - \tilde{\Phi}_0(t)\| \leq \epsilon\theta dt. \tag{4.43}$$

Next, we prove (4.43) using the triangle inequality. Note that the Hamiltonian operator $H_0(t)$ is separable in the sense that

$$H_0(t) = \sum_{k=1}^{d} H_0^{(k)}(t), \tag{4.44}$$

where for $k = 1, \ldots, d$, the operator $H_0^{(k)}(t)$ is given by

$$H_0^{(k)}(t) \coloneqq \frac{1}{\lambda(t)} \left(-\frac{1}{2} \frac{\partial^2}{\partial x_k^2} \right) + \lambda(t) w(x_k). \tag{4.45}$$

If we denote the ground state of $H_0^{(k)}(t)$ as $|\phi_k(t)\rangle$, the separability implies that the ground state of $H_0(t)$ is given by the product of the 1D ground states $\Phi_0(t) = \prod_{k=1}^{d} \phi_k(t)$. Similarly, the solution

$|\Psi_0^\epsilon(t)\rangle$ is a product state given by $\Psi_0^\epsilon(t) = \prod_{k=1}^d \psi_k^\epsilon(t)$, where each $\psi_k^\epsilon(t)$ solves the 1D Schrödinger equation $i\epsilon \frac{\mathrm{d}}{\mathrm{d}t} |\psi_k^\epsilon(t)\rangle = H_0^{(k)}(t) |\psi_k^\epsilon(t)\rangle$. By Lemma 4.3.3, we have that

$$\|\psi_k^\epsilon(t) - e^{i\alpha_k(t)}\phi_k(t)\| \leq \epsilon\theta t, \tag{4.46}$$

where each $\alpha_k(t)$ is a real-valued global phase. We define

$$\tilde{\Phi}_0(t) := \prod_{k=1}^d e^{i\alpha_k(t)}\phi_k(t) = e^{i\sum_{k=1}^d \alpha_k(t)} \prod_{k=1}^d \phi_k(t) = e^{i\sum_{k=1}^d \alpha_k(t)}\Phi_0(t). \tag{4.47}$$

Then, by applying the triangle inequality d times, we end up with

$$\|\Psi_0^\epsilon(t) - \tilde{\Phi}_0(t)\| \leq \sum_{k=1}^d \|\psi_k^\epsilon(t) - e^{i\alpha_k(t)}\phi_k(t)\| \leq \epsilon\theta dt. \tag{4.48}$$

□

4.4 Quantum Algorithms and Complexity Analysis

4.4.1 Main Results

Suppose that we fix a one-dimensional model problem $w(x)$ that satisfies Definition 4.3.2 (e.g., an asymmetric double-well function) and an integer $d \geq 1$. With w and d, we construct a class of optimization instances $\mathscr{C}(w, d)$ as in Definition 4.1.1. Given an instance $f \in \mathscr{C}(w, d)$, we consider a Quantum Hamiltonian Descent algorithm with parameters δ, λ_f, η and η.

Recall that a solution $\boldsymbol{x}$ is called a δ-approximate solution if $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta$, where $\boldsymbol{x}^*$ is the (unique) global minimizer of the optimization instance f.

Algorithm 4.1: Quantum Hamiltonian Descent

1 Specify positive parameters δ, λ_f, ϵ, and η. Let $t_f = \log(\lambda_f)$ and $\lambda(t) = e^{2t-t_f}$. For $0 \le t \le t_f$, we define the QHD Hamiltonian over Ω,

$$H(t) = \frac{1}{\lambda(t)}\left(-\frac{1}{2}\nabla^2\right) + \lambda(t) f.$$

2 Prepare the ground state of $H(0)$, denoted by $|\Phi(0)\rangle$.

3 Use $|\Phi(0)\rangle$ as the initial state and simulate the quantum dynamics governed by the Schrödinger equation,

$$i\epsilon \frac{\mathrm{d}}{\mathrm{d}t} |\Psi^\epsilon(t)\rangle = H(t) |\Psi^\epsilon(t)\rangle , \tag{4.49}$$

up to an additive error η, i.e., a final state $|\Psi^\epsilon_{\mathrm{sim}}(t_f)\rangle$ is obtained such that $\|\Psi^\epsilon_{\mathrm{sim}}(t_f) - \Psi^\epsilon(t_f)\| \le \eta$.

4 Measure the final state $|\Psi^\epsilon_{\mathrm{sim}}(t_f)\rangle$ using the position quadrature $\hat{\boldsymbol{x}}$ and return a d-dimensional vector $\boldsymbol{x}$.

Theorem 4.4.1. *Given a well-formed asymmetric double well w (see Definition 4.3.2) and a fixed integer d, let $f \in \mathscr{C}(w, d)$ be an optimization instance. For a sufficiently small $\delta > 0$, we specify the choices of parameters:*

$$\lambda_f \coloneqq \frac{\beta(2d + 6\log(3))}{\delta^2},\ \epsilon \coloneqq \frac{1}{18 d\theta \log(\lambda_f)},\ \eta \coloneqq \frac{1}{18}, \tag{4.50}$$

where β and θ are constants that only depend on w. Then, Algorithm 4.1 produces a δ-approximate global solution to f using

$$\mathcal{O}\left(\frac{d^3}{\delta^2} \cdot \frac{\log^2(\frac{d^2}{\delta^2})}{\log\log(\frac{d^2}{\delta^2})}\right)$$

queries to O_f and additional

$$\mathcal{O}\left(\frac{d^3}{\delta^2}\left(d + \log^{2.5}\left(\frac{d}{\delta^2}\right)\right)\frac{\log^2(\frac{d^2}{\delta^2})}{\log\log(\frac{d^2}{\delta^2})}\right)$$

1- and 2-qubit gates.

4.4.2 Initial State Preparation

An important step in Algorithm 4.1 is to prepare the ground state of the initial Hamiltonian $H(0)$. Since the function f grows as $\Theta(|x|^4)$, we can choose a large enough constant M such that the quantum dynamics in $\mathbb{R}^d$ can be faithfully simulated over a compact domain $\Omega \coloneqq [-M, M]^d$ (with periodic boundary conditions). We denote $V(\boldsymbol{x})$ as the truncated objective function $f(\boldsymbol{x})$ on Ω. Then, it is sufficient to prepare the ground state of the operator

$$H(0) = \lambda_f \left(-\frac{1}{2}\nabla^2 \right) + \frac{1}{\lambda_f} V.$$

Since we choose $\lambda_f \gg 1$, the potential operator V/λ_f can be regarded as a perturbative term in the initial Hamiltonian $H(0)$. Meanwhile, the ground state of the kinetic operator $\left(-\frac{1}{2}\nabla^2\right)$ over the periodic domain Ω is a constant function that is easily prepared on a quantum computer as a uniform superposition state. Therefore, we may prepare the initial state of $H(0)$ from the ground state of the kinetic operator via a quantum adiabatic evolution.

More concretely, we can simulate a time-dependent Schrödinger equation with the following Hamiltonian,

$$\tilde{H}(t) = -\frac{1}{2}\nabla^2 + \frac{a(t)}{\lambda_f^2} V,$$

where $a(t)\colon [0, T] \to [0, 1]$ is a monotonically increasing function such that $a(0) = 0$, $a(T) = 1$. Note that the ground state of $\tilde{H}(0)$ is a constant function and that of $\tilde{H}(T)$ is the desired initial state in QHD. To ensure the quantum state stays in the ground-energy subspace, the evolution time T usually

depends on a polynomial of $1/\Delta$, where Δ is the minimal spectral gap of $\tilde{H}(t)$ for $0 \leq t \leq T$. For any $t \in [0, T]$, the potential term $\frac{a(t)}{\lambda_f^2} V$ is perturbative so we can derive a constant lower bound for the spectral gap using the standard perturbation theory [56]. This means the adiabatic evolution time T is independent of the problem dimension d.

The query complexity of this quantum evolution is typically proportional to the simulation time T and the L^∞-norm of the perturbation V/λ_f^2. As V is the sum of d identity 1D model problems $w(x) \sim |x - x^*|^4$, we have $\|V\|_\infty \sim dM^4$. Together with $\lambda_f \sim d$, we have that $\|V\|_\infty/\lambda_f^2 \sim M^4/d$, where M is a large constant. It turns out that the initial state of $H(0)$ is prepared with $\mathcal{O}(1/d)$ queries to O_f and an additional $\mathcal{O}(d)$ 1- and 2-qubit gates. Both the query complexity and the gate complexity are significantly lower than the upper bounds claimed in Theorem 4.4.1.

Also, we would like to remark that using the ground state of the kinetic operator $\left(-\frac{1}{2}\nabla^2\right)$ (i.e., a constant function) does not appear to harm the overall performance of QHD. In Section 4.4.4, we conduct an experiment of QHD using the uniform superposition state as the initial state. Numerical results suggest the same convergence rate as predicted by our theoretical argument.

4.4.3 Proof of Theorem 4.4.1

Our proof of Theorem 4.4.1 relies on the following lemma on the ground state of the Schrödinger operator $\hat{H}(\lambda)$. This lemma gives an estimate on the tail probability of the ground state. In particular, we show that the ground state of $\hat{H}(\lambda)$ describes a quantum particle whose position is a sub-Gaussian random variable. The complete proof of Lemma 4.4.2 can be found in Section B.2.2.

Lemma 4.4.2 (Sub-Gaussian ground state)**.** *Let w be a well-formed asymmetric double well. Let $\phi_\lambda(x)$*

be the ground state of the Hamiltonian

$$\hat{H}(\lambda) = \frac{1}{\lambda}\left(-\frac{1}{2}\nabla^2\right) + \lambda w(x).$$

Then, for any $s \in \mathbb{R}$, we have

$$\mathbb{E}_{X\sim|\phi_\lambda|^2}\left[e^{s(X-x^*)}\right] \leq e^{s^2\sigma^2/2}, \tag{4.51}$$

where $\sigma^2 = \beta/\lambda$, and β is a constant that only depends on w.

Also, to leverage the quantum simulation algorithm in Theorem 4.2.4, we need to rescale the time in the Schrödinger equation (4.49). In the following lemma, we give the rescaled quantum dynamics whose final state is precisely the same as in (4.49). The proof of Lemma 4.4.3 is given in Section B.3.

Lemma 4.4.3 (Time rescaling)**.** *The final state $|\Psi^\epsilon(t_f)\rangle$ of the quantum dynamics in Algorithm 4.1 can be obtained by simulating the following Schrödinger equation for $\frac{1}{\epsilon\lambda_f} \leq s \leq \frac{\lambda_f}{\epsilon}$,*

$$i\frac{\mathrm{d}}{\mathrm{d}s}|\Psi(s)\rangle = \frac{1}{2}\left[-\frac{1}{2}\nabla^2 + \varphi(s)f\right]|\Psi(s)\rangle, \tag{4.52}$$

where

$$\varphi(s) \coloneqq \frac{1}{\left(-\epsilon s + \lambda_f + 1/\lambda_f\right)^2} \in [1/\lambda_f^2, \lambda_f^2]. \tag{4.53}$$

Now, we are ready to prove Theorem 4.4.1. We define $B_\delta \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta\}$, i.e.,

the δ-ball entered at $\boldsymbol{x}^*$. Let $1_{B_\delta}(\boldsymbol{x})$ be the indicator function of the set B_δ, i.e.,

$$1_{B_\delta}(\boldsymbol{x}) = \begin{cases} 1 & (\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta) \\ 0 & (\text{otherwise}). \end{cases} \tag{4.54}$$

Then, for a quantum particle described by a quantum state $|\Psi\rangle$, its position is a random variable $X \in \mathbb{R}^d$ such that $X \sim |\Psi|^2$. The success probability can be evaluated through the expectation,

$$\Pr_{X \sim |\Psi|^2}[\|X - \boldsymbol{x}^*\| \leq \delta] = \langle \Psi | \, 1_{B_\delta}(\boldsymbol{x}) \, | \Psi \rangle = \int_{\mathbb{R}^d} 1_{B_\delta}(x) |\Psi(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x}. \tag{4.55}$$

Proof. Let $|\Phi(t_f)\rangle$ be the ground state of the final Hamiltonian $H(t_f) = \frac{1}{\lambda_f}\left(-\frac{1}{2}\nabla^2\right) + \lambda_f f$. The modulus square of the quantum state, $|\Phi(t_f)|^2$, gives a probability density on $\mathbb{R}^d$. We denote $X \in \mathbb{R}^d$ as a random vector following the distribution $|\Phi(t_f)|^2$. Due to our construction of the optimization instances, it is clear that $Y \coloneqq U^\top(X - \boldsymbol{x}^*)$ is a d-dimensional random vector with independent coordinates. Moreover, by Lemma 4.4.2,

$$\mathbb{E}\left[e^{r^\top Y}\right] \leq e^{\|r\|^2\sigma^2/2} \tag{4.56}$$

for all $r \in \mathbb{R}^d$, where $\sigma^2 = \beta/\lambda$ and β is a constant that only depends on w. Then, by Theorem B.2.3, we have that

$$\Pr\left[\|Y\|^2 > d\sigma^2 + 2\sqrt{cd\sigma^4} + 2\sigma^2 c\right] \leq e^{-c}. \tag{4.57}$$

We choose $c = 2\log(3)$ (so $e^{-c} = 1/9$) so that

$$d\sigma^2 + 2\sqrt{cd\sigma^4} + 2\sigma^2 c \leq \frac{\beta(2d + 6\log(3))}{\lambda_f}.$$

If specify the parameter λ_f as in Theorem 4.4.1, it follows that

$$\Pr[\|X_1 - \boldsymbol{x}^*\| > \delta] = \Pr[\|Y\| > \delta] \leq \frac{1}{9}. \tag{4.58}$$

Let $|\Psi^\epsilon(t_f)\rangle$ be the final state of the dynamics (4.49). If we choose ϵ as in Theorem 4.4.1, by Proposition 4.3.4, we have that

$$\|\Psi^\epsilon(t_f) - \tilde{\Phi}(t_f)\| \leq \frac{1}{18}, \tag{4.59}$$

where $\tilde{\Phi}(t_f)$ only differs from the ground state $|\Phi(t_f)\rangle$ by a global phase so $|\Phi(t_f)|^2$ and $|\tilde{\Phi}(t_f)|^2$ give exactly the same probability distribution.

Now, if we simulate the quantum dynamics in Algorithm 4.1 up to an additive error $\eta = 1/18$, by the triangle inequality and (4.59), we have that

$$\|\Psi^\epsilon_{\text{sim}}(t_f) - \tilde{\Phi}(t_f)\| \leq \frac{1}{9}. \tag{4.60}$$

Then, we deduce from Lemma B.3.1 that

$$\left|\langle\Psi^\epsilon_{\text{sim}}(t_f)|\, 1_{B_\delta}\, |\Psi^\epsilon_{\text{sim}}(t_f)\rangle - \left\langle\tilde{\Phi}(t_f)\right|\, 1_{B_\delta}\, \left|\tilde{\Phi}(t_f)\right\rangle\right| \leq \frac{2}{9}. \tag{4.61}$$

Since $\tilde{\Phi}(t_f)$ describes the same probability distribution as $\Phi(t_f)$, we combine (4.58) and (4.61) to yield the desired estimate,

$$\Pr_{X\sim|\Psi^{\epsilon}_{\mathrm{sim}}(t_f)|^2}\left[\|X-\boldsymbol{x}^*\|>\delta\right]\leq\frac{1}{3}. \tag{4.62}$$

It remains to figure out the overall complexity of the quantum simulation. Since the function f grows as $\Theta(|x|^4)$, we can choose a large enough constant M such that the quantum dynamics in $\mathbb{R}^d$ can be faithfully simulated over a compact domain $\Omega := [-M, M]^d$ (with periodic boundary conditions). In what follows, we let $V(\boldsymbol{x})$ be the truncated objective function $f(\boldsymbol{x})$ on Ω.

The initial state of the quantum simulation can be prepared using the procedures in Section 4.4.2. The query/gate complexity of the initial state preparation is at most $\mathcal{O}(d)$.

By Lemma 4.4.3, simulating the quantum simulation in Algorithm 4.1 is equivalent to simulating the effective quantum dynamics (4.52). We use the quantum algorithm in Theorem 4.2.4 to simulate (4.52). For $s \in J$, the time-dependent potential is $V(s, x) = \varphi(s)V(x)$, so we have

$$\|V\|_{\infty,1}=\left(\int_{1/\epsilon\lambda_f}^{\lambda_f/\epsilon}\varphi(s)\,\mathrm{d}x\right)\|V\|_\infty\leq\frac{\lambda_f}{\epsilon}\|V\|_\infty=27\theta\|V\|_\infty d\lambda_f\log(\lambda_f)\leq\mathcal{O}\left(\frac{d^3}{\delta^2}\right). \tag{4.63}$$

Note that $\|V\|_\infty \leq CdM^4$ for an absolute constant $C > 0$. Similarly, we can prove that the time-dependent potential function $V(x, s)$ is Lipschitz in s and the Lipschitz constant is

$$L=\max_{s\in J}\|\dot{V}\|_\infty\leq 2\frac{\lambda_f^3}{27d\theta\lambda(\lambda_f)}\|V\|_\infty\leq\mathcal{O}\left(\frac{d^3}{\delta^6}\right). \tag{4.64}$$

Therefore, by Theorem 4.2.4, the quantum simulation task can be implemented with

$$\mathcal{O}\left(\frac{d^3}{\delta^2}\cdot\frac{\log^2(\frac{d^3}{\delta^2})}{\log\log(\frac{d^3}{\delta^2})}\right) \tag{4.65}$$

queries to the quantum evaluation oracle O_f and additional

$$\mathcal{O}\left(\frac{d^3}{\delta^2}\left(d+\log^{2.5}\left(\frac{d^3}{\delta^4}\right)\right)\frac{\log^2(\frac{d^3}{\delta^2})}{\log\log(\frac{d^3}{\delta^2})}\right) \tag{4.66}$$

1- and 2-qubit gates. □

4.4.4 Numerical Simulation for 1D QHD

As complementary to the theoretical result, we numerically simulate QHD for the 1D model problem $f(x) = x^4 - (x - 1/32)^2 - c$ (where $c \approx -0.296$ so $f(x) \geq 0$). By (4.57), the tail estimate of the final ground state $|\Phi(t_f)\rangle$ is exponentially small in the parameter c,

$$\Pr_{X\sim|\Phi(t_f)|^2}\left[\|X-x^*\| > \frac{\beta(2+3c)}{\lambda_f}\right] \leq e^{-c}. \tag{4.67}$$

If we choose $c = \frac{1}{3}\left(\frac{\delta\lambda_f}{\beta} - 2\right)$ for a fixed $\delta > 0$, the failure probability can be expressed as a function of λ_f,

$$\Pr_{X\sim|\Phi(t_f)|^2}\left[\|X-x^*\| > \delta\right] \leq \mathcal{O}\left(e^{-\frac{\delta}{3\beta}\lambda_f}\right). \tag{4.68}$$

Meanwhile, Theorem 4.4.1 suggests that we need to maintain $\epsilon \sim 1/\log(\lambda_f)$ to make sure that the final state in the quantum simulation is ϵ-close to the ground state $|\Phi(t_f)\rangle$.

In our numerical experiment, we choose $\epsilon = \frac{1}{10\log(\lambda_f)}$ and simulate the Schrödinger equation (4.49) for $0 \le t \le t_f = \log(\lambda_f)$. We use the uniform superposition state as the initial state $|\Phi(0)\rangle$. When t reaches the final time $t = t_f$, we stop the quantum simulation and obtain the final state $|\Psi^{\epsilon}_{\text{sim}}(t_f)\rangle$. For some λ_f, we compute the failure probability by

$$\Pr_{X\sim|\Psi^{\epsilon}_{\text{sim}}(t_f)|^2}[\|X - x^*\| > \delta] = 1 - \langle\Psi^{\epsilon}_{\text{sim}}(t_f)|\, 1_{B_\delta}(x)\, |\Psi^{\epsilon}_{\text{sim}}(t_f)\rangle\,,$$

where we fix $\delta = 0.1$. As shown in (4.68), the failure probability should decay exponentially as the parameter λ_f increases.

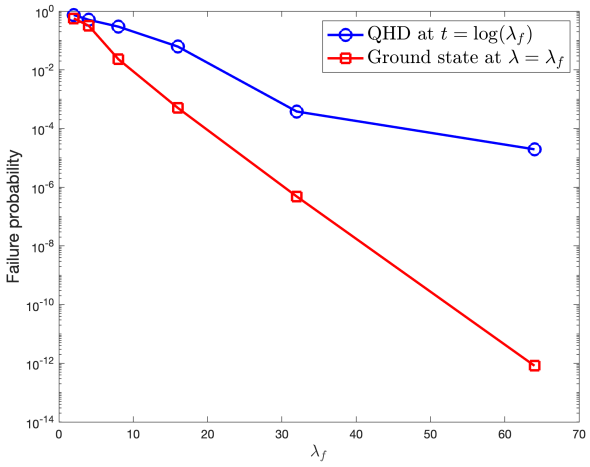


Figure 4.3: Numerical experiment for 1D QHD.

In Figure 4.3, we show the failure probability of QHD as a function of λ_f, see the blue curve. As a reference, we also plot the failure probability suggested by the ground state of $H(t_f)$, see the red curve. As predicted by our theoretical argument, both curves decrease exponentially fast with respect to the parameter λ_f. Nevertheless, a significant gap between these two curves shows that the failure probability of QHD is higher than the exact final ground state. This might be a consequence of the adiabatic approximation error and the inexact choice of the initial state.

4.5 Empirical Study of Classical Algorithms

In this section, we perform an empirical study to evaluate the performance of classical optimization algorithms for nonconvex optimization instances. We consider 6 state-of-the-art classical optimization algorithms (covering four major categories in the optimization literature):

1. Annealing-based and Monte-Carlo algorithms:

    - Dual annealing [71], which combines the generalization of CSA (Classical Simulated Annealing) and FSA (Fast Simulated Annealing);
    - Basin-hopping [72], which conceptually transforms the objective function into a collection of interpenetrating staircases (basins) by local optimization methods, and then explores these basins by a canonical Monte Carlo simulation that perturbs current coordinates and randomly accepts or rejects the new coordinates based on the corresponding basin value.

2. Gradient method: Canonical gradient descent with Gaussian noise added to the current coordinates in each step. This method is called perturbed gradient descent or stochastic gradient descent (SGD) in the literature. We stick to the latter hereafter.

3. Newton and quasi-Newton methods:

    - Interior-point method (Ipopt [73]);
    - Sequential quadratic programming (SQP) [74].

4. Branch-and-bound algorithm: Gurobi [75], which is a popular industrial-level optimization solver.

4.5.1 Methodology

Problem Instances. The problem instances are introduced in Section 4.3.3. For a given positive integer d and a d-by-d orthogonal matrix U, we define a problem instance:

$$\text{minimize} \quad F_U(\boldsymbol{x}),$$

where $\boldsymbol{x} \in \mathbb{R}^d$, $F_U(\boldsymbol{x}) := F(U\boldsymbol{x})$ and $F(\boldsymbol{x}) = F(x_1, \ldots, x_d) := \sum_{k=1}^{d} w(x_k)$. Here, we choose $w(x) = x^4 - \left(x - \frac{1}{32}\right)^2 - c$ with $c := \min_{x \in \mathbb{R}} \left\{x^4 - \left(x - \frac{1}{32}\right)^2\right\} \approx -0.296$. The constant c is introduced such that $\min_{\boldsymbol{x}} F_U(\boldsymbol{x}) = 0$.

Time-to-solution. *Time-to-solution (TTS)* is a standard performance metric for randomized optimization algorithms [32]. Suppose that a randomized algorithm solves a problem instance with success probability p and average runtime t (over its internal randomness), its TTS with success probability p_0 is defined as

$$T_{p_0} := Rt, \quad \text{where } R := \left\lceil \frac{\ln(1 - p_0)}{\ln(1 - p)} \right\rceil. \tag{4.69}$$

Intuitively, T_{p_0} is the expected time spent for R consecutive runs of the algorithm on a problem instance to guarantee that the overall success probability is at least p_0. In the literature, TTS is often evaluated for a fixed problem instance. In our case, however, the optimization instances are also randomly generated. To reflect the average performance of algorithms when applied to a class of randomly generated optimization instances, we introduce the notion of *average TTS*. Suppose that $\mathcal{A}(\theta)$ is an algorithm (parameterized by some adjustable parameter θ) and f are problem instances randomly

drawn from a set $\mathscr{C}$. We define the success probability $p(\theta)$ and average runtime $t(\theta)$ as follows:

$$p(\theta) \coloneqq \Pr[\mathcal{A}(\theta) \text{ solves a problem instance } f], \tag{4.70}$$

$$t(\theta) \coloneqq \mathbb{E}[\text{the time required by } \mathcal{A}(\theta) \text{ to halt}]. \tag{4.71}$$

Similar to the standard definition, we define the average TTS of $\mathcal{A}(\theta)$:

$$T_{p_0}(\theta) \coloneqq \left\lceil \frac{\ln(1-p_0)}{\ln(1-p(\theta))} \right\rceil \cdot t(\theta). \tag{4.72}$$

In practice, $p(\theta)$ can be estimated by counting the successful events out of a large number of trials, each with a different optimization instance. The average runtime t will be estimated similarly. In what follows, for simplicity, we will refer to "average TTS" as TTS if not stated otherwise.

TTS for Algorithms with Adjustable Parameters. For an optimization algorithm $\mathcal{A}(\theta)$, we can fine-tune the adjustable parameter θ (e.g., number of iterations, learning rate, etc.) to yield better performance. It is therefore challenging to justify which adjustable parameter θ we should choose in the estimation of TTS. To address this issue, we adopt the same approach proposed by Rønnow et al. [32] in their study comparing quantum and classical algorithms. We regard an algorithm parameterized by θ as an algorithm family and define its TTS (for a given dimensionality d) by the minimal TTS in the family over all possible θ. Admittedly, we can only try a finite number of θ in reality. Therefore, we pick the parameter range such that it is more likely to cover the optimal choice(s).

Parameter setup. We conduct all of our experiments below on the Zaratan cluster [138] if not stated otherwise. For a fixed dimensionality d and parameter θ, we run an optimization algorithm multiple

times on instances F_U with Haar-random U. We regard an algorithm successfully solves a problem instance "minimize $F_U(\boldsymbol{x})$" if it returns a solution $\boldsymbol{y}$ such that $F_U(\boldsymbol{y}) \leq C/2$, where $C \approx 0.088$ is the (globally) second lowest local minima of $F_U(\boldsymbol{x})$. The success probability (i.e., p) of an algorithm $\mathcal{A}(\theta)$ is evaluated as the fraction of successful events out of a large number of tested instances drawn from $\mathcal{C}(w, d)$. We will not report corresponding TTS if less than 5 success runs are detected since in this case the corresponding estimation of p is likely unreliable. Each run is single-threaded and assigned one CPU core. Runtime for estimating t is measured by CPU core time used but not wall time elapsed. We choose the overall success probability threshold $p_0 = 0.99$.

4.5.2 Annealing-Based and Monte-Carlo Algorithms

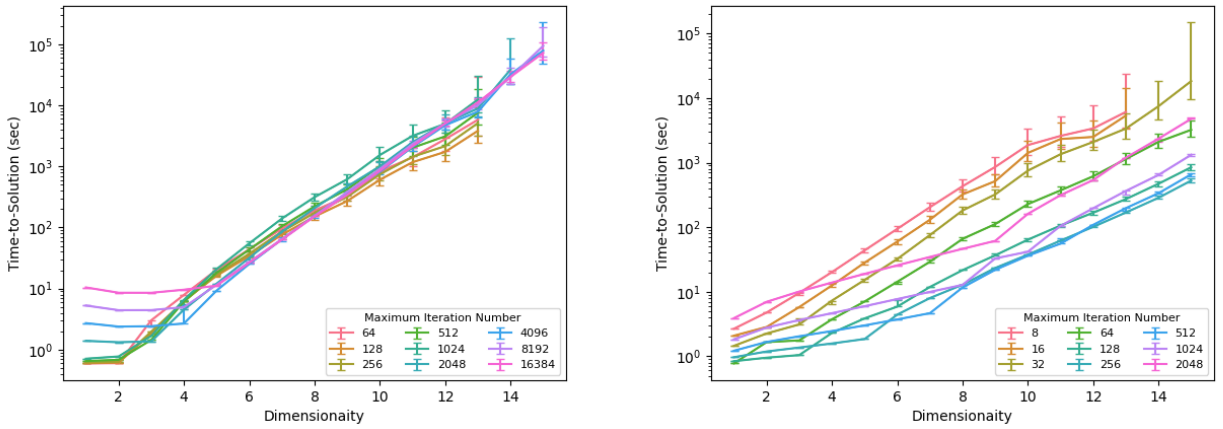
We tested two stochastic optimization algorithms, dual annealing [71] and basin-hopping [72], using their SciPy [139] implementation in Python. An important adjustable parameter θ of these algorithms is the maximal iteration number. The range of θ, together with other important problem parameters, is presented in the following table.

	Range of d	Range of θ	Number of Runs[5]
Dual Annealing	$\{1, 2, \ldots, 15\}$	$\{2^6, 2^7, \ldots, 2^{14}\}$	10000
Basin-Hopping	$\{1, 2, \ldots, 15\}$	$\{2^3, 2^4, \ldots, 2^{11}\}$	10000

Figure 4.4 shows the scaling of two algorithms. Each curve in a plot represents the TTS (parameterized by a specific θ) as a function of the dimension$d$. Note that the "true" TTS of an algorithm is

[5]The number of runs for *each* possible combination of experiment parameters.

the lower envelope of all the TTS curves, each parameterized by a different θ. It is clear that TTS for both algorithms scale exponentially in dimensionality d.



(a) Dual annealing

(b) Basin-hopping

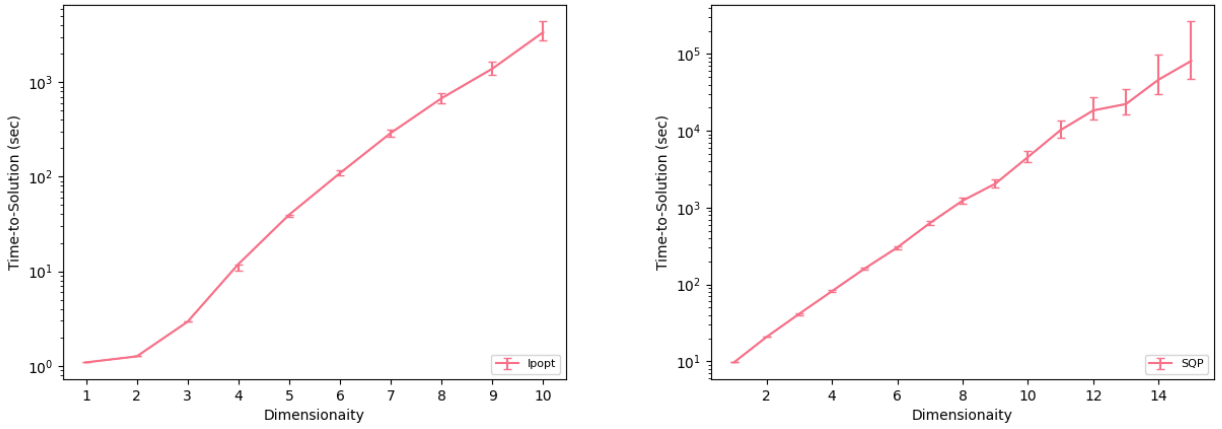
Figure 4.4: TTS scaling of dual annealing and basin-hopping. Confidence intervals are calculated at the 95% level. Results suggest that TTS for both algorithms (lower envelopes of curves) scale exponentially in dimensionality d.

4.5.3 Newton and Quasi-Newton Methods

We also tested two optimization algorithms based on Newton and quasi-Newton methods, Ipopt [73] and SQP (sequential quadratic programming) [74] in its SciPy [139] implementation. The backend linear solver for Ipopt is MA27 from Coin-HSL Archive. The Ipopt experiment was conducted on a consumer laptop (Intel Core i7-8750H CPU 2.20GHz) due to compatibility issues. We do not find an adjustable parameter in these algorithms. Some important experiment parameters are listed below.

	Range of d	Number of Runs[5]
Ipopt	$\{1, 2, \ldots, 10\}$	12000
SQP	$\{1, 2, \ldots, 15\}$	100000

Figure 4.5 shows the scaling of the TTS of the two algorithms. Both scale exponentially in dimensionality d.



(a) Ipopt (b) SQP

Figure 4.5: TTS scaling of Ipopt and SQP. Confidence intervals are calculated at the 95% level. Both scale exponentially in dimensionality d.

4.5.4 Stochastic Gradient Descent

Note that QHD is closely related to classical gradient descent [77]. It is thus natural to ask how gradient descent and its variants perform in our instances. Nevertheless, the fact that $F_U(\boldsymbol{x})$ has 2^d local minima immediately rules out the possibility that deterministic gradient-based algorithms have good performance. Therefore, we focus on stochastic gradient descent (SGD) in this section.

When applying SGD, it is crucial to choose a correct learning rate schedule as it directly determines the performance of SGD. Although we have some crude empirical rules (e.g., the learning rate should be decaying with time), there is little prior knowledge that we can leverage to choose good learning rate schedules for our instances. As a result, we choose two types of schedules of particular interest: (1) constant learning rate; and (2) the learning rate schedules that correspond to the time-dependent functions in QHD. Below we will explain how one can connect the learning rate schedule in SGD with the time-dependent functions in QHD.

Algorithm 4.2: Stochastic Gradient Descent

Input: F_U minimization objective
T total evolution time
$s : [0, T] \to \mathbb{R}_+$ effective learning rate schedule
$s_{\max}$ maximal learning rate
$k \in \mathbb{Z}_+ \cup \{\infty\}$ maximal number of rounds with no smaller objective observed

$t \leftarrow 0$; `// current time`
Draw $x \in \mathbb{R}^d$ from $\mathcal{N}(\mathbf{0}, d^2)$; `// initial guess`
while $t < T$ **do**
  $s_{\text{cur}} \leftarrow s(t)$; `// current effective learning rate`
  if $s_{\text{cur}} < s_{\max}$ **then**
    $x \leftarrow x - s_{\text{cur}}(\nabla F_U(x) + \xi)$ with $\xi \sim \mathcal{N}(\mathbf{0}, 1)$ `// descent`
    $t \leftarrow t + s_{\text{cur}}$
  else
    $x \leftarrow x - s_{\max}(\nabla F_U(x) + \xi)$ with $\xi \sim \mathcal{N}(\mathbf{0}, s_{\text{cur}}/s_{\max})$ `// noise scaled`
    $t \leftarrow t + s_{\max}$
  update minimal $F_U(x)$ observed, and
  break if it is not updated for consecutive k rounds;
output minimal $F_U(x)$ observed with corresponding x;

In Algorithm 4.2, we give the specific version of SGD we used in our experiment. It is worth noting that there are two major differences between Algorithm 4.2 and vanilla SGD: (1) the *effective* learning rate s_{cur} may not be the actual learning rate used if it exceeds $s_{\max}$, and (2) the learning rate schedule is a function of an abstract time t rather than simply a round number. The reason we made these modifications is that by doing so, our SGD will correspond to the following SDE (stochastic

differential equation) [140]:

$$\mathrm{d}x_t = -\nabla f(x_t)\mathrm{d}t + \sqrt{s(t)}\mathrm{d}W_t, \tag{4.73}$$

where W_t is a standard Brownian motion. A recent work by Liu, Su, and Li [76] generalizes this idea and articulates the connection between stochastic differential equations and the Schrödinger Equation. They find that the role of the learning rate s in (4.73) is comparable with the *quantum learning rate* h in the Schrödinger equation,

$$i\frac{\mathrm{d}}{\mathrm{d}t}\Psi = \left(h^2\left(-\frac{1}{2}\nabla^2\right) + f(x)\right)\Psi. \tag{4.74}$$

If we allow the parameter h to be time-dependent, (4.74) turns out to be nothing but QHD in disguise: let $t^\star := 1/2\lambda(t)$ in (4.49) from Algorithm 4.1. (4.49) becomes

$$i\epsilon\frac{\mathrm{d}}{\mathrm{d}t^\star}\left|\Psi^\epsilon(t^\star)\right\rangle = H(t^\star)\left|\Psi^\epsilon(t^\star)\right\rangle, \quad \text{where } H(t^\star) = \frac{1}{(2t^\star)^2}\left(-\frac{1}{2}\nabla^2\right) + f(x). \tag{4.75}$$

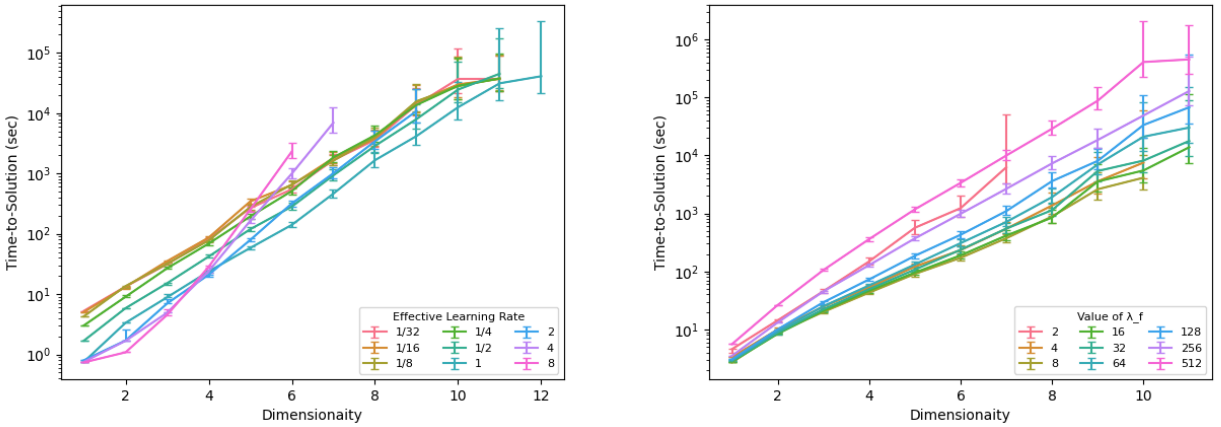
And the range of $t^\star$ is $[1/2\lambda_f, \lambda_f/2]$. The quantum learning rate is then $s = 1/2t^\star$. Translating back to SGD we get the following learning rate schedule $s_{\lambda_f}\colon [0, \lambda_f/2 - 1/2\lambda_f] \to \mathbb{R}_+$,

$$s_{\lambda_f}(t) = \frac{1}{2\left(t + 1/2\lambda_f\right)}.$$

We run the SGD experiments using the following parameters.

	Range of d	Choices of s	$s_{\max}$	k	Number of Runs[5]
SGD	$\{1, 2, \ldots, 15\}$	$\{2^{-5}, 2^{-4}, \ldots, 2^{3}\}$	0.01	1000	10000
SGD	$\{1, 2, \ldots, 15\}$	s_{λ_f} with $\lambda_f \in \{2^1, 2^2, \ldots, 2^9\}$	0.01	∞	10000

Figure 4.6 shows the scaling of SGD for two types of learning rate schedules. Here, we regard the learning rate s as a major adjustable parameter that affects the performance of the algorithm. In each subfigure, a curve represents the TTS scaling for a learning rate s. Again, the true TTS scaling of SGD should be the lower envelope of all curves parameterized by various s. Clearly, the TTS of SGD scales exponentially in dimensionality d in both settings.



(a) Fixed learning rate

(b) QHD-type learning rate

Figure 4.6: TTS scaling of SGD under two types of learning rate schedules respectively. Confidence intervals are calculated at the 95% level. It is clear that TTS in both settings (lower envelopes of curves) scale exponentially in dimensionality d. In some curves, the data points are missing because we do not report the TTS if less than 5 success runs are detected.

4.5.5 Gurobi

We also conduct an experiment to test the performance of Gurobi [75], an industrial-level optimization software based on the branch-and-bound algorithm. Note that our optimization instances are degree-4 polynomials, while Gurobi only accepts quadratic objectives. The standard resolution is to introduce auxiliary variables to represent higher-degree polynomial terms. For example, the following quartic programming problem,

$$\text{minimize} \quad x_1 x_2 x_3 x_4, \tag{4.76}$$

$$\text{where} \quad \boldsymbol{x} \in \mathbb{R}^4, \tag{4.77}$$

is equivalent to a Quadratically Constrained Quadratic Program (QCQP),

$$\text{minimize} \quad X_{12} X_{34}, \tag{4.78}$$

$$\text{where} \quad X_{12} = x_1 x_2, X_{34} = x_3 x_4. \tag{4.79}$$

Similarly, to reformulate the degree-4 nonconvex optimization instance as a QCQP, we need to introduce $\Theta(d^2)$ auxiliary variables X_{ij} for $i, j \in \{1, \ldots, d\}$ and we add corresponding constraints $X_{ij} = x_i x_j$.

In our experiment, the adjustable parameter θ of Gurobi is the *work limit*, which limits the total *work* expended in a run. The *work* metric, introduced by Gurobi, is proportional to the CPU time spent. However, the pre-factor significantly depends on the hardware and the model (problem instance type) that is being solved. In our setting, a rough estimate suggests that 1 work unit corresponds to

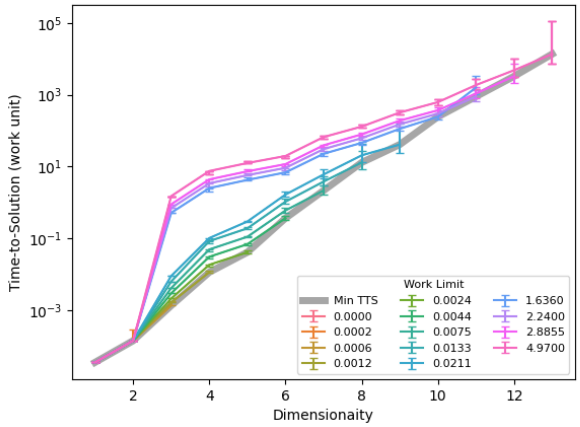


Figure 4.7: TTS scaling of Gurobi which is roughly exponential in dimensionality d. Confidence intervals are calculated at the 95% level. 1 work unit is proportional to 1 second of CPU time. In some curves, the data points are missing because we do not report the TTS if less than 5 success runs are detected.

10–100 seconds of CPU time.

Besides the *work limit*, we also set a maximal timeout of 30 minutes for Gurobi. If Gurobi does not solve a problem instance within 30 minutes, we regard it fails for this instance and use the current work limit as the runtime. Below, we list our experiment parameters.

	Range of d	Time Limit	Number of Runs[5]
Gurobi	$\{1, 2, \ldots, 15\}$	1800 sec	3000

Figure 4.7 shows the TTS scaling of Gurobi. Each TTS curve is measured for a Gurobi program with a fixed θ. In total, 10 TTS curves are depicted. Each curve uses a work limit parameter θ that yields the optimal performance in at least one dimension $d \in \{1, \ldots, 13\}$. Some data points are missing in the plot as we do not report the TTS with an extremely small success probability (i.e., number of successful events less than 5). The lower envelope of these TTS curves (see the gray curve) gives an estimated scaling of the runtime of Gurobi. The TTS of Gurobi essentially scales

exponentially in dimensionality d.

Chapter 5: Computational Complexity of Schrödinger Operators

5.1 Introduction

Simulating quantum-physical systems is one of the primary applications of quantum computers [6, 141, 142]. The well-known Schrödinger equation describes the evolution of a quantum system:

$$i\frac{\mathrm{d}}{\mathrm{d}t}\left|\Psi(t)\right\rangle = \widehat{H}\left|\Psi(t)\right\rangle,$$

where $|\Psi(t)\rangle$ is the quantum wave function and $\widehat{H}$ is the system Hamiltonian. For many quantum-mechanical systems—such as atoms, molecules, and solids [143, 144]—$\widehat{H}$ takes the form of a *Schrödinger operator*,

$$\widehat{H} = -\Delta + V,$$

where $\Delta = \sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2}$ is the Laplace operator and $V\colon \mathbb{R}^n \to \mathbb{R}$ is a potential function on n-dimensional Euclidean space. This Hamiltonian describes a continuous-variable quantum system with n degrees of freedom. For example, when $V(x)$ is a quadratic function, the Hamiltonian represents a system with n coupled quantum harmonic oscillators. For simplicity, we refer to n as the system size. The study of the Schrödinger operator is a fundamental topic in quantum mechanics, with applications to quantum chemistry [145, 146], quantum field theories [147, 148], and continuous-variable quantum information [149, 150].

Due to the central role of Schrödinger operators in quantum physics, their properties have been extensively studied by the mathematical physics community since the 1950s [127, 151–155]. Recently, the theory of Schrödinger operators has found a wide range of applications in computer science, offering insights into classical and quantum optimization algorithms. Shi, Su, and Jordan [140] highlighted

the intriguing connections between Schrödinger operators and the stochastic gradient descent algorithm. Schrödinger operators have also emerged in the quantum computing literature as a powerful tool for designing novel quantum optimization algorithms [51–53, 76–81] that are capable of leveraging the continuous structure of the objective function in the real space in the most natural way. This is in sharp contrast to traditional approaches like Quantum Adiabatic Algorithm or Quantum Approximate Optimization Algorithm, which use qudit Hamiltonians [33].

Given the emerging importance of Schrödinger operators in computer science, it is thus natural to ask the following questions:

What is the computational power of Schrödinger operators? Is it necessary to build a universal quantum computer to simulate them?

At first glance, the answers to these questions may seem straightforward, as it is widely believed that general quantum-mechanical systems cannot be efficiently simulated using classical computers [156, 157]. However, there exist efficient classical algorithms to simulate and/or compute certain physical properties of seemingly quantum processes, such as non-interacting fermions [158–160], quantum circuits consisting of Clifford gates [161], and ferromagnetic transverse field Ising model [162]. Until rigorous theoretical evidence is provided, we cannot dismiss the possibility that Schrödinger operators may admit efficient classical simulation that does not necessarily work for universal quantum computation.

On the other hand, researchers have shown that many systems are indeed difficult to simulate under the notion of BQP-completeness from computational complexity theory. In brief, if simulating a system is BQP-complete, then it is as hard as universal quantum computation. Several quantum systems have been shown to be BQP-complete, including scattering in quantum field theory [163],

Hubbard models [164, 165], and almost all 2-local interactions on 2D square lattices [166]. In other words, there is no efficient classical algorithm for any of these systems at all—unless there is one not only working for the specific system but *all* possible quantum systems, i.e., $\mathsf{BPP} = \mathsf{BQP}$.[1]

Besides the simulation of quantum dynamical processes, estimating the ground energy for a given quantum system (Hamiltonian) is another fundamental task in quantum computing, central to the field of Quantum Hamiltonian Complexity [58]. Over the past decades, researchers have identified the computational complexity of a variety of classes of Hamiltonians. For many classes, we have strong evidence that they are intractable by results that pinpoint the *exact* complexity classes they belong to. An early example is the class of k-local Hamiltonians which can be written as a sum of terms that act on at most k qubits. Kitaev's seminal result shows that the task is QMA-complete for 5-local Hamiltonians [169], where QMA is the quantum analog of NP [170]. Subsequent works improved this result by adding progressively stricter restrictions: the task remains QMA-complete for general 2-local Hamiltonians [171], 2-local Hamiltonians on square lattices [172], and most 2-local Hamiltonians on square lattices with a single fixed type of 2-qubit interactions [166, 173]; If we do not restrict the topology, complexity classification results can be obtained for 2-local Hamiltonians [174, 175]. Other classes of Hamiltonians have also been considered, such as Hamiltonians in one dimension [176, 177], Hamiltonians with no sign problem (a.k.a. stoquastic Hamiltonians) [178–180], $\mathcal{O}(1)$-local Hamiltonians with succinct ground states [181] or guiding states [182–184], Fermi-Hubbard [164] and Bose-Hubbard models [165].

In this chapter, we study the computational complexity of the system governed by a time-

[1] We note that a recent line of work [167, 168] has introduced the complexity class CVBQP to capture the power of polynomial-time continuous-variable (bosonic) quantum computing. Early evidence suggests that CVBQP may exceed BQP, but it remains open whether this suggests (i) an advantage of continuous-variable quantum computing, (ii) merely current limitations of proof techniques, or (iii) overly permissive assumptions in CVBQP's definition that do not reflect realistic bosonic circuits.

independent Schrödinger operator $\widehat{H} = -\Delta + V$. Specifically, we are interested in the hardness of two tasks: (i) simulating the dynamics under $\widehat{H}$ and (ii) estimating the ground energy of $\widehat{H}$, as both tasks are fundamental in physics for a better understanding of corresponding systems [6, 142]. For technical convenience, we restrict the domain of wave functions from $\mathbb{R}^n$ to a finite hypercube $[-1, 1]^n$. We also impose additional constraints on the regularity and complexity of the potential function $V(\boldsymbol{x})$, see Definition 5.3.9. These constraints are necessary for the practical and physical relevance of our computational model. Technically, any restriction only *strengthens* our computational hardness results.

5.1.1 Overview of Results

Simulating the Schrödinger operator is BQP-hard. We show that the task of simulating the dynamics generated by the Schrödinger operator $\widehat{H} = -\Delta + V$ is BQP-hard (Theorem 5.4.2) even if V only consists of 2-body interactions. We also have strong evidence showing that this task itself is in BQP (Theorem 5.3.10).[2] As a consequence, quantum algorithms involving Schrödinger dynamics [51–53, 76–81] cannot be efficiently dequantized by a general-purpose simulation algorithm, unless BPP $=$ BQP.

We prove the hardness result by reducing the simulation of the (time-independent) transverse field Ising model (TIM) to the task of simulating Schrödinger operators. As an interesting byproduct, we also demonstrate that simulating TIM is BQP-complete (Theorem 5.2.6), filling a conceptual gap in Hamiltonian complexity theory, as it has long been believed that TIM cannot be efficiently simulated using classical methods. A perhaps more interesting interpretation is that implementing *time-independent* TIM is powerful enough for universal quantum computing, which is particularly rel-

[2]The reason we did not obtain a BQP-completeness result is mainly technical—we assume a Dirichlet boundary condition for wave functions but the simulation algorithm of [89] works under a periodic boundary condition. We guess that a quantum Galerkin method should be able to efficiently simulate the Schrödinger dynamics with a Dirichlet boundary condition, but a formal treatment is beyond the scope of this thesis.

evant to quantum utility given that there are quantum computers based on TIM architecture such as D-Wave [185].

Determining the ground energy of the Schrödinger operator is StoqMA-complete. We show that for any potential V that involves 2- or higher-order interactions, estimating the ground energy of the Schrödinger operator $\widehat{H} = -\Delta + V$ is StoqMA-complete (Theorem 5.3.3 and Theorem 5.4.1). The complexity class StoqMA captures the computational hardness of finding the ground energy for local Hamiltonians with *no sign problem* which arise in natural physical systems [178–180, 186], and can be placed between the classes NP and QMA. Furthermore, StoqMA is equal to NP under plausible derandomization assumptions [187].[3] Unlike the simulation task, the ground energy problem of Schrödinger operators seems easier[4] than that for general local qubit Hamiltonians, a standard QMA-complete problem.

Implications for bosonic and fermionic Schrödinger operators. In this chapter, we do *not* impose bosonic or fermionic exchange symmetry on the underlying Hilbert space $L^2(\mathbb{R}^n)$ associated with the Schrödinger operator $\widehat{H} = -\Delta + V$. Physically speaking, each of the n degrees of freedom is considered distinguishable. When the particles described by $\widehat{H}$ are indistinguishable bosons or fermions, one must also impose the corresponding exchange symmetry on both the wavefunction $\Psi(r_1, \ldots, r_n)$

[3] One such assumption is that there exists a language in NE $\cap$ coNE with non-deterministic circuit complexity $2^{\Omega(n)}$, which implies AM = NP [188,189]. Therefore, NP $\subseteq$ StoqMA $\subseteq$ AM = NP, where the inclusions NP $\subseteq$ StoqMA $\subseteq$ AM is shown in [178, 179].

[4] This is because StoqMA $\subseteq$ SBP $\subseteq$ AM [178, 179]. If StoqMA = QMA we will have a radical implication BQP $\subseteq$ QMA $\subseteq$ AM $\subseteq$ PH given that it is not true relative to some oracle [190].

and the potential $V(r_1, \ldots, r_n)$:

$$\Psi(\ldots, r_i, \ldots, r_j, \ldots) = \pm\Psi(\ldots, r_j, \ldots, r_i, \ldots),$$

$$V(\ldots, r_i, \ldots, r_j, \ldots) = V(\ldots, r_j, \ldots, r_i, \ldots),$$

where the $+$ sign corresponds to bosons and the $-$ sign to fermions. Recall that one of our results is that estimating the ground energy of Schrödinger operators is in StoqMA (Theorem 5.3.3). Since the ground state of a *bosonic* Schrödinger operator remains unchanged when treating each boson as distinguishable [191, Corollary 3.1], our result immediately implies that the ground energy problem for bosonic Schrödinger operators is in StoqMA. In contrast, the ground energy problem for fermionic Schrödinger operators describing Fermi-Hubbard models is known to be QMA-hard [164].[5]

For simulation of Schrödinger operators, it is straightforward that our BQP simulation algorithm (Theorem 5.3.10) works for both bosonic and fermionic systems. In constrast, our BQP-hardness result (Theorem 5.4.2) does not have a direct implication for bosonic or fermionic Schrödinger operators because the potential function in our construction does not bear the required symmetry.

5.1.2 Technical Contributions

The fact that the Schrödinger operator is defined in an infinite-dimensional Hilbert space and is unbounded (in the sense of operator 2-norm) introduces several major technical difficulties, including that (i) most existing quantum hardness results primarily rely on computational models with finite local degrees of freedom (e.g., qubit Hamiltonians), and (ii) many standard tools in Hamiltonian complexity

[5]Strictly speaking, the potential function V in the Schrödinger operator $\widehat{H} = -\Delta + V$ considered by [164] does not satisfy the regularity assumptions we require in this chapter (see Theorem 5.3.3), due to the singular electron-electron repulsion term $\frac{1}{|r_i - r_j|}$ in V. However, this difference may not be fundamental: the potential can be regularized by imposing a $\mathrm{poly}(n)$ energy cutoff, which likely leaves the low-energy behavior essentially unchanged [192].

theory, such as the Schrieffer–Wolff transformation [193], no longer work in the infinite-dimensional setting. In this section, we briefly discuss our main technical contributions to address the aforementioned questions. These results may also be of independent interest in future research.

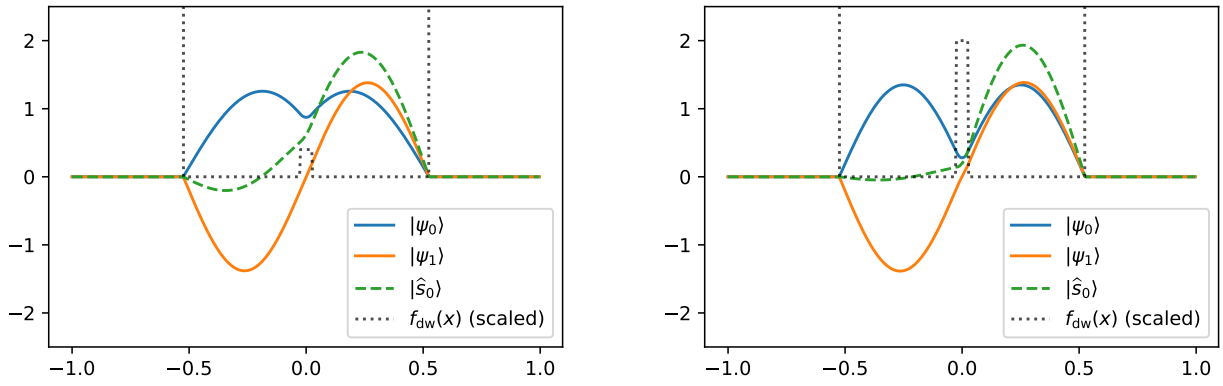
5.1.2.1 Encoding Qubit States into Bosonic Modes

The key step behind our BQP- and StoqMA-hardness results is a perturbative reduction (Theorem 5.4.3) from transverse field Ising model (TIM) Hamiltonians to Schrödinger operators. To be more specific, we embed the whole spectrum of any TIM Hamiltonian H into the low-energy subspace of a Schrödinger operator $\widehat{H}$ in an efficient way. This embedding allows us to retain all physical properties of H, including spectral correspondence and simulation correspondence.[6]

The high-level idea of our perturbative reduction is to map $|+\rangle$ and $|-\rangle$ to the ground state $|\psi_0\rangle$ and the first excited state $|\psi_1\rangle$ of a bosonic mode governed by some Hamiltonian $\widehat{X} \coloneqq -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + f_{\mathrm{dw}}$, where f_{dw} is a one-dimensional symmetric double-well potential. The correspondence allows us to embed Pauli X in the TIM Hamiltonian as $c\widehat{X}$ for some c, because $|\pm\rangle$ can be seen as the ground state and the first excited state of X. Furthermore, $|0\rangle$ (resp., $|1\rangle$) is mapped to $|\widehat{s}_0\rangle = \frac{1}{\sqrt{2}}(|\psi_0\rangle + |\psi_1\rangle)$ (resp., $|\widehat{s}_1\rangle = \frac{1}{\sqrt{2}}(|\psi_0\rangle - |\psi_1\rangle)$). Note that $|\widehat{s}_1\rangle$ will be the reflection of $|\widehat{s}_0\rangle$ across the y-axis due to the symmetry of f_{dw}. Choosing an appropriate f_{dw} can make $|\widehat{s}_0\rangle$ (resp., $|\widehat{s}_1\rangle$) a wave packet that is approximately confined in the region $x > 0$ (resp., $x < 0$); See Figure 5.1. The geometric property of $|\widehat{s}_0\rangle$ and $|\widehat{s}_1\rangle$ will thus make the embedding of a Z_iZ_j Pauli term in the TIM Hamiltonian much easier: it is easy to imagine that (a smooth version of) the diagonal operator $\mathrm{sgn}(x_1x_2)$ acting on $|\widehat{s}_a\rangle\,|\widehat{s}_b\rangle$ will approximately resemble Z_1Z_2 acting on $|a\rangle\,|b\rangle$ for $a, b \in \{0, 1\}$, where $\mathrm{sgn}(\cdot)$ denotes the sign

[6] This resembles the definition of "Hamiltonian simulation" in the work of Cubitt, Montanaro and Piddock [173]. But we cannot adopt their formal definition in this chapter as theirs concerns reductions among finite dimensional systems consisting of qudits.

function.



(a) $s = 100$ (b) $s = 500$

Figure 5.1: First two eigenfunctions $\psi_0(x), \psi_1(x)$ and their linear combination $\widehat{s}_0(x) = (\psi_0(x) + \psi_1(x))/\sqrt{2}$ for the operator $\widehat{X} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + f_{\mathrm{dw}}$ with parameter $w = 0.05$ and $\ell = 0.5$, where f_{dw} is defined in Section 5.4.1.

It is worth noting that our construction is different from the well-known GKP code [194]. In the GKP code, the logcial codespace encoding $|0\rangle$ and $|1\rangle$ is the simultaneous $+1$ eigenspace of two displacement operators, while the encoding subspace we use in this chapter is spanned by the first two eigenstates of a Schrödinger operator with a non-quadratic potential field.

We believe that our idea can be generalized to embed other Hamiltonians into differential operators which are important in the fields of partial differential equations and functional analysis.

5.1.2.2 Substitute for Schrieffer–Wolff Theory

When converting the idea in the previous section into an actual proof, a substitute for Schrieffer–Wolff theory (Lemma 5.4.31) is required to control the behavior of $e^{-i\widehat{H}t}$ when $\widehat{H}$ is approximately block-diagonal. This is because the original Schrieffer–Wolff theory [193] works with a finite-dimensional Hilbert space while our case involves an infinite-dimensional Hilbert space. Unlike the standard Schrieffer–Wolff theory, our Lemma 5.4.31 does not depend on the spectral gap between two diagonal

blocks of a Hamiltonian. Therefore, our argument may be generalized and reused in future research if (i) the underlying Hilbert space is infinite-dimensional and/or (ii) the gap between two blocks cannot be easily controlled.

5.1.3 Organization

The rest of the paper is organized as follows. We introduce hardness results for TIM in Section 5.2, which includes hardness of time-independent simulation and ground energy estimation. Section 5.3 presents BQP and StoqMA membership proofs for Schrödinger operators. Finally, we prove the BQP- and StoqMA-hardness results of Schrödinger operators in Section 5.4.

5.2 Complexity of the Transverse Field Ising Model

In this section we discuss the complexity of TIM Hamiltonians. To be more concrete, Theorem 5.2.4 and 5.2.6 are featured respectively for characterizing the complexity of (i) estimating the ground energy and (ii) time-independent simulation for TIM Hamiltonians which are defined as follows.

Definition 5.2.1. (TIM Hamiltonian) An n-qubit *TIM Hamiltonian* is that of a transverse field Ising model taking the following form:

$$H = \sum_{1\leq u\leq n} (h_u X_u + g_u Z_u) + \sum_{1\leq u,v\leq n} g_{u,v} Z_u Z_v,$$

where $h_u, g_u, g_{u,v}$ are real coefficients and X_u, Z_u are Pauli operators acting on the uth qubit.

Let us introduce the definition of stoquastic Hamiltonian and complexity class StoqMA for later

use.

Definition 5.2.2. (k-local termwise-stoquastic Hamiltonian) We call a Hamiltonian H that acts on the n-qubit Hilbert space $\mathbb{C}^{2^n}$ *k-local termwise-stoquastic* if it can be written as

$$H = \sum_j H_j,$$

where each H_j acts nontrivially on at most k qubits, and when written in the computational basis, H_j's off-diagonal entries are nonpositive.

Definition 5.2.3. (StoqMA [179], informal) The complexity class StoqMA is an extension of the classical class MA where the verifier can accept quantum states as a proof instead of a bitstring in the definition of MA. The verifier can then apply *classical* reversible circuits on the state and measure the first qubit in the $\{|+\rangle, |-\rangle\}$ basis. It accepts the proof iff the measurement outcome is '+'.

The prefix "Stoq" in StoqMA comes from the fact that determining the ground energy of a k-local termwise-stoquastic Hamiltonian is a complete problem for StoqMA as long as $k \geq 2$ [178, 179]. This resembles the fact that determining the ground energy of a general k-local Hamiltonian is QMA-complete for $k \geq 2$ [171].

The following result from Brayvi and Hastings [195] characterizes the difficulty of estimating the ground energy for a general TIM Hamiltonian.

Theorem 5.2.4 ([195]). *Determining the ground energy of TIM Hamiltonian is StoqMA-complete. More specifically, estimating within error $1/\operatorname{poly}(n)$ the smallest eigenvalue of the TIM Hamiltonian*

$$H = \sum_{1 \leq u \leq n} (h_u X_u + g_u Z_u) + \sum_{1 \leq u,v \leq n} g_{u,v} Z_u Z_v,$$

with $|h_u|, |g_u|, |g_{u,v}| \leq \mathrm{poly}(n)$ *given explicitly, is complete for the complexity class* StoqMA *under* P reductions.

Now let us turn to the complexity of simulating a TIM Hamiltonian in a time-independent fashion. The problem is not settled in previous literature explicitly to our best knowledge. However, it turns out that the simulation task is BQP-complete by virtue of a few reductions and techniques from previous works.

Definition 5.2.5. We call a family of Hamiltonians acting on n-qubit Hilbert spaces $\mathbb{C}^{2^n}$ BQP-complete if the following problem is BQP-complete under P reductions:

- **Input.** An integer n, a positive $t \leq \mathrm{poly}(n)$, the description of a Hamiltonian H from the family, a product state $|\psi\rangle \in \mathbb{C}^{2^n}$ with each component being $\mathcal{O}(1)$-qubit and a projector $M = |\mu\rangle\langle\mu| \otimes I$ where $|\mu\rangle$ is an $\mathcal{O}(1)$-qubit state.
- **Output.** `Yes` if $\mathrm{Tr}\{M|\phi\rangle\langle\phi|\} > 2/3$ or `No` if $\mathrm{Tr}\{M|\phi\rangle\langle\phi|\} < 1/3$ given the promise that one must be the case, where $|\phi\rangle \coloneqq e^{-iHt}|\psi\rangle$.

Theorem 5.2.6. *TIM Hamiltonians with coefficients bounded in* $\mathrm{poly}(n)$ *are* BQP*-complete.*

Proof. Obviously it should be in BQP since general local Hamiltonian simulation is in BQP. We focus on proving its BQP-hardness below. By Theorem 42 of [173, SI] we know that $XX + ZZ$ Hamiltonians are BQP-complete.[7] To be more specific, the family of Hamiltonians consist of H taking the shape

$$H = \sum_{1 \leq i,j \leq n} (a_{ij} X_i X_j + b_{ij} Z_i Z_j),$$

[7] In fact, one can even restrict the topology of such Hamiltonians on a 2D square lattice and the BQP-completeness will remain unchanged by virtue of [166]. From that result, a similar $XX + YY$ family also finds its application in verification of quantum simulators [196].

where $|a_{ij}|, |b_{ij}| \leq \mathrm{poly}(n)$. Now we employ the technique from [197, 198] to embed the spectrum of H into a 3-local termwise-stoquastic matrix (in the computational basis). Write $H = -\sum_k H_k$ where H_k are 2-local real matrices of size 2^n. Construct $\widetilde{H_k}$ of size 2^{n+1} from H_k by the following replacement rule for *matrix elements* of H_k:

$$x \mapsto \begin{cases} \left(\begin{smallmatrix} x & 0 \\ 0 & x \end{smallmatrix}\right), & x \geq 0, \\ \left(\begin{smallmatrix} 0 & -x \\ -x & 0 \end{smallmatrix}\right), & x < 0. \end{cases}$$

Note that $-\widetilde{H_k}$ is 3-local if we interpret it as acting on $n+1$ qubits and it is clearly stoquastic. Moreover, it is easy to verify that $\widetilde{H_k} = H_k \otimes |-\rangle\langle -| + |H_k| \otimes |+\rangle\langle +|$ where $|\cdot|$ applies entrywise. Hence we can construct the following 3-local stoquastic Hamiltonian:

$$\widetilde{H} := -\sum_k \widetilde{H_k} = H \otimes |-\rangle\langle -| + \left(-\sum_k |H_k|\right) \otimes |+\rangle\langle +| .$$

Therefore, the evolution of $|\psi\rangle$ under Hamiltonian H can be simulated by the evolution of $|\psi\rangle|-\rangle$ under $\widetilde{H}$. We conclude that 3-local termwise stoquastic Hamiltonians are also BQP-complete. Finally, as detailed in Section 5.4 of [173, SI], 3-local termwise-stoquastic Hamiltonians can be efficiently simulated by TIM Hamiltonians. This is achieved by chaining local simulation protocols from [178] and [195], which reduce 3-local termwise-stoquastic Hamiltonians to 2-local ones, and 2-local ones to TIM Hamiltonians, respectively. □

5.3 Algorithms for Schrödinger Operators

From now on we will focus on the Schrödinger operator $\widehat{H} = -\Delta + V$ in the *hypercube* D with V being a $\mathcal{O}(1)$-*body potential*. The hypercube constraint enforces a Dirichlet boundary condition on all wave functions such that they must vanish on the boundary of D. The only exception is that in Section 5.3.2 we will consider a periodic boundary condition instead of a Dirichlet one.

Definition 5.3.1. (Hypercube) An n-dimensional hypercube is $D \coloneqq [-1, 1]^n$. Throughout this chapter we use D to denote this specific hypercube unless noted otherwise.

Definition 5.3.2. (k-body potential) A potential function $V : D \to \mathbb{R}$ is called k-*body* if it only consists of k-body interaction: In other words, it can be written as the sum of functions that depends on at most k coordinates. For example, the Coulomb potential of n electrons on a straight line,

$$V(r_1, \ldots, r_n) = \sum_{1 \le i < j \le n} \frac{1}{|r_i - r_j|},$$

is a 2-body potential.

The main results in this section are Theorem 5.3.3 and Theorem 5.3.10, which address the upper bound of the complexity of (i) estimating the ground energy and (ii) time-independent simulation for Schrödinger operators, respectively.

5.3.1 Estimating the Ground Energy is in StoqMA

Theorem 5.3.3. *Determining the ground energy of Schrödinger operators with $\mathcal{O}(1)$-body potential is in* StoqMA. *More specifically, let $V : D \to \mathbb{R}$ be a $\mathcal{O}(1)$-body potential function with C^1-norm*

bounded within $\mathrm{poly}(n)$.[8] *Then, estimating within error* $1/\mathrm{poly}(n)$ *the smallest eigenvalue of the Schrödinger operator* $H = -\Delta + V$ *is in the complexity class* StoqMA.

Proof. First, we use the vanilla finite difference method to discretize H. Write $D = D_1 \times D_2 \times \cdots \times D_n$ with $D_j := [-1, 1]$ for $j = 1, 2, \ldots, n$. Fix a stepsize δ. Let $\widetilde{D}$ be the discretized version of D: $\widetilde{D} = D \cap \{(j_1\delta, j_2\delta, \ldots, j_n\delta) : j_1, \ldots, j_n \in \mathbb{Z}\}$ and similarly $\widetilde{D}_i := D_i \cap \{j\delta : j \in \mathbb{Z}\}$. Obviously $\widetilde{D} = \widetilde{D}_1 \times \cdots \times \widetilde{D}_n$. Suppose $V = \sum_j V_j$ is the decomposition of the $\mathcal{O}(1)$-body potential V. Define $\widetilde{V}_j := \sum_{\boldsymbol{x} \in \widetilde{D}} V_j(\boldsymbol{x}) \left|\boldsymbol{x}\right\rangle \left\langle \boldsymbol{x}\right|$. Define $\widetilde{T}_j$ to be a tri-diagonal matrix that discretizes $-\frac{\partial^2}{\partial x_j^2}$:

$$\widetilde{T}_j := \delta^{-2} \left(2I - \sum_{\substack{x,y \in \widetilde{D}_j \\ |x-y|=\delta}} |x\rangle \langle y| \right).$$

Let

$$\widetilde{H} := \sum_{1 \le j \le n} \widetilde{T}_j + \sum_j \widetilde{V}_j.$$

Now by a standard analysis of this finite difference method [199] we have $\left|\lambda_{\min}(H) - \lambda_{\min}(\widetilde{H})\right| \le \delta^{\Omega(1)}$.

Next, we invoke the unary embedding method from [63] for $\widetilde{H}$. Intuitively, we will embed the Hilbert space of a qudit $\mathbb{C}^{m+1}$ into that of m qubits $\mathbb{C}^{2^m}$ by the correspondence $U : |j\rangle \mapsto |0^j 1^{m-j}\rangle$ for $j = 0, 1, \ldots, m$. Define $\mathcal{S}$ to be the embedding subspace $\mathrm{span}\{|0^j 1^{m-j}\rangle\}_j$. For any $(m+1)$-dimensional real Hermitian matrix $A = \sum_{0 \le i,j \le m} \alpha_{i,j} |i\rangle \langle j|$ we define its embedding to be

$$\sigma(A) = \left(\alpha_{0,0} I_0 + \frac{1}{2} \sum_{0 \le j \le m-1} (\alpha_{j-1,j-1} - \alpha_{j,j})(I_j - Z_j) \right) + \sum_{0 \le j < k \le m} X_{k-1} \otimes \cdots \otimes X_{j+1} \otimes \alpha_{j,k} X_j.$$

[8] i.e., the function has supremum norm $\mathrm{poly}(n)$ and it is $\mathrm{poly}(n)$-Lipschitz.

It is straightforward to verify that $\sigma(A) = UAU^\dagger$ in $\mathcal{S}$ [63]. Furthermore, we have:

- $\sigma(I)$ is the identity.
- $\sigma(A)$ is 1-local if A is tri-diagonal.
- $\sigma(A)$ is diagonal (resp., stoquastic) if A is diagonal (resp., stoquastic).

Now, we want to embed the discretized Schrödinger operator $\widetilde{H} = \sum \widetilde{T}_j + \sum \widetilde{V}_j$. Define

$$\widehat{T}_j \coloneqq \sigma(I)^{\otimes j} \otimes \sigma(\widetilde{T}) \otimes \sigma(I)^{\otimes n-j-1} \text{ and } \widehat{V}_j \coloneqq \sigma(\widetilde{V}_j) \otimes \sigma(I)^{\otimes n-k}$$

assuming $\widetilde{V}_j$ acts nontrivially on the first k qudits. This naturally generalizes to a definition for general $\widehat{V}_j$. Let $\widehat{H} \coloneqq \sum \widehat{T}_j + \sum \widehat{V}_j$. By our construction, $\widehat{H} = U^{\otimes n} \widetilde{H} U^{\dagger \otimes n}$ in $\mathcal{S}^{\otimes n}$ and $\widehat{H}\mathcal{S}^{\otimes n} \subseteq \mathcal{S}^{\otimes n}$.

Finally, we introduce a "penalty" diagonal Hamiltonian $\widetilde{Q}$ that separates its lowest eigenspace $\mathcal{S}^{\otimes n}$ from the remaining eigenspaces:

$$\widehat{Q} \coloneqq \sum_{1 \le i \le n} I^{\otimes i-1} \otimes \left(2I - \sum_{0 \le j \le m-2} Z_j Z_{j+1} + Z_0 - Z_{m-1} \right) \otimes I^{n-i}.$$

It can be verified that its spectral gap is $\lambda_1(\widehat{Q}) - \lambda_0(\widehat{Q}) = 4$ and $\lambda_0(\widehat{Q}) = \lambda_{\min}(\widehat{Q}) = 0$.

Now, consider the spectrum of the operator

$$\widehat{H}_\star \coloneqq \widehat{H} + c\widehat{Q},$$

which is a 2-local termwise-stoquastic Hamiltonian. To analyze the smallest eigenvalue of $\widehat{H}_\star$, note that it is block-diagonal regarding subspaces $\mathcal{S}^{\otimes n}$ and $(\mathcal{S}^{\otimes n})^\perp$ because (i) $\widehat{Q}$ has the minimal eigenvalue 0 with the eigenspace $\mathcal{S}^{\otimes n}$ and (ii) $\widehat{H}\mathcal{S}^{\otimes n} \subseteq \mathcal{S}^{\otimes n}$. The smallest eigenvalue of $\widehat{H}_\star$ is just that of $\widetilde{H}$

because $\widehat{H}_\star = \widehat{H} + c \cdot 0 = U^{\otimes n} \widetilde{H} U^{\dagger \otimes n}$ when restricted on the subspace $\mathcal{S}^{\otimes n}$. On the other hand, for any $|\psi\rangle \in (\mathcal{S}^{\otimes n})^\perp$ we have

$$\left\langle \psi | \widehat{H}_\star | \psi \right\rangle = \left\langle \psi | \widehat{H} | \psi \right\rangle + c \left\langle \psi | \widehat{Q} | \psi \right\rangle \geq -\left\| \widehat{H} \right\| + c \cdot 4 \geq -p(n) + 4c.$$

where $p(n)$ is some polynomial in n by the naive estimation of $\widehat{H}$. Choosing $c = p(n)$ we are guarantee that $\left\langle \psi | \widehat{H}_\star | \psi \right\rangle > p(n) > \lambda_{\min}(\widetilde{H})$. Hence the smallest eigenvalue of $\widehat{H}_\star$ on subspace $(\mathcal{S}^{\otimes n})^\perp$ is greater than that on $\mathcal{S}^{\otimes n}$. Therefore, $\lambda_{\min}(\widehat{H}_\star) = \lambda_{\min}(\widetilde{H})$, which is in turn $1/\operatorname{poly}(n)$-close to $\lambda_{\min}(H)$ by choosing a sufficiently small $\delta = 1/\operatorname{poly}(n)$. □

5.3.2 Time-Independent Simulation is in BQP

To show that simulation of time-independent Schrödinger equation with periodic boundary condition is in BQP, we simply invoke the *pseudospectral algorithm* of Childs et al. [89, Theorem 8] with a refined analysis [77, Theorem 4]. Let $D := [0, 1]^n$ in this section for convenience. We first recall the notion of *mesh* and *Fourier series*.

Definition 5.3.4 (Mesh)**.** An n-dimensional mesh of order m is the discretization of the hypercube $[0, 1)^n$, written as $\Gamma_m^n := \{0, \frac{1}{2m+1}, \ldots, \frac{2m}{2m+1}\}^n$.

Definition 5.3.5 (Fourier series)**.** Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Write $\boldsymbol{k} = (k_1, \ldots, k_n) \in \mathbb{Z}^n$. A *Fourier series of order* m is of the form

$$f(\boldsymbol{x}) = \sum_{-m \leq k_1, \ldots, k_n \leq m} c_{\boldsymbol{k}} \prod_{1 \leq j \leq n} \exp(2\pi i k_j x_j).$$

Note that $f(\boldsymbol{x})$ is periodic with translation lattice $\mathbb{Z}^n$. Furthermore, f is uniquely determined by its

value on the mesh Γ_m^n.

Definition 5.3.6 (Pseudospectral algorithm [89])**.** The purpose of this algorithm is to simulate the time-dependent Schrödinger equation

$$i\frac{\partial}{\partial t}\left|\Psi(\boldsymbol{x},t)\right\rangle = \left(-\Delta + V(\boldsymbol{x},t)\right)\left|\Psi(\boldsymbol{x},t)\right\rangle$$

from $t = 0$ to $t = T$ over the hypercube D with a periodic boundary condition. The algorithm maintains the approximation of $\Psi(\boldsymbol{x},t)$, denoted as $\widetilde{\Psi}(\boldsymbol{x},t)$, on the mesh Γ_m^n in the form of a quantum state

$$\left|\widetilde{\Psi}(t)\right\rangle \coloneqq \sum_{\boldsymbol{x}\in\Gamma_m^n} \widetilde{\Psi}(\boldsymbol{x},t)\left|\boldsymbol{x}\right\rangle.$$

- **Input.** Dimension n; Truncation number m; Evolution time T; The *quantum* input state $\left|\widetilde{\Psi}(0)\right\rangle$ such that $\widetilde{\Psi}(\cdot,0) = \Psi(\cdot,0)$; An oracle O_V which returns the value of $V(\boldsymbol{x},t)$ given $\boldsymbol{x}$ and t; Two metadata oracles O_{inv} and O_{norm} (See [89, Sec. 3.2]) which act trivially if V is independent on t.
- **Output.** A state $\left|\widetilde{\Psi}(T)\right\rangle$.

Theorem 5.3.7 ([89])**.** *Suppose $V(\boldsymbol{x},t)\colon D\times[0,T]\to\mathbb{R}$ is bounded, smooth in $\boldsymbol{x}$ and t, and periodic in $\boldsymbol{x}$. Moreover, we assume that the initial wave function $\Psi(\boldsymbol{x},0)$ is periodic on D and V is L-Lipschitz in t. Define $\|V\|_{\infty,1} \coloneqq \int_0^T \|V(\cdot,t)\|_\infty \,\mathrm{d}t$. Then, choosing $m \gg \log\eta$ can guarantee that*

$$\left|\widetilde{\Psi}(\boldsymbol{x},T) - \Psi(\boldsymbol{x},T)\right| \le \eta$$

for every $\boldsymbol{x}\in\Gamma_m^n$. The cost of the algorithm includes the following:

1. *Queries to the quantum evaluation oracle O_V, O_{inv} and O_{norm}:* $\mathcal{O}\left(\|V\|_{\infty,1}\frac{\log(\|V\|_{\infty,1}/\eta)}{\log\log(\|V\|_{\infty,1}/\eta)}\right)$,

2. *1- and 2-qubit gates:*

$$\mathcal{O}\left(\|V\|_{\infty,1}\left(\mathrm{polylog}\left(T/\eta\right)+\log^{2.5}\left(L\|V\|_{\infty,1}/\eta\right)+d\log m\right)\frac{\log(\|V\|_{\infty,1}/\eta)}{\log\log(\|V\|_{\infty,1}/\eta)}\right).$$

Remark 5.3.8. Compared to the original statement in [77, Theorem 4], we made the dependence on m explicit and replaced the $\mathrm{poly}(z)$ by a $\mathrm{polylog}(T/\eta)$ in Theorem 5.3.7 where V is able to be computed with z bits of precision. This can be justified as follows. Setting $z := c\log(T/\eta)$ is sufficient for obtaining the value of V with error at most $\mathcal{O}(c\eta/T)$. By Duhamel's principle we have

$$\begin{aligned}&\exp_{\mathcal{T}}(\int_0^t -iA(s)\mathrm{d}s)-\exp_{\mathcal{T}}(\int_0^t -iB(s)\mathrm{d}s)\\ =&-i\int_0^t\left(\exp_{\mathcal{T}}\left(\int_s^t -iA(\tau)\mathrm{d}\tau\right)(A(s)-B(s))\exp_{\mathcal{T}}\left(\int_0^s -iB(\tau)\mathrm{d}\tau\right)\right)\mathrm{d}s.\end{aligned}$$

Let $A(t)$ be the ideal Schrödinger operator $-\Delta+V(\boldsymbol{x},t)$ and $B(t)$ be one with V truncated with error $\mathcal{O}(c\eta/T)$ as discussed above. We immediately obtain that the truncation incurs at most $\mathcal{O}(c\eta/T)\cdot t=\mathcal{O}(c\eta)$ error in $\Psi(\boldsymbol{x},t)$.

Let us formally define the simulation problem before proceeding to a proof.

Definition 5.3.9. SCHRÖDINGERSIM is a decision problem with its input and output listed as below.

- **Input.** Dimension n; Evolution time $t\leq\mathrm{poly}(n)$; A classical $\mathrm{poly}(n)$-time explicit oracle to compute $V(\boldsymbol{x})$ where $V:D\to\mathbb{R}$ is a smooth potential function with its C^1-norm bounded within $\mathrm{poly}(n)$;[9] $\mathcal{O}(n)$ many of classical $\mathrm{poly}(n)$-time explicit oracles to compute

[9] i.e., the function has supremum norm $\mathrm{poly}(n)$ and it is $\mathrm{poly}(n)$-Lipschitz.

$\mathcal{O}(1)$-dimensional smooth wave functions $\psi_j(\boldsymbol{x})$ and $\mu(\boldsymbol{x})$ with their C^1-norms bounded within $\mathrm{poly}(n)$.

- **Output.** `Yes` if $\mathrm{Tr}\{M\,|\phi\rangle\,\langle\phi|\} > 2/3$ or `No` if $\mathrm{Tr}\{M\,|\phi\rangle\,\langle\phi|\} < 1/3$ given the promise that one must be the case, where

$$|\phi\rangle := e^{-i(-\Delta+V)t}\,|\psi\rangle := e^{-i(-\Delta+V)t}\left(\bigotimes_j |\psi_j\rangle\right) \text{ and } M := |\mu\rangle\,\langle\mu| \otimes I.$$

We will have two variations of the problem depending on which boundary condition of wave functions we are considering: the *Dirichlet* boundary condition and the *periodic* boundary condition.

Theorem 5.3.10. SCHRÖDINGERSIM *with periodic boundary condition is in* BQP.

Proof. We describe the polynomial-time quantum algorithm below. Let the truncation parameter m be determined later. We prepare the state $\left|\widetilde{\Psi}(0)\right\rangle$ in the pseudospectral algorithm as the approximation of $\bigotimes_j |\psi_j\rangle$ on the mesh Γ_m^n:

$$\left|\widetilde{\Psi}(0)\right\rangle := \bigotimes_j \left|\widetilde{\psi_j}\right\rangle := \bigotimes_j \left(\sum_{\boldsymbol{x}\in\Gamma_m^{n_j}} \psi_j(\boldsymbol{x})\,|\boldsymbol{x}\rangle\right),$$

where n_j is the dimension of wave function $\psi_j(\cdot)$. This can be done via standard quantum black-box state preparation with $\mathcal{O}\big(nm^{\mathcal{O}(1)}\big)$ oracle calls to values of $\psi_j(\cdot)$ [200]. Similarly, we construct the circuit U such that $U\,|0\rangle = |\widetilde{\mu}\rangle$, where $|\widetilde{\mu}\rangle$ is the mesh approximation to $|\mu\rangle$ just like $\left|\widetilde{\psi_j}\right\rangle$ to $|\psi_j\rangle$. The implementation of U uses at most $\mathcal{O}\big(m^{\mathcal{O}(1)}\big)$ oracle calls to values of $\mu(\cdot)$. We run the pseudospectral algorithm to obtain the final state $\left|\widetilde{\Psi}(t)\right\rangle$ by feeding it the explicit oracle to $V(\cdot)$. The runtime is $\mathcal{O}(t\,\mathrm{polylog}(t)\,\mathrm{poly}(n)\log m) = \mathcal{O}(\mathrm{poly}(n)\log m)$ if we set the error η to be e^{-n}. It is obvious

that the probability of measuring $U^{-1}\left|\widetilde{\Psi}(t)\right\rangle$ in the computational basis and then getting result 0 is exactly $\mathrm{Tr}\left\{\widetilde{M}\left|\widetilde{\Psi}(t)\right\rangle\left\langle\widetilde{\Psi}(t)\right|\right\}$, where $\widetilde{M}=|\widetilde{\mu}\rangle\langle\widetilde{\mu}|$. Now, $\left|\widetilde{\Psi}(t)\right\rangle$ is the approximation of $|\Psi(t)\rangle = e^{-i(-\Delta+V)t}\left|\widetilde{\Psi}(0)\right\rangle$ with pointwise error at most e^{-n}, and $\left|\left\langle\widetilde{\mu}|\widetilde{\Psi}(0)\right\rangle - \langle\mu|\psi\rangle\right| \leq \mathcal{O}(\sqrt{n}L^2/m)$ if $\psi(\cdot)=\prod_j \psi_j(\cdot)$ and $\mu(\cdot)$ are L-Lipschitz. Therefore,

$$\left|\mathrm{Tr}\left\{\widetilde{M}\left|\widetilde{\Psi}(t)\right\rangle\left\langle\widetilde{\Psi}(t)\right|\right\} - \mathrm{Tr}\{M\,|\phi\rangle\langle\phi|\}\right| \leq \mathcal{O}\left(\frac{\sqrt{n}L^2}{m}+e^{-n}\right).$$

We conclude the proof by that $L\leq \mathrm{poly}(n)$ from the C^1-norm conditions of $\psi_j(\cdot)$ and $\mu(\cdot)$, and hence some $m\leq \mathrm{poly}(n)$ will make the error above smaller than $1/100$ when $n\to\infty$. □

5.4 Hardness of Schrödinger operators

The two main theorems of this section, Theorem 5.4.1 and 5.4.2, establish the hardness of Schrödinger operators in terms of the complexity of finding the ground energy and time-independent simulation. Their correctness follow from (i) Theorem 5.4.3, a perturbative reduction from Schrödinger operators to TIM Hamiltonians and (ii) the hardness results of TIM Hamiltonians, namely Theorem 5.2.4 and 5.2.6.

The rest of this section is organized as follows: In Section 5.4.1 we provide the essential idea behind Theorem 5.4.3 by motivating a perturbative reduction in the one-dimensional case. In Section 5.4.4 we present the proof of Lemma 5.4.4, which serves as our main lemma that plays the central role in showing the correctness of the main perturbative reduction (Theorem 5.4.3). We need some preparation before that: The first ingredient is the semiclassical analysis of Schrödinger operators with V being double-well functions in Section 5.4.2. Then we introduce some perturbation theory tools in Section 5.4.3.

Theorem 5.4.1 (StoqMA-hardness)**.** *Determining the ground energy of Schrödinger operators with 2-body potential is StoqMA-hard. More specifically, let $V : [-1,1]^n \to \mathbb{R}$ be a 2-body potential function with C^1-norm bounded within $\mathrm{poly}(n)$. Then, estimating within error $1/\mathrm{poly}(n)$ the smallest eigenvalue of the Schrödinger operator $H = \Delta + V$ is StoqMA-hard.*

Theorem 5.4.2 (BQP-hardness)**.** SCHRÖDINGERSIM *with Dirichlet boundary condition is BQP-hard even if V is 2-body.*

Theorem 5.4.3 (Perturbative reduction)**.** *Let $\varepsilon := 1/P(n)$ and $t < P(n)$ for some polynomial $P(\cdot)$. For any TIM Hamiltonian*

$$H = \sum_{1 \le u \le n} (a_u X_u + b_u Z_u) + \sum_{1 \le u,v \le n} b_{u,v} Z_u Z_v$$

such that $|a_u|, |b_u|, |b_{u,v}| \le P(n)$, there exists a Schrödinger operator

$$\widehat{H} = -\Delta + V$$

and a $\mathrm{poly}(n)$-time computable $C \le \mathrm{poly}(n)$ satisfying the following:

(i) The potential $V : [-1,1]^n \to \mathbb{R}$ is 2-body, smooth, $\mathrm{poly}(n)$-time computable and has C^1-norm bounded within $\mathrm{poly}(n)$.

(ii) $\left| C\lambda_k(\widehat{H}) - \lambda_k(H) \right| \le \varepsilon$ for every $k \in \{0, 1, \ldots, 2^n - 1\}$.

(iii) $\left\| W^\dagger P_{\mathcal{S}} e^{-iC\widehat{H}t} W - e^{-iHt} \right\| \le \varepsilon$, where $W := \bigotimes_{1 \le k \le n} (|\psi_0^k\rangle\langle -| + |\psi_1^k\rangle\langle +|)$ shifts basis and $\mathcal{S}$ is the image of W. $|\psi_0^k\rangle$ and $|\psi_1^k\rangle$ are 1-dimensional. Moreover, wave functions $\psi_j^k(\cdot)$ are smooth, $\mathrm{poly}(n)$-time computable and have C^1-norm bounded within $\mathrm{polylog}(n)$.

Proof. Without loss of generality assume $a_u \geq 0$ because otherwise one can conjugate H with Z_u and this conjugation simply flip the uth qubit in the $\{|+\rangle, |-\rangle\}$ basis. Now compare the theorem statement of Theorem 5.4.3 and that of Lemma 5.4.4: the only obstacle from directly invoking Lemma 5.4.4 is that it has an additional assumption $a_u \geq 1/M_2$ for every u. The assumption can be safely removed by Lemma 5.4.5 which uses a perturbative reduction. □

Lemma 5.4.4. *Fix $M_1, M_2, \varepsilon_1 > 0$ and $t \geq 0$. For any TIM Hamiltonian*

$$H = \sum_{1 \leq u \leq n} (a_u X_u + b_u Z_u) + \sum_{1 \leq u,v \leq n} b_{u,v} Z_u Z_v$$

such that $|a_u|, |b_u|, |b_{u,v}| \leq M_1$ and $a_u > 1/M_2$, there exists a Schrödinger operator

$$\widehat{H} = -G^\star \Delta + V$$

and a polynomial $P \leq \mathrm{poly}(n, M_1, M_2, 1/\varepsilon_1, t)$ satisfying the following:

(i) $1 \leq G^\star \leq P$ *is* $\mathrm{poly}(n)$*-time computable and the potential* $V : [-1,1]^n \to \mathbb{R}$ *is* 2*-body, smooth,* $\mathrm{poly}(n)$*-time computable and has* C^1*-norm bounded within* P*.*

(ii) $\left|\lambda_k(\widehat{H}) - \lambda_k(H)\right| \leq \varepsilon_1$ *for every* $k \in \{0, 1, \ldots, 2^n - 1\}$*.*

(iii) $\left\|W^\dagger P_{\mathcal{S}} e^{-i\widehat{H}t} W - e^{-iHt}\right\| \leq \varepsilon_1$*, where* $W := \bigotimes_{1 \leq k \leq n} (|\psi_0^k\rangle\langle -| + |\psi_1^k\rangle\langle +|)$ *shifts basis and* $\mathcal{S}$ *is the image of* W*.* $|\psi_0^k\rangle$ *and* $|\psi_1^k\rangle$ *are* 1*-dimensional. Moreover, wave functions* $\psi_j^k(\cdot)$ *are smooth,* $\mathrm{poly}(n)$*-time computable and have* C^1*-norm bounded within* $\mathrm{polylog}(P)$*.*

Lemma 5.4.5. *Fix $M_1, M_2 > 0$. For any TIM Hamiltonian*

$$H = \sum_{1 \le u \le n} (a_u X_u + b_u Z_u) + \sum_{1 \le u,v \le n} b_{u,v} Z_u Z_v$$

such that $|a_u|, |b_u|, |b_{u,v}| \le M_1$ and $a_u \ge 0$, there exists another TIM Hamiltonian

$$\widetilde{H} = \sum_{1 \le u \le n} (\widetilde{a}_u X_u + \widetilde{b}_u Z_u) + \sum_{1 \le u,v \le n} \widetilde{b}_{u,v} Z_u Z_v$$

with all coefficients being $\mathrm{poly}(n)$-time computable, which satisfies the following:

(i) $|\widetilde{a}_u|, \left|\widetilde{b}_u\right|, \left|\widetilde{b}_{u,v}\right| \le \max\{M_1, 1/M_2\}$ *and* $\widetilde{a}_u \ge 1/M_2$ *for any* u.

(ii) $\left|\lambda_k(\widetilde{H}) - \lambda_k(H)\right| \le n/M_2$ *for every* $k \in \{0, 1, \ldots, 2^n - 1\}$.

(iii) $\left\|e^{-i\widetilde{H}t} - e^{-iHt}\right\| \le nt/M_2$.

Proof. We simply choose $\widetilde{a}_u := \min\{a_u, 1/M_2\}$, $\widetilde{b}_u := b_u$ and $\widetilde{b}_{u,v} := b_{u,v}$. Note that $|\widetilde{a}_u - a_u| \le 1/M_2$. Therefore $\left\|H - \widetilde{H}\right\| \le n/M_2$ and hence (ii) and (iii) follow by Weyl's inequality (Lemma 5.4.28) and Lemma 5.4.30, respectively. □

5.4.1 Proof Idea

We explain the rough idea with a one-dimensional toy case (i.e., embedding a qubit into a single bosonic mode). Suppose we want to simulate the TIM Hamiltonian $H = aX + bZ$ with some Schrödinger operator $\widehat{H}$. A natural approach will be to first assume that the reduction map $\sigma : H \mapsto \widehat{H}$ is *linear* for a "reasonable" range of a and b. Let us first consider what $\widehat{Z} := \sigma(Z)$ to choose: Obviously it should be the linear combination of $\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ and a diagonal potential function f since these are

only two allowed terms in a Schrödinger operator $\hat{H}$. Now, many choices are possible but there is one promising candidate being the sign function $\hat{Z} = \mathrm{sgn}(x)$ with two advantages: (i) this choice easily generalizes to higher dimensions and (ii) for any wave function $\psi(x)$ staying in $x > 0$ (resp., $x < 0$) we have $\hat{Z}\left|\psi\right\rangle = \left|\psi\right\rangle$ (resp. $\hat{Z}\left|\psi\right\rangle = -\left|\psi\right\rangle$), which resembles the fact $Z\left|0\right\rangle = \left|0\right\rangle$ (resp., $Z\left|1\right\rangle = -\left|1\right\rangle$). Sticking with this choice, the next step is to find an $\widehat{X} \coloneqq \sigma(X)$ that is compatible with the choice of $\widehat{Z}$ in the sense that we can find wave functions $\widehat{s_0}(x)$ and $\widehat{s_1}(x)$ such that

$$\widehat{X}(\alpha\left|\widehat{s_0}\right\rangle + \beta\left|\widehat{s_1}\right\rangle) \approx \alpha\left|\widehat{s_1}\right\rangle + \beta\left|\widehat{s_0}\right\rangle,$$

$$\widehat{Z}(\alpha\left|\widehat{s_0}\right\rangle + \beta\left|\widehat{s_1}\right\rangle) \approx \alpha\left|\widehat{s_0}\right\rangle - \beta\left|\widehat{s_1}\right\rangle.$$

In other words, $\widehat{Z}, \widehat{X}, \left|\widehat{s_0}\right\rangle, \left|\widehat{s_1}\right\rangle$ approximately retain the algebraic relation among $Z, X, \left|0\right\rangle, \left|1\right\rangle$. We already know that any $\widehat{s_0}(x)$ (resp., $\widehat{s_1}(x)$) that concentrates in the region $x < 0$ (resp., $x > 0$) will fit the bill for $\widehat{Z}$ from the previous discussion about $\widehat{Z}$. Therefore we assume that they have this concentration property from now on. Alas, it turns out that we cannot find a compatible $\widehat{X}$ unless the error hidden in $\approx$ can be a constant.[10] However, note that what we actually wanted to do is maintaining the *spectral* property from H to $\hat{H}$. Therefore an $\hat{X}$ which acts on $\left|\widehat{s_0}\right\rangle, \left|\widehat{s_1}\right\rangle$ like $c_1 X + c_2 I$ acting on $\left|0\right\rangle, \left|1\right\rangle$ for some c_1, c_2 will suffice. Another way to see this is that time-independent simulation under Hamiltonian $c_1 H + c_2 I$ is equivalent to that under H modulo phase difference and time dilation. In light of this, we can heuristically find $\widehat{X}$ by considering the spectral property of $c_1 X + c_2 I$: its eigenstates are $\left|0\right\rangle + \left|1\right\rangle$ and $\left|0\right\rangle - \left|1\right\rangle$. This means a valid $\widehat{X}$ should have eigenfunctions that look like $\widehat{s_0}(x) + \widehat{s_1}(x)$ and $\widehat{s_0}(x) - \widehat{s_1}(x)$. Since we want to embed H to the low-energy subspace of $\widehat{H}$ we prefer these two eigenstates being the ground state and the first excited state of $\widehat{X}$. Recall the concentration

[10] But we need the error to be at least smaller than $1/n$ because error adds up when generalizing from 1 dimension to n dimensions.

property we assumed on $\widehat{s_0}(x)$ and $\widehat{s_1}(x)$ and a good candidate of $\widehat{X}$ emerges: $\widehat{X} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + f_{\mathrm{dw}}$ where f_{dw} is a double square well potential parameterized by $\ell, w, s > 0$ which is defined as follows.

$$f_{\mathrm{dw}}(x) \coloneqq \begin{cases} +\infty, & x \in (\ell + w/2, +\infty) \\ 0, & x \in [w/2, \ell + w/2] \\ s, & x \in [0, w/2) \\ f_{\mathrm{dw}}(-x), & x < 0 \end{cases}$$

As we can see in Figure 5.1, if s is sufficiently large, the first two eigenfunctions of $\widehat{X} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + f_{\mathrm{dw}}$ are approximately $\widehat{s_0}(x) \pm \widehat{s_1}(x)$ for $\widehat{s_0}(x)$ being a partial sine wave in $x > 0$ and $\widehat{s_1}(x) \coloneqq \widehat{s_0}(-x)$.

It turns out that a cleaner approach is to *define* $|\widehat{s_0}\rangle$ and $|\widehat{s_1}\rangle$ from $\widehat{X}$: Fix $\widehat{X} \coloneqq -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + f_{\mathrm{dw}}$ with the ground state $|\psi_0\rangle$ and the first excited state $|\psi_1\rangle$. Let $|\widehat{s_0}\rangle \coloneqq \frac{1}{\sqrt{2}}(|\psi_0\rangle + |\psi_1\rangle)$ and $|\widehat{s_1}\rangle \coloneqq \frac{1}{\sqrt{2}}(|\psi_0\rangle - |\psi_1\rangle)$. We now have that $\widehat{X}$ acting on $\{|\widehat{s_0}\rangle, |\widehat{s_1}\rangle\}$ exactly resembles $\frac{\delta}{2}X + c_2 I$ acting on $\{|0\rangle, |1\rangle\}$ for some c_2 where δ is the spectral gap of $\widehat{X}$. The remaining problem is how well $\widehat{Z} \coloneqq \mathrm{sgn}(x)$ resembles Z in the same manner. It is easy to see from Figure 5.1 that increasing s will make $|\widehat{s_0}\rangle$ concentrate more in the region $x > 0$ and hence the relation $\widehat{Z}|\widehat{s_0}\rangle \approx |\widehat{s_0}\rangle$ is approximated with smaller error. However, this comes with a cost of reducing the spectral gap δ. Note that we want δ to be not too small because our reduction reads $X \mapsto \frac{2}{\delta}\widehat{X}$. An exponentially small δ means the potential function hiding in the Schrödinger operator $\frac{2}{\delta}\widehat{X}$ must be exponentially large, which will make the whole reduction inefficient in the sense that the operator we reduced to involves prohibitively large energy. Therefore, we need to pick an appropriate s such that $\left\| \widehat{Z}|\widehat{s_0}\rangle - |\widehat{s_0}\rangle \right\|$ is small enough but δ is not too small. Studying the relation between δ and s can be regarded as doing semiclassical analysis on the operator $-h^2\Delta + f_{\mathrm{dw}}$ with $s = 1$: this operator is clearly equivalent to $-\Delta + f_{\mathrm{dw}}$ with

$s = 1/h^2$. This is why we will invoke results from semiclassical analysis in Section 5.4.2 for the actual construction of $\widehat{X}$ in Section 5.4.4.

The idea above can be naturally generalized to higher dimensions. Note that we used potential functions that are not continuous in the construction of the reduction $H \mapsto \widehat{H}$ above, and we need smooth potentials in the actual construction of the Schrödinger operator.

5.4.2 One-Dimensional Double-Well Function

The goal of this section is to establish a proof for a technical theorem (Theorem 5.4.23) which will be used later. It characterizes the spectral property of a Schrödinger operator with the potential being a one-dimensional double-well function.

Definition 5.4.6 (Sobolev Space)**.** Let M be a finite-dimensional compact connected Riemannian manifold or an open subset of such a manifold.[11] Define $H_0^2(M) \coloneqq W_0^{2,2}(M)$ as the Sobolev space regarding L^2-norm and 2nd order derivatives, which consists of functions $f : M \to \mathbb{R}$ vanishing on the boundary ∂M.

Operator of Interest. Throughout this section, we focus on analyzing the property of Schrödinger operator $P \coloneqq -h^2\Delta + V$ when $h \to 0$. This is called *semi-classical analysis* in literature. Here, we are only interested in the case that V is a fixed *double-well function* defined as follows.

Definition 5.4.7 (Double-Well Function)**.** We call $V : [-1, 1] \to \mathbb{R}$ a *double-well function* if it satisfies the following:

1. V is symmetric, i.e., $V(-x) = V(x)$ for any x.

[11] Most M encountered later in this section are simply $[-1, 1]$ and its open subsets. The general definition is put here for the sake of completeness.

2. V is smooth on $[-1, 1]$.

3. There exists some $a \in (0, 1)$ such that V has exactly two minima located at a and $-a$ with $V(a) = V(-a) = 0$ and $V''(a) = V''(-a) > 0$.

We restrict the domain of P to be functions in $H_0^2([-1, 1])$. All kets introduced later in this section are *real*-valued functions. For why this works, note that e.g. P is a real differential operator and hence all its eigenfunctions can be made real.

The so-called *Agmon distance* is essential in semi-classical analysis:

Definition 5.4.8 (Agmon Distance)**.** Let M be a finite-dimensional compact connected Riemannian manifold. Let $V : M \to \mathbb{R}$ be a potential function and $E \in \mathbb{R}$ be some specific energy of interest. The *Agmon distance* between $x, y \in M$ is

$$d_{V,E}(x, y) \coloneqq \inf_{\gamma} \int_{\gamma} \sqrt{\max\{V(z) - E, 0\}} \mathrm{d}z,$$

where γ runs over all piecewise C^1 paths connecting x and y. The Agmon distance between a point $x \in M$ and a subset $Y \in M$ follows naturally:

$$d_{V,E}(x, U) \coloneqq \inf_{y \in Y} d_{V,E}(x, y).$$

We define $S_0 \coloneqq d_{V,0}(a, -a)$ as it will make appearances throughout. Here V is the fixed double-well function and a and $-a$ are two minima of V (Refer to Definition 5.4.7).

The notion of *Dirichlet realization* will be helpful when we study local behavior of a Schrödinger operator:

Definition 5.4.9 (Dirichlet Realization)**.** Let $P \coloneqq -h^2\Delta + V$ be a Schrödinger operator with $V : M \to \mathbb{R}$, where M is a finite-dimensional compact connected Riemannian manifold. Let $\Omega \subseteq M$ be a bounded open set. P_Ω is defined as the restriction of P over Ω. To be more specific, P has domain $H_0^2(\Omega)$.

Definition 5.4.10 (Ball)**.** Let M be a Riemannian manifold. Let $d(\cdot,\cdot)$ be a distance function on M. Define $B_d(x,r) \coloneqq \{z \in M \mid d(z,x) \le r\}$ and $\mathring{B}_d(x,r) \coloneqq \{z \in M \mid d(z,x) < r\}$.

Recall that our target is to analyze the behavior of P when $h \to 0$. As such, the kth eigenvalue of P tends to 0 for any fixed k. Therefore we need the following notion to keep track of those eigenvalues we are interested in when $h \to 0$.

Definition 5.4.11 (Eigenvalue Interval, $I(h)$)**.** $I(h)$ is defined as a closed interval in $\mathbb{R}$ that depends on the value of h. We require that $I(h) \to \{0\}$ when $h \to 0$.

Note that we have not yet specified a concrete form of $I(h)$ but only its general properties. In the following, precise definitions of $I(h)$ will be given prior to its use; otherwise the stated results should work for any $I(h)$ satisfying Definition 5.4.11.

Definition 5.4.12 (Local Eigenvalues and Eigenstates)**.** Fix a small $\eta > 0$. Define intervals $L, R \subseteq [-1,1]$ as follows:

$$L \coloneqq \mathring{B}_d(-a, 3S_0) - B_d(a,\eta) - [a, +\infty]$$

$$R \coloneqq \mathring{B}_d(a, 3S_0) - B_d(-a,\eta) - [-\infty, -a]$$

where $d \coloneqq d_{V,0}$. Let $J = L$ or $J = R$. Denote by $\mu_{J,1}, \mu_{J,2}, \ldots, \mu_{J,m_J}$ eigenvalues[12] of P_J in interval

[12] Multiplicity is included here, i.e., degenerate eigenvalues appear multiple times in the list.

$I(h)$ and by $|\phi_{J,1}\rangle, |\phi_{J,2}\rangle, \ldots, |\phi_{J,m_J}\rangle$ their corresponding normalized eigenstates that constitute an orthogonal system. Note that P_L is just the reflection of P_R across the y-axis.

Definition 5.4.13 (Hidden constants in big O notation)**.** The subscript(s) c in the notation $\mathcal{O}_c(\cdot)$ means that the hidden constant depends on c.

Lemma 5.4.14. *Let $P \coloneqq -h^2\Delta + V$ be a Schrödinger operator with some double-well function V. Let $\mathcal{F}$ be the subspace spanned by eigenstates of P whose eigenvalues are in $I(h)$. The following holds for sufficiently small $h > 0$ if there exists $r(h) \geq \Omega(e^{-o(1/h)})$ such that P_L and P_R have no eigenvalues $r(h)$-close to $I(h)$ from outside, i.e., no eigenvalues in $(I(h) + B_d(0, r(h))) - I(h)$:*

(i) There exists a bijection $\mathfrak{b}$ mapping eigenvalues of P in $I(h)$ to eigenvalues of P_L and P_R in $I(h)$. Furthermore, $|\mathfrak{b}(\lambda) - \lambda| \leq \mathcal{O}_\eta\left(e^{-\frac{S_0 - 2\eta - o(1)}{h}}\right)$.

(ii) For each eigenstate $|\phi_{J,k}\rangle$ ($J \in \{L, R\}$), there exists a $|v_{J,k}\rangle \in \mathcal{F}$ such that $\||v_{J,k}\rangle - |\phi_{J,k}\rangle\| \leq \mathcal{O}_\eta\left(e^{-\frac{S_0 - \varepsilon(\eta)}{h}}\right)$ where $\varepsilon(\eta)$ is some function such that $\lim_{\eta \to 0} \varepsilon(\eta) = 0$. $v_{L,k}(\cdot)$ is $v_{R,k}(\cdot)$ reflected across the y-axis. Furthermore, $\{|v_{J,k}\rangle\}$ are almost orthogonal to each other in the sense that $|\langle v_{J,k}|v_{J',k'}\rangle| \leq \mathcal{O}_\eta\left(e^{-\frac{S_0 - \varepsilon(\eta)}{h}}\right)$ for any $(J,k) \neq (J',k')$.

Remark 5.4.15. Lemma 5.4.14 is a special case of Proposition A.3 and the following discussion in Liu et al. [76]. Refer to [201, Section 6.c] and [202, Section 4.2] for more complete treatments. To be more specific, first, it is easy to verify that V satisfies Assumption A.3 and A.4 in [76]. Therefore the first part of Lemma 5.4.14 directly follows from Proposition A.3. Note that our assumption in Lemma 5.4.14 does not ask for spectral property of P like Proposition A.3 does in [76]: the assumption is unnecessary as pointed out in Remark 4.2.3 of [202]. Readers may have noticed that the definition of R (similar for L) in [76] should be $R \coloneqq \mathring{B}_d(a, 3S_0) - B_d(-a, \eta)$, which is a disjoint union of two intervals

(for small enough η) while ours discards the interval on which $V(x)$ has a nonzero lower bound that depends on η. This difference can be justified by that we are interested in the case when $h \to 0$ throughout: P restricted on the other interval cannot have nontrivial eigenfunction with eigenvalue in $I(h)$ since $I(h) \to \{0\}$ and $V(x)$ on that interval is bounded below by a positive value. For the second part of Lemma 5.4.14, the orthogonality comes from Eq. (189) and (190) of [76], while the bound of $\| |v_{J,k}\rangle - |\phi_{J,k}\rangle \|$ can be derived from Eq. (187) of [76] and the decay of ϕ_α, ψ_α given in p. 59 of [201].

Lemma 5.4.16 ([201, Proposition 6.4]). *Let M be a finite-dimensional compact connected Riemannian manifold. Fix an smooth $V : M \to \mathbb{R}$. Define the energy well $U := \{\boldsymbol{x} \mid V(\boldsymbol{x}) \le 0\}$. Let $h \to 0$ and assume that $u = u(\boldsymbol{x}; h)$ is an eigenfunction of $P := -h^2\Delta + V$ with eigenvalue $o(1)$. Then,*

$$\left\| e^{\frac{d_{V,E}(\boldsymbol{x},U)}{h}} u \right\| := \sqrt{\int_M \left| e^{\frac{d_{V,E}(\boldsymbol{x},U)}{h}} u(\boldsymbol{x}) \right|^2 \mathrm{d}\boldsymbol{x}} = \mathcal{O}\left(e^{\frac{o(1)}{h}}\right).^{13}$$

Lemma 5.4.17 ([56, Theorem 11.3]). *Fix $H(\lambda) := -\Delta + \lambda^2 V$ with $V : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following:*

(i) $V \in C^3(\mathbb{R}^n)$; $V \ge 0$; $\liminf_{|\boldsymbol{x}|\to\infty} V(\boldsymbol{x}) = +\infty$.

(ii) V has a single, nondegenerate zero at $\boldsymbol{x} = \mathbf{0}$, i.e., the Hessian of V at $\mathbf{0}$, denoted by A, is positive semidefinite.

Let $K := -\Delta + \frac{1}{2}\boldsymbol{x}^\intercal A \boldsymbol{x}$. We list eigenvalues of K, including multiplicity, as $e_0 \le e_1 \le e_2 \le \cdots$. Similarly, label eigenvalues of $H(\lambda)$ as $e_0(\lambda) \le e_1(\lambda) \le e_2(\lambda) \le \cdots$. Then, for any k we have

$$e_k(\lambda) = \lambda e_k + \mathcal{O}_k\left(\lambda^{\frac{4}{5}}\right), \text{ as } \lambda \to +\infty.$$

[13] The orginal theorem statement in [201] is that for every ε, one have for small enough $h > 0$ depending on ε such that the left-hand side is no more than $C_\varepsilon e^{\varepsilon/h}$ where C_ε is a constant that depends on ε. This clearly induces the statement in Lemma 5.4.16 by letting $\varepsilon \to 0$.

Corollary 5.4.18. *Under the same assumption in Lemma 5.4.17 with $H(h) \coloneqq -h^2\Delta + V$, we have*

$$e_k(h) = he_k + \mathcal{O}_k\left(h^{\frac{6}{5}}\right), \text{ as } h \to 0.$$

Lemma 5.4.19 ([203]). *Fix $H(\lambda) \coloneqq -\Delta + \lambda^2 V$ with $V : \mathbb{R} \to \mathbb{R}$ that satisfies the following:*

(i) V is a nonnegative continuous function such that $V(x)$ takes its absolute minimum zero at some $x = a$ and $x = b$ for some $a, b \in \mathbb{R}$.

(ii) $\liminf_{|\boldsymbol{x}|\to\infty} V(x) > 0$.

Denote by $e_{n-1}(k)$ the nth smallest eigenvalue of $H(\lambda)$. Fix an arbitrary n, we have

$$\liminf_{k\to\infty} \frac{1}{k} \ln(e_n(k) - e_{n-1}(k)) \geq -\int_a^b \sqrt{V(x)}\mathrm{d}x.$$

Lemma 5.4.20 ([56, Theorem 12.3]). *Fix $H(\lambda) \coloneqq -\Delta + \lambda^2 V$ with $V : \mathbb{R} \to \mathbb{R}$ that satisfies the following:*

(i) V is symmetric, i.e., $V(x) = V(-x)$.[14]

(ii) $V \in C(\mathbb{R}^n)$; $V \geq 0$; $\liminf_{|x|\to\infty} V(x) = +\infty$.

(iii) $V(x)$ has two nondegenerate minima at some $x = -a$ and $x = a$ for some $a \in \mathbb{R}$.

Denote by $e_{n-1}(k)$ the nth smallest eigenvalue of $H(\lambda)$. Fix an arbitrary n, we have

$$\limsup_{k\to\infty} \frac{1}{k} \ln(e_1(k) - e_0(k)) \leq -\int_a^b \sqrt{V(x)}\mathrm{d}x.$$

[14] This induces Assumption (A3) in the description of Theorem 12.3 in [56].

Combining Lemma 5.4.19 and 5.4.20 we immediately have the following:

Corollary 5.4.21. *Under the same assumption in Lemma 5.4.20 with $H(h) \coloneqq -h^2\Delta + V$, we have*

$$e_1(k) - e_0(k) \subseteq \Omega\left(e^{-\frac{S_0+o(1)}{h}}\right) \cap \mathcal{O}\left(e^{-\frac{S_0-o(1)}{h}}\right), \quad S_0 \coloneqq d_{V,0}(a,b).$$

Remark 5.4.22. Corollary 5.4.18 (resp., Corollary 5.4.21) can be extended to the case where V is defined only on a compact or a bounded open subset $M \subseteq \mathbb{R}^n$ (resp., $M \subseteq \mathbb{R}$) which is connected: in such a case we require eigenfunctions to vanish on the boundary of M and we no longer need any assumption for $\liminf_{|x|\to\infty} V(x)$. To understand this, it is important to note that the spectrum will be unchanged if we extend V to $\widehat{V}$, where

$$\widehat{V}(\boldsymbol{x}) = \begin{cases} V(\boldsymbol{x}), & x \in M; \\ +\infty, & x \notin M. \end{cases}$$

Now that $\widehat{V}$ is defined on $\mathbb{R}^n$, the only obstacle for applying Corollary 5.4.18 is that $\widehat{V}$ is not continuous, let alone smooth. This can be tackled by a standard method that constructs a smooth function that is sufficiently close to $\widehat{V}$.

Theorem 5.4.23. *Let $E_0 \le E_1 \le E_2 \le \dots$ be eigenvalues of $P \coloneqq -h^2\Delta + V$ with a double-well function V and $\{|\psi_k\rangle\}_{k\in\mathbb{N}}$ their corresponding eigenstates.*

(i) $E_1 - E_0 \subseteq \Omega\left(e^{-\frac{S_0+o(1)}{h}}\right) \cap \mathcal{O}\left(e^{-\frac{S_0-o(1)}{h}}\right)$.

(ii) $E_2 - E_1 \ge \Omega(h)$.

(iii) There exists a $c \in \{\pm 1\}$ such that $|\widehat{s_0}\rangle \coloneqq \frac{1}{\sqrt{2}}\left(|\psi_0\rangle + c\,|\psi_1\rangle\right)$ is bounded away from $x < a$. To

be more specific, the following inequality holds:

$$\forall x_0 \in (-a, a), \quad \int_{-1}^{x_0} |\widehat{s_0}(x)|^2 \mathrm{d}x \leq \mathcal{O}\left(e^{-(2-o(1))\frac{d(a,x_0)}{h}}\right),$$

where $d \coloneqq d_{V,0}$. *A similar inequality holds for* $|\widehat{s_1}\rangle \coloneqq \frac{1}{\sqrt{2}}(|\psi_0\rangle - c\,|\psi_1\rangle)$ *since it is just* $|\widehat{s_0}\rangle$ *reflected across the y-axis.*

Proof of Theorem 5.4.23. We prove (i), (ii) and (iii) in order.

(i). This comes from directly applying Corollary Corollary 5.4.21 with Remark 5.4.22.

(ii). Recall that P_L is just the reflection of P_R across the y-axis. By Corollary 5.4.18 and Remark 5.4.22, the eigenvalues of P_L and P_R are

$$e_k(h) = (2k+1)qh + \mathcal{O}_k\left(h^{6/5}\right), \quad k = 0, 1, 2, \ldots,$$

where $q \coloneqq V''(-a) = V''(a)$. Therefore, by setting interval $I(h)$ to be $[\frac{1}{2}e_0(h), \frac{1}{2}(e_1(h)+e_2(h))]$ it satisfies the condition in Lemma 5.4.14. Lemma 5.4.14(i) tells us $E_{2k+1/2\pm1/2} = e_k(h) \pm \mathcal{O}_\eta\left(e^{-\frac{S_0-2\eta-o(1)}{h}}\right)$. Therefore, (ii) is obtained by picking a sufficiently small η.

(iii). Similar to what we have done for (ii), applying Lemma 5.4.14 with $I(h) = [\frac{1}{2}e_0(h), \frac{1}{2}(e_0(h) + e_1(h))]$ we obtain a symmetric pair $|v_J\rangle \coloneqq |v_{J,0}\rangle$ for $J \in \{L, R\}$ from symmetric pair $|\phi_J\rangle \coloneqq |\phi_{J,0}\rangle$. Note that $|v_J\rangle \in \mathcal{F} = \mathrm{span}(|\psi_0\rangle, |\psi_1\rangle)$. By varying η from Lemma 5.4.14(ii) we have the following

holds for any $\varepsilon > 0$:

$$\begin{aligned} \|v_J\| \in \|\phi_J\| \pm \|v_J - \phi_J\| \in 1 \pm \mathcal{O}_\varepsilon\left(e^{-\frac{S_0-\varepsilon}{h}}\right), \\ |\langle v_L | v_R \rangle| \leq \mathcal{O}_\varepsilon\left(e^{-\frac{S_0-\varepsilon}{h}}\right). \end{aligned} \tag{5.1}$$

For convenience we use notation $\mathcal{O}(r)$ to represent $\mathcal{O}_\varepsilon\left(e^{-\frac{S_0-\varepsilon}{h}}\right)$ below. Now, due to the symmetry between $|v_L\rangle$ and $|v_R\rangle$ we can assume for some $\alpha, \beta \in \mathbb{R}$ that

$$\begin{aligned} |v_L\rangle &= \alpha\,|\psi_0\rangle - \beta\,|\psi_1\rangle\,, \\ |v_R\rangle &= \alpha\,|\psi_0\rangle + \beta\,|\psi_1\rangle\,, \end{aligned} \tag{5.2}$$

because $\psi_0(\cdot)$ is symmetric and $\psi_1(\cdot)$ antisymmetric regarding $x = 0$. Plugging Eq. 5.2 into Eq. 5.1 we get $\alpha^2 + \beta^2 \in 1 \pm \mathcal{O}(r)$ and $\alpha^2 - \beta^2 \in \pm\mathcal{O}(r)$. Therefore,

$$|\alpha|, |\beta| \in \frac{1}{\sqrt{2}} \pm \mathcal{O}(r)\,.$$

Without loss of generality let $\alpha \in \frac{1}{\sqrt{2}} \pm \mathcal{O}(r)$. Therefore $|v_R\rangle = \alpha\,|\psi_0\rangle + c|\beta|\,|\psi_1\rangle$ for some $c \in \{\pm 1\}$, which implies $|\widehat{s_0}\rangle - |v_R\rangle \in \pm\mathcal{O}(r)\,|\psi_0\rangle \pm \mathcal{O}(r)\,|\psi_1\rangle$.

Claim 5.4.24. $\||\widehat{s_0}\rangle - |v_R\rangle\| = \mathcal{O}(r)$.

The next claim is immediate from Lemma 5.4.14(ii).

Claim 5.4.25. $\||v_R\rangle - |\phi_R\rangle\| = \mathcal{O}(r)$.

Now that $|\phi_R\rangle$ is clearly a good estimate of $|\widehat{s_0}\rangle$. Our plan is to show the inequality in (iii) with $|\widehat{s_0}\rangle$ replaced by $|\phi_R\rangle$. Recall that $|\phi_R\rangle$ is the ground state of P_R. Note that we are able to apply Lemma 5.4.16 on P_R despite R not being a finite-dimensional compact Riemannian manifold: It does the work by naturally extending the domain of V in Lemma 5.4.16 from R to $\overline{R}$ and nothing

will change effectively. Set $U = \{a\}$ in Lemma 5.4.16. The eigenvalue for $|\phi_R\rangle$ is $\Theta(h) \leq o(1)$ as discussed in the proof of (ii). Set $d \coloneqq d_{V,0}$ as usual. Extend the definition of $\phi_R(x)$ by letting $\phi_R(x) = 0$ for $x \notin R$. We have

$$\begin{aligned}
\int_{-1}^{x_0} |\phi_R(x)|^2 \mathrm{d}x &\leq e^{-2\frac{d(a,x_0)}{h}} \int_{-1}^{x_0} \left| e^{\frac{d(a,x)}{h}} \phi_R(x) \right|^2 \mathrm{d}x && (d(a,x) \geq d(a,x_0)) \\
&\leq e^{-2\frac{d(a,x_0)}{h}} \int_{R} \left| e^{\frac{d(a,x)}{h}} \phi_R(x) \right|^2 \mathrm{d}x \\
&\leq e^{-2\frac{d(a,x_0)}{h}} \mathcal{O}\left(e^{\frac{o(1)}{h}} \right) && \text{(Lemma 5.4.16)} \\
&\leq \mathcal{O}\left(e^{-2(1-o(1))\frac{d(a,x_0)}{h}} \right).
\end{aligned}$$

Combining the above with Claim 5.4.24 and 5.4.25 we obtain

$$\sqrt{\int_{-1}^{x_0} |\widehat{s_0}(x)|^2 \mathrm{d}x} \leq \sqrt{\int_{-1}^{x_0} |\phi_R(x)|^2} + \mathcal{O}(r) \leq \mathcal{O}\left(e^{-(1-o(1))\frac{d(a,x_0)}{h}} \right) + \mathcal{O}_\varepsilon\left(e^{-\frac{S_0-\varepsilon}{h}} \right)$$

from triangle inequality

$$\sqrt{\int_{-1}^{x_0} |\widehat{s_0}(x)|^2 \mathrm{d}x} - \sqrt{\int_{-1}^{x_0} |\phi_R(x)|^2} \leq \sqrt{\int_{-1}^{x_0} |\widehat{s_0}(x) - \phi_R(x)|^2 \mathrm{d}x} \leq \|\widehat{s_0}(x) - \phi_R(x)\| \leq \mathcal{O}(r).$$

Note that the latter term $\mathcal{O}_\varepsilon\left(e^{-\frac{S_0-\varepsilon}{h}} \right)$ can be absorbed into the former term by choosing sufficiently small ε, since $d(a,x_0) < d(a,-a) = S_0$ from i) $x_0 \in (-a,a)$ and that ii) $\widehat{V}(x) > 0$ for any $x \in \mathbb{R} - \{a\}$. □

5.4.3 Perturbation Theory

We first introduce an alternative way to identify eigenvalues for operators that may not have purely discrete spectrum.

Definition 5.4.26. Let H be a self-adjoint operator that is bounded from below, i.e., $H \succeq cI$ for some c. For $k \in \mathbb{N}$, define

$$\lambda_k(H) \coloneqq \sup_{|\phi_0\rangle,\ldots,|\phi_{k-1}\rangle \in \mathrm{dom}(H)} \inf_{\substack{|\psi\rangle \in \mathrm{dom}(H) \\ \||\psi\rangle\| = 1 \\ |\psi\rangle \perp |\phi_0\rangle,\ldots,|\phi_{k-1}\rangle}} \langle\psi|H|\psi\rangle.$$

For notational convenience, We also define $\lambda_{\min}(H) \coloneqq \lambda_0(H)$ and $\lambda_{\max}(H) \coloneqq \lambda_{m-1}(H)$ where $m \coloneqq \dim(H)$.

The $\lambda_k(H)$ above are basically eigenvalues due to the following lemma:

Lemma 5.4.27 (Min-Max Principle [152, Theorem XIII.1])**.** *$\lambda_k(H)$ is the $(k+1)$th smallest eigenvalue counting multiplicity if it is below the bottom of the essential spectrum of H. Otherwise, $\lambda_k(H)$ is the bottom of the essential spectrum.*

Lemma 5.4.28 (Weyl's Inequality)**.** *Let A and B be self-adjoint, semi-bounded operators of the same domain. Then, for any k we have*

$$|\lambda_k(A) - \lambda_k(B)| \leq \|A - B\|.$$

Proof. By Definition 5.4.26,

$$\begin{aligned}
\lambda_k(A) &= \sup_{|\phi_0\rangle,\ldots,|\phi_{k-1}\rangle\in\mathrm{dom}(A)} \inf_{\substack{|\psi\rangle\in\mathrm{dom}(A)\\ \||\psi\rangle\|=1\\ |\psi\rangle\perp|\phi_0\rangle,\ldots,|\phi_{k-1}\rangle}} \langle\psi|A|\psi\rangle \\
&= \sup\inf\left(\langle\psi|B|\psi\rangle + \langle\psi|(A-B)|\psi\rangle\right) \\
&\geq \sup\left(\inf\langle\psi|B|\psi\rangle + \inf\langle\psi|(A-B)|\psi\rangle\right) \\
&\geq \sup\left(\inf\langle\psi|B|\psi\rangle - \|A-B\|\right) \\
&= \lambda_k(B) - \|A-B\|.
\end{aligned}$$

Switching the role of A and B in the above we get $\lambda_k(B) \geq \lambda_k(A) - \|B-A\|$, which concludes the proof. □

Lemma 5.4.29. *Let H be a (possibly infinite-dimensional) Hamiltonian written as $H = \left(\begin{smallmatrix} A & R^\dagger \\ R & B \end{smallmatrix}\right)$ w.r.t. subspaces $\mathcal{S}$ and $\mathcal{S}^\perp$ such that $m := \dim(\mathcal{S}) < \infty$. Suppose $\Delta := \lambda_{\min}(B) - \lambda_{\max}(A) > 0$. Then, for any $k \in \{0, 1, \ldots, m-1\}$,*

$$|\lambda_k(H) - \lambda_k(A)| \leq \|R\|.$$

Proof. Define $H_0 := H - \left(\begin{smallmatrix} 0 & R^\dagger \\ R & 0 \end{smallmatrix}\right)$. Applying Lemma 5.4.28 we have

$$|\lambda_k(H) - \lambda_k(H_0)| \leq \left\|\left(\begin{smallmatrix} 0 & R^\dagger \\ R & 0 \end{smallmatrix}\right)\right\| = \|R\|$$

for any $k \in \mathbb{N}$. It remains to show that $\lambda_k(H_0) = \lambda_k(A)$ for $k \in \{0, 1, \ldots, m-1\}$: Note that the spectrum of $H_0 = \left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$ is the union of those of A and B. And the assumption $\Delta > 0$ guarantees that the whole spectrum of B is above that of A. Hence $\lambda_k(H_0) = \lambda_k(A)$ for $k \in \{0, 1, \ldots, m-1\}$. □

Lemma 5.4.30. *Let A and B be two self-adjoint operators. Then,*

$$\left\|e^{-iAt} - e^{-iBt}\right\| \leq \|A - B\|t.$$

Proof. By Duhamel's Principle or differentiating $e^{-i(A-B)t}$ we have

$$e^{-iAt} - e^{-iBt} = -i \int_0^t e^{-iA(t-s)}(A - B)e^{-iBs}\mathrm{d}s.$$

The inequality follows by taking norms of both sides. □

The following lemma serves as a substitute for Schrieffer–Wolff theory to control the behavior of e^{-iHt} when H is approximately block-diagonal.

Lemma 5.4.31. *Let H be a (possibly infinite-dimensional) Hamiltonian written as $H = \left(\begin{smallmatrix} A & R^\dagger \\ R & B \end{smallmatrix}\right)$ w.r.t. subspaces $\mathcal{S}$ and $\mathcal{S}^\perp$. For any $t > 0$ we have*

$$\left\|P_{\mathcal{S}} e^{-iHt} P_{\mathcal{S}} - e^{-iAt}\right\| \leq \frac{2\sqrt{2}}{3} \left(\|R\|t\right)^{3/2}.$$

Proof. The proof largely follows that of Theorem 10 in [63] with the exception that we cannot use the vanilla Schrieffer–Wolff theory to prove Claim 5.4.32 because our case involves unbounded (and hence infinite-dimensional) operators.

Let $H_0 = P_{\mathcal{S}} H P_{\mathcal{S}} = \mathrm{diag}(A, \mathbf{0})$, and we define:

$$\mathscr{A}(t) = P_{\mathcal{S}} e^{-iHt} P_{\mathcal{S}}, \quad \mathscr{B}(t) = P_{\mathcal{S}} e^{-iH_0 t} P_{\mathcal{S}}.$$

We also define the $\mathscr{E}(t) \coloneqq \mathscr{A}(t) - \mathscr{B}(t)$. Note that $\mathscr{E}(0) = 0$. We find that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{A}(t) = P_{\mathcal{S}}\left(-iH\right)\left(P_{\mathcal{S}} + P_{\mathcal{S}^\perp}\right)e^{-iHt}P_{\mathcal{S}} = -iH_0\mathscr{A}(t) - iRe^{-iHt}P_{\mathcal{S}},$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{B}(t) = P_{\mathcal{S}}\left(-iH_0\right)e^{-iH_0t}P_{\mathcal{S}} = -iH_0\mathscr{B}(t).$$

Therefore, we have $\dot{\mathscr{E}}(t) = \dot{\mathscr{A}}(t) - \dot{\mathscr{B}}(t) = -iH_0\mathscr{E}(t) - iRe^{-iHt}P_{\mathcal{S}}$. By Duhamel's principle we represent the function $\mathscr{E}(t)$ as the time integral:

$$\mathscr{E}(t) = -i\int_0^t \exp(-iH_0(t-\tau))R\exp(-iH\tau)P_{\mathcal{S}}\,\mathrm{d}\tau,$$

so we can estimate the error (recall that $R = P_{\mathcal{S}}HP_{\mathcal{S}^\perp}$):

$$\begin{aligned}\|\mathscr{E}(t)\| &\leq \int_0^t \|P_{\mathcal{S}}HP_{\mathcal{S}^\perp}\exp(-iH\tau)P_{\mathcal{S}}\|\,\mathrm{d}\tau \leq \int_0^t \|RP_{\mathcal{S}^\perp}\exp(-iH\tau)P_{\mathcal{S}}\|\,\mathrm{d}\tau \\ &\leq \int_0^t \|R\|\left(\sqrt{2\|R\|\tau}\right)\,\mathrm{d}\tau = \frac{2\sqrt{2}}{3}\left(\|R\|t\right)^{3/2}.\end{aligned}$$

In the second step, we use $P_{\mathcal{S}^\perp} = P_{\mathcal{S}^\perp}^2$. In the second to last step, we invoke Claim 5.4.32.

Claim 5.4.32. $\|P_{\mathcal{S}^\perp}\exp(-iHt)P_{\mathcal{S}}\| \leq \sqrt{2\|R\|t}$.

Claim 5.4.32 is equivalent to a $2\|R\|t$ upper bound of $\|P_{\mathcal{S}^\perp}\exp(-iHt)\left|\psi_0\right\rangle\|^2$ for every normalized state $\left|\psi_0\right\rangle \in \mathcal{S}$. We prove such a bound below. Define $\left|\psi\right\rangle \coloneqq \left|\psi(t)\right\rangle \coloneqq \exp(-iHt)\left|\psi_0\right\rangle$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{\mathcal{S}}\left|\psi\right\rangle = -iP_{\mathcal{S}}H\left|\psi\right\rangle$$

since $\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = -iH|\psi\rangle$. Therefore,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\|P_{\mathcal{S}}|\psi\rangle\|^2 &= 2\,\mathrm{Re}\big\{(-iP_{\mathcal{S}}H|\psi\rangle)^\dagger(P_{\mathcal{S}}|\psi\rangle)\big\} \\
&= -2\,\mathrm{Im}\{\langle\psi|HP_{\mathcal{S}}|\psi\rangle\} \\
&= -2\,\mathrm{Im}\{\langle\psi|P_{\mathcal{S}}HP_{\mathcal{S}}|\psi\rangle + \langle\psi|P_{\mathcal{S}^\perp}HP_{\mathcal{S}}|\psi\rangle\} \\
&= -2\,\mathrm{Im}\big\{\big\langle\psi|R^\dagger|\psi\big\rangle\big\}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\|P_{\mathcal{S}^\perp}\exp(-iHt)|\psi_0\rangle\|^2 &= \|P_{\mathcal{S}^\perp}|\psi(t)\rangle\|^2 \\
&= 1 - \|P_{\mathcal{S}}|\psi(t)\rangle\|^2 \\
&= \|P_{\mathcal{S}}|\psi(0)\rangle\|^2 - \|P_{\mathcal{S}}|\psi(t)\rangle\|^2 \\
&= \int_t^0 \left(\frac{\mathrm{d}}{\mathrm{d}\tau}\|P_{\mathcal{S}}|\psi(\tau)\rangle\|^2\right)\mathrm{d}\tau \\
&= \int_t^0 -2\,\mathrm{Im}\big\{\big\langle\psi(\tau)|R^\dagger|\psi(\tau)\big\rangle\big\}\mathrm{d}\tau \\
&\leq 2\|R\|t.
\end{aligned}$$

□

5.4.4 Proof of the Main Lemma

This section is dedicated to the proof of Lemma 5.4.4.

Lemma 5.4.33 (restated). *Fix $M_1, M_2, \varepsilon_1 > 0$ and $t \geq 0$. For any TIM Hamiltonian*

$$H = \sum_{1 \leq u \leq n} (a_u X_u + b_u Z_u) + \sum_{1 \leq u,v \leq n} b_{u,v} Z_u Z_v$$

such that $|a_u|, |b_u|, |b_{u,v}| \leq M_1$ and $a_u > 1/M_2$, there exists a Schrödinger operator

$$\widehat{H} = -G^\star \Delta + V$$

and a polynomial $P \leq \mathrm{poly}(n, M_1, M_2, 1/\varepsilon_1, t)$ satisfying the following:

(i) $1 \leq G^\star \leq P$ *is* $\mathrm{poly}(n)$*-time computable and the potential* $V : [-1,1]^n \to \mathbb{R}$ *is* 2*-body, smooth,* $\mathrm{poly}(n)$*-time computable and has* C^1*-norm bounded within* P*.*

(ii) $\left|\lambda_k(\widehat{H}) - \lambda_k(H)\right| \leq \varepsilon_1$ *for every* $k \in \{0, 1, \ldots, 2^n - 1\}$*.*

(iii) $\left\|W^\dagger P_{\mathcal{S}} e^{-i\widehat{H}t} W - e^{-iHt}\right\| \leq \varepsilon_1$*, where* $W := \bigotimes_{1 \leq k \leq n} (\left|\psi_0^k\right\rangle \langle -| + \left|\psi_1^k\right\rangle \langle +|)$ *shifts basis and* $\mathcal{S}$ *is the image of* W*.* $\left|\psi_0^k\right\rangle$ *and* $\left|\psi_1^k\right\rangle$ *are* 1*-dimensional. Moreover, wave functions* $\psi_j^k(\cdot)$ *are smooth,* $\mathrm{poly}(n)$*-time computable and have* C^1*-norm bounded within* $\mathrm{polylog}(P)$*.*

1D Building Blocks. Fix $f_{\mathrm{dw}}(x) := (x - 1/2)^2 (x + 1/2)^2$ to be a quartic function with minima $x = \pm x^\star := \pm 1/2$.[15] It is easy to verify that f_{dw} is a double-well function that satisfies the requirement of V in Theorem 5.4.23. Define the operator $\widehat{X}(h)$, which depends on a parameter $h > 0$, to be

$$\widehat{X}(h) := C_h \left(-h^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} + f_{\mathrm{dw}} \right),$$

[15] Other choices of $f_{\mathrm{dw}}(x)$ should work in principle but it is trickier to assert the *quantitative* C^1-boundedness of its eigenfunctions, which will be mentioned soon after.

where C_h is chosen such that the spectral gap of $\widehat{X}(h)$ is 2. Let $G_\circ \coloneqq e^{\frac{S_0}{h}}$ where

$$S_0 \coloneqq d_{f_{\mathrm{dw}},0}(x^\star, -x^\star) = 1/6.$$

Let $E_0 \le E_1 \le E_2 \le \cdots$ be eigenvalues of $\widehat{X}(h)$ and $|\psi_0\rangle, |\psi_1\rangle, |\psi_2\rangle, \ldots$ be corresponding eigenstates given h. It is well known that eigenfunctions of an elliptic operator, of which Schrödinger operator is a special case, is smooth if the operator itself is smooth (See Theorem 3 and 6 of [118, Sec. 6.3]). Therefore $\psi_j(\cdot)$ are smooth since $\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ and f_{dw} are smooth. By analytic methods such as WKB approximation it is straightforward that $\psi_j(\cdot)$ has C^1-norm $h^{-\mathcal{O}(j)} \le \log^{\mathcal{O}(j)}(G_\circ)$ for any j.

Claim 5.4.33. *$C_h \in \mathcal{O}\left(G_\circ^{1+o(1)}\right) \cap \Omega\left(G_\circ^{1-o(1)}\right)$, $E_2 - E_1 \ge \Omega\left(G_\circ^{1-o(1)}\right)$ and $E_1 - E_0 = 2$.*

Proof. Theorem 5.4.23(i) and (ii) give $E_1 - E_0 \in \Omega\left(e^{-\frac{S_0+o(1)}{h}}\right) \cap \mathcal{O}\left(e^{-\frac{S_0-o(1)}{h}}\right)$ and $E_2 - E_1 \ge \Omega(h)$ if $C_h = 1$ in the definition of $\widehat{X}$. Now that C_h is chosen such that $E_1 - E_0 = 2$, therefore $C_h \in \mathcal{O}\left(e^{\frac{S_0+o(1)}{h}}\right) \cap \Omega\left(e^{\frac{S_0-o(1)}{h}}\right)$. Hence $E_2 - E_1 \ge \Omega\left(he^{\frac{S_0-o(1)}{h}}\right)$. We conclude with the definition of $G_\circ$. □

Now, if we write $G = C_h h^2$, it is immediate that $G \in \mathcal{O}\left(G_\circ^{1+o(1)}\right) \cap \Omega\left(G_\circ^{1-o(1)}\right)$ by Claim 5.4.33. This means $h^2 C_h \to +\infty$ when $h \to 0$. Note that the implicit map $h \mapsto C_h$ is continuous. Therefore there always exists an h satisfying $G = C_h h^2$ given any sufficiently large G. We can define $\hat{X}(G) \coloneqq \hat{X}(h)$ by choosing such an h. We also write the $(j+1)$th eigenpair of $\widehat{X}(G)$ as $(E_j^G, |\psi_j^G\rangle)$ if we want to highlight the value of G. Note that

$$\widehat{X}(G) = -G\frac{\mathrm{d}^2}{\mathrm{d}x^2} + C_h f_{\mathrm{dw}}.$$

Let us rewrite Claim 5.4.33 in terms of G:

Claim 5.4.34 (Property of $\widehat{X}$)**.** $C_h \in \mathcal{O}\big(G^{1+o(1)}\big) \cap \Omega\big(G^{1-o(1)}\big)$, $E_2 - E_1 \geq \Omega\big(G^{1-o(1)}\big)$ *and* $E_1 - E_0 = 2$.

Next, define the diagonal operator (in the position basis) $\widehat{Z} \coloneqq \widehat{Z}(w)$, which depends on a parameter $w > 0$, to be

$$\widehat{Z}(w) \coloneqq \mathrm{sgn}_w, \quad \text{i.e.,} \quad (\widehat{Z} \cdot f)(x) \coloneqq \mathrm{sgn}_w(x) f(x).$$

where $\mathrm{sgn}_w : \mathbb{R} \to [-1, 1]$ is a fixed smooth $(2/w)$-Lipschitz function such that $\mathrm{sgn}_w(x) = \mathrm{sgn}(x)$ for $|x| \geq w$:

$$\mathrm{sgn}_w(x) \coloneqq \frac{r\left(\frac{1+x/w}{2}\right) - r\left(\frac{1-x/w}{2}\right)}{r\left(\frac{1+x/w}{2}\right) + r\left(\frac{1-x/w}{2}\right)}, \quad \text{where} \quad r(x) \coloneqq \begin{cases} e^{-1/x} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

We know that the Pauli Z written as a matrix in the eigenbasis of Pauli X is $|+\rangle\langle -| + |-\rangle\langle +|$. Let us show the similar property for $\widehat{Z} \coloneqq \widehat{Z}(w)$ and $\widehat{X}$: Denote P_- as the projector onto the "low-energy" space $\mathcal{T}_- \coloneqq \mathrm{span}(|\psi_0\rangle, |\psi_1\rangle)$ and P_+ as the projector onto the "high-energy" space $\mathcal{T}_+ \coloneqq \mathcal{T}_-^{\perp} = \mathrm{span}(|\psi_2\rangle, |\psi_3\rangle, \dots)$. We will prove that

$$P_- \widehat{Z} P_- \approx |\psi_0\rangle\langle\psi_1| + |\psi_1\rangle\langle\psi_0|, \quad P_+ \widehat{Z} P_- \approx \mathbf{0}.$$

Claim 5.4.35 (Property of $\widehat{Z}$)**.** *There exists a function* $\varepsilon(w)$ *satisfying* $\lim_{w\to 0} \varepsilon(w) = 0$ *such that the following holds:*

(i) $P_- \widehat{Z} P_- = (1 \pm \mathcal{O}(r))(|\psi_0\rangle\langle\psi_1| + |\psi_1\rangle\langle\psi_0|) \pm \mathcal{O}(r)\,|\psi_0\rangle\langle\psi_0| \pm \mathcal{O}(r)\,|\psi_1\rangle\langle\psi_1|$,

(ii) $\left\| P_+ \widehat{Z} P_- \right\| \leq \mathcal{O}(r)$,

(iii) $\left\| \widehat{Z} \right\| = 1$,

where $r \coloneqq \frac{1}{\sqrt{G^{1-\varepsilon(w)}}}$ *and* $|\psi_0\rangle, |\psi_1\rangle$ *are eigenstates of* $\hat{X}(G)$.

Proof. We focus on (i) and (ii) since (iii) is obvious from the definition of $\widehat{Z}$. To begin, recall from Theorem 5.4.23(iii) that there exists a $|\widehat{s_0}\rangle \coloneqq \frac{1}{\sqrt{2}}(|\psi_0\rangle + c\,|\psi_1\rangle)$ where $c \in \{\pm 1\}$ such that it is bounded away from $x < x^\star$. To be more specific, plug $x_0 \coloneqq w$ into Theorem 5.4.23(iii):

$$\int_{-1}^{w} |\widehat{s_0}(x)|^2 \mathrm{d}x \leq \mathcal{O}\left(e^{-(2-o(1))\frac{d(x^\star, w)}{h}}\right) = \mathcal{O}\left(\frac{1}{G^{1-\varepsilon(w)}}\right),$$

where $d \coloneqq d_{f_{\mathrm{dw}},0}$. The last equality above comes from the choice of h in the definition of $\widehat{X}$ such that $e^{\frac{S_0}{h}} = G_\circ$ and $G \in \mathcal{O}\left(G_\circ^{1+o(1)}\right) \cap \Omega\left(G_\circ^{1-o(1)}\right)$, and the fact that i) $d(x^\star, -x^\star) = 2d(x^\star, 0)$ from the symmetry of f_{dw} and ii) $d(x^\star, w) \to d(x^\star, 0)$ when $w \to 0$ since $f_{\mathrm{dw}}(x) > 0$ for any $x > 0$. Therefore,

$$\begin{aligned}
\left\langle \widehat{s_0} | \widehat{Z} | \widehat{s_0} \right\rangle &= \int_{-1}^{1} \mathrm{sgn}_w(x) \widehat{s_0}(x)^2 \\
&= \int_{-1}^{w} \mathrm{sgn}_w(x) \widehat{s_0}(x)^2 + \int_{w}^{1} \mathrm{sgn}_w(x) \widehat{s_0}(x)^2 \\
&\geq \int_{-1}^{w} (-1) \cdot \widehat{s_0}(x)^2 + \int_{w}^{1} \mathrm{sgn}_w(x) \widehat{s_0}(x)^2 \\
&= -\mathcal{O}\left(\frac{1}{G^{1-\varepsilon(w)}}\right) + \int_{w}^{1} \widehat{s_0}(x)^2 \\
&\geq 1 - \mathcal{O}\left(\frac{1}{G^{1-\varepsilon(w)}}\right).
\end{aligned}$$

A similar argument gives us

$$\left\langle \widehat{s_1} | \widehat{Z} | \widehat{s_1} \right\rangle \leq -1 + \mathcal{O}\left(\frac{1}{G^{1-\varepsilon(w)}}\right),$$

where $|\widehat{s_1}\rangle \coloneqq \frac{1}{\sqrt{2}}(|\psi_0\rangle - c\,|\psi_1\rangle)$. On the other hand, note that $\left\|\widehat{Z}\,|\widehat{s_0}\rangle\right\| \le 1$ since $\left\|\widehat{Z}\right\| = 1$. We have

$$\left\langle \widehat{s_1}|\widehat{Z}|\widehat{s_0}\right\rangle = \left\||\widehat{s_1}\rangle\langle\widehat{s_1}|\,\widehat{Z}\,|\widehat{s_0}\rangle\right\| \le \left\|(I - |\widehat{s_0}\rangle\langle\widehat{s_0}|)\widehat{Z}\,|\widehat{s_0}\rangle\right\| \le \sqrt{1 - \left(1 - \mathcal{O}\left(\frac{1}{G^{1-\varepsilon(w)}}\right)\right)^2} = \mathcal{O}\left(\frac{1}{\sqrt{G^{1-\varepsilon(w)}}}\right),$$

where the last "$\le$" follows from the Pythagorean theorem on the triangle consisting of $\widehat{Z}\,|\widehat{s_0}\rangle$ and its projections $|\widehat{s_0}\rangle\langle\widehat{s_0}|\,\widehat{Z}\,|\widehat{s_0}\rangle$ and $(I - |\widehat{s_0}\rangle\langle\widehat{s_0}|)\widehat{Z}\,|\widehat{s_0}\rangle$. The same argument can be used again to obtain the $\mathcal{O}\left(\frac{1}{\sqrt{G^{1-\varepsilon(w)}}}\right)$ upper bound for $\left\|P_+\widehat{Z}\,|\widehat{s_0}\rangle\right\|$ and $\left\|P_+\widehat{Z}\,|\widehat{s_1}\rangle\right\|$. Using the estimate above we conclude the proof. □

Construction of $\widehat{H}$. Let $G^\star$ be a parameter to be determined at the very end. Define $G_u \coloneqq G^\star/a_u$ for $u = 1, \ldots, n$. For notational convenience also define $G_{\min} = \min_{1\le u\le n} G_u$. We have $G_{\min} \ge G_\star/M_1$ by the assumption of coefficients $\{a_u\}$. Recall that

$$H = \sum_{1\le u\le n}(a_u X_u + b_u Z_u) + \sum_{1\le u,v\le n} b_{u,v} Z_u Z_v.$$

We construct the corresponding Schrödinger operator $\widehat{H}\colon H^2(\mathbb{R}^n) \to H^2(\mathbb{R}^n)$ as

$$\widehat{H} \coloneqq \widehat{H_X} + \widehat{H_Z},$$

where

$$\widehat{H_X} \coloneqq \sum_{1\le u\le n} a_u \widehat{X_u}(G_u),$$
$$\widehat{H_Z} \coloneqq \sum_{1\le u\le n} b_u \widehat{Z_u} + \sum_{1\le u,v\le n} b_{u,v}\widehat{Z_u}\widehat{Z_v},$$

in which $\widehat{X_k}(G_k) \coloneqq I^{\otimes k-1} \otimes \widehat{X}(G_k) \otimes I^{\otimes n-k}$ is $\widehat{X}(G_k)$ acting on the kth coordinate; similarly, $\widehat{Z_k}$ is $\widehat{Z}$ acting on the kth coordinate and $\widehat{Z_j}\widehat{Z_k}$ is two $\widehat{Z}$ acting on the jth and the kth coordinates simultaneously. Let P_-^G be the projector onto the space $\mathcal{T}_-^G \coloneqq \operatorname{span}(\left|\psi_0^G\right\rangle, \left|\psi_1^G\right\rangle)$ and P_+^G be the projector onto the space $\mathcal{T}_+^G \coloneqq (\mathcal{T}_-^G)^\perp = \operatorname{span}(\left|\psi_2^G\right\rangle, \left|\psi_3^G\right\rangle, \dots)$, where $\{\left|\psi_k^G\right\rangle\}_{k\in\mathbb{N}}$ are eigenstates of $\widehat{X}(G)$. We omit the superscript "G" when it is clear from the context. The reason why we can relate H and $\widehat{H}$ is that the "low-energy" part of $\widehat{H}$ is basically H. To be more specific, define $\mathcal{S} \coloneqq \bigotimes_{1\le k\le n} \mathcal{T}_-^{G_k}$ as the "low-energy" subspace:

Claim 5.4.36. *Define unitary* $W \coloneqq \bigotimes_{1\le k\le n}(\left|\psi_0^{G_k}\right\rangle \langle -| + \left|\psi_1^{G_k}\right\rangle \langle +|)$ *that shifts* $\{|-\rangle, |+\rangle\}$ *basis to* $\{\left|\psi_0^{G_k}\right\rangle, \left|\psi_1^{G_k}\right\rangle\}$ *basis. Then,*

$$\left\|\widehat{H}_{\mathcal{S}} - W(H + cI)W^\dagger\right\| \le \mathcal{O}\big(n^2 M_1 r\big),$$

where $\widehat{H}_{\mathcal{S}}$ *is* $P_{\mathcal{S}}\widehat{H}P_{\mathcal{S}}$ *restricted on* $\mathcal{S}$*, and* $c \coloneqq (E_0+1)\sum_{1\le u\le n} a_u$ *and* $r \coloneqq \frac{1}{\sqrt{G_{\min}^{1-\varepsilon(w)}}}$ *with* $\varepsilon(\cdot)$ *defined in Claim 5.4.35.*

Proof. We decompose H as $H = H_X + H_Z$ the same way for the decomposition $\widehat{H} \coloneqq \widehat{H_X} + \widehat{H_Z}$. By definition we have $P_-\widehat{X}P_- = E_0 |\psi_0\rangle\langle\psi_0| + E_1 |\psi_1\rangle\langle\psi_1|$. Note that $E_1 - E_0 = 2$ by Claim 5.4.34. Therefore,

$$P_{\mathcal{S}}\widehat{H_X}P_{\mathcal{S}} = W\left(H_X + I(E_0+1)\sum_{1\le u\le n} a_u\right)W^\dagger. \tag{5.3}$$

Similarly, by Claim 5.4.35(i), we have

$$\left\|P_{\mathcal{S}}\widehat{H_Z}P_{\mathcal{S}} - WH_ZW^\dagger\right\| \le \mathcal{O}(r)\left(\sum_{1\le u\le n} b_u + \sum_{1\le u,v\le n} b_{u,v}\right) \le \mathcal{O}\big(n^2 M_1 r\big). \tag{5.4}$$

Combining Eq. 5.3 and Eq. 5.4 we obtain the bound. □

Now our goal is to apply Lemma 5.4.29 on $\widehat{H}$ to argue that the "high-energy" part of $\widehat{H}$ does not matter (much). Before that, we need to estimate relevant parameters $R \coloneqq P_{\mathcal{S}^\perp}\widehat{H}P_{\mathcal{S}}$ and $\Delta \coloneqq \lambda_{\min}(\widehat{H}_{\mathcal{S}^\perp}) - \lambda_{\max}(\widehat{H}_{\mathcal{S}})$, where $\widehat{H}_{\mathcal{S}^\perp}$ is $P_{\mathcal{S}^\perp}\widehat{H}P_{\mathcal{S}^\perp}$ restricted on $\mathcal{S}^\perp$.

Claim 5.4.37 (Estimate of $\|R\|$)**.** *Inherit r from Claim 5.4.36. We have $\left\|P_{\mathcal{S}^\perp}\widehat{X}_k P_{\mathcal{S}}\right\|$, $\left\|P_{\mathcal{S}^\perp}\widehat{Z}_k P_{\mathcal{S}}\right\|$ and $\left\|P_{\mathcal{S}^\perp}(\widehat{Z}_j\widehat{Z}_k)P_{\mathcal{S}}\right\|$ are bounded by $\mathcal{O}(r)$. Consequently, $\|R\| = \left\|P_{\mathcal{S}^\perp}\widehat{H}P_{\mathcal{S}}\right\| \le \mathcal{O}(n^2 M_1 r)$.*

Proof. Without loss of generality assume $k = 1$ and $j = 2$. First, $P_{\mathcal{S}^\perp}\widehat{X}_1 P_{\mathcal{S}} = 0$ because $\mathcal{S}$ is spanned by the tensor of eigenstates of $\widehat{X}$. Next we examine $P_{\mathcal{S}^\perp}\widehat{Z}_1 P_{\mathcal{S}}$: Claim 5.4.35(ii) implies

$$\left\|(P_+ \otimes I^{\otimes n-1})\widehat{Z_1}P_{\mathcal{S}}\right\| \le \left\|(P_+ \otimes I^{\otimes n-1})\widehat{Z_1}(P_- \otimes I^{\otimes n-1})\right\| = \mathcal{O}(r)\,. \tag{5.5}$$

Write $P_{\mathcal{S}^\perp} = P_+ \otimes I^{\otimes n-1} + P_- \otimes (I^{\otimes n-1} - P_-^{\otimes n-1})$ and the only thing left to estimate is $(P_- \otimes (I^{\otimes n-1} - P_-^{\otimes n-1}))\widehat{Z_1}P_{\mathcal{S}}$:

$$\begin{aligned}(P_- \otimes (I^{\otimes n-1} - P_-^{\otimes n-1}))\widehat{Z_1}P_{\mathcal{S}} &= (P_- \otimes (I^{\otimes n-1} - P_-^{\otimes n-1}))(\widehat{Z} \otimes I^{\otimes n-1})(P_- \otimes P_-^{\otimes n-1}) \\ &= (P_-\widehat{Z}P_-) \otimes ((I^{\otimes n-1} - P_-^{\otimes n-1})I^{\otimes n-1}P_-^{\otimes n-1}) \\ &= (P_-\widehat{Z}P_-) \otimes 0 = 0.\end{aligned} \tag{5.6}$$

Therefore $\left\|P_{\mathcal{S}^\perp}\widehat{Z}_1 P_{\mathcal{S}}\right\| \le \mathcal{O}(r)$ from Eq. 5.5 and 5.6. Last we estimate $P_{\mathcal{S}^\perp}(\widehat{Z}_1\widehat{Z}_2)P_{\mathcal{S}}$: Use Claim 5.4.35(ii)

again we have

$$\left\|(P_- \otimes P_- \otimes I^{\otimes n-2})(\widehat{Z_1}\widehat{Z_2})P_{\mathcal{S}}\right\| = \mathcal{O}(r^2)$$

$$\left\|(P_- \otimes P_+ \otimes I^{\otimes n-2})(\widehat{Z_1}\widehat{Z_2})P_{\mathcal{S}}\right\| = \mathcal{O}(r)$$

$$\left\|(P_+ \otimes P_- \otimes I^{\otimes n-2})(\widehat{Z_1}\widehat{Z_2})P_{\mathcal{S}}\right\| = \mathcal{O}(r)$$

Write $P_{\mathcal{S}^\perp} = (P_- \otimes P_- + P_- \otimes P_+ + (P_+ \otimes P_-) \otimes I^{\otimes n-2} + P_- \otimes P_- \otimes (I^{\otimes n-2} - P_-^{\otimes n-2})$. Similar to the argument in Eq. 5.6 we have

$$(P_- \otimes P_- \otimes (I^{\otimes n-2} - P_-^{\otimes n-2}))(\widehat{Z_1}\widehat{Z_2})P_{\mathcal{S}} = 0.$$

□

Claim 5.4.38 (Estimate of Δ)**.** $\Delta \coloneqq \lambda_{\min}(\widehat{H}_{\mathcal{S}^\perp}) - \lambda_{\max}(\widehat{H}_{\mathcal{S}}) \geq \Omega\left(\frac{G_{\min}^{1-o(1)}}{M_2}\right)$ *if there exists a* $\delta > 0$ *such that* $G_{\min} \geq (n^2 M_1 M_2)^{1+\delta}$ *always hold.*

Proof. Recall that $\mathcal{S} \coloneqq \bigotimes_{1\leq k\leq n} \mathcal{T}_-^{G_k}$ and $\mathcal{T}_-^G = \mathrm{span}(\left|\psi_0^G\right\rangle, \left|\psi_1^G\right\rangle)$. Without loss of generality let a_1 be the smallest among a_u. We immediately get

$$\lambda_{\min}(P_{\mathcal{S}^\perp}\widehat{H_X}P_{\mathcal{S}^\perp}) = a_1 E_2 + (a_2 + \cdots + a_n)E_0,$$

$$\lambda_{\max}(P_{\mathcal{S}}\widehat{H_X}P_{\mathcal{S}}) = (a_1 + \cdots + a_n)E_1.$$

Note that $\widehat{H_Z}$ is bounded since $\left\|\widehat{Z}_u\right\| = \left\|\widehat{Z}_u\widehat{Z}_v\right\| = 1$:

$$\left\|\widehat{H_Z}\right\| \leq \sum_{1\leq u\leq n} |b_u| + \sum_{1\leq u,v\leq n} |b_{u,v}|.$$

Therefore by Weyl's inequality (Lemma 5.4.28),

$$\begin{aligned}
&\lambda_{\min}(P_{\mathcal{S}^\perp}\widehat{H}P_{\mathcal{S}^\perp}) - \lambda_{\max}(P_{\mathcal{S}}\widehat{H}P_{\mathcal{S}}) \\
&\geq \left(\lambda_{\min}(P_{\mathcal{S}^\perp}\widehat{H_X}P_{\mathcal{S}^\perp}) - \left\|P_{\mathcal{S}^\perp}\widehat{H_Z}P_{\mathcal{S}^\perp}\right\|\right) - \left(\lambda_{\max}(P_{\mathcal{S}}\widehat{H}P_{\mathcal{S}}) + \left\|P_{\mathcal{S}}\widehat{H_Z}P_{\mathcal{S}}\right\|\right) \\
&\geq a_1(E_2 - E_1) - (a_2 + \cdots + a_n)(E_1 - E_0) - 2\left(\sum_{1\leq u\leq n}|b_u| + \sum_{1\leq u,v\leq n}|b_{u,v}|\right) \\
&\geq \frac{1}{M_2}(E_2 - E_1) - (n-1)M_1(E_1 - E_0) - 2n(n+1)M_1.
\end{aligned}$$

We conclude the proof with $E_1 - E_0 = 2$ and $E_2 - E_1 = \Omega\big(G^{1-o(1)}\big)$ from Claim 5.4.34. □

We are ready to apply Lemma 5.4.29 on $\widehat{H}$ to obtain the following claim.

Claim 5.4.39.

$$\left|\lambda_k(H) - \lambda_k(\widehat{H} - cI)\right| \leq \mathcal{O}\big(n^2 M_1 r\big)\,, \tag{5.7}$$

where $c \coloneqq (E_0+1)\sum_{1\leq u\leq n} a_u$ *and* r *is defined in Claim 5.4.36.*

Proof. Claim 5.4.37 and 5.4.38 give $\|R\| \leq \mathcal{O}(n^2 M_1 r)$ and $\Delta \geq \Omega\Big(G_{\min}^{1-o(1)}/M_2\Big) > 0$. Therefore by Lemma 5.4.29, the low-energy spectrum of $\widehat{H}$ matches that of $\widehat{H}_{\mathcal{S}}$ modulo some negligible error. To be specific,

$$\left|\lambda_k(\widehat{H}) - \lambda_k(\widehat{H}_{\mathcal{S}})\right| \leq \|R\| \leq \mathcal{O}\big(n^2 M_1 r\big) \tag{5.8}$$

for every $k \in \{0, 1, \ldots, 2^n - 1\}$. Recall that Claim 5.4.36 says $\left\|\widehat{H}_{\mathcal{S}} - W(H + cI)W^\dagger\right\| \leq \mathcal{O}(n^2 M_1 r)$. Applying Weyl's inequality (Lemma 5.4.28) it converts to

$$\left|\lambda_k(\widehat{H}_{\mathcal{S}}) - (\lambda_k(H) + c)\right| \leq \mathcal{O}\big(n^2 M_1 r\big)\,. \tag{5.9}$$

Here we used the fact that the spectrum of $W(H+cI)W^\dagger$ is that of H shifted by c since W is unitary. The claim follows from Eq. 5.8 and 5.9. □

Claim 5.4.40.

$$\left\| e^{-iHt} - W^\dagger P_{\mathcal{S}} e^{-i(\widehat{H}-cI)t} W \right\| \leq \mathcal{O}\big(n^3 (M_1 r t)^{3/2}\big) .$$

Proof. Claim 5.4.37 gives $\|R\| \leq \mathcal{O}(n^2 M_1 r)$. Therefore by Lemma 5.4.31 we have

$$\left\| P_{\mathcal{S}} e^{-i\widehat{H}t} P_{\mathcal{S}} - e^{-i\widehat{H}_{\mathcal{S}} t} \right\| \leq \mathcal{O}\Big((\|R\| t)^{3/2} \Big) = \mathcal{O}\big(n^3 (M_1 r t)^{3/2}\big) .$$

Plug the bound $\left\| \widehat{H}_{\mathcal{S}} - W(H+cI)W^\dagger \right\| \leq \mathcal{O}(n^2 M_1 r)$ from Claim 5.4.36 into Lemma 5.4.30:

$$\left\| e^{-i\widehat{H}_{\mathcal{S}} t} - e^{-ict} W e^{-iHt} W^\dagger \right\| = \left\| e^{-i\widehat{H}_{\mathcal{S}} t} - e^{-i(W(H+cI)W^\dagger)t} \right\| \leq \mathcal{O}\big(n^2 M_1 r t\big) .$$

□

Note that we can eliminate c in Eq. 5.7 by replacing $\widehat{H}$ with $\widehat{H} - c$. Finally, let us determine $G^\star$:

- Claim 5.4.38 requires $G_{\min} \geq (n^2 M_1 M_2)^{1+\delta}$ for some $\delta > 0$.
- To conclude the proof of (ii) from Eq. 5.7, we straightforwardly need $n^2 M_1 r \ll \varepsilon_1$. Recall that $r \coloneqq \frac{1}{\sqrt{G_{\min}^{1-\varepsilon(w)}}}$ from Claim 5.4.35, which can be satisfied if $G_{\min} \geq \left(\frac{n^2 M_1}{\varepsilon_1}\right)^{2+\delta}$ for some $\delta > 0$ depending on w.
- Similar to the above, we need $n^3 (M_1 r t)^{3/2} \ll \varepsilon_1$ to prove (iii), which can be satisfied if $G_{\min} \geq \left(\frac{n^4 M_1^2 t^2}{\varepsilon_1^{4/3}}\right)^{1+\delta}$.

Therefore we can always choose $G^\star$ such that $G^\star \leq \mathrm{poly}(n, M_1, M_2, 1/\varepsilon_1, t)$ since $G^\star \leq G_{\min} M_1$.

Chapter 6: Provable (Sub)exponential Speedup of QAA and QHD

6.1 Introduction

Quantum Adiabatic Algorithm (QAA), or *adiabatic quantum optimization*, introduced by [31], has emerged as a distinct paradigm for solving various classical optimization problems using quantum mechanics. Formally, an adiabatic optimization algorithm is specified by an initial Hamiltonian H_{M} (conventionally called the *mixer*), a classical discrete optimization problem $f\colon \{0,1\}^n \to \mathbb{R}$, and an interpolation path $H(t)$ determined by a schedule function $s\colon [0,T] \to [0,1]$ satisfying $s(0) = 0$ and $s(T) = 1$, such that the system Hamiltonian at time $t \in [0,T]$ is

$$H(t) = (1 - s(t))H_{\mathrm{M}} + s(t)\sum_x f(x)\,|x\rangle\langle x|\,. \tag{6.1}$$

The algorithm operates by gradually evolving a quantum system from the *mixer* Hamiltonian $H(0) = H_{\mathrm{M}}$, whose ground state is easy to prepare, to the final Hamiltonian $H(T) = \sum_x f(x)\,|x\rangle\langle x|$, whose ground state encodes the solution to the optimization problem f. By ensuring that this evolution occurs slowly following $s(t)$, the system remains in its ground state by virtue of the quantum adiabatic theorem [57, 59], and eventually converges to global minima of f by construction. This simple, yet natural, physics-inspired quantum algorithm has led to extensive empirical and experimental studies, as well as several important theoretical insights (notably [60] and [61]); see [62] for a comprehensive review.

However, precisely characterizing the power of adiabatic quantum optimization remains a significant challenge [61, 62, 204], and only partial progress in relaxed settings is known.

- *Relaxing Optimization.* If the final Hamiltonian $H(T)$ is allowed to have nonzero entries off-

diagonal (instead of the diagonal Hamiltonian in (6.1)), then adiabatic computation is BQP-complete when $H(0)$ and $H(T)$ are local [61]. The reduction in [61] from quantum circuits to adiabatic evolution can also be applied to existing quantum circuits for Simon's problem [205] or Forrelation [190, 206] in the black-box setting to obtain an exponential separation.

It is, however, unclear how to design the final Hamiltonian to solve a *general* classical optimization problem f. In contrast, adiabatic quantum optimization uses the Hamiltonian (6.1), which has a more stringent format that simply enforces $H(T) = \sum_x f(x) \, |x\rangle\langle x|$.

- *Relaxing Adiabaticity or Interpolation.* If the interpolation path is allowed to have a certain diabatic transition [207] or to take a detour to some intermediate Hamiltonian H'_{M} [82], then it has (sub)exponential speedup against classical algorithms in the black-box query model.

  In particular, the Hamiltonian path in [82] is a concatenation of linear interpolation paths $H_{\mathrm{M}} \longrightarrow H'_{\mathrm{M}}$ and $H'_{\mathrm{M}} \longrightarrow \sum_x f(x) \, |x\rangle\langle x|$. Both H_{M} and f are very simple, and the source of quantum speedup is H'_{M}, which encodes the adjacency matrix of a carefully designed weighted graph. While these results show the potential of exponential speedups for corresponding computational models, they are unrelated to the way we use adiabatic evolution to solve discrete optimization problems, i.e., (6.1).

- *Relaxing Speedup.* If we sacrifice the demand for superpolynomial speedup, then indeed quadratic speedup against classical algorithms is known, for example, by adiabatic Grover search [60].

  Unfortunately, to our best knowledge, this represents the state-of-the-art separation and it remains a possibility that adiabatic quantum optimization can be dequantized with merely quadratic overhead, even in the black-box query model.

The above story naturally motivates the following question:

What provable speedup can adiabatic quantum optimization offer compared to classical algorithms?

Similarly, the same question can be asked for QHD:

What provable speedup can QHD offer compared to classical algorithms?

Our Result. The first main contribution of this chapter is a (sub)exponential separation in the black-box query model for adiabatic quantum optimization, where (i) the mixer H_{M} in (6.1) can be as simple as a *transverse field Ising* (TFI) Hamiltonian that is a linear combination of single-qubit Pauli-X and two-qubit Pauli-Z operators and (ii) the adiabatic schedule $s(\cdot)$ in (6.1) is simply linear. This immediately rules out the possibility of efficient black-box dequantization of adiabatic quantum optimization with polynomial overhead.

Theorem 6.1.1 (Informal version of Theorem 6.4.1)**.** *For any $n \geq 1$, there exists a family of discrete optimization problems with objective functions $f : \{0,1\}^n \to [0,1]$ such that the following holds.*

- *Any classical algorithm requires $\exp\left(n^{\Omega(1)}\right)$ queries to $f(\cdot)$ to find an x such that $f(x) - f(x^*) \leq 1/\operatorname{poly}(n)$, where x^* is the global minimizer of f.*
- *There exists a uniformly constructible TFI Hamiltonian H_{TFI}, which depends only on n and has an efficiently preparable ground state, such that adiabatic quantum optimization*

$$i\frac{\mathrm{d}}{\mathrm{d}t}\left|\psi(t)\right\rangle = H(t)\left|\psi(t)\right\rangle, \quad H(t) \coloneqq \left(1-\frac{t}{T}\right)H_{\mathrm{TFI}} + \frac{t}{T}\sum_{x\in\{0,1\}^n} f(x)\left|x\right\rangle\!\left\langle x\right|, \quad t \in [0,T], \tag{6.2}$$

finds the global minimizer x^* with $T = \operatorname{poly}(n)$. Moreover, its digital implementation uses $\operatorname{poly}(n)$ queries to $f(\cdot)$ and has $\operatorname{poly}(n)$ gate complexity.

At a high level, we prove Theorem 6.1.1 by showing that the computational power of adiabatic quantum optimization (6.1)—with H_{M} chosen as some TFI Hamiltonian and $s(t)$ chosen to be linear—is strong enough to efficiently simulate *any* adiabatic evolution whose time-dependent Hamiltonian $K(t)$ is stoquastic and sparse along the way and diagonal at the end. This is rather surprising because, compared to such a general $K(t)$, the time-dependent Hamiltonian of adiabatic quantum optimization we used is far more constrained: (i) it is simply a linear interpolation between the initial and final Hamiltonians, and (ii) its only off-diagonal terms are a sum of single-qubit Pauli-X operators.

As an application, combining Theorem 6.1.1 with techniques adapted from Chapter 5 that relate the TFI model to the Schrödinger operator, it yields a provable (sub)exponential quantum advantage for QHD:

Theorem 6.1.2 (Informal version of Theorem 6.10.2)**.** *There exists a family of continuous functions $f : \mathcal{X} \to [0, 1]$ with a bounded box domain $\mathcal{X} \subset \mathbb{R}^n$ such that, for any n, the following holds:*

- *Any classical algorithm requires $\exp\left(n^{\Omega(1)}\right)$ queries to $f(\cdot)$ to find an x such that $f(x) - f(x^*) \leq 1/\operatorname{poly}(n)$, where x^* is the global minimizer of f.*
- *There exists an auxiliary function $g(x)$, a time-dependent function $\nu(t) = e^{-\Theta(\sqrt{t})}$, both of which only depend on n. The QHD evolution, i.e.,*

$$i\frac{\mathrm{d}}{\mathrm{d}t}\left|\psi(t)\right\rangle = TH(t)\left|\psi(t)\right\rangle, \quad H(t) \coloneqq -\Delta + tf(x) + \nu(t)g(x), \quad t \in [0, t_{\mathrm{f}}], \tag{6.3}$$

finds the global minimizer x^* with $T, t_{\mathrm{f}} = \operatorname{poly}(n)$ from an efficiently preparable ground state of $H(0)$. Its digital implementation takes $\operatorname{poly}(n)$ queries to $f(\cdot)$ and has $\operatorname{poly}(n)$ gate complexity.

We note that the specific dynamics in eq. (6.3) deviates slightly from eq. (2.14) due to a quick-

vanishing term $\nu(t)g(x)$. It can be interpreted as some auxiliary information to boost the performance of QHD's dynamics in the early stage.

Technical Contribution. Our construction stems from the (sub)exponential oracle separation by Gilyén–Hastings–Vazirani (GHV) for adiabatic quantum computing with no sign problem [82]. Despite of being an adiabatic evolution, GHV's construction of the adiabatic path requires a detour to some hard intermediate Hamiltonian and concerns with a graph theoretical problem (instead of an optimization problem) inspired by the seminal Glued Tree problem [208].

Our main technical contribution is a new series of *efficient* reductions—based on perturbation theory [193, 209]—that map adiabatic paths (i.e., time-dependent Hamiltonians) from one to another where the latter adiabatic path can faithfully simulate the former one while being closer to the desired shape (6.2). In particular, our reductions preserve the spectral gap of adiabatic paths, thereby guaranteeing the easiness for quantum adiabatic evolution. On the other hand, the *efficiency* of our reductions preserves the computational hardness for classical algorithms; that is, any algorithm that efficiently solves the reduced problem (induced by the reduced adiabatic path) would naturally give rise to an efficient algorithm to the original problem (induced by the original adiabatic path).

For clarity, we only present the construction that reduces the GHV adiabatic path to (6.2) in this chapter. Our techniques can be extend to more general cases and can be of independent interests.

Related Works. Exponential quantum speedup has been heuristically suggested in early examples like Shor's factorization algorithm [210] and simulation of quantum dynamics [141, 211]. Provable separations are known in the more restricted black-box query model [82, 190, 205, 206, 208, 212], where the input is provided to both quantum and classical algorithms as query oracles. See [213] for a

more comprehensive survey.

In terms of optimization algorithms that have potential to bring exponential speedup, Quantum Approximate Optimization Algorithm (QAOA) [35] is a gate-based alternative to adiabatic quantum optimization that offers the same near-term feasibility for experimental implementation [214] while being more accessible for in-depth analysis [35, 215–218]. For more details, we refer interested readers to [33, 34]. Very recently, the Decoded Quantum Interferometry (DQI) algorithm [12] introduced a novel approach to quantum optimization and identified problems where DQI achieves better approximation ratio than all known efficient classical algorithms. Nevertheless, rigorously establishing a superpolynomial separation either in the black-box query model or conditioned on reasonable complexity-theoretic assumptions for QAOA or DQI still remains challenging.

The *perturbative reduction* framework used in this chapter simulates a target Hamiltonian within the low-energy subspace of a simpler Hamiltonian by introducing mediator qubits and large energy penalties. This general idea has been extensively applied in the field of Hamiltonian complexity [58] since its introduction by [171] to prove the QMA-completeness of 2-local Hamiltonians. Representative works include [172, 173, 178, 193, 195, 219, 220], among many others. These works typically introduce new perturbative reductions and gadgets as their main technical innovation, and this chapter is no exception.

6.2 Technical Overview

We give a high-level technical overview of the proofs of Theorem 6.1.1 and 6.1.2 in this section. Figure 6.1 summarizes our overall proof structure.

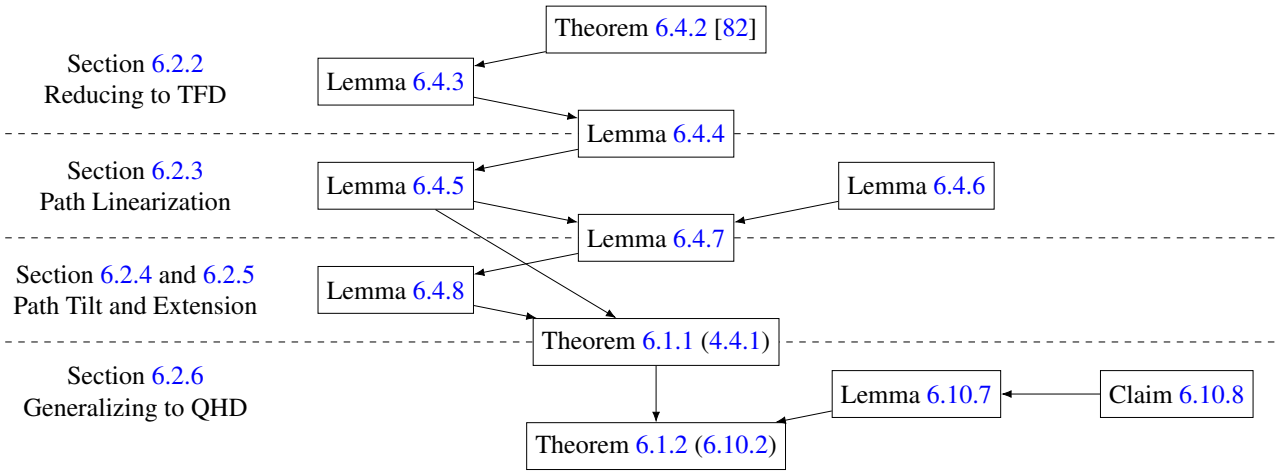


Figure 6.1: Overview diagram of the proof structure illustrating the logical dependencies among the main lemmas and theorems. The dashed lines separate major steps, which are covered in the corresponding subsections of Section 6.2.

6.2.1 GHV Adiabatic Path

GHV's construction considers an entrance-to-exit problem over oracularized graphs where the entrance is given and the exit information is encoded by the graph. It is similar to the Glued Tree problem [208] where exponential speedup is achievable by quantum walk, which is although a non-adiabatic evolution. With additional modifications of the oracularized graphs, GHV managed to solve the entrance-to-exit problem with the following adiabatic path:

$$H_{\text{init}} \longrightarrow H_{\text{graph}} \longrightarrow H_{\text{final}}. \tag{6.4}$$

The path satisfies that (i) $H_{\text{init}} \propto -|0\rangle\langle 0|$ (the given entrance) and $H_{\text{final}} \propto -|u\rangle\langle u|$ for some computational basis state $|u\rangle$ (the desired exit), and (ii) H_{graph} is sparse and stoquastic, referring to a graph whose vertexes are labeled by n-bit strings (i.e., the graph has size 2^n) and which can be accessed via an adjacency list oracle. The minimum spectral gap along the adiabatic path remains at least $1/\operatorname{poly}(n)$ relative to the norms of H_{init}, H_{graph}, and H_{final}, enabling an efficient quantum adiabatic

evolution (i.e., $T = \mathrm{poly}(n)$) that locates $|u\rangle$ with $\mathrm{poly}(n)$ queries. However, any classical algorithm finding $|u\rangle$ must make (sub)exponentially many queries to the adjacency list.

6.2.2 Perturbative Reductions to TFD Hamiltonians (Lemma 6.4.3 and 6.4.4)

Note that the entire GHV adiabatic path $H_{\text{init}} \longrightarrow H_{\text{graph}} \longrightarrow H_{\text{final}}$ consists solely of stoquastic sparse Hamiltonians, since any intermediate Hamiltonian is a linear combination of H_{init}, H_{graph}, and H_{final}. In contrast, for any $t \in [0, 1]$, our final target $H(t)$ in Theorem 6.1.1 is always a *transverse field diagonal (TFD)* Hamiltonian of the form

$$-\sum_i a_i X_i + D, \tag{6.5}$$

where each X_i is a Pauli-X operator on the ith qubit and D is diagonal.

Therefore, a natural next step is to consider some *reduction* from a stoquastic sparse Hamiltonian to a TFD Hamiltonian, which will later be applied *pointwise* to the GHV adiabatic path. The concrete meaning of reduction here is formalized in terms of the *Hamiltonian simulation* framework [173, 195], where perturbative reduction is usually the primary tool.

Our reduction is performed by introducing an intermediate stoquastic *hypercube* Hamiltonian. An n-qubit hypercube Hamiltonian is defined such that its *interaction graph* is a subgraph of the n-dimensional hypercube (including self-loops). Here, the interaction graph of a Hamiltonian H is obtained by interpreting H as a weighted adjacency matrix.

- By introducing a novel edge-subdivision perturbative gadget that allows the amplitude to efficiently tunnel through a *polynomially* long chain, we can alter the interaction graph of a stoquastic sparse Hamiltonian so that it embeds into a hypercube by carefully (re)labeling origi-

nal nodes and gadget nodes. This yields a second-order perturbative reduction from stoquastic sparse Hamiltonians to stoquastic hypercube Hamiltonians (Lemma 6.4.3); see Figure 6.2 for an illustration.

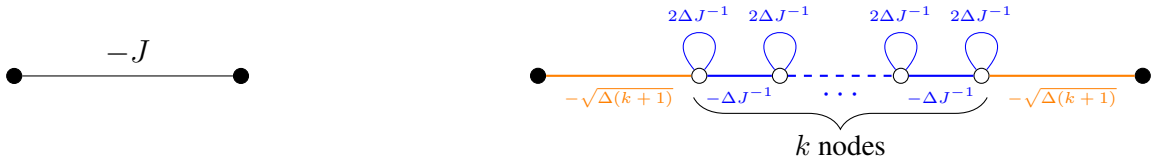


Figure 6.2: An illustration of how we reduce an edge of weight $-J$ ($0 < J \ll 1$) in a stoquastic sparse Hamiltonian to a gadget in a stoquastic hypercube Hamiltonian, where Δ is a parameter being sufficiently large. Nodes represent computational basis states. Edges with different colors correspond to different components[1] from the second-order perturbative reduction; see Lemma 6.4.3 for details. Note that k can be *polynomially* large since it only introduces a $\sqrt{k+1}$ overhead on weights.

- Drawing inspiration from the reduction mapping transverse field Ising (TFI) Hamiltonians to hard-core dimer Hamiltonians in [195], we reduce stoquastic hypercube Hamiltonians to TFD Hamiltonians, via another second-order perturbative reduction (Lemma 6.4.4); see Figure 6.3 for an illustration.



Figure 6.3: An illustration of how we reduce an edge of weight $-J$ ($0 < J \ll 1$) in a stoquastic hypercube Hamiltonian to a gadget in a TFD Hamiltonian, where Δ is a parameter that is sufficiently large. Nodes represent computational basis states. Edges with different colors correspond to different components[1] from the second-order perturbative reduction; see Lemma 6.4.4 for details.

[1] Orange edges come from $\Delta^{1/2}V_{\text{main}}$, and blue edges come from ΔH_0 in the proof of Lemma 6.4.3 and Lemma 6.4.4; the precise edge weights there differ slightly for technical reasons.

Our reduction from stoquastic sparse Hamiltonians to TFD Hamiltonians is general: it applies to any stoquastic sparse Hamiltonian, not solely those from the GHV adiabatic path. See Proposition C.2.1 for the formal statement.

6.2.3 Linearization of Adiabatic Path (Lemma 6.4.5, 6.4.6, and 6.4.7)

Now, suppose that we directly apply the one-to-one reduction—from a stoquastic sparse Hamiltonian to a TFD Hamiltonian as described in Section 6.2.2—to each Hamiltonian along the piecewise-linear GHV adiabatic path $H_{\text{init}} \longrightarrow H_{\text{graph}} \longrightarrow H_{\text{final}}$. The resulting path can be written as a time-dependent TFD Hamiltonian

$$K(t) = -a \sum_i X_i + D(t), \tag{6.6}$$

[2]

where $t \in [0, 1]$, and $D(t)$ is diagonal. However, $K(t)$ is no longer piecewise-linear as the reduction above is not a linear mapping. Therefore, we introduce a *linearization* technique to further reduce $K(t)$ to a linear adiabatic path $H(t)$ (Lemma 6.4.7).

While not piecewise-linear, $K(t)$ is still continuous in t. Hence it can be well approximated by a piecewise-linear path $K_0 \longrightarrow K_1 \longrightarrow \cdots \longrightarrow K_\ell$ for some sufficiently large ℓ (Lemma 6.4.5). To better illustrate the idea, we set $\ell = 2$ and discuss how to linearize $K_0 \longrightarrow K_1 \longrightarrow K_2$ below.

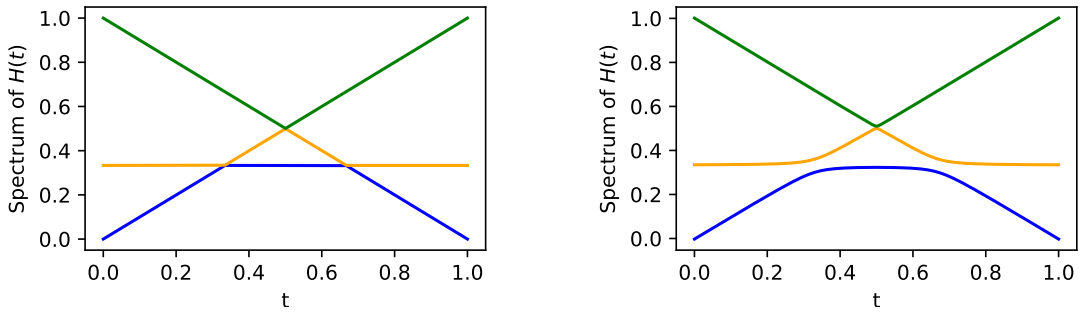
Consider a toy system with Hilbert space $\mathbb{C}^3$ governed by a time-dependent Hamiltonian $A(t)$ that is linear in $t \in [0, 1]$. We construct $A(t)$ such that it gradually shifts the ground state from $|0\rangle$ to

[2]Note that we can guarantee that the coefficients for every X_i are the same, compared to the most general form in (6.5), although the proof idea remains valid even if the coefficients *were* different.

$|1\rangle$, and finally to $|2\rangle$:

$$A(t) = \begin{bmatrix} t & -\lambda & 0 \\ -\lambda & 1/3 & -\lambda \\ 0 & -\lambda & 1-t \end{bmatrix}, \tag{6.7}$$

where $\lambda > 0$ is sufficiently small. The intuition behind this construction is that if λ *were* 0, $A(t)$ would be diagonal and its ground state would change abruptly from $|0\rangle$ to $|1\rangle$ and then to $|2\rangle$ (see Figure 6.4a). Instead, a small but nonzero λ guarantees that the transition is smooth and that the minimum spectral gap of $A(t)$ is $\Theta(\lambda)$ (see Figure 6.4b).



(a) $\lambda = 0$ (b) $\lambda = 0.03$

Figure 6.4: Spectrum of $A(t)$ for different λ values. A small but nonzero λ perturbs the spectrum, ensuring an $\Theta(\lambda)$ spectral gap throughout and making the instantaneous ground state transition of $A(t)$ smooth.

Now, construct the linear adiabatic path

$$H(t) := \Delta \cdot (I \otimes A(t)) + \sum_{\tau=0}^{2} K_\tau \otimes |\tau\rangle\langle\tau| \tag{6.8}$$

for a sufficiently large Δ. By first-order perturbation theory, the degenerate ground space of $I \otimes A(t)$ is perturbed by the minor term $\sum_\tau K_\tau \otimes |\tau\rangle\langle\tau|$, and $H(t)$ turns out to be *effectively* a linear combination of K_0 and K_1 (or K_1 and K_2) when restricted to the perturbed subspace, because the ground state of $A(t)$ is a linear combination of $|0\rangle$ and $|1\rangle$ (or $|1\rangle$ and $|2\rangle$). Thus a reduction from $K_0 \longrightarrow K_1 \longrightarrow K_2$

to $H(t)$ is established.

A remaining issue is that the $H(t)$ obtained in (6.8) is not a TFD Hamiltonian. To resolve this, we adapt the unary embedding method from [63] to transform our toy system $A(t)$ into a transverse field Ising (TFI) model (Lemma 6.4.6). With this modification, along with a slight adjustment to (6.8), $H(t)$ becomes a time-dependent TFD Hamiltonian that is linear in t.

It is worth noting that the second tensor product component in $H(t)$ can be interpreted as acting on a *clock register*. Clock registers were first introduced in [169] to prove QMA-hardness of the local Hamiltonian problem. In that work, the authors also used unary representation to reduce the interaction order (i.e., from $\Theta(\log n)$-local to 5-local) of the target Hamiltonian. Note that, however, our construction is entirely distinct from theirs, aside from sharing a similar concept of clock register.

Our linearization technique is general, and thus can be applied to any time-dependent Hamiltonian $K(t)$, not only to $K_0 \longrightarrow K_1 \longrightarrow K_2$ discussed above. Furthermore, $H(t)$ will be stoquastic if $K(t)$ is stoquastic. See Proposition C.2.2 for the formal statement.

6.2.4 Path Tilt and Extension (Lemma 6.4.8)

The technique in Section 6.2.3 turns out to provide a linear adiabatic path

$$H(t) \coloneqq H_X + H_{ZZ} + H_Z(t) + D,^{3} \qquad t \in [-1, 1] \tag{6.9}$$

that encodes the GHV adiabatic path. Here, H_X, H_{ZZ}, and D are independent of t and denote, respectively, a sum of single-qubit Pauli-X, a sum of one- and two-qubit Pauli-Z, and a diagonal matrix; moreover, $H_Z(t)$ is linear in t and is a sum of single-qubit Z terms. While being linear in t, (6.9) does

[3] Formally, $H_X = \Delta_{\mathsf{C}} I_{2^{n_{\mathsf{B}}}} \otimes X_{\mathsf{L}} + X_{\mathsf{B}} \otimes I_{2^{\ell}}$, $H_{ZZ} = \Delta_{\mathsf{C}} I_{2^{n_{\mathsf{B}}}} \otimes Q_{\mathsf{L}}$, $H_Z(t) = \Delta_{\mathsf{C}} I_{2^{n_{\mathsf{B}}}} \otimes Z_{\mathsf{L}}(t)$, and $D = D_{\mathsf{C}}$ in (6.42).

not match the form required in Theorem 6.1.1, where $H(t)$ is expected to begin with a TFI Hamiltonian and end at a diagonal Hamiltonian.

Our final step is to introduce a small parameter $0 < \eta \ll 1$ and tilt the path by defining

$$H'(t) \coloneqq (1 - t\eta)H_X + H_{ZZ} + H_Z(t) + (1 + t\eta)D. \tag{6.10}$$

It is obvious that for sufficiently small η, $H'(t)$ retains the essential spectral properties of $H(t)$. We then extend $H'(t)$ to the larger domain $t \in [-1/\eta, 1/\eta]$. This way, the Hamiltonian begins with $H'(-1/\eta) = 2H_X + H_{ZZ} + H_Z(-1/\eta)$ and ends at $H'(1/\eta) = H_{ZZ} + H_Z(1/\eta) + 2D$, matching the structure presented in Theorem 6.1.1.

This construction, however, introduces technical challenges. Although the spectral behavior of $H'(t)$ is controlled for $t \in [-1, 1]$, it is nontrivial to guarantee that the spectral gap for $t \in [-1/\eta, -1) \cup (1, 1/\eta]$ is large enough, a requirement for efficient adiabatic evolution. To resolve this, we have to explicitly analyze $H'(t)$ and refine the reductions in Section 6.2.2 and Section 6.2.3 so that the resulting paths there exhibit additional structural properties for later use.

6.2.5 Putting Everything Together (Theorem 6.1.1)

To summarize, we begin with the *stoquastic sparse* adiabatic path in [82]; then reduce it to certain *transverse field diagonal* path which is not linear; then we linearize this path to be of the form $-\sum_i a_i X_i + D(t)$ where $D(t)$ is diagonal; finally we slightly tilt this path and extend both its endpoints to arrive at a linear path that starts with a *transverse field Ising* Hamiltonian and ends at some *diagonal* Hamiltonian.

During these reductions, the adiabatic path not only gains more desirable structures, but also

preserves the inverse polynomial spectral gap. This enables efficient simulation via adiabatic quantum computing to locate the ground state of the final diagonal Hamiltonian (i.e., minimizer of the final optimization problem).

On the other hand, our reductions are black-box based on the original GHV adiabatic path. Hence any classical query to our new adiabatic path can be converted efficiently into classical queries to the original path, and the ground state of our final Hamiltonian corresponds directly to the ground state of the final Hamiltonian in the GHV adiabatic path. As a result, an efficient classical query algorithm for our adiabatic path can also solve the original problem, which is therefore ruled out by the classical hardness of finding the ground state of the ending Hamiltonian established in [82].

6.2.6 Generalization to Continuous Space (Theorem 6.1.2)

Theorem 6.1.1 establishes an exponential quantum advantage for unconstrained binary optimization problems in an oracular setting. We now extend this construction and demonstrate that an exponential quantum advantage persists in continuous domain $\mathbb{R}^n$ (Theorem 6.1.2).

Our construction is inspired by an intricate connection between two seemingly disparate objects: *TFI Hamiltonians* and *Schrödinger operators*, i.e., an unbounded operator of the form $H = -\Delta + V$ with Δ the Laplace operator and V a continuous function in $\mathbb{R}^n$. Despite their structural differences, Chapter 5 reveals that Schrödinger operators are at least as computationally powerful as TFI Hamiltonians. Their key observation is that a TFI Hamiltonian (consisting of 1-qubit Pauli X and up to 2-qubit Pauli Z operators) can be approximately simulated by a Schrödinger operator by the following

replacement rule:

$$Z \mapsto \widehat{Z} = \operatorname{sgn}(x), \text{[4]} \quad -\frac{1}{\Lambda} \cdot X \mapsto \widehat{X} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \lambda^2 f_{\mathrm{dw}}(x), \quad f_{\mathrm{dw}}(x) = \left(x^2 - \frac{1}{4}\right)^2, \tag{6.11}$$

for carefully chosen parameters $\lambda = \Theta(\log n)$ and $\Lambda = \Lambda(\lambda) = \operatorname{poly}(n)$.

In physics, the operator $\widehat{X}$ represents an anharmonic oscillator with a symmetric double-well potential $f_{\mathrm{dw}}(x)$. Due to the symmetry, the first two eigenstates (denoted by $|\chi_0\rangle$ and $|\chi_1\rangle$, respectively) of $\widehat{X}$ exhibit symmetric and anti-symmetric ordering, as illustrated in Figure 6.5a. It is easy to see that $|\chi_0\rangle$ and $|\chi_1\rangle$ are analogous to $|+\rangle$ and $|-\rangle$ for the Pauli-X operator, as they respectively form the ground and first excited states of $\widehat{X}$. Now, consider the linear combinations of the states $|\chi_0\rangle$ and $|\chi_1\rangle$:

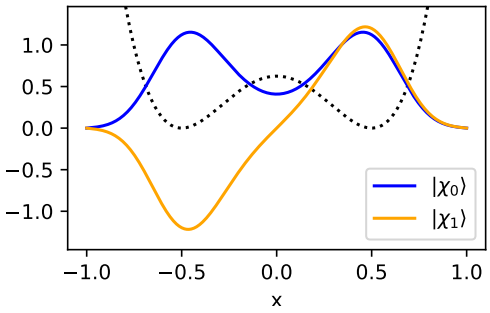
$$|\widehat{0}\rangle \coloneqq \frac{1}{\sqrt{2}}\left(|\chi_0\rangle + |\chi_1\rangle\right), \quad |\widehat{1}\rangle \coloneqq \frac{1}{\sqrt{2}}\left(|\chi_0\rangle - |\chi_1\rangle\right). \tag{6.12}$$

A crucial property of $|\widehat{0}\rangle$ is that its wave function is concentrated in the region $x \geq 0$ with error $\Lambda^{-\Omega(1)}$, as shown in Figure 6.5b. Therefore, we have $\widehat{Z}|\widehat{0}\rangle \approx |\widehat{0}\rangle$ and similarly $\widehat{Z}|\widehat{1}\rangle \approx -|\widehat{1}\rangle$, which justifies the notations $|\widehat{0}\rangle$ and $|\widehat{1}\rangle$ as they serve as analogues of the standard qubit states $|0\rangle$ and $|1\rangle$ in the context of $\widehat{Z}$ and $\widehat{X}$.

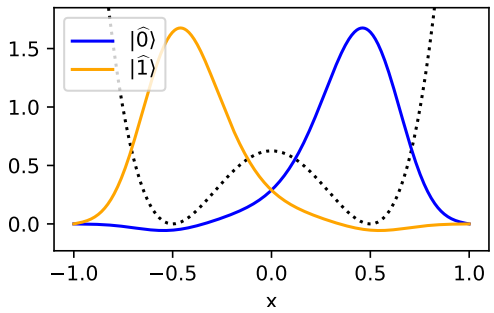
Building on the established mapping between TFI Hamiltonians and Schrödinger operators, we develop a new technical result (Lemma 6.10.7) that allows us to embed a TFD Hamiltonian $H = \sum_u h_u X_u + D$ into a continuous-space Schrödinger operator, where D represents an abstract n-qubit diagonal Hamiltonian. The idea is to keep the replacement rule $-X \mapsto \Lambda\widehat{X}$ from (6.11) while introducing a new mapping

$$D \mapsto \widehat{D}, \quad \widehat{D}\colon [-1,1]^n \to \mathbb{R}. \tag{6.13}$$

[4] Actually, Chapter 5 uses a smoothed version of $\operatorname{sgn}(x)$, but we stick to $\operatorname{sgn}(x)$ here for simplicity.



(a) Low-energy eigenstates of $\widehat{X}$



(b) States $|\widehat{0}\rangle$ and $|\widehat{1}\rangle$

Figure 6.5: Visualizations of $|\chi_0\rangle, |\chi_1\rangle$ regarding $\widehat{X}$ and their linear combination $|\widehat{0}\rangle, |\widehat{1}\rangle$ defined in (6.12). The parameter for $\widehat{X}$ is chosen to be $\lambda = 30$. The potential field f_{dw} (amplified 10 times) is plotted in a dotted line for reference.

In Figure 6.6, we illustrate the construction of $\widehat{D}$ from D. The function values in the four orthants correspond to the diagonal elements of D, and the function equals 0 on the boundary between orthants. From the concentration property of $|\widehat{0}\rangle$ and $|\widehat{1}\rangle$ it is natural to expect that $\widehat{D}(|\widehat{x}_1\rangle|\widehat{x}_2\rangle\cdots|\widehat{x}_n\rangle) \approx \langle x_1\cdots x_n|D|x_1\cdots x_n\rangle(|\widehat{x}_1\rangle|\widehat{x}_2\rangle\cdots|\widehat{x}_n\rangle)$, justifying the notation $\widehat{D}$.

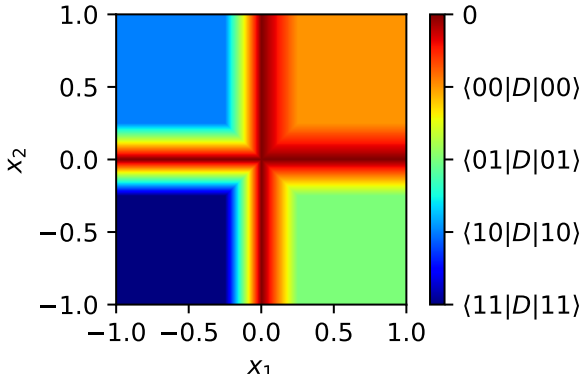


Figure 6.6: The Construction of $\widehat{D}$ from a diagonal Hamiltonian D with parameter $w = 1/4$ in Definition 6.10.4. It is efficient in the sense that a query to $\widehat{D}(\cdot)$ can be implemented by one query to D.

While the construction of $\widehat{D}$ appears natural, the error analysis of Chapter 5 (including, but not limited to, the use of "≈" in the last paragraph) does not extend directly. Their approach only provides a trivial error estimate of $\Theta(2^n\Lambda^{-\Omega(1)}) = 2^{(1-o(1))n} \gg 1$, because the nonlocal nature of D forces its

representation as a sum of Z-tensor products to involve up to 2^n terms. Thus, it is not immediately clear that the $\widehat{D}$ construction works. In contrast, we develop a novel error analysis technique that achieves an error bound of $\mathrm{poly}(n)\Lambda^{-\Omega(1)}$ by leveraging a newfound *symmetry* property of $|\chi_0\rangle$ and $|\chi_1\rangle$, articulated in (6.212).

Exploiting the correspondence between TFD Hamiltonians and Schrödinger operators, we prove a one-to-one mapping between a linear TFD adiabatic path ($H_{\mathrm{TFI}} \longrightarrow D$) from Theorem 6.1.1 and a path of Schrödinger operators:

$$H(t) = -\Delta + tf(x) + \nu(t)g(x), \quad \forall t \geq 0, \tag{6.14}$$

where f is a real-valued objective function on a box-shaped subset of $\mathbb{R}^n$, $g(x)$ is an auxiliary function independent of f, and $\nu(t) = e^{-\Theta(\sqrt{t})}$.

Similar to the linear adiabatic path, the initial ground state of $H(0)$ can be prepared as a product state, and $H(t)$ always admits a $1/\mathrm{poly}(n)$ spectral gap. Moreover, it is obvious from (6.14) that the global minimizer of f is encoded in the ground state of $H(t)$ as $t \to \infty$ and can be thus efficiently extracted by computational basis measurement. This leads to an efficient quantum algorithm for minimizing f by simulating the Schrödinger operator evolution in an adiabatic regime.

On the other hand, an oracle to f can be efficiently implemented with an oracle to D. Therefore any efficient classical quantum algorithm finding the global minimum of f can also finds the minimal entry in D, violating the classical hardness established in Theorem 6.1.1.

Our construction of hard continuous instances from hard discrete instances is general: it applies to any discrete objective functions that admit quantum speedups through adiabatic quantum optimization with the mixer being TFD Hamiltonians. See Proposition C.2.4 for the formal statement.

6.3 Preliminaries

For two binary string x, y of equal length, denote by $\mathrm{dist}(x, y)$ the Hamming distance between x and y.

Oracles. Our classical hardness and quantum efficiency will be proved in the *oracle* model, a.k.a, the *block-box* model. The oracles in this chapter will be classical functions that allow both classical and quantum access.

Formally, let $f\colon \{0,1\}^n \to \{0,1\}^m$ be a function, where we assume without loss generality that the input and output of f are binary strings. The oracle O_f provides efficient query access to f:

- in the classical case, O_f directly evaluates f on the classical query, i.e., $O_f(x) = f(x)$ for any $x \in \{0,1\}^n$;
- in the quantum case, O_f handles queries in superposition and evaluates f in superposition, i.e., $O_f \left|x\right\rangle_{\mathcal{A}} \left|y\right\rangle_{\mathcal{B}} = \left|x\right\rangle_{\mathcal{A}} \left|y \oplus f(x)\right\rangle_{\mathcal{B}}$ where $\mathcal{A}$ is the query register and $\mathcal{B}$ is the answer register; and $O_f \sum_{x,y} \alpha_{x,y} \left|x\right\rangle \left|y\right\rangle = \sum_{x,y} \alpha_{x,y} \left|x\right\rangle \left|y \oplus f(x)\right\rangle$ for any $\{\alpha_{x,y}\}_{x,y}$.

We say O is an oracle if $O = O_f$ for some function f. We emphasize that the query oracle O_f ignores the complexity of implementing f.

Fact 6.3.1. *The following matrix norm inequalities are standard.*

(i) $\|A\| \geq \|A\|_{\max}$, *where* $\|A\|_{\max} \coloneqq \max_{i,j}\{|A_{ij}|\}$ *is the* max norm *of* A.

(ii) $\|A\| \leq \sqrt{\|A\|_1 \|A\|_\infty}$, *where* $\|A\|_1 \coloneqq \max_j \sum_i |A_{ij}|$ *is the induced* 1*-norm and* $\|A\|_\infty \coloneqq \max_i \sum_j |A_{ij}|$ *is the induced* ∞*-norm. In particular,* $\|A\| \leq \|A\|_1 = \|A\|_\infty$ *for Hermitian* A.

6.3.1 Hamiltonian Path

We only work with Hermitian Hamiltonian in this chapter. We use ground state to refer to a *normalized* eigenvector of the Hamiltonian corresponding to its minimal eigenvalue.

A Hamiltonian is *stoquastic* [178] if it does not have a sign problem, i.e., with respect to some fixed basis, it has real entries and all off-diagonal entries are non-positive. The fixed basis will be the computational basis throughout this chapter. While our focus is on stoquastic Hamiltonians, some of our results work for general Hamiltonians.

Hamiltonian path. A *Hamiltonian path* is a time-dependent Hamiltonian $H(t)$ defined on the interval $[t_0, t_1]$. The spectral gap of the path is defined as the infimum of the spectral gaps of $H(t)$ for all $t \in [t_0, t_1]$. Similarly, the norm of the path is given by the supremum of the operator norms of $H(t)$ for all $t \in [t_0, t_1]$. The path is called L-Lipschitz if for any $t, t' \in [t_0, t_1]$, we have

$$\frac{\|H(t) - H(t')\|}{|t - t'|} \leq L. \tag{6.15}$$

Piecewise-linear Hamiltonian path. We use the notation

$$H_1 \longrightarrow H_2 \longrightarrow \cdots \longrightarrow H_\ell \tag{6.16}$$

to describe a time-dependent Hamiltonian $H(t)$ for $t \in [0, 1]$, defined as

$$H(t) \coloneqq (1 - \{\tau\})H_{\lfloor \tau \rfloor} + \{\tau\}H_{\lfloor \tau \rfloor + 1}, \tag{6.17}$$

where $\tau \coloneqq (\ell - 1)t + 1$ ranges over $[1, \ell]$, $\lfloor \tau \rfloor$ is the floor function, and $\{\tau\} = \tau - \lfloor \tau \rfloor$ represents the fractional part of τ.

The following definition extends the notion of an *interaction graph* from a static Hamiltonian to a time-dependent Hamiltonian.

Definition 6.3.2 (Interaction graph and interaction constraint)**.** Given a Hamiltonian H, its *interaction graph* is defined as follows. The vertex set V contains all computational basis states and the edge set E contains an edge (x, y) iff $\langle x|H|y\rangle \neq 0$.[5] For a time-dependent Hamiltonian $H(t)$ and a graph $G = (V, E)$, we say that $H(t)$ has an *interaction constraint* G if the interaction graph for $H(t)$ is a spanning subgraph of G for every t.

Sparsity and oracle access. Recall that a Hamiltonian H is s*-sparse* if it has at most s nonzero entries per row and per column in the computational basis. In other words, the interaction graph associated with H has a maximum degree at most s. Therefore, we call a time-dependent Hamiltonian $H(t)$ *explicitly* s*-sparse* if there exists an associated interaction constraint G with maximum degree at most s.

We say that $H(t)$ is accessible via oracle access if there exist oracles that provide efficient access to its structure:

- INTERACTION CONSTRAINT ORACLE. There exists an interaction constraint G of $H(t)$, such that given a basis state index x, this oracle returns the adjacency list of x in G.
- MATRIX ELEMENT ORACLE. Given t and indices x, y, this oracle returns the value of the matrix element $\langle x|H(t)|y\rangle$.

[5] Note that self-loop is allowed in E.

Despite being nonlocal, any n-qubit explicitly $\mathrm{poly}(n)$-sparse Hamiltonian can be efficiently simulated given oracle access [221, 222]. Moreover, efficient simulation remains feasible when the Hamiltonian is time-dependent [223].

Pauli operators. We denote the standard single-qubit Pauli operators by X, Y, and Z. For multi-qubit systems, we use X_u, Y_u, and Z_u to represent these operators acting specifically on qubit u.

Hamiltonian with special structures. A *transverse field Ising* (TFI) Hamiltonian is a Hamiltonian of the form

$$H = \sum_{u\in[n]} (h_u X_u + g_u Z_u) + \sum_{1\le u<v\le n} g_{u,v} Z_{u,v}, \tag{6.18}$$

where $h_u, g_u, g_{u,v} \in \mathbb{R}$.[6] We call H a *transverse field diagonal* (TFD) Hamiltonian if it has the form

$$H = \sum_{u\in[n]} h_u X_u + D, \tag{6.19}$$

where $h_u \in \mathbb{R}$, and D is diagonal. More generally, we say H is a *hypercube* Hamiltonian if it has an interaction constraint being the n-dimensional hypercube (including self-loops). Specifically, this means that $\langle x|H|y\rangle = 0$ holds for all $x, y \in \{0,1\}^n$ with $\mathrm{dist}(x,y) > 1$. Note that all TFI Hamiltonians are TFD ones, and all TFD Hamiltonians are hypercube ones.

[6] Some definitions of the TFI model exclude the single-qubit Z terms. Nevertheless, including these terms does not alter the computational power of the model [224, Eq. (6)].

6.3.2 Hamiltonian Simulation

In this chapter, we use the term *Hamiltonian simulation*[7] [173, 195] to describe the process of simulating one quantum many-body system using another. Such a simulation involves encoding a Hamiltonian H within another Hamiltonian H', enabling the second system to replicate *all* physical behaviors of the first. We formally introduce this notion below.

Definition 6.3.3 (Hamiltonian simulation [173, 195])**.** Let H be a Hamiltonian acting on a Hilbert space $\mathcal{H}$ of dimension N. A Hamiltonian H_{sim} on Hilbert space $\mathcal{H}_{\mathrm{sim}}$ and an isometry (encoding) $\mathcal{E} : \mathcal{H} \to \mathcal{H}_{\mathrm{sim}}$ are said to simulate H with error $(\varepsilon_{\mathrm{enc}}, \varepsilon)$ if there exists an isometry $\widetilde{\mathcal{E}} : \mathcal{H} \to \mathcal{H}_{\mathrm{sim}}$ such that

(i) The image of $\widetilde{\mathcal{E}}$ coincides with the *low-energy subspace* $\mathcal{L}_N(H_{\mathrm{sim}})$, which is spanned by the eigenvectors of H_{sim} associated with its N smallest eigenvalues.

(ii) $\|H - \widetilde{\mathcal{E}}^\dagger H_{\mathrm{sim}} \widetilde{\mathcal{E}}\| \leq \varepsilon$.

(iii) $\|\mathcal{E} - \widetilde{\mathcal{E}}\| \leq \varepsilon_{\mathrm{enc}}$.

Remark 6.3.4. If $\varepsilon_{\mathrm{enc}} = \varepsilon$, we say that $(H_{\mathrm{sim}}, \mathcal{E})$ simulates H *with error* ε for convenience.

Lemma 6.3.5 (Ground state simulation [195])**.** *Suppose H has a non-degenerate ground state $|g\rangle$ separated from excited states by a spectral gap δ. Suppose $(H_{\mathrm{sim}}, \mathcal{E})$ simulates H with error $(\varepsilon_{\mathrm{enc}}, \varepsilon)$. Then the spectral gap of H_{sim} is at least $\delta - 2\varepsilon$. In addition, H_{sim} has a non-degenerate ground state $|g_{\mathrm{sim}}\rangle$ and*

$$\|\mathcal{E}|g\rangle - |g_{\mathrm{sim}}\rangle\| \leq \varepsilon_{\mathrm{enc}} + O(\delta^{-1}\varepsilon). \tag{6.20}$$

[7] Note that this notion of Hamiltonian simulation should not be confused with the more general use of the term in quantum computing, which typically refers to quantum algorithms designed to approximate the unitary dynamics e^{-iHt} of a given Hamiltonian H [211, 225].

6.3.3 Perturbation Theory

We will use the following standard matrix analysis results.

Fact 6.3.6 (Weyl's inequality, see e.g., [226])**.** *Let A, B be Hermitian matrices. Then for each k, the kth smallest eigenvalue of $A + B$ differs from the kth smallest eigenvalue of A by at most $\|B\|$. As a result, the spectral gap of $A + B$ is at least the spectral gap of A minus $2\|B\|$.*

Fact 6.3.7 (Davis-Kahan $\sin\theta$ theorem, see e.g., [209, Theorem 2])**.** *Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian*[8] *matrices. Let $d \geq 1$ be an integer. Define δ to be the difference between the $(d+1)$th and the dth smallest eigenvalue of A; and assume $\delta > 0$. Let $V \in \mathbb{C}^{n\times d}$ (resp., $\widetilde{V} \in \mathbb{C}^{n\times d}$) have orthonormal columns corresponding to eigenvectors of the first d smallest eigenvalues of A (resp., $A + B$). Then there exists a unitary transform $U \in \mathbb{C}^{d\times d}$ such that*

$$\left\|VU - \widetilde{V}\right\|_{\mathrm{F}} \leq \frac{2^{3/2} \cdot \sqrt{d} \cdot \|B\|}{\delta}. \tag{6.21}$$

Fact 6.3.8 (Property of the Kronecker sum)**.** *Let $A_1, A_2, \ldots, A_n$ be Hermitian matrices. Their Kronecker sum*

$$A := \sum_{i\in[n]} I^{\otimes(i-1)} \otimes A_i \otimes I^{\otimes(n-i)} \tag{6.22}$$

satisfies $\sigma(A) = \sigma(A_1) + \cdots + \sigma(A_n)$, where $\sigma(H)$ denotes the spectrum of H (as a set) and $+$ denotes the Minkowski sum.

Next, we quote perturbative results regarding Hamiltonian simulation.

[8] [209, Theorem 2] states only for real symmetric matrices, but the analysis naturally generalizes to Hermitian matrices.

Taylor series for perturbative reduction. Let a finite-dimensional Hilbert space $\mathcal{H}_{\text{sim}}$ be decomposed into a direct sum

$$\mathcal{H}_{\text{sim}} = \mathcal{H}_- \oplus \mathcal{H}_+. \tag{6.23}$$

Let $P_\pm$ be the projector onto $\mathcal{H}_\pm$. For any linear operator O on $\mathcal{H}_{\text{sim}}$, define

$$O_{--} = P_- O P_-, \quad O_{-+} = P_- O P_+, \quad O_{+-} = P_+ O P_-, \quad O_{++} = P_+ O P_+. \tag{6.24}$$

The operator O is called block-diagonal if $O_{-+} = 0$ and $O_{+-} = 0$.

Let H_0 and V be Hermitian on $\mathcal{H}_{\text{sim}}$ such that H_0 is block-diagonal and $(H_0)_{--} = 0$. Moreover, $(H_0)_{++}$ has all eigenvalues no less than 1. Consider a perturbed Hamiltonian

$$H = \Delta H_0 + V. \tag{6.25}$$

Lemma 6.3.9 ([195]). *Assume $\Delta > 16\|V\|$. Then there exists a* Schrieffer-Wolff transformation *U [193] such that the following holds. $UHU^\dagger$ is block-diagonal with regard to $\mathcal{H}_-$ and $\mathcal{H}_+$. $(UHU^\dagger)_{++}$ has all eigenvalues greater than those of $(UHU^\dagger)_{--}$. Furthermore,*

$$(UHU^\dagger)_{--} = V_{--} - \Delta^{-1} V_{-+} H_0^{-1} V_{+-} + O(\Delta^{-2}\|V\|^3). \tag{6.26}$$

Suppose now that H_{target} is a fixed target Hamiltonian acting on some Hilbert space $\mathcal{H}_{\text{target}}$ and $\mathcal{E}\colon \mathcal{H}_{\text{target}} \to \mathcal{H}_{\text{sim}}$ is some fixed isometry (encoding) such that $\text{Im}\,(\mathcal{E}) = \mathcal{H}_-$. Define the *logical*

target Hamiltonian $\overline{H}_{\text{target}}$ acting on $\mathcal{H}_-$ as

$$\overline{H}_{\text{target}} = \mathcal{E} H_{\text{target}} \mathcal{E}^\dagger. \tag{6.27}$$

Lemma 6.3.10 (First-order reduction [195])**.** *Suppose Hermitian operators* H_0, V *satisfy that* H_0 *is block-diagonal,* $(H_0)_{--} = \zeta \cdot P_-$ *for some* $\zeta \in \mathbb{R}$,[9] $(H_0)_{++}$ *has all eigenvalues at least* $\zeta + \Delta$*, and*

$$\|\overline{H}_{\text{target}} - (H_0)_{--} - (V)_{--}\| \leq \varepsilon/2. \tag{6.28}$$

If $\Delta \geq O(\varepsilon^{-1}\|V\|^2 + \varepsilon_{\text{enc}}^{-1}\|V\|)$*, then* $H_{\text{sim}} = H_0 + V$ *and* $\mathcal{E}$ *above simulate* H_{target} *with error* $(\varepsilon_{\text{enc}}, \varepsilon)$*.*

Lemma 6.3.11 (Second-order reduction [195])**.** *Suppose Hermitian operators* $H_0, V_{\text{main}}, V_{\text{extra}}$ *satisfy that* H_0 *is block-diagonal,* $(H_0)_{--} = 0$*,* $(H_0)_{++}$ *has all eigenvalues at least* 1*,* V_{extra} *is block-diagonal,* $(V_{\text{main}})_{--} = 0$*, and*

$$\|\overline{H}_{\text{target}} - (V_{\text{extra}})_{--} + (V_{\text{main}})_{-+} H_0^{-1} (V_{\text{main}})_{+-}\| \leq \varepsilon/2. \tag{6.29}$$

Let $\Lambda = \max\{\|V_{\text{main}}\|, \|V_{\text{extra}}\|\}$*. If* $\Delta \geq O(\varepsilon^{-2}\Lambda^6 + \varepsilon_{\text{enc}}^{-2}\Lambda^2)$*, then* $H_{\text{sim}} = \Delta H_0 + \Delta^{1/2} V_{\text{main}} + V_{\text{extra}}$ *and* $\mathcal{E}$ *above simulate* H_{target} *with error* $(\varepsilon_{\text{enc}}, \varepsilon)$*.*

6.4 Proof Outline for Theorem 6.1.1

We provide the detailed proof outline of our main result Theorem 6.1.1 here.

Theorem 6.4.1 (Formal version of Theorem 6.1.1)**.** *For any* n*, there exist an* n*-qubit TFI Hamiltonian*

[9] The $\zeta = 0$ case is proved in [195] and it is trivial to generalize to arbitrary ζ by subtracting $\zeta \cdot I$ from H_0.

H_{TFI} and a family of n-qubit diagonal Hamiltonians D such that the linear Hamiltonian path $H(t)$ ($t \in [0,1]$) of the form

$$H_{\mathrm{TFI}} \longrightarrow D \tag{6.30}$$

has exponential quantum advantage given oracle access to D (and an explicit description of H_{TFI}):

(i) Any classical algorithm requires at least $\exp\left(n^{\Omega(1)}\right)$ queries to find $u \in \{0,1\}^n$ that minimizes $\langle u|D|u\rangle$, with success probability greater than $\exp\left(-n^{\Omega(1)}\right)$.

(ii) $H(t)$ has spectral gap $n^{\Omega(1)}$, norm $\mathrm{poly}(n)$, and the ground state of $H(0) = H_{\mathrm{TFI}}$ is $n^{-\Omega(1)}$-close to $|+\rangle^{\otimes(n-\ell)}|0\rangle^{\otimes\ell}$ for some ℓ. Consequently, by Lemma C.1.1, simulating the Schrödinger dynamics

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = TH(t)|\psi(t)\rangle\,, \quad |\psi(0)\rangle = |+\rangle^{\otimes(n-\ell)}|0\rangle^{\otimes\ell}\,, \quad T = \mathrm{poly}(n) \tag{6.31}$$

yields a final state $|\psi(1)\rangle$ that is $n^{-\Omega(1)}$-close to $|u\rangle$ with query and gate complexity $\mathrm{poly}(n)$.

Our proof of Theorem 6.4.1 starts with the following breakthrough construction due to Gilyén, Hastings, and Vazirani [82], which is a family of stoquastic Hamiltonian paths that allows efficient adiabatic quantum computation (see Theorem 6.4.2 (ii) and Lemma C.1.1) but is provably hard for classical query algorithms (Theorem 6.4.2 (i)).

Theorem 6.4.2 ([82]). *Let $\mathfrak{n}$ be an arbitrary positive integer and let $\mathfrak{m} = \mathfrak{n}^{16/5-o(1)}$. There exists a family of $\mathfrak{n}$-qubit stoquastic Hamiltonian paths $H_{\mathrm{GHV}}(t)$ that are explicitly $\Theta(\mathfrak{m})$-sparse, of the form*

$$H_{\text{init}} \longrightarrow H_{\text{graph}} \longrightarrow H_{\text{final}}, \tag{6.32}$$

where $H_{\text{init}} = -\mathfrak{m}\,|0^{\mathfrak{n}}\rangle\langle 0^{\mathfrak{n}}|$ *and* $H_{\text{final}} = -\mathfrak{m}\,|u\rangle\langle u|$ *for some* $u \in \{0,1\}^{\mathfrak{n}}$*, such that the following holds given oracle access to* $H(t)$*:*

(i) Any classical algorithm requires at least $\exp\left(\mathfrak{n}^{1/5-o(1)}\right)$ *queries to find* u *with probability greater than* $\exp\left(-\mathfrak{n}^{1/5-o(1)}\right)$.

(ii) $H(t)$ *has spectral gap* $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $O(\mathfrak{m})$.

We reserve $\mathfrak{n}, \mathfrak{m}$ as the parameter choices from Theorem 6.4.2 for frequent reference.

Given the *sparse* Hamiltonian path in Theorem 6.4.2, we first amend it to have *hypercube interaction* while preserving some simplicity of the start and the end Hamiltonian. The resulting Hamiltonian path of this reduction is formalized in Lemma 6.4.3 and proved in Section 6.5. We remark that Lemma 6.4.3 is not a (piecewise) linear path of Hamiltonians but a curve of Hamiltonians with bounded Lipschitz condition.

Lemma 6.4.3. *Let* $n_{\mathsf{A}} = \Theta(\mathfrak{m})$*. There exists a family of* n_{A}*-qubit stoquastic hypercube Hamiltonian paths* $H(t)$ *for* $t \in [0,1]$*, such that the following holds given oracle access to* $H(t)$*:*

(i) $H(0) = -\mathfrak{m}\sum_{i\in[n_{\mathsf{A}}]} X_i$*,* $H(1) = -\mathfrak{m}\,|u\rangle\langle u|$ *for some* $u \in \{0,1\}^{n_{\mathsf{A}}}$*, and* $H(t)$ *is* $\mathrm{poly}(\mathfrak{n})$*-Lipschitz.*

(ii) Any classical algorithm requires at least $\exp\left(\mathfrak{n}^{1/5-o(1)}\right)$ *queries to find* u *with probability greater than* $\exp\left(-\mathfrak{n}^{1/5-o(1)}\right)$.

(iii) $H(t)$ *has spectral gap* $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $\mathrm{poly}(\mathfrak{n})$.

With the hypercube structure of Hamiltonian paths from Lemma 6.4.3, we use another reduction to further fix it to be Hamiltonian paths consisting of TFD Hamiltonians. Then the Hamiltonian path

evolution only changes the diagonal terms. This is formalized in Lemma 6.4.4 below and proved in Section 6.6.

Lemma 6.4.4. *Let $n_{\mathsf{B}} \coloneqq \Theta(\mathfrak{m})$. There exist a parameter $\Delta_{\mathsf{B}} \in [\mathfrak{m}, \mathrm{poly}(\mathfrak{m})]$ and a family of n_{B}-qubit stoquastic Hamiltonian paths $H(t)$ for $t \in [0,1]$ of the form*

$$H(t) = X_{\mathsf{B}} + D(t) \quad \textit{and} \quad X_{\mathsf{B}} \coloneqq -\Delta_{\mathsf{B}} \sum_{i \in [n_{\mathsf{B}}]} X_i, \tag{6.33}$$

where $D(t)$ is diagonal, such that the following holds given oracle access to $D(t)$:

(i) $D(0) = 0$, $D(1) = -3\Delta_{\mathsf{B}} n_{\mathsf{B}} \, |u\rangle\langle u|$ for some $u \in \{0,1\}^{n_{\mathsf{B}}}$, and $D(t)$ is $\mathrm{poly}(\mathfrak{n})$-Lipschitz.

(ii) Any classical algorithm requires at least $\exp\big(\mathfrak{n}^{1/5-o(1)}\big)$ queries to find u with probability greater than $\exp\big(-\mathfrak{n}^{1/5-o(1)}\big)$.

(iii) $H(t)$ has spectral gap $\Omega(\sqrt{\mathfrak{m}})$ and norm $\mathrm{poly}(\mathfrak{n})$.

One evident difference between Lemma 6.4.4 and Theorem 6.4.1 is that the Hamiltonian path in Lemma 6.4.4 is not a straight line but a complicated curve. To address this issue, we first approximate the curve by piecewise linear paths as the following Lemma 6.4.5, the validity of which is guaranteed by the Lipschitz condition in Lemma 6.4.4.

Lemma 6.4.5. *$H(t)$ in Lemma 6.4.4 can be replaced with Hamiltonian paths of the form*

$$X_{\mathsf{B}} + D_0 \longrightarrow X_{\mathsf{B}} + D_1 \longrightarrow \cdots \longrightarrow X_{\mathsf{B}} + D_\ell, \tag{6.34}$$

where $\ell = \mathrm{poly}(\mathfrak{n})$.

Proof. For each $k = 0, 1, \ldots, \ell$, define $D_k = D(k/\ell)$, where $D(t), t \in [0,1]$ comes from Lemma 6.4.4. It remains to prove the spectral gap and norm condition for (6.34). Fix arbitrary $k \in [\ell]$ and $\alpha \in [0,1]$. Define

$$\widetilde{H} = X_{\mathsf{B}} + \alpha \cdot D_{k-1} + (1-\alpha) \cdot D_k = \alpha \cdot H\left(\frac{k-1}{\ell}\right) + (1-\alpha) \cdot H\left(\frac{k}{\ell}\right), \tag{6.35}$$

where $H(t), t \in [0,1]$ comes from Lemma 6.4.4. To bound the operator norm, we simply observe that

$$\left\|\widetilde{H}\right\| \leq \max\{\|H((k-1)/\ell)\|, \|H(k/\ell)\|\} \leq \mathrm{poly}(\mathfrak{n}). \tag{6.36}$$

For the spectral gap, notice that

$$\begin{aligned}\left\|\widetilde{H} - H(k/\ell)\right\| &\leq \|H((k-1)/\ell) - H(k/\ell)\| = \|D((k-1)/\ell) - D(k/\ell)\| \qquad \text{(since } \alpha \in [0,1]\text{)} \\ &\leq \mathrm{poly}(\mathfrak{n})/\ell \leq 1. \quad \text{(by Lemma 6.4.4 (i) and setting } \ell = \mathrm{poly}(\mathfrak{n}) \text{ sufficiently large)}\end{aligned}$$

Hence by Fact 6.3.6, the spectral gap of $\widetilde{H}$ is at least the spectral gap of $H(k/\ell)$ minus 2, which is still $\Omega(\sqrt{\mathfrak{m}})$ by Lemma 6.4.4 (iii). □

To further linearize the piecewise linear path from Lemma 6.4.5, we take a slight detour to describe a toy TFI Hamiltonian path that sequentially shifts ground states over the computational basis. Later we will use this property to embed turning points in Lemma 6.4.5.

TFI Hamiltonian that shifts ground states. Define $|\boldsymbol{j}\rangle := \left|1^j 0^{\ell-j}\right\rangle$ and let $\varepsilon_{\mathsf{L}}, \Delta_{\mathsf{L}} > 0$ be parameters to be determined. We start with an ℓ-qubit time-dependent TFI Hamiltonian $H_{\mathsf{L}}(t)$ that is *linear* in $t \in \mathbb{R}$. This Hamiltonian path will gradually shift the ground state from $|\mathbf{0}\rangle$ to $|\mathbf{1}\rangle, |\mathbf{2}\rangle, \ldots, |\boldsymbol{\ell}\rangle$ in a

sequential order.

Formally, $H_{\mathsf{L}}(t) = Q_{\mathsf{L}} + X_{\mathsf{L}} + Z_{\mathsf{L}}(t)$ where $X_{\mathsf{L}} = -\sum_{i\in[\ell]} X_i$ and

$$Q_{\mathsf{L}} = \Delta_{\mathsf{L}} \left((\ell-1)I + Z_1 - Z_\ell - \sum_{i\in[\ell-1]} Z_i Z_{i+1} \right), \quad Z_{\mathsf{L}}(t) = \frac{1}{\varepsilon_{\mathsf{L}}} \sum_{i\in[\ell]} (i-t)(I - Z_i). \tag{6.37}$$

The properties of $H_{\mathsf{L}}(t)$ is listed in the following Lemma 6.4.6 and proved in Section 6.7. We emphasize that Lemma 6.4.6 (iii) is the key property that allows us to shift ground states. We also remark that Lemma 6.4.6 (iv) is an extra benefit that will become useful in a later reduction (Lemma 6.4.8).

Lemma 6.4.6. *Assume $\ell \geq 1$ and $\varepsilon_{\mathsf{L}} \in (0, 1/10]$. There exists a $\Delta_{\mathsf{L}} = \mathrm{poly}(\ell, 1/\varepsilon_{\mathsf{L}})$ such that the following holds for $t \in \mathbb{R}$:*

(i) $\|H_{\mathsf{L}}(t)\| \leq (|t|+1)\,\mathrm{poly}(\ell, 1/\varepsilon_{\mathsf{L}})$.

(ii) The spectral gap of $H_{\mathsf{L}}(t)$ is at least $1/2$.

(iii) Let $|\psi(t)\rangle$ be a ground state of $H_{\mathsf{L}}(t)$. If $t \in [0, \ell+1]$, then $|\psi(t)\rangle$ is $O(\varepsilon_{\mathsf{L}})$-close to a normalized state that is a linear combination of $|\boldsymbol{a}\rangle$ and $|\boldsymbol{b}\rangle$, where $a := \max\{\lfloor t - 1/2 \rfloor, 0\}$ and $b := \min\{\lceil t - 1/2 \rceil, \ell\}$.

(iv) For any $\lambda \in [0, 2]$, the Hamiltonian

$$H_{\mathsf{L}}(\lambda, t) := Q_{\mathsf{L}} + \lambda X_{\mathsf{L}} + Z_{\mathsf{L}}(t) \tag{6.38}$$

satisfies that its spectral gap is at least $1/2$ and its ground state is $O(\varepsilon_{\mathsf{L}})$-close to $|\mathbf{0}\rangle$ (resp., $|\boldsymbol{\ell}\rangle$) if $t \leq 0$ (resp., $t \geq \ell + 1$).

Given the toy system $H_{\mathsf{L}}(t)$, we return to our current piecewise linear Hamiltonian path $H(t)$ from Lemma 6.4.5 and embed all its turning points into $H_{\mathsf{L}}(t)$. This addresses another issue that Lemma 6.4.5 is not a straight path.

Formally, for $(D_0, D_1, \ldots, D_\ell)$ in Lemma 6.4.5, we define the $(n_{\mathsf{B}}+\ell)$-qubit Hamiltonian $H_{\mathsf{C}}(t)$ as follows:

$$H_{\mathsf{C}}(t) \coloneqq \Delta_{\mathsf{C}} \cdot I_{2^{n_{\mathsf{B}}}} \otimes H_{\mathsf{L}}(t) + X_{\mathsf{B}} \otimes I_{2^\ell} + D_{\mathsf{C}}, \quad D_{\mathsf{C}} \coloneqq \sum_{0 \le i \le \ell} D_i \otimes |\boldsymbol{i}\rangle\langle\boldsymbol{i}|\,, \tag{6.39}$$

where Δ_{C} is a parameter to be determined.

Observe that $H_{\mathsf{C}}(t)$ is linear in t since $H_{\mathsf{L}}(t)$ is linear. In addition, we show that it simulates the Hamiltonian path from Lemma 6.4.5. The following Lemma 6.4.7 is proved in Section 6.8 using first-order perturbative reduction and properties in Lemma 6.4.6.

Lemma 6.4.7. *Assume* $\varepsilon_{\mathsf{L}} \le 1/\operatorname{poly}(\mathfrak{n})$. *There exist*

$$\Delta_{\mathsf{L}} = \operatorname{poly}(\mathfrak{n}, 1/\varepsilon_{\mathsf{L}}) \quad \textit{and} \quad 10\|D_{\mathsf{C}}\| \le \Delta_{\mathsf{C}} \le \operatorname{poly}(\mathfrak{n}) \tag{6.40}$$

such that the Hamiltonian path

$$H_{\mathsf{C}}(0) \longrightarrow H_{\mathsf{C}}(\ell+1) \tag{6.41}$$

has spectral gap $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $\operatorname{poly}(\mathfrak{n}, 1/\varepsilon_{\mathsf{L}})$.

By expanding the definition of $H_{\mathsf{C}}(t)$, we have

$$H_{\mathsf{C}}(t) = \Delta_{\mathsf{C}} \cdot I_{2^{n_{\mathsf{B}}}} \otimes H_{\mathsf{L}}(t) + X_{\mathsf{B}} \otimes I_{2^\ell} + D_{\mathsf{C}} \tag{6.42}$$

$$= \Delta_{\mathsf{C}} \cdot I_{2^{n_{\mathsf{B}}}} \otimes (Q_{\mathsf{L}} + X_{\mathsf{L}} + Z_{\mathsf{L}}(t))) + X_{\mathsf{B}} \otimes I_{2^{\ell}} + D_{\mathsf{C}}. \tag{6.43}$$

Note that Q_{L} and $Z_{\mathsf{L}}(t)$ are classical Ising Hamiltonians that consist of 1- and 2-local Pauli-Z's. On the other hand, X_{L} and X_{B} are 1-local X-type Hamiltonian; and D_{C} is just diagonal (not necessarily local). Comparing these with our target Theorem 6.4.1, we need to remove D_{C} from the starting Hamiltonian and remove $X_{\mathsf{L}}, X_{\mathsf{B}}$ from the terminating Hamiltonian. This is achieved by slightly tilting and then extending the Hamiltonian path $H_{\mathsf{C}}(t)$.

Path tilt and extension. Let $\eta > 0$ be a parameter to be determined later. We slightly tilt the $X_{\mathsf{L}}, X_{\mathsf{B}}$ and D_{C} component of $H_{\mathsf{C}}(t)$ as the following Hamiltonian path $H_{\mathsf{C}'}$:

$$H_{\mathsf{C}'}(t) \coloneqq \Delta_{\mathsf{C}} \cdot I_{2^{n_{\mathsf{B}}}} \otimes (Q_{\mathsf{L}} + (1-t\eta)X_{\mathsf{L}} + Z_{\mathsf{L}}(t)) + (1-t\eta) \cdot X_{\mathsf{B}} \otimes I_{2^{\ell}} + (1+t\eta) \cdot D_{\mathsf{C}}. \tag{6.44}$$

Observe that if $t = -1/\eta$, then $H_{\mathsf{C}'}(t)$ is a TFI Hamiltonian (i.e., D_{C} is removed); and if $t = 1/\eta$, then $H_{\mathsf{C}'}(t)$ is diagonal (i.e., $X_{\mathsf{L}}, X_{\mathsf{B}}$ are removed). While this is precisely what Theorem 6.4.1 needs, we still need to verify that $H_{\mathsf{C}'}(t)$ makes sense for quantum adiabatic computing. This is formalized in the following Lemma 6.4.8 and proved in Section 6.9.

Lemma 6.4.8. *There exist* $\varepsilon_{\mathsf{L}} = 1/\operatorname{poly}(\mathfrak{n})$, $\Delta_{\mathsf{L}} \leq \operatorname{poly}(\mathfrak{n})$,

$$10\|D_{\mathsf{C}}\| \leq \Delta_{\mathsf{C}} \leq \operatorname{poly}(\mathfrak{n}) \quad \textit{and} \quad 1/\operatorname{poly}(\mathfrak{n}) \leq \eta \leq 1/(10\|D_{\mathsf{C}}\| + \ell) \tag{6.45}$$

such that the Hamiltonian path

$$H_{\mathsf{C}'}(-1/\eta) \longrightarrow H_{\mathsf{C}'}(1/\eta) \tag{6.46}$$

has spectral gap $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $\mathrm{poly}(\mathfrak{n})$.

To put everything together and prove Theorem 6.4.1, we simply write out the Hamiltonian path in Lemma 6.4.8.

Proof of Theorem 6.4.1. Let $n \coloneqq n_{\mathsf{B}} + \ell = \mathrm{poly}(\mathfrak{n})$. Since $\mathfrak{n}$ is arbitrary from Theorem 6.4.2, our construction holds for infinitely many n, which can be easily extended to all positive integer n by some dummy padding argument.

Starting Hamiltonian H_{TFI}. Recall (6.44) and Lemma 6.4.6. For the starting Hamiltonian, we have

$$H_{\mathrm{TFI}} \coloneqq H_{\mathsf{C}'}(-1/\eta) = \Delta_{\mathsf{C}} \cdot I_{2^{n_{\mathsf{B}}}} \otimes H_{\mathsf{L}}(2, -1/\eta) + 2X_{\mathsf{B}} \otimes I_{2^{\ell}}. \tag{6.47}$$

Since $H_{\mathsf{L}}(\lambda, t)$ and X_{B} are both TFI Hamiltonian, we know that H_{TFI} is indeed a TFI Hamiltonian.

In addition, by Lemma 6.4.6 (iv), we know that the ground state of $H_{\mathsf{L}}(2, -1/\eta)$ (and hence $\Delta_{\mathsf{C}} H_{\mathsf{L}}(2, -1/\eta)$) is $O(\varepsilon_{\mathsf{L}})$-close to $|\mathbf{0}\rangle = |0^{\ell}\rangle$. Since $X_{\mathsf{B}} = -\Delta_{\mathsf{B}} \sum_{i \in [n_{\mathsf{B}}]} X_i$ from Lemma 6.4.4, we know that the ground state of $2X_{\mathsf{B}}$ is $|+\rangle^{n_{\mathsf{B}}}$. As a result, the ground state of H_{TFI} is $O(\varepsilon_{\mathsf{L}})$-close to $|+\rangle^{n_{\mathsf{B}}} |0^{\ell}\rangle$. As $\varepsilon_{\mathsf{L}} = 1/\mathrm{poly}(\mathfrak{n}) = 1/\mathrm{poly}(n)$, this, together with the spectral gap condition and norm condition from Lemma 6.4.8, verifies Theorem 6.4.1 (ii).

Ending Hamiltonian D. Now for the Hamiltonian in the end, we have

$$\begin{aligned} D \coloneqq H_{\mathsf{C}'}(1/\eta) &= \Delta_{\mathsf{C}} \cdot I_{2^{n_{\mathsf{B}}}} \otimes H_{\mathsf{L}}(0, 1/\eta) + 2D_{\mathsf{C}} && \text{(by (6.44))} \\ &= \Delta_{\mathsf{C}} \cdot I_{2^{n_{\mathsf{B}}}} \otimes (Q_{\mathsf{L}} + Z_{\mathsf{L}}(1/\eta)) + 2 \sum_{0 \le i \le \ell} D_i \otimes |\boldsymbol{i}\rangle\langle\boldsymbol{i}| . && \text{(by (6.39))} \end{aligned}$$

Recall from (6.37) that Q_{L} and $Z_{\mathsf{L}}(t)$ are diagonal, and each D_i is diagonal. Hence D is indeed a diagonal Hamiltonian. Moreover, it is obvious from the above that a query to D can be implemented by a query to D_i for some $i = 0, 1, \ldots, \ell$.

Classical hardness. Finally we prove Theorem 6.4.1 (i). To this end, it is sufficient to show that any classical algorithm, which makes q queries to D and finds its ground state, can be turned into a classical algorithm $\mathcal{A}$ that makes q queries to $\{D_j\}_{j=0,1,\ldots,\ell}$ in Lemma 6.4.5 and finds the target u in Lemma 6.4.5, because Lemma 6.4.5 (ii) then asserts that q must be exponentially large.

In fact, the construction of $\mathcal{A}$ is straightforward because of (i) the implementation of the oracle to D using oracles to D_j, as discussed in the previous paragraph, and (ii) the fact that the ground state of D is exactly $|u, \boldsymbol{\ell}\rangle$, since every other computational basis state have larger energy—this is implicit in the proof of Lemma 6.4.8, and we include an explicit argument here for completeness:

- For any $u' \neq u$, we have

$$\langle u', \boldsymbol{\ell}|D|u', \boldsymbol{\ell}\rangle - \langle u, \boldsymbol{\ell}|D|u, \boldsymbol{\ell}\rangle = 2\langle u'|D_\ell|u'\rangle - 2\langle u|D_\ell|u\rangle = 6\Delta_{\mathsf{B}} n_{\mathsf{B}} > 0. \tag{6.48}$$

- For any v and $0 \leq j < \ell$, we have

$$\begin{aligned}
&\langle v, \boldsymbol{j}|D|v, \boldsymbol{j}\rangle - \langle v, \boldsymbol{\ell}|D|v, \boldsymbol{\ell}\rangle && (6.49)\\
&= \left(2\langle v|D_j|v\rangle + \Delta_{\mathsf{C}}\langle \boldsymbol{j}|Z_{\mathsf{L}}(1/\eta)|\boldsymbol{j}\rangle\right) - \left(2\langle v|D_\ell|v\rangle + \Delta_{\mathsf{C}}\langle \boldsymbol{\ell}|Z_{\mathsf{L}}(1/\eta)|\boldsymbol{\ell}\rangle\right) && (6.50)\\
&= \frac{\Delta_{\mathsf{C}}(\ell - j)(2/\eta - \ell - j - 1)}{\varepsilon_{\mathsf{L}}} + 2\left(\langle v|D_j|v\rangle - \langle v|D_\ell|v\rangle\right) && \text{(by (6.37))}\\
&> 1 + 4\|D_{\mathsf{C}}\| - 2\|D_{\mathsf{C}}\| > 0. && \text{(by (6.45))}
\end{aligned}$$

- For any v and $\tau \notin \{\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{\ell}\}$, define w as the Hamming weight of τ. We have

$$\langle v, \tau | D | v, \tau \rangle - \langle v, \boldsymbol{w} | D | v, \boldsymbol{w} \rangle \tag{6.51}$$

$$= \Delta_{\mathsf{C}} \left(\langle \tau | Z_{\mathsf{L}}(1/\eta) | \tau \rangle + \langle \tau | Q_{\mathsf{L}} | \tau \rangle \right) - 2 \langle v | D_w | v \rangle - \Delta_{\mathsf{C}} \left(\langle \boldsymbol{w} | Z_{\mathsf{L}}(1/\eta) | \boldsymbol{w} \rangle + \langle \boldsymbol{w} | Q_{\mathsf{L}} | \boldsymbol{w} \rangle \right) \tag{6.52}$$

$$> \Delta_{\mathsf{C}} \left(\langle \tau | Q_{\mathsf{L}} | \tau \rangle - \langle \boldsymbol{w} | Q_{\mathsf{L}} | \boldsymbol{w} \rangle \right) - 2 \langle v | D_w | v \rangle$$

(since $\langle \tau | Z_{\mathsf{L}}(1/\eta) | \tau \rangle > \langle \boldsymbol{w} | Z_{\mathsf{L}}(1/\eta) | \boldsymbol{w} \rangle$ from (6.37))

$$\geq \Delta_{\mathsf{C}} \cdot 1 - 2 \| D_{\mathsf{C}} \| \geq 0, \tag{6.53}$$

  where we also use (6.37) to derive $\langle \tau | Q_{\mathsf{L}} | \tau \rangle \geq 1 + \langle \boldsymbol{w} | Q_{\mathsf{L}} | \boldsymbol{w} \rangle$ for the last inequality (see also Claim 6.7.1 for a detailed calculation).

In summary, for each $|v, \tau\rangle$, let w be the Hamming weight of τ and we have

$$\langle v, \tau | D | v, \tau \rangle \geq \langle v, \boldsymbol{w} | D | v, \boldsymbol{w} \rangle \geq \langle v, \boldsymbol{\ell} | D | v, \boldsymbol{\ell} \rangle \geq \langle u, \boldsymbol{\ell} | D | u, \boldsymbol{\ell} \rangle . \tag{6.54}$$

In addition, at least one of the inequality becomes strict if $|v, \tau\rangle \neq |u, \boldsymbol{\ell}\rangle$. This means that $|u, \boldsymbol{\ell}\rangle$ is the ground state of D. □

6.5 Reduction I: Proof of Lemma 6.4.3

The goal of this section is to prove Lemma 6.4.3 (restated below).

Lemma 6.4.3. *Let $n_{\mathsf{A}} = \Theta(\mathfrak{m})$. There exists a family of n_{A}-qubit stoquastic hypercube Hamiltonian paths $H(t)$ for $t \in [0, 1]$, such that the following holds given oracle access to $H(t)$:*

(i) $H(0) = -\mathfrak{m} \sum_{i \in [n_{\mathsf{A}}]} X_i$, $H(1) = -\mathfrak{m} \, |u\rangle\langle u|$ *for some* $u \in \{0,1\}^{n_{\mathsf{A}}}$, *and* $H(t)$ *is* $\mathrm{poly}(\mathfrak{n})$-*Lipschitz.*

(ii) Any classical algorithm requires at least $\exp\left(\mathfrak{n}^{1/5-o(1)}\right)$ *queries to find* u *with probability greater than* $\exp\left(-\mathfrak{n}^{1/5-o(1)}\right)$.

(iii) $H(t)$ *has spectral gap* $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $\mathrm{poly}(\mathfrak{n})$.

The proof consists of two major steps. In the first step, we apply a second-order perturbative reduction (Lemma 6.3.11) to transform the Hamiltonian path in Theorem 6.4.2 into a Hamiltonian path composed of stoquastic hypercube Hamiltonians. In the second step, we append it with additional paths at both the beginning and end, to ensure that the final Hamiltonian path starts with $-\mathfrak{m} \sum_{i \in [n_{\mathsf{A}}]} X_i$ and terminates at $-\mathfrak{m} \, |u\rangle\langle u|$, as stated in Lemma 6.4.3 (i).

Organization. The first step for Lemma 6.4.3 is presented in Section 6.5.1. The second step is in Section 6.5.2. Then in Section 6.5.3 we put two steps together and prove Lemma 6.4.3. The missing proofs from this section are deferred to Section 6.5.4.

Notation. We (re)define some notation. Let $K(t)$ for $t \in [0,1]$ be the $\mathfrak{n}$-qubit Hamiltonian path H_{GHV} in Theorem 6.4.2. Write $K(1) = -\mathfrak{m} \, |w\rangle\langle w|$, and reserve the notation u such that it is consistent with the statement of Lemma 6.4.3. Without loss of generality, assume that $G = (V, E)$ is the interaction constraint of $K(t)$. In particular, G is an s-regular graph with self-loops and $s = \Theta(\mathfrak{m})$. Let M be an upper bound on $\|K(t)\|$ and $M = O(\mathfrak{m})$.

6.5.1 First Step – Sparse Hamiltonians to Hypercube Hamiltonians

Let $n_{\mathsf{A}} \coloneqq 3\mathfrak{n} + 2s$. Below we show how to reduce $K(t)$ to an n_{A}-qubit stoquastic hypercube Hamiltonian $H'(t)$ such that $H'(t)$ simulates $K(t)$ in the sense of Definition 6.3.3. This reduction is one-to-one and thus for convenience we write $K \coloneqq K(t)$ and $H' \coloneqq H'(t)$.

Note that for hypercube Hamiltonians, non-zero entries can only exist between nodes of distance at most one. To obtain this for $(x, y) \in E$ which may have arbitrary $\mathrm{dist}(x, y)$, we will use a path $P_{x,y}$ to link $(x, y) \in E$ such that $P_{x,y}$ only consists of hypercube transition edges. To describe $P_{x,y}$, we define an order for all $\mathfrak{n}$-bit binary strings and the rank function among them.

Definition 6.5.1 (Order and rank)**.** For any $x, y \in \{0, 1\}^{\mathfrak{n}}$, we interpret them as $\mathfrak{n}$-bit binary numbers and compare them using "<","=", and ">" accordingly. For each $x \in \{0, 1\}^{\mathfrak{n}}$, we list all $y > x$ with $(x, y) \in E$ in the increasing order and define $\mathrm{rk}(x, y)$ to be the rank of y in this list. Similarly we list all $y < x$ with $(x, y) \in E$ in the decreasing order and define $\mathrm{rk}(x, y)$ to be the rank of y in this list.

Fix an $(x, y) \in E$ pair satisfying $x < y$. Let $\ell \coloneqq \mathrm{dist}(x, y) \leq \mathfrak{n}$ and consider the sequence

$$x =: z_0, \quad z_1, \quad z_2, \quad \ldots, \quad z_{\ell-1}, \quad z_\ell \coloneqq y \tag{6.55}$$

that connects x and y, where each z_i, z_{i+1} differs by exactly one bit (i.e., they have Hamming distance 1). For uniqueness, we choose $z_0, z_1, \ldots, z_\ell$ to be the lexicographically smallest. While this is already a valid hypercube transition path of $\mathfrak{n}$-bit binary strings connecting x and y, for the sake of later reduction, we pad it into a valid hypercube transition path of $(n_{\mathsf{A}} = 3\mathfrak{n} + 2s)$-bit binary strings of a *fixed* length of $2\mathfrak{n} + 3$.

Definition 6.5.2 (Path $P_{x,y}$)**.** Define $r_1 = \mathrm{rk}(x,y), r_2 = \mathrm{rk}(y,x)$ and set $e_{r_1} = 0^{r_1-1}10^{s-r_1}, e_{r_2} = 0^{r_2-1}10^{s-r_2}$. We connect $(3\mathfrak{n}+2s)$-bit strings $(x,x,e_{r_1},0^s,0^{\mathfrak{n}})$ and $(y,y,0^s,e_{r_2},0^{\mathfrak{n}})$ with a path $P_{x,y}$ of length $2\mathfrak{n}+3$, where indexes of consecutive nodes differ by one bit. In $P_{x,y}$, we connect following nodes in order:

- $(x,x,e_{r_1},0^s,0^{\mathfrak{n}}), (x,z_1,e_{r_1},0^s,0^{\mathfrak{n}}), \ldots, (x,z_{\ell-1},e_{r_1},0^s,0^{\mathfrak{n}}), (x,y,e_{r_1},0^s,0^{\mathfrak{n}})$; ($\ell+1$ nodes)
- $(x,y,e_{r_1},0^s,0^{\mathfrak{n}-1}1), (x,y,e_{r_1},0^s,0^{\mathfrak{n}-2}1^2), \ldots, (x,y,e_{r_1},0^s,0^{\ell}1^{\mathfrak{n}-\ell})$; ($\mathfrak{n}-\ell$ nodes)
- $(x,y,e_{r_1},e_{r_2},0^{\ell}1^{\mathfrak{n}-\ell})$; (one node)
- $(x,y,0^s,e_{r_2},0^{\ell}1^{\mathfrak{n}-\ell}), (x,y,0^s,e_{r_2},0^{\ell+1}1^{\mathfrak{n}-\ell-1}), \ldots, (x,y,0^s,e_{r_2},0^{\mathfrak{n}-1}1)$; ($\mathfrak{n}-\ell$ nodes)
- $(x,y,0^s,e_{r_2},0^{\mathfrak{n}}), (z_1,y,0^s,e_{r_2},0^{\mathfrak{n}}), \ldots, (z_{l-1},y,0^s,e_{r_2},0^{\mathfrak{n}}), (y,y,0^s,e_{r_2},0^{\mathfrak{n}})$. ($\ell+1$ nodes)

For convenience we label the nodes in $P_{x,y}$ by $v_{x,y,i}$ for $i \in [2\mathfrak{n}+3]$ following the listed order above.

To establish the reduction using Lemma 6.3.11, we denote by $\mathcal{K}$ and $\mathcal{H}$ the Hilbert spaces associated with K and H' respectively. Define the encoding isometry $\mathcal{E}\colon \mathcal{K} \to \mathcal{H}$ by

$$\mathcal{E}\,|x\rangle = |x,x,0^s,0^s,0^{\mathfrak{n}}\rangle \quad \text{for } x \in \{0,1\}^{\mathfrak{n}}. \tag{6.56}$$

Observe that the image of $\mathcal{E}$ is n_{A}-bit binary strings, consistent with paths in Definition 6.5.2.

Let $\mathcal{H}_- \coloneqq \mathrm{Im}\,\mathcal{E}$ and we write $\mathcal{H} = \mathcal{H}_- \oplus \mathcal{H}_+$. In light of Lemma 6.3.11, our plan is to construct

$$H' \coloneqq \Delta H_0 + \Delta^{1/2} V_{\text{main}} + V_{\text{extra}}, \tag{6.57}$$

for some $\Delta, H_0, V_{\text{main}}, V_{\text{extra}}$ such that $V_{\text{extra}} - (V_{\text{main}})_{-+} H_0^{-1} (V_{\text{main}})_{+-}$ is close to $\overline{K}$.

Entries of K with small offsets. Let $\varepsilon \coloneqq 1/2$. For any $x, y \in \{0,1\}^{\mathfrak{n}}$ with $(x,y) \in E$, define

$$a_{x,y} \coloneqq \langle x|K|y\rangle - \frac{\varepsilon}{2s}. \tag{6.58}$$

Informally, the value $a_{x,y}$ simply records the (x,y)th entry of K with a small offset $\varepsilon/(2s)$ for non-degeneracy: since K is stoquastic, for any $x \neq y$, we have $K_{x,y} \leq 0$ and thus $a_{x,y} \leq -\varepsilon/(2s)$.

Construction of H_0 and V_{main}. Initially, all matrix elements of H_0, V_{main} are set to zero, and we modify specific entries as needed.

Set $(H_0)_{++}$ as the identity mapping. Now for each $(x,y) \in E$ with $x < y$, recall the path $P_{x,y} = (v_{x,y,1}, \ldots, v_{x,y,2\mathfrak{n}+3})$ from Definition 6.5.2. We modify the following matrix elements of H_0 and V_{main}, while their corresponding transposed entries are set accordingly due to Hermiticity:

- Define
$$\alpha \coloneqq \frac{M+1}{4\sin^2\left(\frac{\pi}{4\mathfrak{n}+8}\right)} \leq O(\mathfrak{n}^2 M). \tag{6.59}$$
- $\langle x, x, 0^s, 0^s, 0^{\mathfrak{n}}|V_{\text{main}}|v_{x,y,1}\rangle = \langle y, y, 0^s, 0^s, 0^{\mathfrak{n}}|V_{\text{main}}|v_{x,y,2\mathfrak{n}+3}\rangle = -\sqrt{\alpha(2\mathfrak{n}+4)}$.
- $\langle v_{x,y,i}|H_0|v_{x,y,i+1}\rangle = \alpha \cdot a_{x,y}^{-1}$ for $i \in [2\mathfrak{n}+2]$.
- $\langle v_{x,y,i}|H_0|v_{x,y,i}\rangle = -2\alpha \cdot a_{x,y}^{-1}$ for $i \in [2\mathfrak{n}+3]$.

Note that $a_{x,y}^{-1}$ is valid due to the small offsets from (6.58).

By some standard calculation of tridiagonal matrices which we defer to Section 6.5.4, we obtain the following properties on H_0 and V_{main}.

Claim 6.5.3. *H_0 is block-diagonal with respect to each $\mathcal{H} = \mathcal{H}_- \oplus \mathcal{H}_+$. $(H_0)_{++}$ has minimal eigenvalue at least* 1. *Moreover,*

$$\|H_0\| \leq O\left(\max_{x,y} \frac{\alpha}{|a_{x,y}|}\right) \tag{6.60}$$

and

$$\langle y, y, 0^s, 0^s, 0^{\mathfrak{n}} | (V_{\text{main}})_{-+} H_0^{-1} (V_{\text{main}})_{+-} | x, x, 0^s, 0^s, 0^{\mathfrak{n}} \rangle = \begin{cases} -a_{x,y}, & x \neq y, \\ -(2\mathfrak{n}+3) \sum_{\substack{y:\, y \neq x \\ (x,y) \in E}} a_{x,y}, & x = y. \end{cases} \tag{6.61}$$

Construction of V_{extra}. At this point, we define V_{extra} as

$$V_{\text{extra}} \coloneqq \sum_x \left(a_{x,x} - (2\mathfrak{n}+3) \sum_{\substack{y:\, y \neq x \\ (x,y) \in E}} a_{x,y} \right) |x, x, 0^s, 0^s, 0^{\mathfrak{n}}\rangle\langle x, x, 0^s, 0^s, 0^{\mathfrak{n}}| \,. \tag{6.62}$$

Then by (6.58) and Claim 6.5.3, we have

$$V_{\text{extra}} - (V_{\text{main}})_{-+} H_0^{-1} (V_{\text{main}})_{+-} = \overline{K - \frac{\varepsilon}{2s} E} = \overline{K} - \frac{\varepsilon}{2s} \overline{E}, \tag{6.63}$$

where $\overline{A} \coloneqq \mathcal{E} A \mathcal{E}^\dagger$ for any A, and E denotes the unweighted adjacency matrix of G (with a slight abuse of notation).

Parameters in the reduction. Note that $\left\|\overline{E}\right\| = \|E\| \leq \|E\|_1 \leq s$ by Fact 6.3.1 (ii). Therefore all preconditions of Lemma 6.3.11 are met. Applying Lemma 6.3.11, we obtain some $\Delta \leq \text{poly}(1/\varepsilon, \Lambda)$

such that $(H', \mathcal{E})$ simulates K with error ε, where

$$\Lambda = \max\{\|V_{\text{main}}\|, \|V_{\text{extra}}\|\} \leq \text{poly}(\mathfrak{n}, s, M). \tag{6.64}$$

Hence we get $\Delta \leq \text{poly}(1/\varepsilon, \mathfrak{n}, s, M)$. For H_0, Claim 6.5.3 gives

$$\|H_0\| \leq O\left(\max_{x,y}\left\{\frac{\alpha}{|a_{x,y}|}\right\}\right) \leq O\left(\frac{\mathfrak{n}^2 M}{\varepsilon/s}\right) \leq \text{poly}(1/\varepsilon, \mathfrak{n}, s, M). \tag{6.65}$$

Combining these estimates, we get $\|H'\| \leq \text{poly}(1/\varepsilon, \mathfrak{n}, s, M) = \text{poly}(\mathfrak{n})$ as $s, M \leq O(\mathfrak{m}) \leq \text{poly}(\mathfrak{n})$ by Theorem 6.4.2.

Note that $(H', \mathcal{E})$ simulates K with error $\varepsilon = 1/2$. By Theorem 6.4.2 (ii) and Lemma 6.3.5, we have the following claim.

Claim 6.5.4. *$H'(t)$ has spectral gap at least $\Omega(\sqrt{\mathfrak{m}})$ and norm $\text{poly}(\mathfrak{n})$.*

By mechanically expanding the construction above, it is not hard to see that $H'(t)$ is Lipschitz. This is stated in the following claim and proved in Section 6.5.4.

Claim 6.5.5. *$H'(t)$ is $\text{poly}(\mathfrak{n})$-Lipschitz.*

6.5.2 Second Step – Cleaner Start and End

Inheriting the notation from Section 6.5.1, we first append a linear path at the end of $H'(1)$. Write $H'(1) = \Delta H_0 + \Delta^{1/2} V_{\text{main}} + V_{\text{extra}}$. Recall that $K(1) = -\mathfrak{m}\,|w\rangle\langle w|$. Let $|u\rangle := \mathcal{E}\,|w\rangle$ where $\mathcal{E}$ is the encoding isometry from (6.56). The path to be appended is

$$H'(1) = \Delta H_0 + \Delta^{1/2} V_{\text{main}} + V_{\text{extra}} \longrightarrow \Delta H_0 + V_{\text{extra}} \longrightarrow -\mathfrak{m}\,|u\rangle\langle u|\,. \tag{P1}$$

We will show that (P1) preserves the spectral gap.

Claim 6.5.6. *The Hamiltonian path (P1) has spectral gap* $\Omega(\mathfrak{m})$ *and norm* $\mathrm{poly}(\mathfrak{n})$. *Consequently, it is* $\mathrm{poly}(\mathfrak{n})$*-Lipschitz by piecewise linearity.*

Proof. The norm bound is evident by Claim 6.5.4 and we focus on the spectral gap.

We start with the first half of (P1). For any $\lambda \in [0,1]$, define $L(\lambda) \coloneqq \Delta H_0 + \lambda \Delta^{1/2} V_{\mathrm{main}} + V_{\mathrm{extra}}$. By Lemma 6.3.11, $(L(\lambda), \mathcal{E})$ simulates the Hamiltonian

$$\mathcal{E}^\dagger \left((V_{\mathrm{extra}})_{--} - \lambda^2 (V_{\mathrm{main}})_{-+} H_0^{-1} (V_{\mathrm{main}})_{+-} \right) \mathcal{E} \tag{6.66}$$

with error $\varepsilon = 1/2$ by the construction in Section 6.5.1. By Lemma 6.3.5, this implies that the spectral gap of $L(\lambda)$ is at least the spectral gap of (6.66) minus 2ε. By (6.62) and (6.58), we have

$$\mathcal{E}^\dagger (V_{\mathrm{extra}})_{--} \mathcal{E} = -\mathfrak{m} \cdot |w\rangle\langle w| + \sum_{x \in \{0,1\}^{\mathfrak{n}}} (s-2)(2\mathfrak{n}+3) \cdot \frac{\varepsilon}{2s} \cdot |x\rangle\langle x| \,, \tag{6.67}$$

$$= -\mathfrak{m} \cdot |w\rangle\langle w| + \frac{\varepsilon (s-2)(2\mathfrak{n}+3)}{2s} \cdot I, \tag{6.68}$$

where we also use the fact that $K(1) = -\mathfrak{m}\,|w\rangle\langle w|$. By Claim 6.5.3, we also have

$$\mathcal{E}^\dagger (V_{\mathrm{main}})_{-+} H_0^{-1} (V_{\mathrm{main}})_{+-} \mathcal{E} = (s-1)(2\mathfrak{n}+3) \sum_{x \in \{0,1\}^{\mathfrak{n}}} \frac{\varepsilon}{2s} \cdot |x\rangle\langle x| + \sum_{\substack{x,y:\ x \neq y \\ (x,y) \in E}} \frac{\varepsilon}{2s} \cdot |x\rangle\langle y| \tag{6.69}$$

$$= \left(\frac{\varepsilon (s-1)(2\mathfrak{n}+3)}{2s} - \frac{\varepsilon}{2s} \right) \cdot I + \frac{\varepsilon}{2s} \cdot E, \tag{6.70}$$

where E denotes the unweighted adjacency matrix of G, with a slight abuse of notation. Thus, (6.66)

has the form

$$-\mathfrak{m}\,|w\rangle\langle w| - \frac{\lambda^2\varepsilon}{2s}\cdot E + c(s,\mathfrak{n},\varepsilon,\lambda)\cdot I, \tag{6.71}$$

where $c(s,\mathfrak{n},\varepsilon,\lambda)$ is some constant depending only on $\lambda, s, \mathfrak{n}, \varepsilon$. Therefore, its spectral gap is at least $\mathfrak{m} - \lambda^2\varepsilon$ by Fact 6.3.6 and $\|E\| \le \|E\|_1 = s$. Together with Lemma 6.3.5, this implies that the first half of the Hamiltonian path (P1) is at least $\mathfrak{m} - \lambda^2\varepsilon - 2\varepsilon \ge \mathfrak{m} - 3\varepsilon = \Omega(\mathfrak{m})$.

For the second half of (P1), note that it is always block-diagonal with respect to $\mathcal{H}_-$ and $\mathcal{H}_+$. The Hamiltonian restricted to $\mathcal{H}_+$ gradually shifts from $\Delta(H_0)_{++}$ to 0; and, recalling (6.71), the restriction to $\mathcal{H}_-$ shifts from $-\mathfrak{m}\,|u\rangle\langle u| + c(s,\mathfrak{n},\varepsilon,0)\cdot I_{--}$ to $-\mathfrak{m}\,|u\rangle\langle u|$. Since $\Delta(H_0)_{++}$ has eigenvalues greater than those of $-\mathfrak{m}\,|u\rangle\langle u| + c(s,\mathfrak{n},\varepsilon,0)\cdot I_{--}$ by Lemma 6.3.11, it follows that the spectral gap is always exactly $\mathfrak{m}$. This completes the proof of Claim 6.5.6. □

Write $H'(0) = \Delta\widetilde{H}_0 + \Delta^{1/2}\widetilde{V}_{\text{main}} + \widetilde{V}_{\text{extra}}$. We append the following Hamiltonian path to the beginning of $H'(t)$:

$$-\mathfrak{m}\,|0^{n_{\mathsf{A}}}\rangle\langle 0^{n_{\mathsf{A}}}| \longrightarrow \Delta\widetilde{H}_0 + \widetilde{V}_{\text{extra}} \longrightarrow \widetilde{H}_0 + \Delta^{1/2}\widetilde{V}_{\text{main}} + \widetilde{V}_{\text{extra}} = H'(0). \tag{P2}$$

Recall that $K(0) = -\mathfrak{m}\,|0^{\mathfrak{n}}\rangle\langle 0^{\mathfrak{n}}|$ and $\mathcal{E}\,|0^{\mathfrak{n}}\rangle = |0^{n_{\mathsf{A}}}\rangle$. By an analogous argument as in Claim 6.5.6, we have the following claim.

Claim 6.5.7. *The Hamiltonian path (P2) has spectral gap* $\Omega(\mathfrak{m})$ *and norm* $\mathrm{poly}(\mathfrak{n})$. *Consequently, it is* $\mathrm{poly}(\mathfrak{n})$*-Lipschitz by piecewise linearity.*

Finally, consider the following Hamiltonian path, which we append in front of (P2):

$$-\mathfrak{m}\sum_{i\in[n_{\mathsf{A}}]} X_i \longrightarrow -\mathfrak{m}\sum_{i\in[n_{\mathsf{A}}]} Z_i \longrightarrow -\mathfrak{m}\,|0^{n_{\mathsf{A}}}\rangle\langle 0^{n_{\mathsf{A}}}|\,. \tag{P3}$$

Claim 6.5.8. *The Hamiltonian path (P3) has spectral gap at least* $\mathfrak{m}$ *and norm* $O(\mathfrak{m})$*. Consequently, it is* $O(\mathfrak{m})$*-Lipschitz by piecewise linearity.*

Proof. The second half of (P3) is always diagonal with spectral gap decreasing linearly from $2\mathfrak{m}$ to $\mathfrak{m}$. For the first half, we apply Fact 6.3.8 and its spectral gap is identical to that of the single-qubit Hamiltonian path $-\mathfrak{m}X \longrightarrow -\mathfrak{m}Z$. By direct calculation, the spectral gap of $\lambda X + (1-\lambda)Z$ is given by $2\sqrt{2\lambda^2 - 2\lambda + 1} \geq \sqrt{2}$. Hence the first half of (P3) has spectral gap at least $\sqrt{2}\mathfrak{m}$. □

6.5.3 Putting Everything Together

Finally we prove Lemma 6.4.3 by combining the hypercube Hamiltonian path in Section 6.5.1 and the additional decoration in Section 6.5.2.

Proof of Lemma 6.4.3. Recall Hamiltonian paths $H'([0,1])$ from Section 6.5.1 and (P3), (P2), (P1) from Section 6.5.2. We sequentially connect (P3), (P2), $H'([0,1])$, and (P1). This forms a Hamiltonian path $H : [0,4] \to \mathcal{H}$. By rescaling the domain linearly from $[0,4]$ to $[0,1]$, we obtain the final path $H(t)$.

By Claim 6.5.4, Claim 6.5.5, Lemma 6.4.3, Claim 6.5.7, and Claim 6.5.8, $H(t)$ has spectral gap $\Omega(\sqrt{\mathfrak{m}})$, norm $\mathrm{poly}(\mathfrak{n})$, and it is $\mathrm{poly}(\mathfrak{n})$-Lipschitz. This verifies Lemma 6.4.3 (i) and Lemma 6.4.3 (iii).

For Lemma 6.4.3 (ii), we note that any query to $\langle x|H(t)|y\rangle$ reduces to queries of the (x,y) entries of $H_0(t')$, $V_{\mathrm{main}}(t')$, and $V_{\mathrm{extra}}(t')$ for some $t' := t'(t)$. By our construction, these can be implemented using $O(1)$, $O(1)$, and $O(s)$ queries, respectively, to the entries of $K(t')$. Consequently, any q-query classical algorithm that finds $|u\rangle$ can also find $|w\rangle$, since $|w\rangle = \mathcal{E}^\dagger |u\rangle$. This transformation results in an $O(qs)$-query classical algorithm with oracle access to $K(t')$. By Theorem 6.4.2 (i), we have $O(qs) \geq \exp\left(\mathfrak{n}^{1/5-o(1)}\right)$. Since $s = \mathrm{poly}(\mathfrak{n})$, this implies that $q \geq \exp\left(\mathfrak{n}^{1/5-o(1)}\right)$. □

6.5.4 Deferred Proofs

Proof of Claim 6.5.3. Observe that every index $|v, v', a, b, c\rangle$ in $\mathcal{H}_+$ is only used at most once in some path, because $v = x$ when $a \neq 0$, and $v' = y$ when $b \neq 0$. This implies that $(H_0)_{++}$ is block-diagonal with respect to each $P_{x,y}$, and thus also block-diagonal on $\mathcal{H}$.

In fact, observe that $(H_0)_{++}$ restricted on (nodes of) $P_{x,y}$ is a $(2\mathfrak{n}+3) \times (2\mathfrak{n}+3)$ tridiagonal matrix of the form

$$-\frac{\alpha}{a_{x,y}} \cdot \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}. \tag{6.72}$$

By standard calculations (see e.g., [227, Page 20]), (6.72) has eigenvalues

$$-\frac{\alpha}{a_{x,y}}\left(2 - 2\cos\left(\frac{k\pi}{2\mathfrak{n}+4}\right)\right) = -\frac{\alpha}{a_{x,y}} \cdot 4\sin^2\left(\frac{k\pi}{4\mathfrak{n}+8}\right) \quad \text{for } k \in [2\mathfrak{n}+3]. \tag{6.73}$$

This readily proves (6.60). In addition, the minimal eigenvalue of (6.72) is

$$\begin{aligned} -\alpha \cdot \frac{1}{a_{x,y}} \cdot 4\sin^2\left(\frac{\pi}{4\mathfrak{n}+8}\right) &= \frac{M+1}{4\sin^2\left(\frac{\pi}{4\mathfrak{n}+8}\right)} \cdot \frac{1}{-a_{x,y}} \cdot 4\sin^2\left(\frac{\pi}{4\mathfrak{n}+8}\right) \quad \text{(by (6.59))} \\ &\geq \frac{M+1}{4\sin^2\left(\frac{\pi}{4\mathfrak{n}+8}\right)} \cdot \frac{1}{M+1} \cdot 4\sin^2\left(\frac{\pi}{4\mathfrak{n}+8}\right) \\ &= 1, \end{aligned} \tag{6.74}$$

which implies that $(H_0)_{++}$ has minimal eigenvalue at least 1. Standard calculations (see e.g., [228,

Lemma 3]) also show that the inverse of (6.72) is

$$-\frac{a_{x,y}}{\alpha}\cdot\frac{1}{2\mathfrak{n}+4}\sum_{i,j\in[2\mathfrak{n}+3]}|i\rangle\langle j|\cdot\min\{i,j\}(2\mathfrak{n}+4-\max\{i,j\}). \tag{6.75}$$

Now we compute $(V_{\text{main}})_{-+}H_0^{-1}(V_{\text{main}})_{+-}$ as follows: For $x<y$, we have

$$\begin{aligned}
&\langle y,y,0^s,0^s,0^{\mathfrak{n}}|(V_{\text{main}})_{-+}H_0^{-1}(V_{\text{main}})_{+-}|x,x,0^s,0^s,0^{\mathfrak{n}}\rangle\\
&=\alpha(2\mathfrak{n}+4)\,\langle v_{x,y,2\mathfrak{n}+3}|H_0^{-1}|v_{x,y,1}\rangle && \text{(by definition of } H_0 \text{ and } V_{\text{main}})\\
&=\alpha(2\mathfrak{n}+4)\left(-\frac{a_{x,y}}{\alpha}\cdot\frac{1}{2\mathfrak{n}+4}\right) && \text{(by (6.75))}\\
&=-a_{x,y}.
\end{aligned} \tag{6.76}$$

On the other hand, for diagonal terms,

$$\begin{aligned}
&\langle x,x,0^s,0^s,0^{\mathfrak{n}}|(V_{\text{main}})_{-+}H_0^{-1}(V_{\text{main}})_{+-}|x,x,0^s,0^s,0^{\mathfrak{n}}\rangle\\
&=\sum_{\substack{y:\,y>x\\(x,y)\in E}}\alpha(2\mathfrak{n}+4)\,\langle v_{x,y,1}|H_0^{-1}|v_{x,y,1}\rangle\\
&\quad+\sum_{\substack{y:\,y<x\\(x,y)\in E}}\alpha(2\mathfrak{n}+4)\,\langle v_{y,x,2\mathfrak{n}+3}|H_0^{-1}|v_{y,x,2\mathfrak{n}+3}\rangle && \text{(by definition of } H_0 \text{ and } V_{\text{main}})\\
&=\sum_{\substack{y:\,y\neq x\\(x,y)\in E}}\alpha(2\mathfrak{n}+4)\left(-\frac{a_{x,y}}{\alpha}\cdot\frac{2\mathfrak{n}+3}{2\mathfrak{n}+4}\right) && \text{(by (6.75))}\\
&=-\sum_{\substack{y:\,y\neq x\\(x,y)\in E}}(2\mathfrak{n}+3)\cdot a_{x,y}.
\end{aligned} \tag{6.77}$$

This completes the proof of Claim 6.5.3. □

Proof of Claim 6.5.5. To analyze the Lipschitz constant of $H'(t)$, consider our reduction $K(t)\mapsto$

$H'(t)$. We know that $K(t)$ is $O(\mathfrak{m})$-Lipschitz since it is piecewise linear and has norm bounded by $O(\mathfrak{m})$ as in Theorem 6.4.2. Therefore, $\|K(t)\|_{\max}$ is also $O(\mathfrak{m})$-Lipschitz due to Fact 6.3.1 (i).

Recall the definition in (6.58) that

$$a_{x,y}(t) = \langle x|K(t)|y\rangle - \varepsilon/(2s). \tag{6.78}$$

This $a_{x,y}(t)$ is $O(\mathfrak{m})$-Lipschitz and $1/a_{x,y}(t)$ is $O(\mathfrak{m}s^2/\varepsilon^2)$-Lipschitz thanks to the small offset.

Now, we analyze the Lipschitz continuity of the individual terms in $H'(t)$.

- Looking into the construction of $H_0(t)$, we observe that $H_0(t)$ is block-diagonal with each block taking value in $\left\{0, 1, -\frac{\alpha}{a_{x,y}(t)}T\right\}$, where T is a constant matrix (6.72) in the proof of Claim 6.5.3. Given that $\alpha \le O(\mathfrak{n}^2 M)$ by (6.59), we conclude that $H_0(t)$ is $O(\mathfrak{n}^2 M\mathfrak{m}s^2/\varepsilon^2)$-Lipschitz.

- $V_{\text{main}}(t)$ has exactly $s-1$ nonzero entries per row, each equal to $-\sqrt{\alpha(2\mathfrak{n}+4)}$ by construction. Thus, it is $O(\mathfrak{n}^{2.5}M)$-Lipschitz.

- $V_{\text{extra}}(t)$ is diagonal, and its Lipschitz constant is $O(sM)$ by its definition in (6.62).

Finally, since $\Delta \le \text{poly}(1/\varepsilon, \mathfrak{n}, s, M)$, we conclude that $H'(t) = \Delta H_0(t) + \Delta^{1/2}V_{\text{main}}(t) + V_{\text{extra}}(t)$ is $\text{poly}(\mathfrak{n}, s, \mathfrak{m}, M, 1/\varepsilon)$-Lipschitz as claimed. □

6.6 Reduction II: Proof of Lemma 6.4.4

In this section we prove Lemma 6.4.4 (restated below).

Lemma 6.4.4. *Let $n_{\mathsf{B}} \coloneqq \Theta(\mathfrak{m})$. There exist a parameter $\Delta_{\mathsf{B}} \in [\mathfrak{m}, \text{poly}(\mathfrak{m})]$ and a family of n_{B}-qubit*

stoquastic Hamiltonian paths $H(t)$ for $t \in [0,1]$ of the form

$$H(t) = X_{\mathsf{B}} + D(t) \quad \textit{and} \quad X_{\mathsf{B}} \coloneqq -\Delta_{\mathsf{B}} \sum_{i \in [n_{\mathsf{B}}]} X_i, \tag{6.33}$$

where $D(t)$ is diagonal, such that the following holds given oracle access to $D(t)$:

(i) $D(0) = 0$, $D(1) = -3\Delta_{\mathsf{B}} n_{\mathsf{B}} \, |u\rangle\langle u|$ *for some* $u \in \{0,1\}^{n_{\mathsf{B}}}$, *and* $D(t)$ *is* $\mathrm{poly}(\mathfrak{n})$*-Lipschitz.*

(ii) Any classical algorithm requires at least $\exp\left(\mathfrak{n}^{1/5-o(1)}\right)$ *queries to find* u *with probability greater than* $\exp\left(-\mathfrak{n}^{1/5-o(1)}\right)$.

(iii) $H(t)$ *has spectral gap* $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $\mathrm{poly}(\mathfrak{n})$.

The proof consists of two major steps. In the first step, we apply a second-order perturbative reduction (Lemma 6.3.11) to transform the Hamiltonian path in Lemma 6.4.3 into a Hamiltonian path composed of Hamiltonians of the form $H'(t) = X_{\mathsf{B}} + D'(t)$, where $D'(t)$ is diagonal and $X_{\mathsf{B}} = -\Delta_{\mathsf{B}} \sum_{i \in [n_{\mathsf{B}}]} X_i$ is some fixed X-type interaction. In the second step, we further decorate the obtained Hamiltonian path by appending additional paths at both its beginning and end, ensuring that the final Hamiltonian path starts with X_{B} and terminates at $X_{\mathsf{B}} - 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \, |u\rangle\langle u|$, as stated in Lemma 6.4.4 (i).

Organization. The first step is formalized in Section 6.6.1 and the second step is in Section 6.6.2. In Section 6.6.3 we put them together and prove Lemma 6.4.4. The missing proofs from this section can be found in Section 6.6.4.

Notation. We (re)define some notation. Let n be n_{A} and let $K(t)$ for $t \in [0,1]$ be the n-qubit Hamiltonian path in Lemma 6.4.3. Write $K(1) = -\mathfrak{m} \, |w\rangle\langle w|$, and reserve the notation u such that

it is consistent with the statement of Lemma 6.4.4. Let M denote an upper bound on $\|K(t)\|$ and $\mathfrak{m} \leq M \leq \mathrm{poly}(n) = \mathrm{poly}(\mathfrak{n})$.

6.6.1 First Step – Hypercube Hamiltonians to TFD Hamiltonians

Let $n_{\mathsf{B}} \coloneqq 2n$. Below we show how to reduce $K(t)$ to an n_{B}-qubit Hamiltonian $H'(t) = X_{\mathsf{B}} + D'(t)$ for any $t \in [0,1]$ such that $H'(t)$ simulates $K(t)$ in the sense of Definition 6.3.3. This reduction is one-to-one and thus for convenience we write $K \coloneqq K(t)$ and $H' \coloneqq H'(t)$.

To establish the proof using Lemma 6.3.11, we denote $\mathcal{K}$ and $\mathcal{H}$ by the Hilbert spaces associated with K and H' respectively. Define the encoding isometry $\mathcal{E} \colon \mathcal{K} \to \mathcal{H}$ by

$$\mathcal{E}\,|x\rangle = |x,x\rangle \quad \text{for } x \in \{0,1\}^n. \tag{6.79}$$

Let $\mathcal{H}_- \coloneqq \operatorname{Im}\mathcal{E}$ and we write $\mathcal{H} = \mathcal{H}_- \oplus \mathcal{H}_+$. In light of Lemma 6.3.11, our plan is to construct

$$H' \coloneqq \Delta H_0 + \Delta^{1/2} V_{\mathrm{main}} + V_{\mathrm{extra}}, \tag{6.80}$$

for some $\Delta, H_0, V_{\mathrm{main}}, V_{\mathrm{extra}}$ such that $V_{\mathrm{extra}} - (V_{\mathrm{main}})_{-+} H_0^{-1} (V_{\mathrm{main}})_{+-}$ is close to $\overline{K}$.

Entries of K with small offsets. Let $\varepsilon \coloneqq 1/2$. For any $x, y \in \{0,1\}^n$ with $\mathrm{dist}(x,y) \leq 1$, define

$$a_{x,y} \coloneqq \langle x|K|y\rangle - \frac{\varepsilon}{2(n+1)}. \tag{6.81}$$

Since K is stoquastic, the small offset ensures that $a_{x,y} < \frac{\varepsilon}{2(n+1)}$ whenever $\mathrm{dist}(x,y) = 1$.

Construction of H_0. We define the diagonal matrix H_0 as follows:

$$\langle x, y|H_0|x, y\rangle = \begin{cases} 0 & x = y, \\ -(M+1)\cdot a_{x,y}^{-1} & \mathrm{dist}(x,y) = 1, \\ (M+1)\left(\mathfrak{m} + \frac{\varepsilon}{2(n+1)}\right)^{-1} \cdot \mathrm{dist}(x,y) & \text{otherwise.} \end{cases} \tag{6.82}$$

We note the following easy observation.

Claim 6.6.1. $(H_0)_{--} = 0$ *and eigenvalues of* $(H_0)_{++}$ *are at least* 1.

Proof. Note that $\mathcal{H}_-$ is spanned by $|x, x\rangle, x \in \{0,1\}^n$. Hence $(H_0)_{--} = 0$. For $x \neq y$, we have

$$a_{x,y} \geq (M+1)\left(M + \frac{\varepsilon}{2(n+1)}\right)^{-1} \geq (M+1)(M+1)^{-1} = 1, \tag{6.83}$$

which means that eigenvalues of $(H_0)_{++}$ are at least 1. □

Construction of V_{main} and V_{extra}. We construct V_{main} and V_{extra} as follows:

$$V_{\mathrm{main}} \coloneqq -\sqrt{\frac{M+1}{2}} \sum_{i\in[2n]} X_i, \quad V_{\mathrm{extra}} \coloneqq \sum_x \left(a_{x,x} - \sum_{y:\ \mathrm{dist}(x,y)=1} a_{x,y}\right) |x,x\rangle\langle x,x| \,. \tag{6.84}$$

By a direct calculation, which we defer to Section 6.6.4, we have the following desired relation.

Claim 6.6.2. $(V_{\mathrm{main}})_{--} = 0$ *and*

$$(V_{\mathrm{extra}})_{--} - (V_{\mathrm{main}})_{-+} H_0^{-1} (V_{\mathrm{main}})_{+-} = \overline{K} - \frac{\varepsilon}{2(n+1)} \cdot \overline{E}, \tag{6.85}$$

where $\overline{A} \coloneqq \mathcal{E} A \mathcal{E}^\dagger$ *for any* A *and* $E \coloneqq \sum_{x,y:\ \mathrm{dist}(x,y)\leq 1} |x\rangle\langle y|$.

Parameters in the reduction. Note that $\left\|\overline{E}\right\| = \|E\| \leq \|E\|_1 = n+1$ by Fact 6.3.1 (ii). Therefore all preconditions of Lemma 6.3.11 are met. Applying Lemma 6.3.11, we have that for any $\Delta \geq \mathrm{poly}(1/\varepsilon, \Lambda)$, $(H', \mathcal{E})$ simulates K with error ε, where

$$\Lambda = \max\{\|V_{\mathrm{main}}\|, \|V_{\mathrm{extra}}\|\} \leq O(nM) = \mathrm{poly}(\mathfrak{n}). \tag{6.86}$$

Hence it suffices to ensure $\Delta \geq \mathrm{poly}(1/\varepsilon, \mathfrak{n})$. For H_0, it is diagonal and

$$\|H_0\| \leq (M+1) \cdot \frac{2(n+1)}{\varepsilon} \cdot n \leq \mathrm{poly}(1/\varepsilon, n, M) = \mathrm{poly}(1/\varepsilon, \mathfrak{n}) \tag{6.87}$$

by (6.82) and (6.81). Combining these estimates we get $\|H'\| \leq \mathrm{poly}(1/\varepsilon, \mathfrak{n}, \Delta)$.

Finally, note that H_0 and V_{extra} are diagonal, while $V_{\mathrm{main}} = -\sqrt{\frac{M+1}{2}} \sum_{i\in[2n]} X_i$. Thus, the Hamiltonian H' can be written in the desired form:

$$H' = X_{\mathsf{B}} + D', \quad X_{\mathsf{B}} \coloneqq -\Delta_{\mathsf{B}} \sum_{i\in[2n]} X_i = \Delta^{1/2} V_{\mathrm{main}}, \quad D' \coloneqq \Delta H_0 + V_{\mathrm{extra}}, \tag{6.88}$$

where we set

$$\Delta_{\mathsf{B}} = \sqrt{\frac{\Delta(M+1)}{2}}. \tag{6.89}$$

Since $\varepsilon = 1/2$, the parameter Δ determined implicitly in (6.89) can be chosen to satisfy the requirement $\Delta \geq \mathrm{poly}(1/\varepsilon, \mathfrak{n})$, provided that $\Delta_{\mathsf{B}} \geq \mathrm{poly}(\mathfrak{n})$.

With the spectral property of $K(t)$ from Lemma 6.4.3 (iii), the fact that $(H', \mathcal{E})$ simulates K with error ε gives us the following claim.

Claim 6.6.3. *$H'(t)$ has spectral gap at least $\Omega(\sqrt{\mathfrak{m}})$ and norm $\mathrm{poly}(\mathfrak{n}, \Delta_{\mathsf{B}})$.*

We can also analyze the Lipschitz condition for $H'(t)$.

Claim 6.6.4. *$H'(t)$ is $\mathrm{poly}(\mathfrak{n}, \Delta_{\mathsf{B}})$-Lipschitz.*

Proof. Consider our reduction $K(t) \mapsto H'(t)$ where $K(t)$ is $\mathrm{poly}(\mathfrak{n})$-Lipschitz by Lemma 6.4.3 (i). Therefore, $\|K(t)\|_{\max}$ is also $\mathrm{poly}(\mathfrak{n})$-Lipschitz by Fact 6.3.1 (i). Hence, recalling (6.81), we know that $a_{x,y}(t)$ is $\mathrm{poly}(\mathfrak{n})$-Lipschitz and $1/a_{x,y}(t)$ is $O(\mathrm{poly}(\mathfrak{n})/\varepsilon^2)$-Lipschitz for any $x, y \in \{0,1\}^n$. Now we analyze the Lipschitz continuity of the individual terms in $H'(t)$.

- $H_0(t)$ is diagonal. Therefore it is $O(M\,\mathrm{poly}(\mathfrak{n})/\varepsilon^2) = \mathrm{poly}(\mathfrak{n})$-Lipschitz by its definition in (6.82).
- V_{main} remains constant with respect to t.
- $V_{\mathrm{extra}}(t)$ is diagonal, and its Lipschitz constant is $O(nM) = \mathrm{poly}(\mathfrak{n})$ by its definition in (6.84).

Finally, since $\Delta \le O({\Delta_{\mathsf{B}}}^2)$ by (6.89), we conclude that $H'(t) = \Delta H_0(t) + \Delta^{1/2} V_{\mathrm{main}} + V_{\mathrm{extra}}(t)$ is $\mathrm{poly}(\mathfrak{n}, \Delta) = \mathrm{poly}(\mathfrak{n}, \Delta_{\mathsf{B}})$-Lipschitz. □

6.6.2 Second Step – Cleaner Start and End

Inheriting the notation from Section 6.6.1, we first append a linear path at the end of $H'(1)$. Based on the construction of $H'(t)$ from Section 6.6.1, we write

$$H'(1) = \Delta H_0 + V \quad \text{and} \quad V \coloneqq \Delta^{1/2} V_{\mathrm{main}} + V_{\mathrm{extra}}. \tag{6.90}$$

Recall that $K(1) = -\mathfrak{m}\,|w\rangle\langle w|$. Let $|u\rangle := \mathcal{E}\,|w\rangle$ where $\mathcal{E}$ is the encoding isometry from (6.79). The path to be appended is

$$H'(1) \longrightarrow H'(1) - 3\Delta_{\mathsf{B}} n_{\mathsf{B}}\,|u\rangle\langle u| \longrightarrow X_{\mathsf{B}} - 3\Delta_{\mathsf{B}} n_{\mathsf{B}}\,|u\rangle\langle u|\,. \tag{P4}$$

We will show that (P4) preserves the spectral gap. The first half uses Taylor expansion of perturbative reduction (Lemma 6.3.9) and the second half is a direct calculation. The proof of Claim 6.6.5 is deferred to Section 6.6.4.

Claim 6.6.5. *Assume $\Delta_{\mathsf{B}} = \mathrm{poly}(\mathfrak{n})$. The Hamiltonian path (P4) has spectral gap $\Omega(n\Delta_{\mathsf{B}})$ and norm $\mathrm{poly}(\mathfrak{n})$. Consequently, it is $\mathrm{poly}(\mathfrak{n})$-Lipschitz by piecewise linearity.*

To decorate the start, we explicitly calculate $H'(0)$. Since $K(0) = -\mathfrak{m}\sum_{i\in[n]} X_i$, we have

$$\begin{aligned}
H'(0) &= X_{\mathsf{B}} + \Delta H_0(0) + V_{\mathrm{extra}}(0) && \text{(by (6.88))}\\
&= X_{\mathsf{B}} + \Delta \sum_{x,y} \alpha\cdot \mathrm{dist}(x,y)\,|x,y\rangle\langle x,y| + \sum_x \beta\,|x,x\rangle\langle x,x| && \text{(by (6.82) and (6.84))}\\
&= X_{\mathsf{B}} + \frac{1}{2}\Delta\alpha\left(nI - \sum_{i\in[n]} Z_i\otimes Z_{n+i}\right) + \beta I_{--}, && \text{(since } Z_i\otimes Z_{n+i}\,|x,y\rangle = (-1)^{x_i+y_i}\,|x,y\rangle\text{)}
\end{aligned}$$

where

$$\alpha = (M+1)\left(\mathfrak{m} + \frac{\varepsilon}{2(n+1)}\right)^{-1} \quad \text{and} \quad \beta = n\mathfrak{m} + \frac{(n-1)\varepsilon}{2(n+1)}. \tag{6.91}$$

We now append the following Hamiltonian path to the beginning of $H'(0)$:

$$X_{\mathsf{B}} \longrightarrow X_{\mathsf{B}} + \frac{1}{2}\Delta\alpha\left(nI - \sum_{i\in[n]} Z_i\otimes Z_{n+i}\right) \longrightarrow H'(0) \tag{P5}$$

Claim 6.6.6. *The Hamiltonian path (P5) has spectral gap* $\Omega(\min\{\mathfrak{m}, \Delta_{\mathsf{B}}\})$ *and norm* $\mathrm{poly}(\mathfrak{n}, \Delta_{\mathsf{B}})$. *Consequently, it is* $\mathrm{poly}(\mathfrak{n}, \Delta_{\mathsf{B}})$*-Lipschitz by piecewise linearity.*

Proof. We have shown that $(H'(0), \mathcal{E})$ simulates $K(0)$ with error $\varepsilon = 1/2$ in Section 6.6.1. Therefore, every Hamiltonian on the second half of the Hamiltonian path (P5), which has format $H'(0) - \lambda\beta I_{--}$, simulates $K(0) - \lambda\beta I$ by the same argument. Since $K(0) = -\mathfrak{m}\sum_{i\in[n]} X_i$ has spectral gap $\mathfrak{m}$, the spectral gap of the second half of (P5) is at least $\mathfrak{m} - 2\varepsilon = \Omega(\mathfrak{m})$ by Lemma 6.3.5.

For the first half of (P5), applying Fact 6.3.8, the spectral gap is identical to that of the Hamiltonian path

$$-\Delta_{\mathsf{B}}(X \otimes I + I \otimes X) \longrightarrow -\Delta_{\mathsf{B}}(X \otimes I + I \otimes X) - \frac{1}{2}\Delta\alpha(Z \otimes Z). \tag{6.92}$$

Let $L(\lambda) \coloneqq -\Delta_{\mathsf{B}}(X \otimes I + I \otimes X) - \lambda \cdot \frac{1}{2}\Delta\alpha(Z \otimes Z)$. Note that $L(\lambda)$ is block-diagonal in the Bell basis $\{|00\rangle \pm |11\rangle, |01\rangle \pm |10\rangle\}$ and its eigenvalues are $\pm\frac{1}{2}\lambda\Delta\alpha$ and $\pm\sqrt{\left(\frac{1}{2}\lambda\Delta\alpha\right)^2 + 4{\Delta_{\mathsf{B}}}^2}$. Therefore, the spectral gap is

$$\begin{aligned}
\Omega\left(\min\left\{\Delta_{\mathsf{B}}, \frac{{\Delta_{\mathsf{B}}}^2}{\lambda\Delta\alpha}\right\}\right) &= \Omega\left(\min\left\{\Delta_{\mathsf{B}}, \frac{\Delta(M+1)/2}{\lambda\Delta(M+1)\left(\mathfrak{m} + \frac{\varepsilon}{2(n+1)}\right)^{-1}}\right\}\right) && \text{(by (6.91) and (6.89))} \\
&= \Omega\left(\min\left\{\Delta_{\mathsf{B}}, \frac{\mathfrak{m}}{\lambda}\right\}\right) = \Omega\left(\min\{\Delta_{\mathsf{B}}, \mathfrak{m}\}\right) && \text{(since } \lambda \in [0,1]\text{)}
\end{aligned}$$

as claimed. □

6.6.3 Putting Everything Together

Finally we prove Lemma 6.4.4 by combining the Hamiltonian path in Section 6.6.1 and the additional decoration in Section 6.6.2.

Proof of Lemma 6.4.4. Recall Hamiltonian paths $H'([0,1])$ from Section 6.6.1 and (P5), (P4) from Section 6.6.2. We sequentially connect (P5), $H'([0,1])$, and (P4). This forms a Hamiltonian path $H : [0,3] \to \mathcal{H}$. By rescaling the domain linearly from $[0,3]$ to $[0,1]$, we obtain the final path $H(t)$.

Set $\Delta_{\mathsf{B}} \in [\mathfrak{m}, \mathrm{poly}(\mathfrak{m})]$ and recall that $\mathfrak{m}, \mathfrak{n}, n$ are polynomially related. By Claim 6.6.3, Claim 6.6.4, Claim 6.6.5, and Claim 6.6.6, $H(t)$ has spectral gap $\Omega(\sqrt{\mathfrak{m}})$, norm $\mathrm{poly}(\mathfrak{n})$, and it is $\mathrm{poly}(\mathfrak{n})$-Lipschitz. This verifies Lemma 6.4.4 (i) and Lemma 6.4.4 (iii).

For Lemma 6.4.4 (ii), we note that any query to $\langle x|H(t)|y\rangle$ reduces to queries of the (x,y) entries of $H_0(t')$ and $V_{\mathrm{extra}}(t')$ for some $t' := t'(t)$. By our construction, these can be implemented using $O(1)$ and $O(n)$ queries, respectively, to the entries of $K(t')$. Consequently, any q-query classical algorithm that finds $|u\rangle$ can also find $|w\rangle$, since $|w\rangle = \mathcal{E}^\dagger |u\rangle$. This transformation results in an $O(qn)$-query classical algorithm with oracle access to $K(t')$. By Lemma 6.4.3 (ii), we obtain $O(qn) \geq \exp\big(\mathfrak{n}^{1/5-o(1)}\big)$. Since $n = n_{\mathsf{A}} = \Theta(\mathfrak{m}) = \mathfrak{n}^{16/5-o(1)}$, this implies that $q \geq \exp\big(\mathfrak{n}^{1/5-o(1)}\big)$. □

6.6.4 Deferred Proofs

Proof of Claim 6.6.2. The first term is obvious by the definition of $\mathcal{H}_-$. For the second term, observe that H_0 is diagonal and hence

$$\langle x,y|(H_0^{-1})_{++}|x,y\rangle = \begin{cases} -a_{x,y} \cdot (M+1)^{-1} & \mathrm{dist}(x,y) = 1, \\ \left(\mathfrak{m} + \frac{\varepsilon}{2(n+1)}\right)(M+1)^{-1}\,\mathrm{dist}(x,y)^{-1} & \mathrm{dist}(x,y) > 1. \end{cases} \tag{6.93}$$

Therefore

$$\langle x,x|(V_{\text{main}})_{-+}H_0^{-1}(V_{\text{main}})_{+-}|x,x\rangle = \frac{M+1}{2}\cdot \sum_{\substack{(x',y')\\ \text{dist}(x,x')+\text{dist}(x,y')=1}} \langle x',y'|(H_0^{-1})_{++}|x',y'\rangle \tag{6.94}$$

$$= -\sum_{y:\ \text{dist}(x,y)=1} a_{x,y}. \tag{6.95}$$

For $\text{dist}(x,y)=1$, we have

$$\langle y,y|(V_{\text{main}})_{-+}H_0^{-1}(V_{\text{main}})_{+-}|x,x\rangle = \frac{M+1}{2}\cdot\big(\langle x,y|H_0^{-1}|x,y\rangle + \langle y,x|H_0^{-1}|y,x\rangle\big) \tag{6.96}$$

$$= -a_{x,y}. \tag{6.97}$$

For $\text{dist}(x,y)>1$, we have

$$\langle y,y|(V_{\text{main}})_{-+}H_0^{-1}(V_{\text{main}})_{+-}|x,x\rangle = 0. \tag{6.98}$$

In summary, we have

$$(V_{\text{extra}})_{--} - (V_{\text{main}})_{-+}H_0^{-1}(V_{\text{main}})_{+-} = \sum_{\substack{(x,y)\\ \text{dist}(x,y)\le 1}} a_{x,y}\,|x,x\rangle\langle y,y| \tag{6.99}$$

$$= \sum_{\substack{(x,y)\\ \text{dist}(x,y)\le 1}} \left(\langle x|K|y\rangle - \frac{\varepsilon}{2(n+1)}\right)|x,x\rangle\langle y,y| \quad \text{(by (6.81))}$$

$$= \overline{K} - \frac{\varepsilon}{2(n+1)}\cdot\overline{E} \tag{6.100}$$

as claimed. □

Proof of Claim 6.6.5. We only analyze the spectral gap of this Hamiltonian path. For the first half of (P4), we define for any $\lambda \in [0,1]$

$$L(\lambda) \coloneqq H'(1) - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| = \Delta H_0 + (V - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right|). \tag{6.101}$$

Recall from (6.89) that $\Delta_{\mathsf{B}} = \sqrt{\frac{\Delta(M+1)}{2}}$ and $M = \mathrm{poly}(\mathfrak{n})$. Setting $\Delta_{\mathsf{B}} = \mathrm{poly}(\mathfrak{n})$ sufficiently large, we have $\Delta > 16\|V - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right|\|$. By Lemma 6.3.9, the low-energy part of $L(\lambda)$ is given by

$$\underbrace{(V_{--} - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right|) - \Delta^{-1} V_{-+} H_0^{-1} V_{+-}}_{G(\lambda)} + O(\Delta^{-2}\|V - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right|\|^3). \tag{6.102}$$

The $O(\cdot)$ term above is $o(1)$ as $\Delta_{\mathsf{B}} = \mathrm{poly}(\mathfrak{n})$. Therefore, the spectral gap of $L(\lambda)$ is $o(1)$-close to the spectral gap of

$$\begin{aligned}
G(\lambda) &= (V_{--} - \Delta^{-1} V_{-+} H_0^{-1} V_{+-}) - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| && (6.103)\\
&= \overline{K} - \frac{\varepsilon}{2(n+1)} \overline{E} - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| && \text{(by Claim 6.6.2)}\\
&= -\mathfrak{m} \cdot \mathcal{E} \left|w\right\rangle\!\left\langle w\right| \mathcal{E}^\dagger - \frac{\varepsilon}{2(n+1)} \overline{E} - \lambda \cdot 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| && \text{(since } K(1) = -\mathfrak{m} \left|w\right\rangle\!\left\langle w\right|)\\
&= -(\mathfrak{m} + 3\lambda\Delta_{\mathsf{B}} n_{\mathsf{B}}) \left|u\right\rangle\!\left\langle u\right| - \frac{\varepsilon}{2(n+1)} \overline{E}. && \text{(since } \mathcal{E} \left|w\right\rangle = \left|u\right\rangle)
\end{aligned}$$

Recall that $E \coloneqq \sum_{x,y:\ \mathrm{dist}(x,y)\le 1} \left|x\right\rangle\!\left\langle y\right|$ and $\left\|\overline{E}\right\| = \|E\| \le n+1$. Therefore, the spectral gap of $L(\lambda)$, i.e., the first half of the Hamiltonian path (P4), is at least $\mathfrak{m} + 3\lambda\Delta_{\mathsf{B}} n_{\mathsf{B}} - O(1) = \Omega(n\Delta_{\mathsf{B}})$ by Fact 6.3.6.

For the second half of (P4), we rewrite it as

$$X_{\mathsf{B}} + \Delta H_0 + V_{\text{extra}} - 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| \longrightarrow X_{\mathsf{B}} - 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| . \tag{6.104}$$

By Fact 6.3.6, its spectral gap is at least $-2\|X_{\mathsf{B}}\|$ plus the spectral gap of the Hamiltonian path

$$\Delta H_0 + V_{\text{extra}} - 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| \longrightarrow -3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right| , \tag{6.105}$$

which is a Hamiltonian path of diagonal Hamiltonians. Let $\lambda \in [0,1]$. By Claim 6.6.1, we know that $(\lambda \Delta H_0)_{--} = 0$ and eigenvalues of $(\lambda \Delta H_0)_{++}$ are always non-negative. By (6.84) and (6.81) and the fact that $\overline{K} = -\mathfrak{m} \left|u\right\rangle\!\left\langle u\right|$, we know that

$$\lambda V_{\text{extra}} = \lambda \cdot \left(-\mathfrak{m} \left|u\right\rangle\!\left\langle u\right| - \frac{\varepsilon(n-1)}{2(n+1)} \sum_x \left|x,x\right\rangle\!\left\langle x,x\right| \right) = \lambda \cdot \left(-\mathfrak{m} \left|u\right\rangle\!\left\langle u\right| - \frac{\varepsilon(n-1)}{2(n+1)} \cdot I_{--} \right) . \tag{6.106}$$

This means the spectral gap of $\lambda \left(\Delta H_0 + V_{\text{extra}}\right) - 3\Delta_{\mathsf{B}} n_{\mathsf{B}} \left|u\right\rangle\!\left\langle u\right|$ (i.e., (6.105)) is $3\Delta_{\mathsf{B}} n_{\mathsf{B}} + \lambda\mathfrak{m}$. Since $\|X_{\mathsf{B}}\| = \left\|-\Delta_{\mathsf{B}} \sum_{i \in n_{\mathsf{B}}} X_i\right\| = \Delta_{\mathsf{B}} n_{\mathsf{B}}$, (6.104) has spectral gap at least $3\Delta_{\mathsf{B}} n_{\mathsf{B}} - 2\Delta_{\mathsf{B}} n_{\mathsf{B}} = \Delta_{\mathsf{B}} n_{\mathsf{B}}$. This verifies the spectral gap of the second half of (P4) and completes the proof of Claim 6.6.5. □

6.7 Toy TFI Hamiltonian: Proof of Lemma 6.4.6

In this section we prove Lemma 6.4.6 (restated below), which constructs a toy TFI Hamiltonian that gradually shifts ground states.

Lemma 6.4.6. *Assume $\ell \geq 1$ and $\varepsilon_{\mathsf{L}} \in (0, 1/10]$. There exists a $\Delta_{\mathsf{L}} = \text{poly}(\ell, 1/\varepsilon_{\mathsf{L}})$ such that the following holds for $t \in \mathbb{R}$:*

(i) $\|H_{\mathsf{L}}(t)\| \le (|t|+1)\,\mathrm{poly}(\ell, 1/\varepsilon_{\mathsf{L}})$.

(ii) *The spectral gap of* $H_{\mathsf{L}}(t)$ *is at least* $1/2$.

(iii) *Let* $|\psi(t)\rangle$ *be a ground state of* $H_{\mathsf{L}}(t)$. *If* $t \in [0, \ell+1]$, *then* $|\psi(t)\rangle$ *is* $O(\varepsilon_{\mathsf{L}})$*-close to a normalized state that is a linear combination of* $|\boldsymbol{a}\rangle$ *and* $|\boldsymbol{b}\rangle$, *where* $a := \max\{\lfloor t - 1/2 \rfloor, 0\}$ *and* $b := \min\{\lceil t - 1/2 \rceil, \ell\}$.

(iv) *For any* $\lambda \in [0, 2]$, *the Hamiltonian*

$$H_{\mathsf{L}}(\lambda, t) := Q_{\mathsf{L}} + \lambda X_{\mathsf{L}} + Z_{\mathsf{L}}(t) \tag{6.38}$$

satisfies that its spectral gap is at least $1/2$ *and its ground state is* $O(\varepsilon_{\mathsf{L}})$*-close to* $|\mathbf{0}\rangle$ *(resp.,* $|\boldsymbol{\ell}\rangle$*) if* $t \le 0$ *(resp.,* $t \ge \ell + 1$*).*

Notation. We recall related definitions here. For each $j = 0, 1, \ldots, \ell$, we define $|\boldsymbol{j}\rangle := \left|1^j 0^{\ell-j}\right\rangle$. Let $\varepsilon_{\mathsf{L}}, \Delta_{\mathsf{L}} > 0$ be parameters to be determined. Define the TFI Hamiltonian $H_{\mathsf{L}}(t) = Q_{\mathsf{L}} + X_{\mathsf{L}} + Z_{\mathsf{L}}(t)$ where $X_{\mathsf{L}} = -\sum_{i\in[\ell]} X_i$ and

$$Q_{\mathsf{L}} = \Delta_{\mathsf{L}} \left((\ell - 1)I + Z_1 - Z_\ell - \sum_{i\in[\ell-1]} Z_i Z_{i+1} \right), \quad Z_{\mathsf{L}}(t) = \frac{1}{\varepsilon_{\mathsf{L}}} \sum_{i\in[\ell]} (i - t)(I - Z_i). \tag{6.107}$$

We first verify the easy norm condition.

Proof of Lemma 6.4.6 (i). By the definition of $H_{\mathsf{L}}(t)$, we have

$$\|H_{\mathsf{L}}(t)\| \le \|Q_{\mathsf{L}}\| + \|X_{\mathsf{L}}\| + \|Z_{\mathsf{L}}(t)\| \tag{6.108}$$

$$\leq \Delta_{\mathsf{L}} \cdot 10\ell + \ell + \frac{2}{\varepsilon_{\mathsf{L}}} \sum_{i\in[\ell]} (i + |t|) \qquad \text{(by (6.107))}$$

$$\leq (|t| + 1)\, \mathrm{poly}(\ell, 1/\varepsilon_{\mathsf{L}}), \tag{6.109}$$

where we use the fact that $\ell \geq 1$, $\varepsilon_{\mathsf{L}} \in (0, 1]$, and $\Delta_{\mathsf{L}} = \mathrm{poly}(\ell, 1/\varepsilon_{\mathsf{L}})$ for the last inequality. □

To verify other properties, we expand $H_{\mathsf{L}}(\lambda, t)$ as follows:

$$H_{\mathsf{L}}(\lambda, t) = Q_{\mathsf{L}} + \lambda X_{\mathsf{L}} + Z_{\mathsf{L}}(t) \tag{6.110}$$

$$= -\lambda \sum_{i\in[\ell]} X_i + \Delta_{\mathsf{L}} \left((\ell - 1)I + Z_1 - Z_\ell - \sum_{i\in[\ell-1]} Z_i Z_{i+1} \right) + \frac{1}{\varepsilon_{\mathsf{L}}} \sum_{i\in[\ell]} (i - t)(I - Z_i) \tag{6.111}$$

$$= -\lambda \sum_{(x,x'):\ \mathrm{dist}(x,x')=1} |x\rangle\langle x'| + \sum_{x\in\{0,1\}^\ell} (\alpha_x + \beta_x)\, |x\rangle\langle x|\,, \tag{6.112}$$

where

$$\alpha_x = \Delta_{\mathsf{L}} \left(\ell - 1 + (-1)^{x_1} - (-1)^{x_\ell} - \sum_{i\in[\ell-1]} (-1)^{x_i + x_{i+1}} \right) \tag{6.113}$$

and

$$\beta_x = \frac{1}{\varepsilon_{\mathsf{L}}} \sum_{i\in[\ell]} (i - t)\, (1 - (-1)^{x_i})\,. \tag{6.114}$$

Note that $H_{\mathsf{L}}(t) = H_{\mathsf{L}}(1, t)$.

Recall that $|\boldsymbol{j}\rangle = \left|1^j 0^{\ell-j}\right\rangle$. Define $\mathcal{I} = \{\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{\ell}\}$. We also note the following easy claim for α_x, β_x, the proof of which is deferred to Section 6.7.4.

Claim 6.7.1. *For $x = \boldsymbol{j} \in \mathcal{I}$, we have $\alpha_x = 0$ and $\beta_x = j(j+1-2t)/\varepsilon_{\mathsf{L}}$. For $x \notin \mathcal{I}$, we have $\alpha_x \geq \Delta_{\mathsf{L}}$ and $|\beta_x| \leq 2(\ell^2 + |t| \cdot \ell)/\varepsilon_{\mathsf{L}}$.*

In the rest of the section, we verify the rest of the properties in Lemma 6.4.6. This is divided into two cases depending on the range of t.

Organization. In Section 6.7.1, we prove Lemma 6.4.6 for the unbounded time regime where $|t|$ is extremely large. In Section 6.7.2, we handle the bounded time regime for reasonably large $|t|$. In Section 6.7.3, we prove some useful spectral results of certain tridiagonal matrices, which are used in Section 6.7.2. Missing proofs from the section can be found in Section 6.7.4.

6.7.1 Unbounded Time Regime

We first handle the unbounded time regime where $|t| \geq 10\ell^2$.

In this case, we define

$$H_{\mathsf{L}}{}'(t) = \sum_{x \in \{0,1\}^\ell} (\alpha_x + \beta_x) |x\rangle\langle x| . \tag{6.115}$$

Since $H_{\mathsf{L}}(\lambda, t) - H_{\mathsf{L}}{}'(t) = -\lambda \sum_{i \in [\ell]} X_i$. we have the following easy fact.

Claim 6.7.2. *If $\lambda \in [0, 2]$, then $\|H_{\mathsf{L}}(\lambda, t) - H_{\mathsf{L}}{}'(t)\| = \lambda\ell$.*

Since $H_{\mathsf{L}}{}'(t)$ is diagonal, its spectrum is also easy to analyze.

Claim 6.7.3. *Assume $\ell \geq 1$ and $\Delta_{\mathsf{L}} \geq 100\ell^4/\varepsilon_{\mathsf{L}}$.*

- *If $t \leq -10\ell^2$, then $H_{\mathsf{L}}{}'(t)$ has spectral gap at least $4\ell/\varepsilon_{\mathsf{L}}$ and $|\mathbf{0}\rangle$ is its ground state.*
- *If $t \geq 10\ell^2$, then $H_{\mathsf{L}}{}'(t)$ has spectral gap at least $4\ell/\varepsilon_{\mathsf{L}}$ and $|\boldsymbol{\ell}\rangle$ is its ground state.*

Proof. By Claim 6.7.1, we first note that for $x \notin \mathcal{I}$, we have

$$\alpha_x + \beta_x \geq \Delta_{\mathsf{L}} - \frac{2(\ell^2 + |t| \cdot \ell)}{\varepsilon_{\mathsf{L}}} \geq \frac{50\ell^4}{\varepsilon_{\mathsf{L}}} \quad \text{for } |t| \leq 10\ell^2. \tag{6.116}$$

Assume $t \leq -10\ell^2$ and apply Claim 6.7.1. For $x = \mathbf{0}$, we have $\alpha_x + \beta_x = 0$. For $x = \boldsymbol{j}$ where $j \in [\ell]$, we have $\alpha_x + \beta_x = j(j+1-2t)/\varepsilon_{\mathsf{L}} \geq 4\ell/\varepsilon_{\mathsf{L}}$. Combining with (6.116), we know that the spectral gap of ${H_{\mathsf{L}}}'(t)$ is at least $4\ell/\varepsilon_{\mathsf{L}}$ and $|\mathbf{0}\rangle$ is its ground state.

Assume $t \geq 10\ell^2$ and the calculation is similar. For $x = \boldsymbol{\ell}$, we have $\alpha_x + \beta_x = \ell(\ell+1-2t)/\varepsilon_{\mathsf{L}} < 0$. For $x = \boldsymbol{j}$ where $j = 0, 1, \ldots, \ell - 1$, we have

$$(\alpha_x + \beta_x) - (\alpha_{\boldsymbol{\ell}} + \beta_{\boldsymbol{\ell}}) = \frac{j(j+1-2t) - \ell(\ell+1-2t)}{\varepsilon_{\mathsf{L}}} \geq \frac{2t - \ell^2 - \ell}{\varepsilon_{\mathsf{L}}} \geq \frac{4\ell}{\varepsilon_{\mathsf{L}}}. \tag{6.117}$$

Combining with (6.116), we know that the spectral gap of ${H_{\mathsf{L}}}'(t)$ is at least $4\ell/\varepsilon_{\mathsf{L}}$ and $|\boldsymbol{\ell}\rangle$ is its ground state. □

Now we can complete the proof of Lemma 6.4.6 in this unbounded time regime. Note that Lemma 6.4.6 (iii) does not apply in this regime and hence we only verify Lemma 6.4.6 (ii) and Lemma 6.4.6 (iv).

Proof of Lemma 6.4.6 (ii) and (iv): the $|t| \geq 10\ell^2$ Case. For Lemma 6.4.6 (ii), by Fact 6.3.6, the spectral gap of $H_{\mathsf{L}}(t) = H_{\mathsf{L}}(1,t)$ is at least the spectral gap of ${H_{\mathsf{L}}}'(t)$ minus $2\|H_{\mathsf{L}}(1,t) - {H_{\mathsf{L}}}'(t)\|$. By Claim 6.7.3 and Claim 6.7.2, this is at least $(4\ell/\varepsilon_{\mathsf{L}}) - 2\ell \geq 2\ell \geq 2$.

For Lemma 6.4.6 (iv), we apply Fact 6.3.7 with Claim 6.7.2 and Claim 6.7.3, which implies that

the ground states of $H_{\mathsf{L}}(\lambda, t)$ and ${H_{\mathsf{L}}}'(t)$ are within distance

$$\frac{2^{3/2} \cdot 1 \cdot 2\ell}{4\ell/\varepsilon_{\mathsf{L}}} \leq \varepsilon_{\mathsf{L}} \tag{6.118}$$

in the ℓ_2 norm. □

6.7.2 Bounded Time Regime

Now we handle the bounded time regime where $|t| \leq 10\ell^2$.

We will show that $H_{\mathsf{L}}(\lambda, t)$ embeds a smaller Hamiltonian of dimension $\ell + 1$. To specify this Hamiltonian, we denote $\mathcal{H}$ by the Hilbert space associated with $H_{\mathsf{L}}(\lambda, t)$ and let $\mathcal{K}$ be the Hilbert space spanned by $|0\rangle, |1\rangle, \ldots, |\ell\rangle$. Define the encoding isometry $\mathcal{E} \colon \mathcal{K} \to \mathcal{H}$ by

$$\mathcal{E} |j\rangle = |\boldsymbol{j}\rangle = \left|1^j 0^{\ell - j}\right\rangle. \tag{6.119}$$

Let $\mathcal{H}_- := \operatorname{Im} \mathcal{E}$ and we write $\mathcal{H} = \mathcal{H}_- \oplus \mathcal{H}_+$. Define Hamiltonian $\widetilde{H_{\mathsf{L}}}(\lambda, t)$ by

$$\widetilde{H_{\mathsf{L}}}(\lambda, t) := -\lambda \sum_{j \in [\ell]} \left(|j-1\rangle\langle j| + |j\rangle\langle j-1|\right) + \frac{1}{\varepsilon_{\mathsf{L}}} \sum_{j=0}^{\ell} \beta_{\boldsymbol{j}} |j\rangle\langle j|. \tag{6.120}$$

We now show by first-order reduction (Lemma 6.3.10) that $H_{\mathsf{L}}(\lambda, t)$ simulates $\widetilde{H_{\mathsf{L}}}(\lambda, t)$.

Lemma 6.7.4. *Assume $|t| \leq 10\ell^2$, $\lambda \in [0, 2]$, and $\Delta_{\mathsf{L}} = \operatorname{poly}(\ell, 1/\varepsilon_{\mathsf{L}})$. Then $H_{\mathsf{L}}(\lambda, t)$ and $\mathcal{E}$ simulate $\widetilde{H_{\mathsf{L}}}(\lambda, t)$ with error ε_{L}.*

Proof. Fix λ, t and shorthand $H_{\mathsf{L}} = H_{\mathsf{L}}(\lambda, t)$, $\widetilde{H_{\mathsf{L}}} = \widetilde{H_{\mathsf{L}}}(\lambda, t)$. In light of Lemma 6.3.10, define

$$V = -\lambda \sum_{\substack{(x,x'):\ \mathrm{dist}(x,x')=1 \\ x\in\mathcal{I} \vee x'\in\mathcal{I}}} |x\rangle\langle x'| + \sum_{x\in\{0,1\}^\ell} \beta_x |x\rangle\langle x| \tag{6.121}$$

and

$$H_0 = \widetilde{H_{\mathsf{L}}} - V = -\lambda \sum_{\substack{(x,x'):\ \mathrm{dist}(x,x')=1 \\ x\notin\mathcal{I} \wedge x'\notin\mathcal{I}}} |x\rangle\langle x'| + \sum_{x\in\{0,1\}^\ell} \alpha_x |x\rangle\langle x| \tag{6.122}$$

$$= -\lambda \sum_{\substack{(x,x'):\ \mathrm{dist}(x,x')=1 \\ x\notin\mathcal{I} \wedge x'\notin\mathcal{I}}} |x\rangle\langle x'| + \sum_{x\notin\mathcal{I}} \alpha_x |x\rangle\langle x| \,. \qquad \text{(by Claim 6.7.1)}$$

Hence H_0 is block-diagonal and $(H_0)_{--} = 0$. In addition by Fact 6.3.6 and Claim 6.7.1, the minimal eigenvalue of $(H_0)_{++}$ is at least

$$\min_{x\notin\mathcal{I}} \alpha_x - 2\lambda\ell \geq \Delta_{\mathsf{L}} - 2\lambda\ell \geq \Delta_{\mathsf{L}}/2, \tag{6.123}$$

where we recall $\lambda \in [0, 2]$ and $\Delta_{\mathsf{L}} = \mathrm{poly}(\ell, 1/\varepsilon_{\mathsf{L}})$ for the last inequality. To apply Lemma 6.3.10, it remains to show $\Delta_{\mathsf{L}}/2 \geq O({\varepsilon_{\mathsf{L}}}^{-1}\|V\|^2)$. To this end, we simply note that

$$\|V\| \leq \lambda\ell + \max_x |\beta_x| \leq \mathrm{poly}(|t|, \lambda, \ell, 1/\varepsilon_{\mathsf{L}}) \qquad \text{(by Claim 6.7.1)}$$

$$= \mathrm{poly}(\ell, 1/\varepsilon_{\mathsf{L}}). \qquad \text{(since } |t| \leq 10\ell^2 \text{ and } \lambda \in [0, 2])$$

This completes the proof of Lemma 6.7.4. □

By Lemma 6.7.4, we aim to study the spectrum of $\widetilde{H_{\mathsf{L}}}(\lambda, t)$, which is essentially the low energy

spectrum of $H_{\mathsf{L}}(\lambda, t)$. Since $\widetilde{H_{\mathsf{L}}}(\lambda, t)$ is a tridiagonal matrix, we need the following result which will be proved in Section 6.7.3.

Lemma 6.7.5. *Assume $0 < \varepsilon_{\mathsf{L}} \le 1/10$ and $\lambda \in [0, 2]$. Then $\widetilde{H_{\mathsf{L}}}(\lambda, t)$ satisfies the following properties.*

(i) For $t \in [0, \ell + 1]$, let $|\psi\rangle$ be a ground state of $\widetilde{H_{\mathsf{L}}}(1, t)$. Then the spectral gap of $\widetilde{H_{\mathsf{L}}}(1, t)$ is at least $3/4$ and $|\psi\rangle$ is $O(\varepsilon_{\mathsf{L}})$-close to a normalized state that is a linear combination of $|a\rangle$ and $|b\rangle$, where $a := \max\{\lfloor t - 1/2 \rfloor, 0\}$ and $b := \min\{\lceil t - 1/2 \rceil, \ell\}$.

(ii) For any $\lambda \in [0, 2]$ and $t \le 0$ (resp., $t \ge \ell + 1$), let $|\psi\rangle$ be a ground state of $\widetilde{H_{\mathsf{L}}}(\lambda, t)$. Then the spectral gap of $\widetilde{H_{\mathsf{L}}}(1, t)$ is at least $3/4$ and $|\psi\rangle$ is $O(\varepsilon_{\mathsf{L}})$-close to $|0\rangle$ (resp., $|\ell\rangle$).

Given Lemma 6.7.5 and Lemma 6.7.4, we prove Lemma 6.4.6 in this bounded time regime. Together with the analysis in the unbounded time regime (Section 6.7.1), this completes the whole proof of Lemma 6.4.6.

Proof of Lemma 6.4.6 (ii), (iii), and (iv): the $|t| \le 10\ell^2$ Case. Recall from Lemma 6.7.4 that $H_{\mathsf{L}}(\lambda, t)$ and $\mathcal{E}$ simulate $\widetilde{H_{\mathsf{L}}}(\lambda, t)$ with error ε_{L}, where $\mathcal{E} \colon |j\rangle \mapsto |\boldsymbol{j}\rangle$.

The spectral gap of $H_{\mathsf{L}}(t) = H_{\mathsf{L}}(1, t)$ is thus at least that of $\widetilde{H_{\mathsf{L}}}(1, t)$ minus $2\varepsilon_{\mathsf{L}}$ by Fact 6.3.6, which is $3/4 - 2\varepsilon_{\mathsf{L}} \ge 1/2$ by Lemma 6.7.5 (i) and (ii), which concludes the proof of Lemma 6.4.6 (ii) and the spectral gap part of Lemma 6.4.6 (iv).

Now we turn to ground states. By Lemma 6.7.5 (i), the ground state of $\widetilde{H_{\mathsf{L}}}(1, t)$ is $O(\varepsilon_{\mathsf{L}})$-close to a normalized state that is a linear combination of $|a\rangle$ and $|b\rangle$. Thus by Lemma 6.3.5, the ground state of $H_{\mathsf{L}}(t) = H_{\mathsf{L}}(1, t)$ is then $(O(\varepsilon_{\mathsf{L}}) + \varepsilon_{\mathsf{L}})$-close to that of $|\boldsymbol{a}\rangle$ and $|\boldsymbol{b}\rangle$, as claimed in Lemma 6.4.6 (iii). Similarly, Lemma 6.7.5 (ii) implies the ground state part of Lemma 6.4.6 (iv). □

6.7.3 Properties of Certain Tridiagonal Matrices

To prove Lemma 6.7.5, we introduce Lemma 6.7.6 and Lemma 6.7.7. Both lemmas analyze the smallest eigenvalues, and corresponding eigenspaces, of the tridiagonal matrix

$$A \coloneqq \begin{bmatrix} p_0 & -\lambda & & & \\ -\lambda & p_1 & -\lambda & & \\ & \ddots & \ddots & \ddots & \\ & & -\lambda & p_{\ell-1} & -\lambda \\ & & & -\lambda & p_\ell \end{bmatrix} \quad \text{where} \quad \lambda > 0 \tag{6.124}$$

but under slightly different assumptions.

Lemma 6.7.6. *Let $|\psi\rangle \in \mathbb{R}^{\ell+1}$ be a ground state of A. Assume for some $i \in \{0, 1, \ldots, \ell\}$ and $C > 0$, we have $p_j \geq p_i + C$ for all $j \in \{0, 1, \ldots, \ell\} \setminus \{i\}$. Then the spectral gap of A is at least $C - 4\lambda$, and $|\psi\rangle$ is $(8\lambda/C)$-close to $\alpha\,|i\rangle$ for some $|\alpha| = 1$.*

Proof. Let $D = \operatorname{diag}(p_0, \ldots, p_\ell)$. Then $A = D + E$ where E contains the off-diagonal terms in A and $\|E\| \leq 2\lambda$. Hence by Fact 6.3.6, the spectral gap of A is lower bounded by the spectral gap of D (which is $-p_i + \min_{j \neq i} p_j$) minus $2 \cdot \|E\| \leq 4\lambda$. Since $p_j \geq p_i + C$, the spectral gap of A is at least $C - 4\lambda$.

Now we argue about $v_i \coloneqq \langle i|\psi\rangle$. First assume without loss of generality $v_i \geq 0$. Then observe that the ground space of D is spanned by $|i\rangle$. Hence by Fact 6.3.7, we have

$$\|\alpha\,|i\rangle - |\psi\rangle\| \leq \frac{2^{3/2}\|E\|}{C} \leq \frac{8\lambda}{C} \tag{6.125}$$

as desired. □

Lemma 6.7.7. *Let $|\psi\rangle \in \mathbb{R}^{\ell+1}$ be a ground state of A. Assume for some $i \in [\ell]$, we have $|p_i - p_{i-1}| \leq \varepsilon$ and $p_j \geq \max\{p_{i-1}, p_i\} + C$ for all $j \in \{0, 1, \ldots, \ell\} \setminus \{i-1, i\}$, where $C \geq 4\lambda + \varepsilon$. Then the spectral gap of A is at least $3\lambda/4$, and $|\psi\rangle$ is $(8\lambda/C)$-close to $\alpha |i-1\rangle + \beta |i\rangle$ for some α, β such that $|\alpha|^2 + |\beta|^2 = 1$.*

Proof. Define matrix E to have the following non-zero entries: $\langle i-2|E|i-1\rangle = \langle i-1|E|i-2\rangle = \langle i|E|i+1\rangle = \langle i+1|E|i\rangle = -\lambda$ and $\langle i-1|E|i\rangle = \langle i|E|i-1\rangle = \lambda$ (omit if indices run out of range). Then we have the following cases:

- if $i = 1$, then E is $\begin{bmatrix} & \lambda & \\ \lambda & & -\lambda \\ & -\lambda & \end{bmatrix}$ padded with zeros, which has $\|E\| = \sqrt{2} \cdot \lambda$;
- if $i = \ell$, then E is $\begin{bmatrix} & -\lambda & \\ -\lambda & & \lambda \\ & \lambda & \end{bmatrix}$ padded with zeros, which has $\|E\| = \sqrt{2} \cdot \lambda$;
- otherwise $2 \leq i \leq \ell - 1$, then E is $\begin{bmatrix} & -\lambda & & \\ -\lambda & & \lambda & \\ & \lambda & & -\lambda \\ & & -\lambda & \end{bmatrix}$ padded with zeros, which has $\|E\| = \frac{\sqrt{5}+1}{2} \cdot \lambda$.

Let $D = A - E$. Then D is block-diagonal of tridiagonal matrices:

$$D = \begin{bmatrix} p_0 & -\lambda & & & & & & & \\ \ddots & \ddots & \ddots & & & & & & \\ & -\lambda & p_{i-2} & & & & & & \\ & & & p_{i-1} & -2\lambda & & & & \\ & & & -2\lambda & p_i & & & & \\ & & & & & p_{i+1} & -\lambda & & \\ & & & & & \ddots & \ddots & \ddots & \\ & & & & & & -\lambda & p_\ell \end{bmatrix}, \tag{6.126}$$

where the middle block $\begin{bmatrix} p_{i-1} & -2\lambda \\ -2\lambda & p_i \end{bmatrix}$ has eigenvalues

$$\mu_1 = \frac{1}{2}\left(p_{i-1} + p_i - \sqrt{16\lambda^2 + (p_{i-1} - p_i)^2}\right) \quad \text{and} \quad \mu_2 = \frac{1}{2}\left(p_{i-1} + p_i + \sqrt{16\lambda^2 + (p_{i-1} - p_i)^2}\right). \tag{6.127}$$

Note that

$$\mu_1 \le \frac{p_{i-1} + p_i}{2} - 2\lambda < \frac{p_{i-1} + p_i}{2} + 2\lambda \le \mu_2 \le \frac{p_{i-1} + p_i}{2} + 2\lambda + \frac{\varepsilon}{2}. \tag{6.128}$$

Whereas the eigenvalues of D outside the middle block are, by Fact 6.3.6, at least $\min_{j \neq i-1,i}\{p_j\} - 2\lambda \ge \max\{p_{i-1}, p_i\} + C - 2\lambda$. Since $C \ge 4\lambda + \varepsilon$, we know that μ_1, μ_2 are the smallest two eigenvalues of D. Hence by Fact 6.3.6 again, the spectral gap of A is lower bounded by the spectral gap of D (which is $\mu_2 - \mu_1 \ge 4\lambda$) minus $2 \cdot \|E\| \le (\sqrt{5} + 1)\lambda$. Since $\sqrt{5} + 1 < 3.25$, the spectral gap of A is at least $3\lambda/4$ as claimed.

To argue about the eigenvector $|\psi\rangle$, notice that $\{|i-1\rangle, |i\rangle\}$ span the ground space of D. Let $|\phi\rangle$ be an eigenvector of A corresponding to the second smallest eigenvalue. By Fact 6.3.7, there is a 2×2 unitary U such that

$$\|R\|_{\mathrm{F}} \leq \frac{2^{3/2} \cdot 2^{1/2} \cdot \|E\|}{C} \leq \frac{2 \cdot (\sqrt{5}+1) \cdot \lambda}{C} \leq \frac{8\lambda}{C}, \tag{6.129}$$

where $R := (|i-1\rangle\langle 1| + |i\rangle\langle 2|)U - (|\psi\rangle\langle 1| + |\phi\rangle\langle 2|)$. Therefore, the desired bound can be obtained via

$$\| \langle 1|U|1\rangle |i-1\rangle + \langle 2|U|1\rangle |i\rangle - |\psi\rangle\|_{\mathrm{F}} = \|R|1\rangle\|_{\mathrm{F}} \leq \|R\|_{\mathrm{F}} \||1\rangle\| = \|R\|_{\mathrm{F}}, \tag{6.130}$$

where the last inequality follows from the sub-multiplicativity of Frobenius norm. Setting $\alpha = \langle 1|U|1\rangle$ and $\beta = \langle 2|U|1\rangle$ completes the proof of Lemma 6.7.7. □

Now we are ready to prove Lemma 6.7.5.

Proof of Lemma 6.7.5. Recall from (6.120) and Claim 6.7.1 that

$$\widetilde{H_{\mathsf{L}}}(\lambda, t) = -\lambda \sum_{j \in [\ell]} (|j-1\rangle\langle j| + |j\rangle\langle j-1|) + \frac{1}{\varepsilon_{\mathsf{L}}} \sum_{j=0}^{\ell} j(j+1-2t) |j\rangle\langle j|. \tag{6.131}$$

We write $p_j := \langle j|\widetilde{H_{\mathsf{L}}}(\lambda, t)|j\rangle$ for $j = 0, 1, \ldots, \ell$ and therefore $\widetilde{H_{\mathsf{L}}}(\lambda, t)$ is exactly A defined in (6.124).

For Item (ii), we consider two cases separately:

- If $t \leq 0$, we apply Lemma 6.7.6 with $(C, i) = (2/\varepsilon_{\mathsf{L}}, 0)$ to get the spectral gap lower bound 12 and that $|\psi\rangle$ is $O(1/\varepsilon_{\mathsf{L}})$-close to $|0\rangle$ since $\varepsilon_{\mathsf{L}} \leq 1/10$ and $\lambda \leq 2$ by assumption.
- If $t \geq \ell + 1$, similarly we apply Lemma 6.7.6 with $(C, i) = (2/\varepsilon_{\mathsf{L}}, \ell + 1)$.

To prove Item (i), we divide t into separate cases and frequently use the following equation:

$$\varepsilon_{\mathsf{L}}(p_i - p_j) = i(i+1-2t) - j(j+1-2t) = (i-j)(i+j+1-2t). \tag{6.132}$$

- Assume $t \in [0, 3/4]$. Then for any $i \geq 1$,

$$\varepsilon_{\mathsf{L}}(p_i - p_0) = i(i+1-2t) \underbrace{\geq}_{\text{set } i=1} 2 - 2t \geq \frac{1}{2}. \tag{6.133}$$

  Hence we can apply Lemma 6.7.6 with $C = 1/(2\varepsilon_{\mathsf{L}})$.

- Assume $t \in [\ell + 1/4, \ell + 1]$. Then for any $i \leq \ell - 1$,

$$\varepsilon_{\mathsf{L}}(p_i - p_\ell) = (\ell - i)(2t - \ell - 1 - i) \underbrace{\geq}_{\text{set } i=\ell-1} 2t - 2\ell \geq \frac{1}{2}. \tag{6.134}$$

  Hence we can apply Lemma 6.7.6 with $C = 1/(2\varepsilon_{\mathsf{L}})$.

- Assume $t \in [i - 1/4, i + 1/4]$ for some integer $i = 1, 2, \ldots, \ell - 1$. Then $|\varepsilon_{\mathsf{L}}(p_i - p_{i-1})| = |2i - 2t| \leq 1/2$. For $j \neq i-1, i$,

$$\varepsilon_{\mathsf{L}}(p_j - p_i) \underbrace{\geq}_{j=i-2 \text{ or } i+1} \min\{2(2t+1-2i), 2(i+1-t)\} \geq 1, \tag{6.135}$$

  and

$$\varepsilon_{\mathsf{L}}(p_j - p_{i-1}) \underbrace{\geq}_{j=i-2 \text{ or } i+1} \min\{2(t+1-i), 2(2i+1-2t)\} \geq 1. \tag{6.136}$$

  Hence we can apply Lemma 6.7.7 with $(\varepsilon, C) = (1/(2\varepsilon_{\mathsf{L}}), 1/\varepsilon_{\mathsf{L}})$.

- Otherwise $t \in [i+1/4, i+3/4]$ for some integer $i = 1, 2, \ldots, \ell-1$. Then for any $j \neq i$,

$$\varepsilon_{\mathsf{L}}(p_j - p_i) \underbrace{\geq}_{j=i-1 \text{ or } i+1} \min\{2(t-i), 2(i+1-t)\} \geq \frac{1}{2}. \tag{6.137}$$

Hence we can apply Lemma 6.7.6 with $C = 1/(2\varepsilon_{\mathsf{L}})$. □

6.7.4 Deferred Proofs

Proof of Claim 6.7.1. We do a direct calculation based on (6.112) and (6.113). If $x = \boldsymbol{j}$, then

$$\sum_{i \in [\ell-1]} (-1)^{x_i + x_{i+1}} = \begin{cases} \ell - 1, & j = 0, \ell, \\ \ell - 3, & j \in [\ell - 1], \end{cases} \quad \text{and} \quad (-1)^{x_1} - (-1)^{x_\ell} = \begin{cases} 0, & j = 0, \ell, \\ -2, & j \in [\ell - 1], \end{cases} \tag{6.138}$$

which implies $\alpha_x = 0$. We also have

$$\sum_{i \in [\ell]} (i - t)\,(1 - (-1)^{x_i}) = \sum_{1 \leq i \leq j} 2(i - t) = j(j + 1 - 2t), \tag{6.139}$$

which implies $\beta_x = j(j + 1 - 2t)/\varepsilon_{\mathsf{L}}$.

Now assume $x \notin \mathcal{I}$. By (6.114), we have

$$|\beta_x| \leq \frac{1}{\varepsilon_{\mathsf{L}}} \sum_{i \in [\ell]} (i + |t|) \cdot 2 = \frac{\ell(\ell+1) + 2|t| \cdot \ell}{\varepsilon_{\mathsf{L}}} \leq \frac{2(\ell^2 + |t| \cdot \ell)}{\varepsilon_{\mathsf{L}}}, \tag{6.140}$$

where we use $\ell \geq 1$ for the last inequality. Regarding α_x, define $s = |\{i \in [\ell - 1] \colon x_i \neq x_{i+1}\}|$. Since $x \notin \mathcal{I}$, we either have (1) $s \geq 2$ or (2) $s = 1$ but $x = \left|0^j 1^{\ell-j}\right\rangle$ for some $1 \leq j \leq \ell - 1$. In case (1), we

have

$$\alpha_x \geq \Delta_{\mathsf{L}} \left(\ell - 1 - 2 - (-s + \ell - 1 - s)\right) \geq 2\Delta_{\mathsf{L}}(s-1) \geq \Delta_{\mathsf{L}}. \tag{6.141}$$

In case (2), we have

$$\alpha_x = \Delta_{\mathsf{L}} \left(\ell - 1 + 1 + 1 - (-s + \ell - 1 - s)\right) = 4\Delta_{\mathsf{L}} \geq \Delta_{\mathsf{L}}. \tag{6.142}$$

Hence we always have $\alpha_x \geq \Delta_{\mathsf{L}}$. □

6.8 Reduction III: Proof of Lemma 6.4.7

We prove Lemma 6.4.7 (restated below) in this section.

Lemma 6.4.7. *Assume* $\varepsilon_{\mathsf{L}} \leq 1/\operatorname{poly}(\mathfrak{n})$. *There exist*

$$\Delta_{\mathsf{L}} = \operatorname{poly}(\mathfrak{n}, 1/\varepsilon_{\mathsf{L}}) \quad \textit{and} \quad 10\|D_{\mathsf{C}}\| \leq \Delta_{\mathsf{C}} \leq \operatorname{poly}(\mathfrak{n}) \tag{6.40}$$

such that the Hamiltonian path

$$H_{\mathsf{C}}(0) \longrightarrow H_{\mathsf{C}}(\ell + 1) \tag{6.41}$$

has spectral gap $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $\operatorname{poly}(\mathfrak{n}, 1/\varepsilon_{\mathsf{L}})$.

The proof idea is to view the shifts of the $\ell + 1$ Hamiltonians as the ground state shifts of $H_{\mathsf{L}}(t)$ in Lemma 6.4.6, which is formalized via a first-order perturbative reduction (Lemma 6.3.10).

Notation. We (re)define some notation. Let $n \coloneqq n_{\mathsf{B}}$, $\Delta \coloneqq \Delta_{\mathsf{C}}$, and $\varepsilon \coloneqq \varepsilon_{\mathsf{L}}$. For $\tau \in \{0,1\}^n$, define

$$K_\tau \coloneqq \begin{cases} X_{\mathsf{B}} + D_j & \tau = \boldsymbol{j} = 1^j 0^{\ell - j} \text{ for some } j, \\ X_{\mathsf{B}} & \text{otherwise.} \end{cases} \tag{6.143}$$

Note that by Lemma 6.4.5 (iii) the Hamiltonian path

$$K_{\mathbf{0}} \longrightarrow K_{\mathbf{1}} \longrightarrow \cdots \longrightarrow K_{\boldsymbol{\ell}} \tag{6.144}$$

has spectral gap at least $\delta \coloneqq \Omega(\sqrt{\mathfrak{m}})$ and norm $\mathrm{poly}(\mathfrak{n})$. Denote $M \coloneqq \max_{\tau \in \{0,1\}^\ell} \|K_\tau\|$ and we have $M \leq \mathrm{poly}(\mathfrak{n})$ by Lemma 6.4.5 (iii) and the bound $\Delta_{\mathsf{B}} \leq \mathrm{poly}(\mathfrak{n})$ in Lemma 6.4.4.

Let t be fixed. We shorthand $H_{\mathsf{L}} \coloneqq H_{\mathsf{L}}(t)$, $H_0 \coloneqq \Delta \cdot I_{2^n} \otimes H_{\mathsf{L}}(t)$, and $V \coloneqq X_{\mathsf{B}} \otimes I_{2^\ell} + D_{\mathsf{C}}$. The Hamiltonian of interest $H_{\mathsf{C}} \coloneqq H_{\mathsf{C}}(t)$ in (6.39) then becomes

$$H_{\mathsf{C}} = H_0 + V, \quad \text{where} \quad H_0 = \Delta \cdot I_{2^n} \otimes H_{\mathsf{L}}, \quad V = \sum_{\tau \in \{0,1\}^\ell} K_\tau \otimes |\tau\rangle\langle\tau| \,. \tag{6.145}$$

We also denote $a \coloneqq \max\{\lfloor t - 1/2 \rfloor, 0\}$ and $b \coloneqq \min\{\lceil t - 1/2 \rceil, \ell\}$, as in Lemma 6.4.6 (iii).

Now we prove Lemma 6.4.7.

Proof of Lemma 6.4.7. Let $\mathcal{H}$ be the Hilbert space associated with H_0, H_{C}, and V. Let $\mathcal{H}_-$ be the ground space of H_0. Let ζ be the ground energy of H_{L} and let $|v\rangle$ be a ground state of H_{L}. By definition, $\mathcal{H}_-$ is spanned by $|\psi\rangle \otimes |v\rangle$ for any n-qubit state $|\psi\rangle$. We explicitly write out $|v\rangle$ as $|v\rangle =$

$\sum_{\tau\in\{0,1\}^\ell} v_\tau \,|\tau\rangle$. Then by Lemma 6.4.6 (iii), we have

$$\sum_{\tau\notin\{\boldsymbol{a},\boldsymbol{b}\}} v_\tau^2 \le O(\varepsilon^2). \tag{6.146}$$

Let $P_- \coloneqq I_{2^n}\otimes|v\rangle\langle v|$ be the projection onto $\mathcal{H}_-$. Then

$$(H_0)_{--} = P_- H_0 P_- = \Delta\zeta\cdot P_- = \Delta\zeta\cdot I_{2^n}\otimes|v\rangle\langle v| \tag{6.147}$$

and

$$V_{--} = P_- V P_- = \left(\sum_{\tau\in\{0,1\}^\ell} v_\tau^2\cdot K_\tau\right)\otimes|v\rangle\langle v|\,. \tag{6.148}$$

Define

$$H_{\text{target}} \coloneqq (H_0)_{--}+V_{--} = \left(\Delta\zeta\cdot I_{2^n}+\sum_\tau v_\tau^2\cdot K_\tau\right)\otimes|v\rangle\langle v| \tag{6.149}$$

$$= \left(\Delta\zeta\cdot I_{2^n}+v_{\boldsymbol{a}}^2K_a+v_{\boldsymbol{b}}^2K_b+\sum_{\tau\notin\{\boldsymbol{a},\boldsymbol{b}\}} v_\tau^2\cdot K_\tau\right)\otimes|v\rangle\langle v|\,. \tag{6.150}$$

Observe that the spectrum of H_{target} is the same to that of $\sum_\tau v_\tau^2\cdot K_\tau$ up to an additive $\Delta\zeta$ shift. In addition, since $K_{\boldsymbol{a}}\longrightarrow K_{\boldsymbol{b}}$ is part of (6.144) and has minimal spectral gap δ, the spectral gap of $\frac{v_{\boldsymbol{a}}^2K_{\boldsymbol{a}}+v_{\boldsymbol{b}}^2K_{\boldsymbol{b}}}{v_{\boldsymbol{a}}^2+v_{\boldsymbol{b}}^2}$ is at least δ. By Fact 6.3.6, the spectral gap of H_{target} is

$$\delta_{\text{target}} \ge \left(v_{\boldsymbol{a}}^2+v_{\boldsymbol{b}}^2\right)\cdot\delta - 2\cdot\left\|\sum_{\tau\notin\{\boldsymbol{a},\boldsymbol{b}\}} v_\tau^2\cdot K_\tau\right\| \tag{6.151}$$

$$\ge \left(v_{\boldsymbol{a}}^2+v_{\boldsymbol{b}}^2\right)\cdot\delta - 2\cdot(1-v_{\boldsymbol{a}}^2-v_{\boldsymbol{b}}^2)\cdot M \qquad \text{(since } \|K_\tau\|\le M\text{)}$$

$$\ge (1-O(\varepsilon^2))\cdot\delta - O(\varepsilon^2 M) \qquad \text{(by (6.146))}$$

$$\geq 0.99 \cdot \delta, \tag{6.152}$$

where we assume $\varepsilon \leq 1/\operatorname{poly}(\mathfrak{n}) \leq O(\sqrt{\delta/M})$ for the last inequality.

Now, write $\mathcal{H} = \mathcal{H}_- \oplus \mathcal{H}_+$ and regard P_- as an encoding isometry from $\mathcal{H}_-$ to $\mathcal{H}$. Let P_+ be the projection onto $\mathcal{H}_+$. Since $\mathcal{H}_-$ is defined to be the ground space of H_0, the minimal eigenvalue of $(H_0)_{++} = P_+ H_0 P_+$ is the second smallest eigenvalue of H_0. Since $H_0 = \Delta \cdot I \otimes H_{\mathsf{L}}$ and by Lemma 6.4.6 (ii), we know that eigenvalues of $(H_0)_{++}$ are greater than those of $(H_0)_{--}$ by at least $\Delta/2$. Hence by Lemma 6.3.10, (H_{C}, P_-) simulates H_{target} with error $1/2$, provided that $\Delta \geq \operatorname{poly}(M)$ as $\|V\| = \max_\tau \|K_\tau\| = M$. By Lemma 6.3.5, the spectral gap of H_{C} is at least $\delta_{\text{target}} - 1 = \Omega(\delta) = \Omega(\sqrt{\mathfrak{m}})$. This verifies the spectral gap property of Lemma 6.4.7.

To check the norm condition of Lemma 6.4.7, we recall Lemma 6.4.6 (i) that $\|H_{\mathsf{L}}\| \leq \operatorname{poly}(\mathfrak{n}, 1/\varepsilon)$ as $t \in [0, \ell + 1]$ and $\ell = \operatorname{poly}(\mathfrak{n})$ from Lemma 6.4.5. Hence

$$\|H_{\mathsf{C}}\| \leq \|H_0\| + \|V\| = \Delta \|H_{\mathsf{L}}\| + M \leq \operatorname{poly}(\mathfrak{n}, 1/\varepsilon) \tag{6.153}$$

as claimed. □

6.9 Reduction IV: Proof of Lemma 6.4.8

We prove Lemma 6.4.8 (restated below) in this section.

Lemma 6.4.8. *There exist* $\varepsilon_{\mathsf{L}} = 1/\operatorname{poly}(\mathfrak{n})$, $\Delta_{\mathsf{L}} \leq \operatorname{poly}(\mathfrak{n})$,

$$10\|D_{\mathsf{C}}\| \leq \Delta_{\mathsf{C}} \leq \operatorname{poly}(\mathfrak{n}) \quad \textit{and} \quad 1/\operatorname{poly}(\mathfrak{n}) \leq \eta \leq 1/(10\|D_{\mathsf{C}}\| + \ell) \tag{6.45}$$

such that the Hamiltonian path

$$H_{\mathsf{C}'}(-1/\eta) \longrightarrow H_{\mathsf{C}'}(1/\eta) \tag{6.46}$$

has spectral gap $\Omega(\sqrt{\mathfrak{m}})$ *and norm* $\mathrm{poly}(\mathfrak{n})$.

The proof of Lemma 6.4.8 is divided into two parts. The first part handles the middle Hamiltonian path $H_{\mathsf{C}'}(0) \longrightarrow H_{\mathsf{C}'}(\ell+1)$, which is essentially Lemma 6.4.7. The second part handles the start $H_{\mathsf{C}'}(-1/\eta) \longrightarrow H_{\mathsf{C}'}(0)$ and the end $H_{\mathsf{C}'}(\ell+1) \longrightarrow H_{\mathsf{C}'}(1/\eta)$, which relies on the first-order perturbative reduction (Lemma 6.3.10) similar to the proof of Lemma 6.4.7.

Notation. We (re)define some notation. Let $t \in [-1/\eta, 1/\eta]$ and let $n \coloneqq n_{\mathsf{B}}$, $\Delta \coloneqq \Delta_{\mathsf{C}}$, and $\varepsilon \coloneqq \varepsilon_{\mathsf{L}}$. Recall from (6.44) that we have

$$H_{\mathsf{C}'}(t) = \Delta I_{2^n} \otimes H_{\mathsf{L}}(1-t\eta, t) + (1-t\eta) X_{\mathsf{B}} \otimes I_{2^\ell} + (1+t\eta) D_{\mathsf{C}}, \tag{6.154}$$

where $H_{\mathsf{L}}(1-t\eta, t) = Q_{\mathsf{L}} + (1-t\eta)X_{\mathsf{L}} + Z_{\mathsf{L}}(t)$ as defined in Lemma 6.4.6 (iv).

For $\tau \in \{0,1\}^n$, define

$$K_\tau(t) \coloneqq \begin{cases} (1-t\eta)X_{\mathsf{B}} + (1+t\eta)D_j & \tau = \boldsymbol{j} \coloneqq 1^j 0^{\ell-j} \text{ for some } j, \\ (1-t\eta)X_{\mathsf{B}} & \text{otherwise.} \end{cases} \tag{6.155}$$

Note that we have

$$D_0 = 0 \quad \text{and} \quad D_\ell = -3\Delta_{\mathsf{B}} n \, |u\rangle\langle u| \tag{6.156}$$

for some $u \in \{0,1\}^n$ and $\Delta_{\mathsf{B}} \in [\mathfrak{m}, \mathrm{poly}(\mathfrak{m})]$ by Lemma 6.4.4 and Lemma 6.4.5. We also note the following easy observation.

Claim 6.9.1. *The spectral gap of $K_0(t)$ is at least $(1-t\eta)\cdot\Delta_{\mathsf{B}}n$. The spectral gap of $K_\ell(t)$ is at least $(1+5t\eta)\cdot\Delta_{\mathsf{B}}n$.*

Proof. Recall that $X_{\mathsf{B}} = -\Delta_{\mathsf{B}}\sum_{i\in[n]}X_i$. The statement follows directly from (6.155) and (6.156). □

We inherit the value of Δ and the dependency $\Delta_{\mathsf{L}} = \mathrm{poly}(\ell, 1/\varepsilon)$ from Lemma 6.4.7. Therefore we only need to show that $\varepsilon, \eta \geq 1/\mathrm{poly}(\mathfrak{n})$ suffices for our purposes.

Claim 6.9.2. *Assume $\varepsilon, \eta \leq 1/\mathrm{poly}(\mathfrak{n})$. $H_{\mathsf{C}'}(t)$ has spectral gap $\Omega(\sqrt{\mathfrak{m}})$ for $t \in [0, \ell+1]$.*

Proof. For $t \in [0, \ell+1]$, we have

$$\|H_{\mathsf{C}'}(t) - H_{\mathsf{C}}(t)\| = \|t\eta\,(D_{\mathsf{C}} - \Delta I_{2^n}\otimes X_{\mathsf{L}} - X_{\mathsf{B}}\otimes I_{2^\ell})\| \leq \eta\cdot\mathrm{poly}(\mathfrak{n}) \leq 1, \tag{6.157}$$

where we set $\eta \leq 1/\mathrm{poly}(\mathfrak{n})$ sufficiently small. Since $\varepsilon \leq 1/\mathrm{poly}(\mathfrak{n})$, we can apply Lemma 6.4.7 to get an $\Omega(\sqrt{\mathfrak{m}})$ spectral gap lower bound for $H_{\mathsf{C}}(t)$. Then the claim follows from Lemma 6.3.5. □

Claim 6.9.3. *Assume $\varepsilon, \eta \leq 1/\mathrm{poly}(\mathfrak{n})$. $H_{\mathsf{C}'}(t)$ has spectral gap $\Omega(\sqrt{\mathfrak{m}})$ for $t \in [-1/\eta, 1/\eta]\setminus[0, \ell+1]$.*

Proof. Fix $t \in [-1/\eta, 1/\eta]\setminus[0, \ell+1]$. Let $a \coloneqq 0$ if $t \leq 0$ and $a \coloneqq \ell$ if $t \geq \ell+1$. We shorthand $K_\tau \coloneqq K_\tau(t)$, $H_{\mathsf{L}} \coloneqq H_{\mathsf{L}}(1-t\eta, t)$, $H_0 \coloneqq \Delta\cdot I_{2^n}\otimes H_{\mathsf{L}}$, and $V \coloneqq (1-t\eta)X_{\mathsf{B}}\otimes I_{2^\ell} + (1+t\eta)D_{\mathsf{C}}$. The Hamiltonian of interest $H_{\mathsf{C}'} \coloneqq H_{\mathsf{C}'}(t)$ in (6.154) then becomes

$$H_{\mathsf{C}'} = H_0 + V, \quad \text{where} \quad H_0 = \Delta\cdot I_{2^n}\otimes H_{\mathsf{L}}, \quad V = \sum_{\tau\in\{0,1\}^\ell} K_\tau(t)\otimes|\tau\rangle\langle\tau|\,. \tag{6.158}$$

Denote $M \coloneqq \max_{\tau\in\{0,1\}^\ell} \|K_\tau\|$ and, by Lemma 6.4.4 and Lemma 6.4.5, we have

$$M \leq (1+t\eta)\,\mathrm{poly}(n) + 2t\eta\Delta_{\mathsf{B}} n \leq \mathrm{poly}(\mathfrak{n}). \tag{6.159}$$

Our analysis here is similar to the proof of Lemma 6.4.7. Let $\mathcal{H}$ be the Hilbert space associated with H_0, $H_{\mathsf{C}'}$, and V. Let $\mathcal{H}_-$ be the ground space of H_0. Let ζ be the ground energy of H_{L} and $|v\rangle$ be a ground state of H_{L}. By definition, $\mathcal{H}_-$ is spanned by $|\psi\rangle \otimes |v\rangle$ for any n-qubit state $|\psi\rangle$. We explicitly write out $|v\rangle = \sum_{\tau\in\{0,1\}^\ell} v_\tau |\tau\rangle$. Then by Lemma 6.4.6 (iv), we have

$$\sum_{\tau\neq \boldsymbol{a}} v_\tau^2 \leq O(\varepsilon^2). \tag{6.160}$$

Let $P_- \coloneqq I_{2^n} \otimes |v\rangle\langle v|$ be the projection onto $\mathcal{H}_-$. Then

$$(H_0)_{--} = P_- H_0 P_- = \Delta\zeta \cdot P_- = \Delta\zeta \cdot I_{2^n} \otimes |v\rangle\langle v| \tag{6.161}$$

and

$$V_{--} \coloneqq P_- V P_- = \left(\sum_{\tau\in\{0,1\}^\ell} v_\tau^2 \cdot K_\tau\right) \otimes |v\rangle\langle v|\,. \tag{6.162}$$

Define

$$H_{\mathrm{target}} \coloneqq (H_0)_{--} + V_{--} = \left(\Delta\zeta \cdot I_{2^n} + \sum_\tau v_\tau^2 \cdot K_\tau\right) \otimes |v\rangle\langle v| \tag{6.163}$$

$$= \left(\Delta\zeta \cdot I_{2^n} + v_{\boldsymbol{a}}^2 K_{\boldsymbol{a}} + \sum_{\tau\neq\boldsymbol{a}} v_\tau^2 \cdot K_\tau\right) \otimes |v\rangle\langle v| \tag{6.164}$$

Observe that the spectrum of H_{target} is the same to that of $\sum_\tau v_\tau^2 \cdot K_\tau$ up to an additive $\Delta\zeta$ shift. In

addition, $K_{\boldsymbol{a}}$ has minimal spectral gap at least $\delta \coloneqq \Delta_{\mathsf{B}} n$ by Claim 6.9.1. By Fact 6.3.6, the spectral gap of H_{target} is

$$\delta_{\mathrm{target}} \geq v_{\boldsymbol{a}}^2 \cdot \delta - 2 \cdot \left\| \sum_{\tau \neq \boldsymbol{a}} v_\tau^2 \cdot K_\tau(t) \right\| \tag{6.165}$$

$$\geq v_{\boldsymbol{a}}^2 \cdot \delta - 2 \cdot (1 - v_{\boldsymbol{a}}^2) \cdot M \qquad \text{(since } \|K_\tau(t)\| \leq M\text{)}$$

$$\geq (1 - O(\varepsilon^2)) \cdot \delta - O(\varepsilon^2 M) \qquad \text{(by (6.160))}$$

$$\geq 0.99 \cdot \delta, \tag{6.166}$$

where we assume $\varepsilon \leq 1/\operatorname{poly}(\mathfrak{n}) \leq O(\sqrt{\delta/M})$ for the last inequality.

Now, write $\mathcal{H} = \mathcal{H}_- \oplus \mathcal{H}_+$ and regard P_- as an encoding isometry from $\mathcal{H}_-$ to $\mathcal{H}$. Let P_+ be the projection onto $\mathcal{H}_+$. Since $\mathcal{H}_-$ is defined to be the ground space of H_0, the minimal eigenvalue of $(H_0)_{++} = P_+ H_0 P_+$ is the second smallest eigenvalue of H_0. Since $H_0 = \Delta \cdot I \otimes H_{\mathsf{L}}$ and by Lemma 6.4.6 (iv), we know that eigenvalues of $(H_0)_{++}$ are greater than those of $(H_0)_{--}$ by at least $\Delta/2$. Hence by Lemma 6.3.10, $(H_{\mathsf{C}'}, P_-)$ simulates H_{target} with error $1/2$, provided that $\Delta \geq \operatorname{poly}(M)$ as $\|V\| = \max_\tau \|K_\tau\| = M$. Recall (6.159) that $M \leq \operatorname{poly}(\mathfrak{n})$. Hence it suffices to assume $\Delta = \operatorname{poly}(\mathfrak{n})$. Finally by Lemma 6.3.5, the spectral gap of $H_{\mathsf{C}'}$ is at least $\delta_{\mathrm{target}} - 1 = \Omega(\delta) = \Omega(\sqrt{\mathfrak{m}})$. This completes the proof of Claim 6.9.3. □

Now we wrap up the analysis of Lemma 6.4.8.

Proof of Lemma 6.4.8. The spectral gap condition follows directly from Claim 6.9.2 and Claim 6.9.3.

To verify the norm condition, we recall (6.44) and perform a direct calculation

$$\|H_{\mathsf{C}'}(t)\| \leq \Delta \cdot \|H_{\mathsf{L}}(1 - t\eta, t)\| + |1 - t\eta| \cdot \|X_{\mathsf{B}}\| + |1 + t\eta| \cdot \|D_{\mathsf{C}}\| \tag{6.167}$$

$$= \mathrm{poly}(\mathfrak{n}, \|H_{\mathsf{L}}(1-t\eta,t)\|, \|D_{\mathsf{C}}\|) \qquad \text{(since } \Delta, \|X_{\mathsf{B}}\| \le \mathrm{poly}(\mathfrak{n})\text{)}$$

$$= \mathrm{poly}(\mathfrak{n}, \|H_{\mathsf{L}}(1-t\eta,t)\|)$$

(since $\|D_{\mathsf{C}}\| = \max_{0\le i\le \ell} \|D_i\| \le \mathrm{poly}(\mathfrak{n})$ by (6.39) and Lemma 6.4.4)

$$= \mathrm{poly}(\mathfrak{n}), \tag{6.168}$$

where we use Lemma 6.4.6 and $\ell = \mathrm{poly}(\mathfrak{n}), \varepsilon \ge 1/\mathrm{poly}(\mathfrak{n})$ for the last inequality. □

6.10 Extension to Continuous Optimization: Theorem 6.1.2

In this section, we prove an exponential quantum-classical separation for continuous optimization. We consider the following box-constrained minimization problem

$$\arg\min_{x\in\mathcal{X}} f(x), \tag{P}$$

where we assume the objective function f is Lipschitz continuous in a box-shaped domain $\mathcal{X} = \{(x_1,\ldots,x_n)\in\mathbb{R}^n\colon\ -J_j \le x_j \le J_j,\ \forall j\in[n]\}$, where $J_j \le \mathrm{poly}(n)$ for any $j\in[n]$. We denote the global minimizer of f in $\mathcal{X}$ as x^*.

This section is organized as follows. In Section 6.10.1, we present the main result of this section, Theorem 6.10.2, whose proof is detailed in Section 6.10.2. The proof is based on Lemma 6.10.7 that reduces TFD Hamiltonians (as in (6.19)) to a family of (unbounded) Hamiltonian operators defined over the box $\mathcal{X}$. Then, in Section 6.10.3, we prove Lemma 6.10.7 by leveraging the properties of the $\widehat{X}$ operator introduced in Chapter 5 and incorporating new insights. Some missing proofs from Section 6.10.3 are deferred to Section 6.10.4 for better readability and flow.

6.10.1 Hamiltonian Formulation of Continuous Optimization

Definition 6.10.1 (Hamiltonian path in real space)**.** Let n be a positive integer and $\mathcal{X}$ is a compact set in $\mathbb{R}^n$. We call a 1-parameter family $\{H(t) : t \geq 0\}$ a Hamiltonian path over $\mathcal{X}$ if for each $t \geq 0$, we have

$$H(t) = -\Delta + V(t, x), \tag{6.169}$$

where $\Delta = \sum_{j\in[n]} \frac{\partial^2}{\partial_{x_j}^2}$ is the Laplace operator and $V(t, x)\colon [0, \infty) \times \mathcal{X} \to \mathbb{R}$ is a time-dependent potential function.

For each fixed t, $H(t)$ is a Hamiltonian operator that acts on quantum states in $\mathcal{X}$. In particular, for a smooth wave function $\Psi\colon \mathcal{X} \to \mathbb{C}$ with unit L^2-norm (i.e., a test function), we have

$$H(t)\Psi(x) = -\Delta\Psi(x) + V(t, x)\Psi(x). \tag{6.170}$$

This is a natural generalization of the finite-dimensional Hamiltonian path (as in Section 6.3.1) to unbounded Hamiltonians. Similar to the finite-dimensional cases, the continuous-space Hamiltonian path $H(t)$ can lead to a polynomial-time quantum algorithm to solve the minimization problem (P) if the following conditions holds: (1) the ground state of $H(0)$ can be efficiently prepared, (2) $H(t)$ admits a polynomial small spectral gap, (3) $\dot{V}(t, x)$ and $\ddot{V}(t, x)$ is polynomially bounded, (4) the minimizer of $f(x^*)$ is "encoded" in the ground state of $H(t)$ for some large t, and (5) the Hamiltonian $H(t)$ can be efficiently simulated. When all the requirements are satisfied, we can implement an adiabatic-type quantum simulation for $H(t)$ and extract the information of x^* by measuring the ground state of $H(t)$ for some large t. The details of this protocol and its computational complexity can be found in Apppendix C.1.

In particular, Definition 6.10.1 encompasses the quantum Hamiltonian descent algorithm as a special case. Let f be an objective function over $\mathcal{X}$. With an appropriate time rescaling, QHD is governed by the evolution of the following quantum Hamiltonian:

$$H(t) = -\Delta + s(t)f(x), \tag{6.171}$$

where $s(t)$ is certain time-dependent functions. In other words, QHD follows a Hamiltonian path over $\mathcal{X}$ with $V(t,x) = s(t)f(x)$, which is separable in t and x.

Theorem 6.10.2 (Formal version of Theorem 6.1.2)**.** *For any n, there exists a family of continuous functions $\mathcal{M}_n = \{f\colon \mathcal{X} \to \mathbb{R}\}$ with a fixed box domain $\mathcal{X} \subset \mathbb{R}^n$ where the diameter of $\mathcal{X}$ is at most $\operatorname{poly}(n)$. For any $f \in \mathcal{M}_n$, we can construct a Hamiltonian path*

$$H(t) \coloneqq -\Delta + V(t,x), \quad V(t,x) \coloneqq \nu(t)g(x) + tf(x), \quad \forall t \geq 0, x \in \mathcal{X} \tag{6.172}$$

where the auxiliary function $g(x)$ and $\nu(t)$ are efficiently computable (see (6.187), (6.188), and (6.201)) and they only depend on n. Moreover, $g(x)$ is bounded within $\pm\operatorname{poly}(n)$, and $\nu(t) = e^{-\Theta(\sqrt{t})}$. The Hamiltonian path $\{H(t)\}_{t\geq 0}$ exhibits an exponential quantum advantage given oracle access to f:

(i) Any classical algorithm requires at least $\exp\bigl(n^{\Omega(1)}\bigr)$ queries to find an x such that $f(x) - f(x^*) \leq 1/\operatorname{poly}(n)$, with success probability greater than $\exp\bigl(-n^{\Omega(1)}\bigr)$.

(ii) Let $\phi(t,x)$ be the instantaneous ground state of $H(t)$ for all $t \geq 0$. The following properties hold for $H(t)$:

- *The initial ground state $\phi(0,x)$ is $n^{-\Omega(1)}$-close to a product state of the form $\bigotimes_{i\in[n]} |\vartheta_i\rangle$,*

with each $|\vartheta_i\rangle$ being efficiently preparable.

- *For any $t \geq 0$, $H(t)$ has spectral gap $1/\operatorname{poly}(n)$, and $\|\dot{H}(t)\|, \|\ddot{H}(t)\| \leq \operatorname{poly}(n)$.*
- *For any $\eta > 0$, there exists $t_{\mathrm{f}} = \operatorname{poly}(n, 1/\eta)$ such that $\langle\phi(t_{\mathrm{f}})|f|\phi(t_{\mathrm{f}})\rangle - f(x^*) \leq \eta/10$.*

Consequently, simulating the Schrödinger dynamics for $0 \leq t \leq t_{\mathrm{f}}$:

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = TH(t)|\psi(t)\rangle, \quad |\psi(0)\rangle = \bigotimes_{i\in[n]} |\vartheta_i\rangle, \quad T = \operatorname{poly}(n), \tag{6.173}$$

and measuring the final state $|\psi(t_{\mathrm{f}})\rangle$ using the position (i.e., computational) basis, we obtain an $x \in \mathcal{X}$ with query and gate complexity $\operatorname{poly}(n, 1/\eta)$ such that $f(x) - f(x^*) \leq \eta$ with probability at least $9/10 - n^{-\Omega(1)}$.

Remark 6.10.3. We refer readers to Proposition C.2.4 for more quantitative properties regarding $f(x)$ and $g(x)$.

The proof of Theorem 6.10.2 is established by constructing a Hamiltonian path from the time-dependent TFD Hamiltonian path

$$H_{\mathrm{TFI}} \longrightarrow H_{\mathrm{prob}} \tag{6.174}$$

in Theorem 6.4.1[10] via a key reduction lemma formulated in Lemma 6.10.7. Here, H_{TFI} denotes a TFD Hamiltonian of the form:

$$H_{\mathrm{TFI}} = -\sum_i a_i X_i + D_{\mathrm{TFI}}, \tag{6.175}$$

and H_{prob} represents a diagonal problem Hamiltonian. As in Theorem 6.4.1, the ground state of H_{TFI} is a simple product state $|+\rangle^{\otimes(n-\ell)}|0\rangle^{\otimes\ell}$ for some ℓ, up to a small error, and the linear Hamiltonian path

[10]Note that we use H_{prob} to refer the original diagonal Hamiltonian D in Theorem 6.4.1. The symbol D is reserved for a different purpose in this section.

(6.174) admits an $n^{\Omega(1)}$ spectral gap.

We now give an explicit construction of the objective function f. The proof of Theorem 6.10.2 and the key reduction lemma, including the details on g and ν, will be provided in Section 6.10.2.

Definition 6.10.4 (Construction of $\widehat{D}$)**.** First, we define a helper function $\theta\colon [-1,1] \to \{0,1\}$:

$$\theta(\xi) \coloneqq \mathbf{1}_{\xi<0}, \quad \xi \in [-1,1],$$

which can be naturally extended to multi-dimensional settings $\theta\colon [-1,1]^n \to \{0,1\}^n$:

$$\theta(\xi) = \theta(\xi_1,\ldots,\xi_n) \coloneqq (\theta(\xi_1),\ldots,\theta(\xi_n)). \tag{6.176}$$

Suppose that D is an arbitrary n-qubit diagonal Hamiltonian. We define a function $\widehat{D}\colon [-1,1]^n \to \mathbb{R}$ by

$$\widehat{D}(\xi_1,\ldots,\xi_n) \coloneqq \min\{1, \xi_{\min}/w\} \cdot \langle\theta(\xi)|D|\theta(\xi)\rangle, \quad \xi_{\min} \coloneqq \min_{i\in[n]}\{|\xi_i|\}, \tag{6.177}$$

where $w \ll 1$ is a universal constant that will be specified later (see Claim 6.10.8 (iii)). When the context is unambiguous, we will also use $\widehat{D}$ to denote the diagonal operator defined by the function $\widehat{D}(\xi)$, i.e., for any test function $\phi\colon [-1,1]^n \to \mathbb{R}$,

$$(\widehat{D}\phi)(\xi) = \widehat{D}(\xi)\phi(\xi).$$

We refer the readers to Figure 6.6 for an intuitive illustration of our construction of the continuous function $\widehat{D}$.

Fact 6.10.5. *The function $\widehat{D}$ satisfies the following properties:*

(i) $\widehat{D}$ *is continuous in* $[-1,1]^n$ *and* $O(\|D\|)$*-Lipschitz.*

(ii) The global minimum of $\widehat{D}$ *is* $\min\{\mu_0(D), 0\}$*, where* $\mu_0(D)$ *denotes the ground energy (i.e., minimal diagonal element) of* D*.*

Rescaling the path in Theorem 6.4.1. Without loss of generality, we assume that all diagonal elements of H_{prob} lie in $[-1, 0)$ for the linear Hamiltonian path from Theorem 6.4.1,

$$H_{\text{TFI}} = -\sum_{i\in[n]} a_i X_i + D_{\text{TFI}} \longrightarrow H_{\text{prob}}, \tag{6.178}$$

by an observation that the path $A \longrightarrow \alpha B + \beta I$ has spectral gap at least $\alpha\delta$ if the path $A \longrightarrow B$ has spectral gap δ for $\alpha \leq 1$. This way, the spectral gap of (6.178) is at least $1/\operatorname{poly}(n)$ and all other claims from Theorem 6.4.1 still hold.

Construction of the objective function f. We define

$$f(\xi) = \widehat{H}_{\text{prob}} + \sum_{i\in[n]} a_i f_{\text{dw}}(\xi_i), \quad f_{\text{dw}}(z) = (z - 1/2)^2(z + 1/2)^2. \tag{6.179}$$

Note that $f_{\text{dw}}(z)$ is a quartic double well function with global minima at $z = \pm 1/2$, as shown in Figure 6.5. In other words, the objective function f is essentially a continuous-space interpolation of the problem Hamiltonian H_{prob}. Due to the double well function f_{dw}, the local minima of f are attained at the points $\{\pm 1/2\}^n$. Since all the diagonal elements in H_{prob} are within $[-1, 0)$, so by Fact 6.10.5 (ii) applied on $\widehat{H}_{\text{prob}}$,

$$\min_{\xi\in[-1,1]^n} f(\xi) = f(\xi^*) = \mu_0(H_{\text{prob}}) < 0, \quad \xi^* = \arg\min_{\xi\in[-1,1]^n} f(\xi). \tag{6.180}$$

The negativity of the global minimum of $f(\xi)$ follows from the fact that all diagonal elements in H_{prob} are negative.

Remark 6.10.6. The actual potential function $f(x)$ in (6.172) is a (spatially) rescaled version of $f(\xi)$. The details can be found in the proof of Theorem 6.10.2; cf. Section 6.10.2.

6.10.2 Proof of the Theorem

Notation and Convention. From here to the end of this section, we adopt the following notational conventions and assumptions. Denote a Hamiltonian acting on position space by a hat, e.g., $\widehat{H}$. All wave functions introduced are *real*-valued.[11] We will frequently use ξ to denote an element in $\mathbb{R}$ or $\mathbb{R}^n$ while reserving x for binary strings in $\{0,1\}^n$. For a state $|\phi\rangle$, we denote its corresponding wave function by $\Psi_\phi(\cdot)$ or $\phi(\cdot)$. Let $\mu_j(\widehat{H})$ be the $(j+1)$th smallest eigenvalue of $\widehat{H}$, counting multiplicity.[12]

First, we present the key reduction lemma. We define an operator $\widehat{X}$ in the interval $[-1,1]$ (with vanishing boundary condition):

$$\widehat{X}(\lambda) = -\frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + \lambda^2 f_{\text{dw}}(\xi), \tag{6.181}$$

parameterized by some $\lambda > 0$. Let $|\chi_0\rangle$ and $|\chi_1\rangle$ be the ground and first-excited state of $\widehat{X}$, respectively; and we assume $\chi_0(\xi), \chi_1(\xi) \geq 0$ for $\xi \geq 0$ by fixing their global phases; see Figure 6.5a. Let $\mathfrak{T} \coloneqq \operatorname{span}(|\chi_0\rangle, |\chi_1\rangle)$ be the low-energy subspace of $\widehat{X}$. Let $\mathfrak{S} \coloneqq \mathfrak{T}^{\otimes n}$ and we define an isometry $\mathcal{E}\colon \mathbb{C}^{2^n} \to \mathfrak{S}$:

$$\mathcal{E} \coloneqq (|\chi_0\rangle\langle +| + |\chi_1\rangle\langle -|)^{\otimes n}, \quad \text{where } |\pm\rangle \coloneqq \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle). \tag{6.182}$$

[11] This assumption is valid because, for example, any Schrödinger operator is a real differential operator, and thus all its eigenfunctions can be chosen to be real.

[12] $\mu_j(\cdot)$ is well defined for any self-adjoint operator that is bounded from below; see Lemma 5.4.27.

Lemma 6.10.7. *Let $\{a_i \geq 0\}_i$ be a set of fixed coefficients and D be an arbitrary n-qubit diagonal Hamiltonian. For any $\lambda > 0$, we define two operators:*

- *The operator $\widehat{H}(\lambda)$ is a Hamiltonian defined over the box $[-1,1]^n \subset \mathbb{R}^n$:*

$$\widehat{H}(\lambda) \coloneqq \sum_{i\in[n]} a_i \widehat{X}_i(\lambda) + \widehat{D}, \quad \text{where } \widehat{X}_i \coloneqq -\frac{\mathrm{d}^2}{\mathrm{d}\xi_i^2} + \lambda^2 f_{\mathrm{dw}}(\xi_i) \text{ for } i \in [n], \tag{6.183}$$

- *The operator $H(\lambda)$ is an n-qubit Hamiltonian:*

$$H(\lambda) \coloneqq -\sum \frac{a_i}{\Lambda(\lambda)} X_i + D, \quad \text{where } \frac{\mathrm{d}^k \Lambda}{\mathrm{d}\lambda^k} = \left(e^{\lambda/6}\right)^{1\pm o_k(1)} \text{ for } k \in \mathbb{N}.^{13} \tag{6.184}$$

Then, there exists $\varepsilon \leq O(\sqrt{n}\Lambda^{-1/6}\|D\|)$ such that the followings hold:

(i) The ground energy of $\widehat{H}(\lambda)$, denoted by $\mu_0(\widehat{H})$, is no more than

$$\mu_0(H) + \sum_{i\in[n]} a_i \cdot \Theta(\lambda) + \varepsilon. \tag{6.185}$$

(ii) The spectral gap of $\widehat{H}(\lambda)$, denoted by $\delta \coloneqq \mu_1(\widehat{H}) - \mu_0(\widehat{H})$, is at least

$$\delta \coloneqq \min\left\{\mu_1(H), \min_{i\in[n]}\{a_i\} \cdot \Theta(\lambda) + \mu_0(D)\right\} - \mu_0(H) - \varepsilon. \tag{6.186}$$

(iii) If $\delta > 0$, the ground state of $\widehat{H}$ is $O(\varepsilon/\delta)$-close to $\mathcal{E}|\psi_0\rangle$, where $|\psi_0\rangle$ denotes the ground state of $H(\lambda)$.

Now, we are ready to prove Theorem 6.10.2. It is worth noting that, in the proof, we use $\widehat{H}(t)$ to

[13] $\Lambda(\lambda)$ is an efficiently computable function being half the inverse of the spectral gap of $\widehat{X}$; see Claim 6.10.8.

denote the Hamiltonian path in the real space to avoid confusion with Lemma 6.10.7, which is slightly different from the theorem statement.

Proof of Theorem 6.10.2. Item (i) follows directly from our construction of the objective function f. Specifically, by the definition as in (6.179), f can be efficiently computed from a given linear Hamiltonian path $H(t)$, assuming oracle access to the n-qubit diagonal Hamiltonian D. Consequently, any classical query algorithm that minimizes f would also effectively solve the discrete optimization problem defined by D, incurring at most a polynomial overhead. Thus, the classical query lower bound established in Theorem 6.4.1 (i) still applies.

Next, we turn to Item (ii), showing the quantum advantage of the Hamiltonian path (6.172). Let a_i, D_{TFI}, and H_{prob} be the same as in a chosen Hamiltonian path $H_{\mathrm{TFI}} \longrightarrow H_{\mathrm{prob}}$. Let $\Lambda(\lambda)$ be the same as in Lemma 6.10.7, and $\lambda_0 > 0$ be a fixed number that will be specified later. We denote $\Lambda_0 = \Lambda(\lambda_0)$ and define

$$g(\xi) \coloneqq \frac{1}{\Lambda_0}\widehat{D}_{\mathrm{TFI}}(\xi) + \lambda_0^2 \sum_i^n a_i f_{\mathrm{dw}}(\xi_i), \tag{6.187}$$

$$\nu(\lambda) \coloneqq \frac{\Lambda_0}{\Lambda(\lambda)}, \quad \varphi(\lambda) \coloneqq \lambda^2 - \nu(\lambda)\lambda_0^2. \tag{6.188}$$

By Lemma 6.10.7, we have that $\nu(\lambda) = O(e^{-c\lambda})$ for some absolute constant $c > 0$. For all $\lambda \geq \lambda_0$, we have $\varphi(\lambda) \geq 0$ and $\varphi(\lambda) = \Theta(\lambda^2)$.

As discussed earlier, the eigenvalues of H_{diag} from (6.178) lie in $[-1, 0)$. Let $f(\xi)$ be the objective function described in (6.179). We define the following Hamiltonian path for any $\lambda > \lambda_0$:

$$\widehat{H}(\lambda) \coloneqq -\sum_i a_i \frac{\partial^2}{\partial_{\xi_i}^2} + \nu(\lambda)g(\xi) + \varphi(\lambda)f(\xi) \tag{6.189}$$

$$= -\sum_i a_i \widehat{X}_i(\lambda) + \left(\frac{1}{\Lambda(\lambda)} \widehat{D}_{\mathrm{TFI}} + \varphi(\lambda) \widehat{H}_{\mathrm{prob}} \right). \tag{6.190}$$

According to Lemma 6.10.7, the spectral properties of the real-space Hamiltonian path $\widehat{H}(\lambda)$ can be studied by investigating the following n-qubit Hamiltonian:

$$H(\lambda) \coloneqq -\frac{1}{\Lambda(\lambda)} \sum_i a_i X_i + \left(\frac{1}{\Lambda(\lambda)} D_{\mathrm{TFI}} + \varphi(\lambda) H_{\mathrm{prob}} \right) \tag{6.191}$$

$$= \frac{1}{\Lambda(\lambda)} H_{\mathrm{TFI}} + \varphi(\lambda) H_{\mathrm{prob}}. \tag{6.192}$$

We note that the ground state of the initial Hamiltonian $H(\lambda_0)$ is equivalent to that of H_{TFI}. By Theorem 6.4.1 (ii), the ground state of H_{TFI} can be approximated by a product state up to an error $n^{-\Omega(1)}$. Therefore, by Lemma 6.10.7 (iii), the ground state of $\widehat{H}(\lambda_0)$ can be efficiently prepared up to an error $n^{-\Omega(1)} + O(\varepsilon/\delta) \leq n^{-\Omega(1)}$, where ε and $\delta = \delta(\lambda)$ are specified later in (6.196).

By Theorem 6.4.1, the Hamiltonian $H(\lambda)$ has a spectral gap:

$$\mu_1(H(\lambda)) - \mu_0(H(\lambda)) \geq \left(\varphi(\lambda) + \frac{1}{\Lambda(\lambda)} \right) \cdot \frac{1}{\mathrm{poly}(n)} = \Theta\left(\frac{\lambda^2}{\mathrm{poly}(n)} \right). \tag{6.193}$$

For notational convenience, we denote $D(\lambda) \coloneqq \Lambda^{-1}(\lambda) D_{\mathrm{TFI}} + \varphi(\lambda) H_{\mathrm{prob}}$. By Weyl's inequality[14], we have

$$\mu_0(D(\lambda)) - \mu_0(H(\lambda)) \geq -\frac{1}{\Lambda(\lambda)} \sum_i a_i \geq -\frac{\mathrm{poly}(n)}{\Lambda(\lambda)}. \tag{6.194}$$

[14]This is the usual Weyl's inequality [226] generalized to semi-bounded self-adjoint operators; a proof is implicit in Lemma 5.4.28.

Since $\Lambda(\lambda)$ decays exponentially fast, we can choose $\lambda_0 = \log(\mathrm{poly}(n))$ so that for $\lambda \geq \lambda_0$,

$$\min_i\{a_i\} \cdot \Theta(\lambda) + \mu_0(D) - \mu_0(H) \geq \Theta(\lambda), \tag{6.195}$$

where we used the fact $\min_i\{a_i\} \geq 1$ by examining (6.46), the explicit form of the path in Theorem 6.4.1. The spectral gap of $\widehat{H}(t)$ can be deduced from the above results and Lemma 6.10.7 (ii):

$$\delta(\lambda) \geq \min\left(\frac{\varphi(\lambda) + \frac{1}{\Lambda(\lambda)}}{\mathrm{poly}(n)}, \Theta(\lambda)\right) - \varepsilon, \quad \varepsilon \leq O(\sqrt{n}\|D\|e^{-c\lambda/6}). \tag{6.196}$$

Since $\|D\|$ is continuous in λ and tends to infinity when $\lambda \to +\infty$, there exists a $\lambda' = \Theta(1)$ such that $\|D(\lambda)\| \geq 1$ for all $\lambda \geq \lambda'$. Therefore, we may choose

$$\lambda_0 = \max\left(1, \lambda', \log(\mathrm{poly}(n))\right), \tag{6.197}$$

such that

$$\delta(\lambda) \geq \frac{1}{\mathrm{poly}(n)} \quad \text{and} \quad \varepsilon \leq \frac{\delta(\lambda)}{n^{\Omega(1)}}, \quad \forall \lambda \geq \lambda_0. \tag{6.198}$$

To recover the Hamiltonian path as described in Theorem 6.10.2, we need to apply both spatial and temporal rescalings. First, we consider the change of (spatial variable):

$$z_i = \frac{1}{\sqrt{a_i}}\xi_i, \quad \forall i \in [n]. \tag{6.199}$$

Then, by the chain rule, $\frac{\partial}{\partial_{\xi_i}} = \frac{\partial z_i}{\partial \xi_i}\frac{\partial}{\partial_{z_i}} = a_i^{-1/2}\frac{\partial}{\partial_{z_i}}$. Now, we consider the rescaled box $\mathcal{X} = \bigtimes_{i=1}^n [-a_i^{-1/2}, a_i^{-1/2}]$

and the new Hamiltonian operator:

$$\widehat{H}(\lambda, z) = -\Delta + \nu(\lambda) g(\xi(z)) + \varphi(\lambda) f(\xi(z)), \quad z \in \mathcal{X}. \tag{6.200}$$

Since all eigenstates can be rescaled in the same way, this new Hamiltonian has a $1/\operatorname{poly}(n)$ spectral gap, as discussed before. Moreover, the diameter of the new box $\mathcal{X}$ is at most $\operatorname{poly}(n)$.

Next, we consider temporal scaling. Note that by $\lambda_0 \geq \log(\operatorname{poly}(n))$ from (6.197), the function $\varphi(\lambda)$ is monotonically increasing in λ with $\varphi(\lambda_0) = 0$. We can thus introduce a new variable:

$$t \coloneqq \varphi^{-1}(\lambda) \in [0, \infty], \quad \lambda \in [\lambda_0, \infty]. \tag{6.201}$$

This change of variable leads to the Hamiltonian path:

$$\widehat{H}(t) = -\Delta + \nu(t) g(z) + t f(z), \tag{6.202}$$

where $\nu(t) = \nu(\varphi^{-1}(\lambda))$. Since $\varphi(\lambda) = \Theta(\lambda^2)$, the function $\nu(t)$ decays at a (sub)exponential rate: $\nu(t) = e^{-\Theta(\sqrt{t})}$. Note that both $\dot{\nu}(t)$ and $\ddot{\nu}(t)$ decay at a (sub)exponential rate that is *independent of* n by (6.188) and (6.184). Moreover, g and f are bounded within $\operatorname{poly}(n)$ by construction. It follows that $\|\dot{\widehat{H}}\|$ and $\|\ddot{\widehat{H}}\|$ is also bounded within $\operatorname{poly}(n)$ for all $t \geq 0$.

Therefore, by Lemma C.1.2, we can simulate the Schrödinger dynamics for $0 \leq t \leq t_{\mathrm{f}}$ and obtain a final state $|\psi(t_{\mathrm{f}})\rangle$ that is $n^{-\Omega(1)}$-close to $|\phi(t_{\mathrm{f}})\rangle$. Define the projector $\widehat{P}$ that projects a wave function onto the subspace supported on the region where $f(\xi)$ is η-close to the global minimum $f(\xi^*)$:

$$(\widehat{P}\psi)(\xi) = \psi(\xi) \cdot \mathbf{1}\{f(\xi) - f(\xi^*) \leq \eta\}. \tag{6.203}$$

Now, we claim $\langle\phi(t_{\mathrm{f}})|f|\phi(t_{\mathrm{f}})\rangle - f(\xi^*) \leq \eta/10$ for some $t_{\mathrm{f}} = \mathrm{poly}(n, 1/\eta)$. Recall that $\lambda = \varphi(t)$. By the positivity of the operator $-\frac{\partial^2}{\partial_\xi^2}$ and Lemma 6.10.7 (i), we have

$$\varphi(\lambda)\langle\phi(t)|f|\phi(t)\rangle \leq \langle\phi(t)|\widehat{H}(t)|\phi(t)\rangle \leq \mu_0(H(\lambda)) + \sum_i a_i \cdot \Theta(\lambda) + \varepsilon, \tag{6.204}$$

where $\varepsilon \leq O(\sqrt{n}\|D\|e^{-c\lambda})$ is exponentially small in λ. Also, we have that $H(\lambda) = -\frac{1}{\Lambda(\lambda)}H_{\mathrm{TFI}} + \varphi(\lambda)H_{\mathrm{prob}}$, and $1/\Lambda(\lambda)$ is again exponentially small in λ. By Weyl's inequality,

$$\langle\phi(t)|f|\phi(t)\rangle \leq \mu_0(H_{\mathrm{prob}}) + \sum_i a_i \cdot \Theta(\lambda^{-1}) + \varepsilon', \tag{6.205}$$

where ε' is exponentially small in λ. By construction, we have $\mu_0(H_{\mathrm{prob}}) = f(\xi^*)$; meanwhile, $\sum_i a_i \leq \mathrm{poly}(n)$ by the norm condition of the path in Theorem 6.4.1 (ii). Therefore, for any $\eta > 0$, we can choose $\lambda' = \mathrm{poly}(n, 1/\eta)$ and $t_{\mathrm{f}} = \varphi^{-1}(\lambda') = \mathrm{poly}(n, 1/\eta)$ such that

$$\langle\phi(t_{\mathrm{f}})|f|\phi(t_{\mathrm{f}})\rangle - f(\xi^*) \leq \eta/10. \tag{6.206}$$

Finally, by invoking Markov's inequality on (6.206), we have

$$\langle\phi(t_{\mathrm{f}})|\widehat{P}|\phi(t_{\mathrm{f}})\rangle = \Pr_{\xi\sim|\phi(t_{\mathrm{f}})|^2}[f(\xi) - f(\xi^*) \leq \eta] \geq 1 - \frac{1}{10} = 0.9. \tag{6.207}$$

As a result,

$$\begin{aligned}
&\Pr_{\xi\sim|\psi(t_{\mathrm{f}})|^2}[f(\xi) - f(\xi^*) \leq \eta] \\
&= \langle\psi(t_{\mathrm{f}})|\widehat{P}|\psi(t_{\mathrm{f}})\rangle
\end{aligned}$$

$$= \langle\phi(t_f)|\widehat{P}|\phi(t_f)\rangle + \langle\phi(t_f)|\widehat{P}(|\psi(t_f)\rangle - |\phi(t_f)\rangle) + (\langle\psi(t_f)| - \langle\phi(t_f)|)\widehat{P}|\psi\rangle$$

$$\geq 0.9 - 2 \cdot \||\psi(t_f)\rangle - |\phi(t_f)\rangle\| \cdot \|P\|$$

$$= 0.9 - n^{-\Omega(1)},$$

as stated in Theorem 6.10.2 (ii). □

6.10.3 Simulating TFD Hamiltonians with Schrödinger Operators

To prove Lemma 6.10.7, we first introduce the required definitions and notation.

Effective computational basis. Recall that $|\chi_0\rangle$ and $|\chi_1\rangle$ are the ground state and the first excited state of $\widehat{X}$ defined in (6.181). We define the following states which serve as analogues of the standard qubit states $|0\rangle$ and $|1\rangle$ in the context of $\widehat{X}$:

$$|\widehat{0}\rangle := \frac{1}{\sqrt{2}}\left(|\chi_0\rangle + |\chi_1\rangle\right) \quad \text{and} \quad |\widehat{1}\rangle := \frac{1}{\sqrt{2}}\left(|\chi_0\rangle - |\chi_1\rangle\right). \tag{6.208}$$

Moreover, for any $x = x_1 x_2 \ldots x_n \in \{0,1\}^n$, define the effective computational state

$$|\widehat{x}\rangle := |\widehat{x}_1\rangle \otimes |\widehat{x}_2\rangle \otimes \cdots \otimes |\widehat{x}_n\rangle. \tag{6.209}$$

Recall the definition of the encoding isometry $\mathcal{E}$ from (6.182). We have $\mathcal{E} = (|\chi_0\rangle\langle +| + |\chi_1\rangle\langle -|)^{\otimes n} = (|\widehat{0}\rangle\langle 0| + |\widehat{1}\rangle\langle 1|)^{\otimes n}$. Therefore, for any $x \in \{0,1\}^n$ it holds that

$$\mathcal{E}|x\rangle = |\widehat{x}\rangle. \tag{6.210}$$

Let $\mathfrak{T} \coloneqq \mathrm{span}(|\chi_0\rangle, |\chi_1\rangle) = \mathrm{span}(|\widehat{0}\rangle, |\widehat{1}\rangle)$ be the low-energy subspace of $\widehat{X}$, and let $\mathfrak{S} \coloneqq \mathfrak{T}^{\otimes n}$ be the space spanned by $\{|\widehat{x}\rangle\}_{x\in\{0,1\}^n}$.

Restricted inner product. We use the following notation for the inner product restricted to the region $\mathcal{I}$:

$$\langle\psi|\phi\rangle_{\mathcal{I}} \coloneqq \int_{\mathcal{I}} \psi(\xi)\phi(\xi)\mathrm{d}\xi. \tag{6.211}$$

We present a key claim about $\widehat{X}$ based on Theorem 5.4.23 and its proof.[15]

Claim 6.10.8 (Properties of $\widehat{X}$)**.** *The following properties hold for $\widehat{X}$:*

(i) $\mu_0(\widehat{X}) = \Theta(\lambda)$, $\mu_1(\widehat{X}) - \mu_0(\widehat{X}) = 2/\Lambda$, *and* $\mu_2(\widehat{X}) - \mu_1(\widehat{X}) = \Theta(\lambda)$, *where* $\Lambda = \Lambda(\lambda)$ *satisfies (6.184).*

(ii) $|\widehat{0}\rangle$ *and* $|\widehat{1}\rangle$ *are symmetric under reflection, i.e.,* $\Psi_{\widehat{0}}(\xi) = \Psi_{\widehat{1}}(-\xi)$. *Consequently,*

$$\langle\widehat{0}|\widehat{1}\rangle_{[0,1]} = \langle\widehat{0}|\widehat{1}\rangle_{[-1,0]} = \frac{1}{2}\langle\widehat{0}|\widehat{1}\rangle = 0. \tag{6.212}$$

(iii) There exists $w \in (0, 1/2)$ *such that the wave function* $\Psi_{\widehat{0}}(\cdot)$ *is concentrated within the interval* $(w, 1]$ *in the sense that*

$$\langle\widehat{0}|\widehat{0}\rangle_{[-1,w]} \le O(\Lambda^{-2/3}), \tag{6.213}$$

Similarly, $\langle\widehat{1}|\widehat{1}\rangle_{[-w,1]} \le O(\Lambda^{-2/3})$ *due to the symmetry from Item (ii).*

Using Claim 6.10.8 and the definition of $\widehat{D}$ from (6.177), we establish the following claims about $\widehat{X}$ and $\widehat{D}$. Their proofs, which involve nontrivial technical arguments for $\widehat{D}$, are deferred to

[15] Note that in Chapter 5, the parameters are given as $\lambda = 1/h$, $S_0 = 1/6$, and $a = \pm 1/2$. Here, S_0 denotes the *Agmon distance* between $-a$ and a with respect to the potential function f_{dw} at energy $E = 0$.

Section 6.10.4.

Claim 6.10.9 ($\widehat{X}$ effectively implements X)**.** *Denote* $\mu_0 \coloneqq \mu_0(\widehat{X})$.

(i) $\widehat{X}_i$ *acts on subspace* $\mathfrak{S}$ *the same way* $-\Lambda^{-1}X_i$ *acts on the* n*-qubit Hilbert space, modulo an overall energy shift* $\mu_0 + \Lambda^{-1}$*:*

$$P_{\mathfrak{S}}\left(\widehat{X}_i - \mathcal{E}\left(-\Lambda^{-1}X_i + \left(\mu_0 + \Lambda^{-1}\right)I\right)\mathcal{E}^\dagger\right)P_{\mathfrak{S}} = 0. \tag{6.214}$$

(ii) $P_{\mathfrak{S}^\perp}\widehat{X}_i P_{\mathfrak{S}} = 0$.

(iii) The minimum eigenvalue of $\widehat{X}_i$ *restricted on subspace* $\mathfrak{S}^\perp$ *is* $\mu_0 + \Theta(\lambda)$*:*

$$\mu_0\left(P_{\mathfrak{S}^\perp}\widehat{X}_i P_{\mathfrak{S}^\perp}\right) = \mu_0 + \Theta(\lambda). \tag{6.215}$$

Claim 6.10.10 ($\widehat{D}$ effectively implements D)**.**

(i) $\left\|P_{\mathfrak{S}}\left(\widehat{D} - \mathcal{E}D\mathcal{E}^\dagger\right)P_{\mathfrak{S}}\right\| \leq O(\sqrt{n}\Lambda^{-1/3}\|D\|)$.

(ii) $\left\|P_{\mathfrak{S}^\perp}\widehat{D}P_{\mathfrak{S}}\right\| \leq O(\sqrt{n}\Lambda^{-1/6}\|D\|)$.

(iii) $\mu_0\left(P_{\mathfrak{S}^\perp}\widehat{D}P_{\mathfrak{S}^\perp}\right) \geq \mu_0(\widehat{D}) = \mu_0(D)$.

Now, express $\widehat{H}$ in block form with respect to the decomposition $\mathfrak{S} \oplus \mathfrak{S}^\perp$ of the Hilbert space $\mathcal{W}([-1,1]^n)$:

$$\widehat{H} = \begin{bmatrix} P_{\mathfrak{S}}\widehat{H}P_{\mathfrak{S}} & P_{\mathfrak{S}}\widehat{H}P_{\mathfrak{S}^\perp} \\ P_{\mathfrak{S}^\perp}\widehat{H}P_{\mathfrak{S}} & P_{\mathfrak{S}^\perp}\widehat{H}P_{\mathfrak{S}^\perp} \end{bmatrix} =: \begin{bmatrix} \widehat{H}_{\mathfrak{S}} & R^\dagger \\ R & \widehat{H}_{\mathfrak{S}^\perp} \end{bmatrix}. \tag{6.216}$$

We are able to analyze $\widehat{H}_{\mathfrak{S}}$, $\widehat{H}_{\mathfrak{S}^\perp}$, and R, as follows.

- **Analysis of $\widehat{H}_{\mathfrak{S}}$.** Define

$$\eta \coloneqq \sum_{i\in[n]} a_i \left(\mu_0(\widehat{X}) + \Lambda^{-1}\right) = \sum_{i\in[n]} a_i \cdot \Theta(\lambda), \tag{6.217}$$

where the last equality follows from Claim 6.10.8 (i). Then, the following holds when restricted to the subspace $\mathfrak{S}$:

$$\widehat{H}_{\mathfrak{S}} - \mathcal{E}(H+\eta I)\mathcal{E}^\dagger \tag{6.218}$$

$$= \widehat{H} - \mathcal{E}(H+\eta I)\mathcal{E}^\dagger \tag{6.219}$$

$$= \underbrace{\sum_{i\in[n]} a_i \widehat{X}_i + \widehat{D}}_{\widehat{H}} - \mathcal{E}\left(\underbrace{-\sum_{i\in[n]} \frac{a_i}{\Lambda} X_i + D}_{H} + \eta I\right)\mathcal{E}^\dagger \tag{6.220}$$

$$= \sum_{i\in[n]} a_i \widehat{X}_i - \mathcal{E}\left(-\sum_{i\in[n]} \frac{a_i}{\Lambda} X_i + \underbrace{\sum_{i\in[n]} a_i \left(\mu_0(\widehat{X}) + \Lambda^{-1}\right)}_{\eta} I\right)\mathcal{E}^\dagger + \left(\widehat{D} - \mathcal{E}D\mathcal{E}^\dagger\right) \tag{6.221}$$

$$= \sum_{i\in[n]} a_i \left(\underbrace{\widehat{X}_i - \mathcal{E}\left(-\frac{1}{\Lambda} X_i + \left(\mu_0(\widehat{X}) + \Lambda^{-1}\right) I\right)\mathcal{E}^\dagger}_{=0 \text{ by Claim 6.10.9 (i)}}\right) + \left(\widehat{D} - \mathcal{E}D\mathcal{E}^\dagger\right) \tag{6.222}$$

$$= \widehat{D} - \mathcal{E}D\mathcal{E}^\dagger. \tag{6.223}$$

Consequently, by Claim 6.10.10 (i),

$$\left\|\widehat{H}_{\mathfrak{S}} - \mathcal{E}(H+\eta I)\mathcal{E}^\dagger\right\| = \left\|P_{\mathfrak{S}}\left(\widehat{D} - \mathcal{E}D\mathcal{E}^\dagger\right)P_{\mathfrak{S}}\right\| \le O(\sqrt{n}\Lambda^{-1/3}\|D\|). \tag{6.224}$$

- **Analysis of $\widehat{H}_{\mathfrak{S}^\perp}$.** We now analyze the smallest eigenvalue of $\widehat{H}_{\mathfrak{S}^\perp}$. Recall that $\widehat{H} = \sum_i a_i \widehat{X}_i + \widehat{D}$. By the property of the Kronecker sum[16] and Claim 6.10.9 (iii), the spectrum[17] of $\sum_i a_i \widehat{X}_i$ is given by

$$\sigma\Big(\sum_{i\in[n]} a_i \widehat{X}_i\Big) = \sigma(a_1\widehat{X}) + \cdots + \sigma(a_n\widehat{X}) = \left\{ \sum_{i\in[n]} a_i \mu_{k_i}(\widehat{X}) \colon k_1, \ldots, k_n \in \mathbb{N} \right\}. \tag{6.225}$$

Since $\mathfrak{S} = \mathfrak{T}^{\otimes n}$, where $\mathfrak{T}$ is spanned by the ground state $|\chi_0\rangle$ and the first excited state $|\chi_1\rangle$ of $\widehat{X}$, any eigenstate of $\sum_i a_i \widehat{X}_i$ not in $\mathfrak{S}$ must correspond to at least one index $k_i \notin \{0,1\}$ in (6.225). Consequently, the minimum eigenvalue among those eigenstates is

$$\sum_{i\in[n]} a_i \mu_0(\widehat{X}) + \min_{i\in[n]}\{a_i\}\Big(\mu_2(\widehat{X}) - \mu_0(\widehat{X})\Big) \geq \eta + \min_{i\in[n]}\{a_i\} \cdot \Theta(\lambda). \tag{6.226}$$

By Weyl's inequality and Claim 6.10.10 (iii), the minimum eigenvalue of $\widehat{H}$ restricted to $\mathfrak{S}^\perp$ is at least

$$\mu_0\Big(P_{\mathfrak{S}} \sum_{i\in[n]} a_i \widehat{X}_i P_{\mathfrak{S}}\Big) + \mu_0(P_{\mathfrak{S}} \widehat{D} P_{\mathfrak{S}}) \geq \eta + \min_{i\in[n]}\{a_i\} \cdot \Theta(\lambda) + \mu_0(D). \tag{6.227}$$

- **Analysis of R.** It is straightforward from Claim 6.10.9 (ii) and Claim 6.10.10 (ii) that

$$\|R\| = \left\| P_{\mathfrak{S}^\perp} \left(\sum_{i\in[n]} a_i \widehat{X}_i + \widehat{D} \right) P_{\mathfrak{S}} \right\| = \left\| \sum_{i\in[n]} a_i \underbrace{P_{\mathfrak{S}^\perp} \widehat{X}_i P_{\mathfrak{S}}}_{=0} + P_{\mathfrak{S}^\perp} \widehat{D} P_{\mathfrak{S}} \right\| \leq O(\sqrt{n}\Lambda^{-1/6}\|D\|). \tag{6.228}$$

Summarizing (6.224), (6.227), and (6.228) we have the subsequent claim.

[16] This is Fact 6.3.8 generalized to (unbounded) self-adjoint operators; see [229, Corollary 7.25].

[17] We use the fact that the spectrum of the Schrödinger operator $\widehat{X}$ is purely discrete; see [79, Remark 18].

Claim 6.10.11 (Bounds on $\widehat{H}_{\mathfrak{S}}, \widehat{H}_{\mathfrak{S}^\perp}, R$)**.** *There exists $\varepsilon \leq O(\sqrt{n}\Lambda^{-1/6}\|D\|)$ such that the following holds:*

(i) *$\widehat{H}_{\mathfrak{S}}$ is $(\varepsilon/2)$-close to $\mathcal{E}(H+\eta I)\mathcal{E}^\dagger$ in operator norm distance.*

(ii) $\mu_0(\widehat{H}_{\mathfrak{S}^\perp}) \geq \eta + \min_i\{a_i\} \cdot \Theta(\lambda) + \mu_0(D)$.

(iii) $\|R\| \leq \varepsilon/4$.

We are now in position to prove Lemma 6.10.7.

Proof of Lemma 6.10.7. Recall from (6.216) that we write $\widehat{H} = \begin{bmatrix} \widehat{H}_{\mathfrak{S}} & R^\dagger \\ R & \widehat{H}_{\mathfrak{S}^\perp} \end{bmatrix}$. The ground energy of $\widehat{H}$ is thus at most the smallest eigenvalue of $\widehat{H}_{\mathfrak{S}}$, which, by Claim 6.10.11 (i), is in turn no more than

$$\mu_0\left(\mathcal{E}(H+\eta I)\mathcal{E}^\dagger\right) + \varepsilon = \mu_0(H) + \eta + \varepsilon = \mu_0(H) + \sum_{i\in[n]} a_i \cdot \Theta(\lambda) + \varepsilon, \tag{6.229}$$

as stated in Item (i).

For Item (ii), observe that the spectral gap of $\widehat{H}$ is at least

$$\min\left\{\mu_1(\widehat{H}_{\mathfrak{S}}), \mu_0(\widehat{H}_{\mathfrak{S}^\perp})\right\} - \mu_0(\widehat{H}_{\mathfrak{S}}) \tag{6.230}$$

$$\geq \min\left\{\underbrace{\mu_1(H) + \eta - \varepsilon/2}_{\text{by Claim 6.10.11 (i)}}, \underbrace{\eta + \min_{i\in[n]}\{a_i\} \cdot \Theta(\lambda) + \mu_0(D)}_{\text{by Claim 6.10.11 (ii)}}\right\} - \left(\underbrace{\mu_0(H) + \eta + \varepsilon/2}_{\text{by Claim 6.10.11 (i)}}\right) \tag{6.231}$$

$$\geq \min\left\{\mu_1(H), \min_{i\in[n]}\{a_i\} \cdot \Theta(\lambda) + \mu_0(D)\right\} - \mu_0(H) - \varepsilon =: \delta, \tag{6.232}$$

where we used Weyl's inequality in (6.231).

Regarding Item (iii), denote by $|\psi\rangle$ the ground state of $\widehat{H}$. We apply Fact 6.3.7[18] with $d = 1$ and

$$A = \begin{bmatrix} \widehat{H}_{\mathfrak{S}} & 0 \\ 0 & \widehat{H}_{\mathfrak{S}^\perp} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & R^\dagger \\ R & 0 \end{bmatrix} \tag{6.233}$$

to obtain

$$\|\mathcal{E}|g\rangle - |\psi\rangle\| \le \frac{2^{3/2}\|B\|}{\delta} = \frac{2^{3/2}\|R\|}{\delta} \le \varepsilon/\delta, \qquad (\|R\| \le \varepsilon \text{ by Claim 6.10.11 (iii)})$$

by noting that the ground state of A coincides with that of $\widehat{H}_{\mathfrak{S}}$ when $\delta > 0$ since $\mu_0(\widehat{H}_{\mathfrak{S}^\perp}) - \mu_0(\widehat{H}_{\mathfrak{S}}) \ge \delta$ from (6.232). □

6.10.4 Deferred Proofs

Proof of Claim 6.10.9. Note that $\widehat{X}_i$ is diagonal on the basis $\{|\chi_0\rangle, |\chi_1\rangle\}^{\otimes n}$ when restricted to subspace $\mathfrak{S}$, since $|\chi_0\rangle$ and $|\chi_1\rangle$ are eigenstates of $\widehat{X}$:

$$[\widehat{X}]_{\{|\chi_0\rangle,|\chi_1\rangle\}} = \begin{bmatrix} \mu_0(\widehat{X}) & 0 \\ 0 & \mu_1(\widehat{X}) \end{bmatrix} \overset{\text{Claim 6.10.8 (i)}}{=\!=} \begin{bmatrix} \mu_0 & 0 \\ 0 & \mu_0 + 2/\Lambda \end{bmatrix}. \tag{6.234}$$

Comparing it to

$$[-\Lambda^{-1}X]_{\{|+\rangle,|-\rangle\}} = \begin{bmatrix} -\Lambda^{-1} & 0 \\ 0 & \Lambda^{-1} \end{bmatrix}, \tag{6.235}$$

we conclude the proof of Item (i). The correctness of Item (ii) is obvious by diagonalizing $\widehat{X}$, and Item (iii) follows from Claim 6.10.8 (i) directly. □

[18] The Davis–Kahan $\sin\theta$ theorem works for unbounded self-adjoint operators; see [230, Section 6].

To prepare for the proof of Claim 6.10.10, we introduce some necessary notation and concepts.

Wave function truncation. In light of Claim 6.10.8 (iii), we define truncated states $|\widetilde{0}\rangle$ and $|\widetilde{1}\rangle$, supported only on the "right" and "left" halves of the domain $[-1,1]$, aligned with where $|\widehat{0}\rangle$ and $|\widehat{1}\rangle$ are concentrated, respectively:

$$\Psi_{\widetilde{0}}(\xi) = \begin{cases} \Psi_{\widehat{0}}(\xi), & \xi \in (0,1], \\ 0, & \xi \in [-1,0], \end{cases} \qquad \Psi_{\widetilde{1}}(\xi) = \begin{cases} \Psi_{\widehat{1}}(\xi), & \xi \in [-1,0), \\ 0, & \xi \in [0,1]. \end{cases} \tag{6.236}$$

Define $|\mathring{0}\rangle \coloneqq |\widehat{0}\rangle - |\widetilde{0}\rangle$ and $|\mathring{1}\rangle \coloneqq |\widehat{1}\rangle - |\widetilde{1}\rangle$. Claim 6.10.8 (iii) can then be relaxed and translated to

$$\langle\mathring{0}|\mathring{0}\rangle = \langle\mathring{1}|\mathring{1}\rangle \le O(\Lambda^{-2/3}). \tag{6.237}$$

Moreover, for any $x = x_1 x_2 \dots x_n \in \{0,1\}^n$, we define

$$|\widetilde{x}\rangle \coloneqq |\widetilde{x}_1\rangle \otimes |\widetilde{x}_2\rangle \otimes \cdots \otimes |\widetilde{x}_n\rangle, \quad |\mathring{x}\rangle \coloneqq |\widehat{x}\rangle - |\widetilde{x}\rangle. \tag{6.238}$$

Note that $|\widehat{x}_i\rangle = |\widetilde{x}_i\rangle + |\mathring{x}_i\rangle$ is an orthogonal decomposition. Hence we have the following orthogonal decomposition:

$$\underbrace{|\widehat{x}_1\rangle|\widehat{x}_2\rangle\cdots|\widehat{x}_n\rangle}_{|\widehat{x}\rangle} = \underbrace{\sum_{i\in[n]} |\widetilde{x}_1\rangle\cdots|\widetilde{x}_{i-1}\rangle|\mathring{x}_i\rangle|\widehat{x}_{i+1}\rangle\cdots|\widehat{x}_n\rangle}_{|\mathring{x}\rangle} + \underbrace{|\widetilde{x}_1\rangle|\widetilde{x}_2\rangle\cdots|\widetilde{x}_n\rangle}_{|\widetilde{x}\rangle}. \tag{6.239}$$

Therefore, by the Pythagorean theorem,

$$\langle \mathring{x}|\mathring{\widetilde{x}}\rangle \le \sum_{i\in[n]} \langle \mathring{x}_i|\mathring{\widetilde{x}}_i\rangle \le O(n\Lambda^{-2/3}), \quad \langle \widetilde{x}|\widetilde{\widetilde{x}}\rangle = 1 - \langle \mathring{x}|\mathring{\widetilde{x}}\rangle \ge 1 - O(n\Lambda^{-2/3}). \tag{6.240}$$

Decomposition of $\widehat{D}$. Recall the definition of $\theta(\cdot)$ from (6.176). To analyze $\widehat{D}$, we decompose it into a principal part and a remainder:

$$\widehat{D} = \widehat{D}_{\mathrm{p}} + \widehat{D}_{\mathrm{r}}, \tag{6.241}$$

where the principal part $\widehat{D}_{\mathrm{p}}$ is defined by

$$\widehat{D}_{\mathrm{p}}(\xi) = \widehat{D}_{\mathrm{p}}(\xi_1, \ldots, \xi_n) \coloneqq \langle x|D|x\rangle, \quad x \coloneqq \theta(\xi). \tag{6.242}$$

We now examine $\widehat{D}_{\mathrm{p}}$.

Claim 6.10.12 (Properties of $\widehat{D}_{\mathrm{p}}$)**.** *The following properties hold for* $\widehat{D}_{\mathrm{p}}$.

(i) $\left\| P_{\mathfrak{S}} \left(\widehat{D}_{\mathrm{p}} - \mathcal{E} D \mathcal{E}^{\dagger} \right) P_{\mathfrak{S}} \right\| \le O(\sqrt{n}\Lambda^{-1/3}\|D\|)$.

(ii) $\left\| P_{\mathfrak{S}^{\perp}} \widehat{D}_{\mathrm{p}} P_{\mathfrak{S}} \right\| \le O(\sqrt{n}\Lambda^{-1/6}\|D\|)$.

Proof. Define intervals $\mathcal{I}_0 \coloneqq (0, 1]$ and $\mathcal{I}_1 \coloneqq [-1, 0)$. Note that the supports of wave functions $\Psi_{\widetilde{0}}(\cdot)$ and $\Psi_{\widetilde{1}}(\cdot)$ are $\mathcal{I}_0$ and $\mathcal{I}_1$, respectively. We further define

$$\mathcal{I}(x) \coloneqq \mathcal{I}_{x_1} \times \mathcal{I}_{x_2} \times \cdots \times \mathcal{I}_{x_n}. \tag{6.243}$$

It is immediate that the support of wave function $\Psi_{\widetilde{x}}(\cdot)$ is $\mathcal{I}(x)$. Also, by definition (6.242), we have

$$\widehat{D}_{\mathrm{p}}(\xi) = \langle x|D|x\rangle \quad \text{for} \quad \xi \in \mathcal{I}(x). \tag{6.244}$$

Therefore,

$$\langle \widetilde{x}|\widehat{D}_{\mathrm{p}}|\widetilde{x}\rangle = \langle x|D|x\rangle\langle \widetilde{x}|\widetilde{x}\rangle. \tag{6.245}$$

For any $x \in \{0,1\}^n$, we have

$$\begin{aligned}
\langle \widehat{x}|\widehat{D}_{\mathrm{p}}|\widehat{x}\rangle &= \underbrace{\langle \widetilde{x}|\widehat{D}_{\mathrm{p}}|\widetilde{x}\rangle + \langle \widetilde{x}|\widehat{D}_{\mathrm{p}}|\mathring{x}\rangle}_{\langle \widetilde{x}|\widehat{D}_{\mathrm{p}}|\widehat{x}\rangle} + \langle \mathring{x}|\widehat{D}_{\mathrm{p}}|\widehat{x}\rangle && \text{(by (6.238))} \\
&= \langle x|D|x\rangle\langle \widetilde{x}|\widetilde{x}\rangle + \langle \widetilde{x}|\widehat{D}_{\mathrm{p}}|\mathring{x}\rangle + \langle \mathring{x}|\widehat{D}_{\mathrm{p}}|\widehat{x}\rangle && \text{(by (6.245))} \\
&\le \langle x|D|x\rangle\langle \widetilde{x}|\widetilde{x}\rangle - 2\cdot \||\mathring{x}\rangle\|\cdot\|D\| && (\|\widehat{D}_{\mathrm{p}}\| = \|D\| \text{ by (6.242)}) \\
&\le \langle x|D|x\rangle(1 - O(n\Lambda^{-2/3})) - \|D\|\cdot O(\sqrt{n}\Lambda^{-1/3}) && \text{(by (6.240))} \\
&= \langle x|D|x\rangle - \|D\|\cdot O(\sqrt{n}\Lambda^{-1/3}). && (\langle x|D|x\rangle \le \|D\|)
\end{aligned}$$

On the other hand, for any $x \neq y$, assume that $x_j \neq y_j$ for some j. We have for any $z \in \{0,1\}^n$ that

$$\langle \widehat{x}|\widehat{y}\rangle_{\mathcal{I}(z)} = \prod_{i\in[n]} \langle \widehat{x}_i|\widehat{y}_i\rangle_{\mathcal{I}_{z_i}} = 0, \tag{6.246}$$

since $\langle \widehat{x}_j|\widehat{y}_j\rangle_{\mathcal{I}_{z_j}} = 0$ by Claim 6.10.8 (ii). Therefore,

$$\begin{aligned}
\langle \widehat{x}|\widehat{D}_{\mathrm{p}}|\widehat{y}\rangle &= \sum_{z\in\{0,1\}^n} \langle \widehat{x}|\widehat{D}_{\mathrm{p}}|\widehat{y}\rangle_{\mathcal{I}(z)} && (6.247) \\
&= \sum_{z\in\{0,1\}^n} \langle z|D|z\rangle\langle \widehat{x}|\widehat{y}\rangle_{\mathcal{I}(z)} && \text{(by (6.244))} \\
&= \sum_{z\in\{0,1\}^n} \langle z|D|z\rangle\cdot 0 = 0. && \text{(by (6.246))}
\end{aligned}$$

Item (i) thus follows from our estimates of $\langle \widehat{x}|\widehat{D}_{\mathrm{p}}|\widehat{x}\rangle$ and $\langle \widehat{x}|\widehat{D}_{\mathrm{p}}|\widehat{y}\rangle$ above.

For Item (ii), let $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |\widehat{x}\rangle$ be an arbitrary state in $\mathfrak{S}$. We first estimate $\left\|\widehat{D}_{\mathrm{p}}|\psi\rangle\right\|$:

$$\begin{aligned}
\left\|\widehat{D}_{\mathrm{p}}|\psi\rangle\right\|^2 &= \sum_{z\in\{0,1\}^n} \langle z|D|z\rangle^2 \langle\psi|\psi\rangle_{\mathcal{I}(z)} && \text{(by (6.244))} \\
&= \sum_z \langle z|D|z\rangle^2 \sum_x \alpha_x^2 \langle\widehat{x}|\widehat{x}\rangle_{\mathcal{I}(z)} && \text{(by (6.246))} \\
&= \sum_z \langle z|D|z\rangle^2 \alpha_z^2 \langle\widehat{z}|\widehat{z}\rangle_{\mathcal{I}(z)} + \sum_z \langle z|D|z\rangle^2 \sum_{x\colon x\neq z} \alpha_x^2 \langle\widehat{x}|\widehat{x}\rangle_{\mathcal{I}(z)} && (6.248) \\
&\le \sum_z \langle z|D|z\rangle^2 \alpha_z^2 + \|D\|^2 \sum_x \alpha_x^2 \sum_{z\colon z\neq x} \langle\widehat{x}|\widehat{x}\rangle_{\mathcal{I}(z)} && (6.249) \\
&= \sum_z \langle z|D|z\rangle^2 \alpha_z^2 + \|D\|^2 \sum_x \alpha_x^2 \langle\mathring{x}|\mathring{x}\rangle && (6.250) \\
&= \left\|D\mathcal{E}^\dagger|\psi\rangle\right\|^2 + \|D\|^2 \cdot O(n\Lambda^{-2/3}), && (6.251)
\end{aligned}$$

where the last (in)equality follows from (6.240), and the equality $\sum_{z\colon z\neq x} \langle\widehat{x}|\widehat{x}\rangle_{\mathcal{I}(z)} \le \langle\mathring{x}|\mathring{x}\rangle$ used in (6.250) comes from the fact that the disjoint union of $\mathcal{I}(z)$ is exactly the support of $\Psi_{\mathring{x}}(\cdot)$, which is in turn simply $\Psi_{\widehat{x}}(\cdot)$ truncated.

For notational convenience, define

$$r \coloneqq \left\|D\mathcal{E}^\dagger|\psi\rangle\right\| = \left\|P_{\mathfrak{S}}\mathcal{E}D\mathcal{E}^\dagger P_{\mathfrak{S}}|\psi\rangle\right\| \le \|D\| \quad \text{and} \quad s \coloneqq \sqrt{n}\Lambda^{-1/3}\|D\|. \tag{6.252}$$

Finally, by the Pythagorean theorem,

$$\begin{aligned}
\left\|P_{\mathfrak{S}^\perp}\widehat{D}_{\mathrm{p}}|\psi\rangle\right\|^2 &= \left\|\widehat{D}_{\mathrm{p}}|\psi\rangle\right\|^2 - \underbrace{\left\|P_{\mathfrak{S}}\widehat{D}_{\mathrm{p}}|\psi\rangle\right\|^2}_{=\left\|P_{\mathfrak{S}}\widehat{D}_{\mathrm{p}}P_{\mathfrak{S}}|\psi\rangle\right\|^2} && (6.253) \\
&= \left\|\widehat{D}_{\mathrm{p}}|\psi\rangle\right\|^2 - \left(\left\|P_{\mathfrak{S}}\mathcal{E}D\mathcal{E}^\dagger P_{\mathfrak{S}}|\psi\rangle\right\| - \left\|P_{\mathfrak{S}}\left(\widehat{D}_{\mathrm{p}} - \mathcal{E}D\mathcal{E}^\dagger\right)P_{\mathfrak{S}}|\psi\rangle\right\|\right) && (6.254)
\end{aligned}$$

$$\leq \left(r^2 + O(s^2)\right) - \left(r - O(s)\right)^2 \qquad \text{(by (6.251), (6.252), and Item (i))}$$

$$= O(rs + s^2) \tag{6.255}$$

$$\leq O(\|D\| \cdot \sqrt{n}\Lambda^{-1/3}\|D\| + n\Lambda^{-2/3}\|D\|^2) \qquad \text{(by (6.252))}$$

$$\leq O(n\Lambda^{-1/3}\|D\|^2), \tag{6.256}$$

as claimed in Item (ii). □

Next, we examine $\widehat{D}_{\mathrm{r}}$. Since $\widehat{D}_{\mathrm{r}} = \widehat{D} - \widehat{D}_{\mathrm{p}}$, by definitions (6.177) and (6.242) we have

$$\widehat{D}_{\mathrm{r}}(\xi) = \widehat{D}_{\mathrm{r}}(\xi_1, \ldots, \xi_n) \coloneqq \min\{0, \xi_{\min}/w - 1\} \cdot \langle x|D|x\rangle, \quad \xi_{\min} \coloneqq \min_{i\in[n]}\{|\xi_i|\}, \quad x \coloneqq \theta(\xi). \tag{6.257}$$

Claim 6.10.13 (Property of $\widehat{D}_{\mathrm{r}}$)**.** $\left\|\widehat{D}_{\mathrm{r}} P_{\mathfrak{S}}\right\| \leq O(\sqrt{n}\Lambda^{-1/3}\|D\|)$.

Proof. Define $\mathcal{J} \coloneqq \mathcal{J}_1 \cup \mathcal{J}_2 \cup \cdots \cup \mathcal{J}_n$, where

$$\mathcal{J}_i \coloneqq [-1, 1]^{i-1} \times [-w, w] \times [-1, 1]^{n-i}. \tag{6.258}$$

It is immediate from (6.257) that $\|\widehat{D}_{\mathrm{r}}\| = \|D\|$ and $\widehat{D}_{\mathrm{r}}(\xi) = 0$ for $\xi \in [-1, 1]^n \setminus \mathcal{J}$. Therefore, for *any* $|\psi\rangle \in \mathcal{W}([-1, 1]^n)$ we have

$$\left\|\widehat{D}_{\mathrm{r}}|\psi\rangle\right\|^2 \leq \left\|\widehat{D}_{\mathrm{r}}^2\right\| \langle\psi|\psi\rangle_{\mathcal{J}} = \|D\|^2 \langle\psi|\psi\rangle_{\mathcal{J}} \leq \|D\|^2 \sum_{i\in[n]} \langle\psi|\psi\rangle_{\mathcal{J}_i}. \tag{6.259}$$

Now, assume $|\psi\rangle \in \mathfrak{S}$ so that we can write $|\psi\rangle = \alpha|\widehat{0}\rangle|\psi_0\rangle + \beta|\widehat{1}\rangle|\psi_1\rangle$ where $\alpha^2 + \beta^2 = 1$. We

have

$$\begin{aligned}
\frac{1}{2}\langle\psi|\psi\rangle_{\mathcal{J}_1} &\le \alpha^2\langle\widehat{0},\psi_0|\widehat{0},\psi_0\rangle_{\mathcal{J}_1} + \beta^2\langle\widehat{1},\psi_1|\widehat{1},\psi_1\rangle_{\mathcal{J}_1} && \text{(triangle inequality with } \langle\cdot|\cdot\rangle_{\mathcal{J}_1}\text{)} \\
&= \alpha^2\langle\widehat{0}|\widehat{0}\rangle_{[-w,w]}\langle\psi_0|\psi_0\rangle + \beta^2\langle\widehat{1}|\widehat{1}\rangle_{[-w,w]}\langle\psi_1|\psi_1\rangle && (6.260) \\
&\le \alpha^2\langle\widehat{0}|\widehat{0}\rangle_{[-1,w]}\cdot 1 + \beta^2\langle\widehat{1}|\widehat{1}\rangle_{[-w,1]}\cdot 1 && (6.261) \\
&\le O(\Lambda^{-2/3}). && \text{(by Claim 6.10.8 (iii))}
\end{aligned}$$

Similar estimates work for every $\langle\psi|\psi\rangle_{\mathcal{J}_i}$ in (6.259). Therefore, we have

$$\langle\psi|\widehat{D}_{\mathrm{r}}^2|\psi\rangle \le \|D\|^2\cdot n\cdot O(\Lambda^{-2/3}) = O(n\Lambda^{-2/3}\|D\|^2), \tag{6.262}$$

as desired. □

With Claim 6.10.12 and 6.10.13, we can now prove Claim 6.10.10.

Proof of Claim 6.10.10. For Item (i), when restricted on subspace $\mathfrak{S}$, we have

$$\left\|\widehat{D}-\mathcal{E}D\mathcal{E}^\dagger\right\| = \left\|\widehat{D}_{\mathrm{p}}+\widehat{D}_{\mathrm{r}}-\mathcal{E}D\mathcal{E}^\dagger\right\| \le \left\|\widehat{D}_{\mathrm{p}}-\mathcal{E}D\mathcal{E}^\dagger\right\| + \left\|\widehat{D}_{\mathrm{r}}\right\| \le O(\sqrt{n}\Lambda^{-1/3}\|D\|), \tag{6.263}$$

where the final bound uses Claim 6.10.12 (i) and Claim 6.10.13.

Similarly, for Item (ii) we have

$$\left\|P_{\mathfrak{S}^\perp}\widehat{D}P_{\mathfrak{S}}\right\| \le \left\|P_{\mathfrak{S}^\perp}\widehat{D}_{\mathrm{p}}P_{\mathfrak{S}}\right\| + \underbrace{\left\|P_{\mathfrak{S}^\perp}\widehat{D}_{\mathrm{r}}P_{\mathfrak{S}}\right\|}_{\le\left\|\widehat{D}_{\mathrm{r}}P_{\mathfrak{S}}\right\|} \le O(\sqrt{n}\Lambda^{-1/6}\|D\|), \tag{6.264}$$

where the last inequality follows from Claim 6.10.12 (ii) and Claim 6.10.13.

Now we consider the statement $\mu_0\left(P_{\mathfrak{S}^\perp}\widehat{D}P_{\mathfrak{S}^\perp}\right) \geq \mu_0(\widehat{D}) = \mu_0(D)$ in Item (iii). The inequality follows from the variational principle: projecting to a subspace can only increase the ground energy. The equality uses the assumption in Lemma 6.10.7 that $\mu_0(D) \leq 0$, together with Fact 6.10.5 (ii). $\square$

Chapter 7: Conclusion

Quantum computing offers the possibility of accelerating important computational tasks beyond the capabilities of classical algorithms. Among the various directions explored in the literature, physics-inspired quantum optimization algorithms have attracted significant attention due to their compatibility with near-term quantum hardware and potential for achieving substantial speedups in optimization problems with practical relevance. This thesis investigated the potential of speedups for two such algorithms, namely Quantum Adiabatic Algorithm (QAA) and Quantum Hamiltonian Descent (QHD), by both empirical and theoretical approaches.

7.1 Summary of Contributions

The primary contributions of this thesis are summarized as follows.

In Chapter 3, we established the global convergence of QHD for three classes of nonsmooth continuous objective functions: convex, strongly convex, and nonconvex. To the best of our knowledge, the nonconvex result is the first rigorous convergence guarantee for nonsmooth continuous optimization for both classical and quantum optimization algorithms. We also conducted complementary numerical evaluations to demonstrate that QHD consistently outperforms classical subgradient methods.

In Chapter 4, we presented evidence of the scalable quantum advantage by QHD. We constructed a family of objective functions with exponentially many local minima and proved that QHD minimizes them in polynomial time. In contrary, a comprehensive empirical study revealed that several off-the-shelf state-of-the-art classical solvers require superpolynomial time to find global minima. Our construction has also inspired subsequent work for investigating quantum–classical separations in optimization using adiabatic QHD [65].

In Chapter 5, we investigated the complexity of Schrödinger operators, since QHD is formu-

lated as real-space Schrödinger dynamics. We proved that simulating time-independent Schrödinger dynamics is BQP-hard and estimating their ground energy is StoqMA-complete under mild technical assumptions. Aside from their independent theoretical interest, a direct consequence of our results is that quantum algorithms based on Schrödinger dynamics (which certainly include QHD) cannot be efficiently dequantized by a general-purpose classical algorithm.

In Chapter 6, we presented the first provable exponential quantum speedups for QHD and QAA by constructing specific objective functions that require exponentially many classical queries but only polynomial quantum resources. To our best knowledge, the QAA result is the first provable superquadratic quantum speedup for QAA in the context of optimization since it was proposed in 2000, and the QHD result is also the first provable superquadratic speedup for the algorithm.

7.2 Future Directions

Despite the progress made in this thesis, many important questions remain open. Below we outline several potential directions for future research.

- **Finding new physics-inspired algorithms.** Further investigation of quantum-mechanical processes and phenomena may lead to new algorithmic ideas. Besides the physics-inspired algorithms mentioned in this thesis, another promising example arises from dissipation in open quantum systems: Recent works have used engineered dissipative dynamics to design quantum algorithms for state preparation and related tasks [231].

- **Characterizing optimization landscapes that permit dramatic speedups.** In Chapter 4 and Chapter 6, we constructed hard instances that separate the performance of QAA/QHD from classical algorithms. However, these instances are artificially engineered to challenge classical algo-

rithms. A more practical direction is to identify *structural* properties of optimization landscapes that enable QAA/QHD to achieve substantial speedups. While the direction appears daunting at first, some initial progress has been made [65].

- **Realizing practical quantum speedups.** Even under idealized assumptions, turning a theoretical quantum advantage into a practical one is still highly nontrivial. As discussed in the introduction, one attractive feature of physics-inspired algorithms is their relative compatibility with near-term quantum hardware without error correction. However, substantial work is still required to translate asymptotic advantages into real-world performance, should asymptotic advantages exist. For example, implementing QHD on Ising-type hardware requires nontrivial theoretical developments in encoding and simulation, as well as careful engineering on realistic devices [63].

- **Developing an intuitive understanding of quantum speedups.** A clearer conceptual understanding of the source of quantum speedups would be valuable for algorithm design. One example is the Quantum Singular Value Transformation (QSVT) framework [232], which has been presented as a unifying abstraction for a broad class of quantum algorithms. In contrast, the mechanisms underlying speedups for QAA and QHD is much less understood. Explanations based solely on quantum tunneling appear incomplete: in some settings, classical quantum Monte Carlo methods reproduce the same tunneling scaling effect (with polynomial overhead) in optimization [233,234]. Developing a comparably robust intuition for physics-inspired quantum optimization therefore remains an important open problem.

Appendix A: Appendices for Chapter 3

A.1 Subgradient Methods

A function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *convex* if for all $0 \le t \le 1$ and any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$, we have

$$f(t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2) \le tf(\boldsymbol{x}_2) + (1-t)f(\boldsymbol{x}_2). \tag{A.1}$$

The gradient of a nonsmooth function f is not well-defined everywhere. Instead, we consider a weaker notion of subgradient. Let $f\colon \mathbb{R}^d \to \mathbb{R}$. We say a vector $\boldsymbol{g} \in \mathbb{R}^d$ is a *subgradient* of f at $\boldsymbol{x}$ if,

$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \boldsymbol{g}^\top(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^d. \tag{A.2}$$

The subgradient of a function f at a fixed point is not unique. The *subdifferential* of f at $\boldsymbol{x} \in \mathbb{R}^d$ is defined as the set of all subgradients of f at $\boldsymbol{x}$, denoted by

$$\partial f(\boldsymbol{x}) \coloneqq \{\boldsymbol{g} \in \mathbb{R}^d \colon f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \boldsymbol{g}^\top(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^d\}. \tag{A.3}$$

We say f is *subdifferentiable* at $\boldsymbol{x}$ if $\partial f(\boldsymbol{x}) \ne \varnothing$. When f is convex and finite over $\mathbb{R}^d$, the subdifferential $\partial f(\boldsymbol{x})$ is a convex, bounded, and closed set for every $\boldsymbol{x}$. The directional derivative $f'(\boldsymbol{x}, \boldsymbol{v})$ at any $\boldsymbol{x}$ along the direction $\boldsymbol{v}$ can be represented by

$$f'(\boldsymbol{x}, \boldsymbol{v}) = \lim_{t \to 0^+} \frac{f(\boldsymbol{x} + t\boldsymbol{v}) - f(\boldsymbol{x})}{t} = \max_{\boldsymbol{g} \in \partial f(\boldsymbol{x})} \boldsymbol{g}^\top \boldsymbol{v}. \tag{A.4}$$

In particular, when f is differentiable at $\boldsymbol{x}$, $\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$ and we recover the following formula from standard calculus,

$$f'(\boldsymbol{x}, \boldsymbol{v}) = \nabla f(\boldsymbol{x})^\top \boldsymbol{v}. \tag{A.5}$$

A point $\boldsymbol{x}^*$ is a local minimizer of a function f (not necessarily convex) if and only if f is subdifferentiable at $\boldsymbol{x}^*$ and $0 \in \partial f(\boldsymbol{x}^*)$.

We say a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *L-Lipschitz* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|. \tag{A.6}$$

If f is differentiable, the gradient of f must be bounded by the Lipschitz constant, i.e., $\|\nabla f\|_\infty \leq L$. A function f is *locally Lipschitz* if for any $\boldsymbol{x}_0 \in \mathbb{R}^d$, there exists a local Lipschitz constant $\ell_{x_0} > 0$ and a radius $\delta > 0$ such that

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| \leq \ell_{x_0}\|\boldsymbol{x} - \boldsymbol{x}_0\|, \quad \forall \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \delta. \tag{A.7}$$

It is well-known that a locally Lipschitz function is differentiable almost everywhere.

Clarke's gradient [235] is a generalization of the notion of subdifferential that enables the analysis of nonconvex and nonsmooth functions through convex analysis. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz function and Ω_f be the set such that f is not differentiable. By the local Lipschitz property, the set Ω_f is of measure 0. For a point $\boldsymbol{x} \in \mathbb{R}^d$, we define the *generalized gradient* (or *Clarke's gradient*) of the function f as follows:

$$\partial_C f(\boldsymbol{x}) \coloneqq \operatorname{conv}\left(\{\boldsymbol{g}\colon \boldsymbol{g} \in \mathbb{R}^d, \nabla f(\boldsymbol{x}_k) \to \boldsymbol{g}, \boldsymbol{x}_k \to \boldsymbol{x}, \boldsymbol{x}_k \notin \Omega_f\}\right), \tag{A.8}$$

where $\mathrm{conv}(A)$ is the *convex hull* of a set A. A point $\boldsymbol{x}^* \in \mathbb{R}^d$ is called a *Clarke stationary point* if $0 \in \partial_C f(\boldsymbol{x}^*)$. Note that Clarke stationarity is a necessary condition for a local optimum (including both minimum and maximum), but the inverse is not true.

The *subgradient algorithm* generalizes the gradient descent algorithm by replacing the negative gradient update at each step with a negative subgradient update:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - s_k g_k, \ \forall k = 0, 1, \cdots \tag{A.9}$$

where $\boldsymbol{x}_0$ is an initial guess, $g_k \in \partial f(\boldsymbol{x}_k)$ is *any* subgradient of f at $\boldsymbol{x}_k$, and $s_k > 0$ is a step size. In contrast to gradient descent, the subgradient algorithm is in general not a descent method (i.e., $\{f(\boldsymbol{x}_k)\}_k$ is not monotonically decreasing). Therefore, it is natural to track the current best solution $\boldsymbol{x}_{\mathrm{best}}^{(k)}$ and best function values $f_{\mathrm{best}}^{(k)}$:

$$\boldsymbol{x}_{\mathrm{best}}^{(k)} \coloneqq \underset{1 \le k \le K}{\arg\min}\, f(\boldsymbol{x}_k), \quad f_{\mathrm{best}}^{(k)} \coloneqq \min_{1 \le k \le K} f(\boldsymbol{x}_k). \tag{A.10}$$

With a constant step size $s > 0$, it has been proven that the subgradient algorithm has the following asymptotic convergence result for a convex function f with a Lipschitz constant L:

$$f_{\mathrm{best}}^{(k)} - f(\boldsymbol{x}^*) \le \frac{\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2}{2sk} + \frac{L^2 s}{2}, \tag{A.11}$$

where $\boldsymbol{x}^*$ is the minimizer of f. In this result, we observe a nonvanishing optimality gap $L^2 s/2 = \Omega(s)$ independent of the iteration number k. This behavior of the subgradient algorithm is radically different from gradient descent, as the optimality gap of the latter algorithm converges to 0 at a rate $\mathcal{O}(k^{-1})$.

In practice, it is more favorable to consider a diminishing sequence of step sizes $\{s_k\}$, which will yield a vanishing optimality gap. In the following theorem, we consider a sequence of step sizes decrease as $k^{-1/2}$, and it turns out that the optimality gap also converges at the same speed.

Theorem A.1.1 ([69]). *Suppose that f is a convex function such that $\|g\| \leq L$ for any $\boldsymbol{x}$ and $g \in \partial f(\boldsymbol{x})$. Let $\boldsymbol{x}^*$ be the minimizer of f, $R \coloneqq \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|$, and choose a step size $s_k = \frac{R}{L\sqrt{k}}$. Then, we have*

$$f_{\text{best}}^{(k)} - f(\boldsymbol{x}^*) \leq RL \cdot \frac{\ln k}{\sqrt{k}}. \tag{A.12}$$

For strongly convex optimization problems, the convergence rate can be improved to $\mathcal{O}(k^{-1})$ provided that we choose a sequence of step size that decay at the same rate:

Theorem A.1.2 ([112]). *Suppose that f is a μ-strongly convex function such that $\|g\| \leq L$ for any $\boldsymbol{x}$ and $g \in \partial f(\boldsymbol{x})$.*[1] *Let $\boldsymbol{x}^*$ be the minimizer of f and choose a step size $s_k = \frac{2}{\mu(k+1)}$. Then, we have*

$$f_{\text{best}}^{(k)} - f(\boldsymbol{x}^*) \leq \frac{2L^2}{\mu} \cdot \frac{1}{k+1}. \tag{A.13}$$

The k-dependence in Theorem A.1.1 and Theorem A.1.2 is tight up to a factor of $\ln k$ for first-order methods: As established in [114], for any k, there exist convex functions (or strongly convex functions, respectively) for which any first-order method fails to achieve an optimality gap $o(1/\sqrt{k})$ (or $o(1/k)$, respectively) using k queries.[2]

[1] This requires that f is restricted on a bounded region in $\mathbb{R}^d$. The subgradient algorithm referenced here is thus the projected subgradient algorithm [112], a natural extension that enforces constraints by projecting iterates onto the convex feasible region.

[2] We write $f = o(g)$ if and only if $f/g \to 0$.

A.2 Analysis of Schrödinger Equations

A.2.1 Spectral Theory of Schrödinger Operators

The standard theory of Schrödinger operators guarantees they have nice spectral property under minimal assumptions:

Lemma A.2.1 ([152, Theorem XIII.47 and XIII.67]). *Let $V\colon \mathbb{R}^d \to \mathbb{R}$ be continuous*[3] *and assume $\lim_{\boldsymbol{x}\to\infty} V(\boldsymbol{x}) = +\infty$. Then, $H = -\Delta + V$ has purely discrete spectrum with a nondegenerate strictly positive ground state. As a consequence, H has a positive* spectral gap *which is defined as the difference between the smallest and the second smallest (counting multiplicities) eigenvalues of H.*

Remark A.2.2. Lemma A.2.1 can be extended to the case where V is defined on $\Omega \coloneqq [-1,1]^d$: in such a case we no longer need any assumption for $\lim_{\boldsymbol{x}\to+\infty} V(\boldsymbol{x})$. To understand this, it is important to note that the spectrum will be unchanged if we extend V to $\widehat{V}$, where

$$\widehat{V}(\boldsymbol{x}) = \begin{cases} V(\boldsymbol{x}), & x \in \Omega; \\ +\infty, & x \notin \Omega. \end{cases} \tag{A.14}$$

Now that $\widehat{V}$ is defined on $\mathbb{R}^d$, the only obstacle for applying Lemma A.2.1 is that $\widehat{V}$ is not continuous. This can be tackled by a standard method that constructs a continuous function that is sufficiently close to $\widehat{V}$.

[3] The continuity assumption can be weakened to $V \in L^2_{\mathrm{loc}}(\mathbb{R}^d)$; See [152].

A.2.2 Regularity Theory of Schrödinger Equations

In this section, we show that QHD preserves the smoothness of the quantum state in the evolution. Specifically, we consider the following time-dependent Schrödinger equation,

$$i\partial_t \Psi(t, \boldsymbol{x}) = \left(-\frac{1}{2\lambda(t)} \Delta + \lambda(t) f(\boldsymbol{x}) \right) \Psi(t, \boldsymbol{x}), \tag{A.15}$$

where $\boldsymbol{x} \in \Omega = [0, 1)^d$ with periodic boundary condition, $f(\boldsymbol{x}) \in L^\infty(\Omega)$, and $\lambda(t) \colon [0, T] \to \mathbb{R}^+$ is a continuous function.

Remark A.2.3. While we consider unconstrained optimization problems in this paper, the minimum of the objective function is often distributed in a compact subset in $\mathbb{R}^d$, which means it is only necessary for an optimization algorithm to search within a finite bounding box. In QHD, this means we can restrict the quantum evolution in a box since the quantum wave function is essentially vanishing outside of this bounding box. Therefore, it is reasonable to model QHD by a Schrödinger equation with periodic boundary conditions (up to proper rescaling of the space).

Definition A.2.4. For any $s > 0$, we define the H^s-norm of a function $u \in L^2(\Omega)$ by

$$\|u\|_{H^s} = \left(\sum_{\boldsymbol{k} \in \mathbb{Z}^d} (1 + 4\pi^2 |\boldsymbol{k}|^2)^s |\hat{u}(\boldsymbol{k})|^2 \mathrm{d}\boldsymbol{k} \right)^{1/2}, \tag{A.16}$$

where $\hat{u}(\boldsymbol{k})$ is the Fourier transform of the function $u(\boldsymbol{x})$,

$$\hat{u}(\boldsymbol{k}) = \int_\Omega u(\boldsymbol{x}) e^{-2i\pi \boldsymbol{x} \cdot \boldsymbol{k}} \, \mathrm{d}\boldsymbol{x}. \tag{A.17}$$

Lemma A.2.5 (Sobolev estimates)**.** *Let* $\Psi(t, \boldsymbol{x})\colon [0, T] \to L^2(\Omega)$ *be the solution to (A.15). Then, for any* $s > 0$ *and* $t \in [0, T]$*, there is a constant* $C_T > 0$ *such that*

$$\|\Psi(t)\|_{H^s} \leq C_T \|\Psi_0\|_{H^s}. \tag{A.18}$$

Proof. Direct calculation shows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\Psi(t)\|_{H^s}^2 = 2\,\mathrm{Re}\{\langle \partial_t \Psi(t), (1-\Delta)^s \Psi(t)\rangle\}, \tag{A.19}$$

where the operator $(1-\Delta)^s$ is defined by

$$(1-\Delta)^s u \coloneqq \mathcal{F}^{-1}\left[\sum_{\boldsymbol{k}\in\mathbb{Z}^d} (1 + 4\pi^2|\boldsymbol{k}|^2)^s \mathcal{F}(u)(\boldsymbol{k})\right], \tag{A.20}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ represents Fourier transform and inverse Fourier transform, respectively. We plug (A.15) into (A.19), and it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\Psi(t)\|_{H^s}^2 = 2\,\mathrm{Re}\left(\left\langle \left(-i\frac{1}{2\lambda(t)}\Delta - i\lambda(t) f\right)\Psi(t), (1-\Delta)^s\Psi(t)\right\rangle\right) \tag{A.21}$$

Note that the self-adjoint operators Δ commutes with $(1-\Delta)^s$ for any $s > 0$, therefore $\langle \Delta\Psi(t), (1-\Delta)^s\Psi(t)\rangle$ is a real number for any t, and

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\Psi(t)\|_{H^s}^2 = 2\lambda(t)\,\mathrm{Re}\left(\langle -if\Psi(t), (1-\Delta)^s\Psi(t)\rangle\right) \leq 2\lambda(t)\|f\|_\infty \|\Psi(t)\|_{H^s}^2, \tag{A.22}$$

where the last step follows from Hölder's inequality. Next, we invoke the Grönwall's inequality and

obtain

$$\|\Psi(t)\|_{H^s}^2 \leq \exp\left(\int_0^t \lambda(s)\|f\|_\infty \mathrm{d}s\right)\|\Psi(0)\|_{H^s}^2. \tag{A.23}$$

Thus, let $C_T \coloneqq \exp\left(\int_0^T \lambda(s)\|f\|_\infty \mathrm{d}s\right)$, we complete the proof of the lemma. □

Lemma A.2.6 (Sobolev embedding theorem)**.** *If $s - k > d/2$, then we have $H^s(\mathbb{R}^d) \hookrightarrow C^k(\mathbb{R}^d)$.*

Proof. This is a standard result in functional analysis, see [236] for details. □

Remark A.2.7. Note that the $\hookrightarrow$ symbol means the embedding of the Sobolev space in a larger space. Here, embedding not only means the Sobolev space H^s is completely included in C^k as a set but also the H^s-norm naturally extends to C^k.

Theorem A.2.8. *Consider the time-dependent Schrödinger equation (A.15) subject to a smooth initial condition $\Psi_0(\boldsymbol{x}) \in C^\infty(\Omega)$. Then, the PDE (A.15) admits a unique solution such that $\Psi(t, \boldsymbol{x}) \in C^\infty(\mathbb{R}^d)$ for all finite $t \geq 0$.*

Proof. Note that the initial state $\Psi_0(\boldsymbol{x}) \in C^\infty(\Omega) \hookrightarrow H^s(\Omega)$ for any $s > 0$. Therefore, any finite time $t \geq 0$, by Lemma A.2.5, we have $\Psi(t) \in H^s(\Omega)$, i.e., the Schrödinger equation preserves the H^s space. Then, by Lemma A.2.6, we have

$$\Psi(t) \in \cap_{s>d/2} H^s(\omega) \subset \cap_{k>0} C^k(\omega) \subset C^\infty(\Omega), \tag{A.24}$$

which concludes the proof. □

A.2.3 Quantum Simulation of Schrödinger Equations with Nonsmooth Potentials

Lemma A.2.9. *Let $\Psi(s,x)$ denote the exact solution of (3.11) and $\widetilde{\Psi}(s,x)$ denote the approximated solution by the Fourier spectral method (truncated up to frequency n).*[4] *We assume that the initial data $\Psi_0(x)$ is periodic and analytic in $x \in \Omega$. Then, for any integer $n \geq 1$, the error from the Fourier spectral method satisfies*

$$\max_{x,t} |\Psi(x,t) - \widetilde{\Psi}(x,t)| \leq 2r^{n/2+1}, \tag{A.25}$$

where $0 < r < \min(\frac{1}{2}, \frac{1}{Ae^{\xi}})$, A is an absolute constant that only depends on $\Psi_0(x)$, and ξ is the same as in (3.12).

Proof. We give the proof in one dimension, as the same argument is readily generalized to arbitrary finite dimensions. We assume the initial data $\Psi_0(x)$ is periodic over $[0, 2\pi]$. The analyticity implies that, for any $x \in [0, 2\pi]$, there is a constant C such that

$$\left|\Psi_0^{(k)}(x)\right| \leq C^{k+1}(k!). \tag{A.26}$$

Therefore, the function $\Psi_0(x)$ admits an analytic continuation in the strip

$$\Gamma_0 = \left\{ z \in \mathbb{C} : |z - x| < \frac{1}{C}, x \in [0, 2\pi] \right\}.$$

According to the discussion in Section A.2.2, the Sobolev norm of the wave function $\Psi(s)$ has an

[4] More details on the Fourier spectral method can be found in Section 2.2 (in particular, Lemma 1) in [89].

exponential growth:

$$\left\|\Psi^{(k)}(s,\cdot)\right\| \leq e^{\xi(s)S}\left\|\Psi_0^{(k)}\right\| \leq e^{\xi(s)S}C^{k+1}(k!), \tag{A.27}$$

where $\xi(s) = \int_0^s \varphi(r)\mathrm{d}r$. This implies that the wave function $\Psi(s,x)$ is periodic and analytic for any finite $s \in [0, s_f]$. The strip on which $\Psi(s,x)$ admits an analytic continuation is

$$\Gamma_s = \left\{ z \in \mathbb{C} : |z - x| < \frac{1}{Ce^{\xi(s)S}}, x \in [0, 2\pi] \right\}. \tag{A.28}$$

Suppose that the function $\Psi(s,x)$ allows an exact, infinite trigonometric polynomial representation (see [89, Lemma 16]),

$$\Psi(s,x) = \sum_{k=0}^{\infty} c_k(s)e^{ikx}. \tag{A.29}$$

Let $\tilde{\Psi}(s,x)$ be the truncated Fourier series up to $k = n/2$, then the error from the Fourier spectral method satisfies

$$|\Psi(s,x) - \tilde{\Psi}(s,x)| \leq \sum_{k=n/2+1}^{\infty} |c_k(s)|. \tag{A.30}$$

Since the function $\Psi(s,x)$ allows an analytic continuation in the strip Γ_s, it follows from the Riemann–Lebesgue lemma that for $k \to \infty$, we have

$$|c_k| \cdot e^{-k\frac{1}{Ce^{\xi S}}} \to 0. \tag{A.31}$$

It turns out that the error from the Fourier spectral method satisfies

$$|\Psi(x,t) - \tilde{\Psi}(x,t)| \leq \sum_{k=n/2+1}^{\infty} |c_k(t)| \leq ACe^{\xi S} e^{-\frac{n/2+1}{Ce^{\xi S}}}. \tag{A.32}$$

Note that we use the approximation $e^{-x} \approx 1 - x$ for small x in the last step. □

Now, we are ready to prove Proposition 3.2.4.

Proof. By Lemma A.2.9, to ensure that the simulation error is bounded by η, we may choose the truncation number

$$n \leq \mathcal{O}\left(e^{\xi S} \log\left(\frac{1}{\eta}\right)\right). \tag{A.33}$$

Plugging this truncation number in Equation (113) in [89], we prove the query complexity in Proposition 3.2.4. It is worth noting that the truncation number n scales exponentially ξS (i.e., the L^1-norm of the time-dependent potential $\varphi(s)f$). This implies that the space complexity using quantum computers is polynomial in ξS. □

A.3 Experiment Details

A.3.1 Test Functions

Below list 12 functions tested in Section 3.3.

1. `WF` (WF) [237]

$$f = \max\left(\frac{1}{2}\left(x_1 + \frac{10x_1}{x_1 + 0.1} + 2x_2^2\right), \frac{1}{2}\left(-x_1 + \frac{10x_1}{x_1 + 0.1} + 2x_2^2\right), \frac{1}{2}\left(x_1 - \frac{10x_1}{x_1 + 0.1} - 2x_2^2\right)\right)$$

tested on $[-10,\ 10]^2$ with globally optimal solution at $f(0,\ 0) = 0$.

2. `CROWNEDCROSS` (Crowned Cross) [238]

$$f = 0.0001 \left(|\sin(x_1)\sin(x_2)\exp\left(100 - \frac{\sqrt{x_1^2 + x_2^2}}{\pi}\right)| + 1 \right)^{0.1}$$

tested on $[-10,\ 15]^2$ with globally optimal solution at $f(0,\ 0) = 0.0001$.

3. `BUKIN06` (Bukin 6) [238]

$$f = 100\sqrt{|x_2 - 0.01x_1^2|} + 0.01|x_1 + 10|$$

tested on $[-15,\ -5] \times [-3,\ 3]$ with globally optimal solution at $f(-10,\ 1) = 0$.

4. `KEANE` (Keane) [239]

$$f = -\frac{|\cos^4(x_1) + \cos^4(x_2) - 2\cos^2(x_1)\cos^2(x_2)|}{\sqrt{x_1^2 + 2x_2^2}}$$

tested on $[10^{-8},\ 10]^2$ with the globally optimal solution at $f(1.60086, 0.468498) = 0.673207$.

5. `SCHWEFEL` (Schwefel) [238]

$$f = 418.9828872724336 - x_1 \sin\left(\sqrt{|x_1|}\right)$$

tested on $[-500,\ 500]$ with globally optimal solution at $f(420.9687474737558) = 0$.

6. `ACKLEY` (Ackley) [238]

$$f = -20\exp\left(-0.2\sqrt{\frac{x_1^2 + x_2^2}{2}}\right) - \exp\left(\frac{\cos(2\pi x_1) + \cos(2\pi x_2)}{2}\right) + 20 + e$$

tested on $[-15,\ 30]^2$ with globally optimal solution at $f(0,\ 0) = 0$.

7. `XINSHEYANG04` (Xinshe Yang 4) [238]

$$f = \left(\sin^2(x_1) + \sin^2(x_2) - \exp\left(-x_1^2 - x_2^2\right)\right)\exp\left(-\sin^2\left(\sqrt{|x_1|}\right) - \sin^2\left(\sqrt{|x_2|}\right)\right)$$

tested on $[-10,\ 10]^2$ with globally optimal solution at $f(0,\ 0) = -1$.

8. `CARROMTABLE` (Carrom Table) [240]

$$f = -\frac{1}{30}\exp\left(\left|2 - \frac{2}{\pi}\sqrt{x_1^2 + x_2^2}\right|\right)\cos^2(x_1)\cos^2(x_2)$$

tested on $[-10,\ 10]^2$ with globally optimal solution at $f = (\pm 9.646157266349,\ \pm 9.646157266349) = -24.1568155165$.

9. `RANA` (Rana) [238]

$$\begin{aligned} f = {} & x_1 \sin\left(\sqrt{|x_2 - x_1 + 1|}\right)\cos\left(\sqrt{|x_2 + x_1 + 1|}\right) \\ & + (x_2 + 1)\sin\left(\sqrt{|x_2 + x_1 + 1|}\right)\cos\left(\sqrt{|x_2 - x_1 + 1|}\right) \end{aligned}$$

tested on $[-500,\ 500]^2$ with globally optimal solution at $f(-300.3376328023,\ 500) = -500.802160296664$.

10. `DROPWAVE` (Drop Wave) [240]

$$f = -\frac{1 + \cos\left(12\sqrt{x_1^2 + x_2^2 + x_3^2}\right)}{2 + 0.5\left(x_1^2 + x_2^2 + x_3^2\right)}$$

tested on $[-5.12,\ 5.12]^3$ with globally optimal solution at $f(0,\ 0,\ 0) = -1$.

11. `LAYEB04` (Layeb 4) [241]

$$f = \sum_{i=1}^{2} (\ln(|x_i x_{i+1}| + 0.001) + \cos(x_1 + x_{i+1}))$$

tested on $[-10,\ 10]^3$ with globally optimal solution at $f(0,\ (2j-1)\pi,\ 0) = 2\ln(0.001) - 2$, where $j \in \{-1,\ 0,\ 1,\ 2\}$.

12. `DAMAVANDI` (Damavandi) [238]

$$f = \left(1 - \left|\frac{\sin(\pi(x_1 - 2))\sin(\pi(x_2 - 2))}{\pi^2(x_1 - 2)(x_2 - 2)}\right|^5\right)\left(2 + (x_1 - 7)^2 + 2(x_2 - 7)^2\right)$$

tested on $[0, 14]^2$ with the globally optimal solution at $f(2 + 10^{-10}, 2 + 10^{-10}) = 0$.

In the numerical experiment (Section 3.3), we further augment these functions with a continuous barrier that steeply diverges *out* of the original domain. This adjustment is made because the discretized QHD algorithm in simulation (see [79, Section 5]) implicitly alters the topology of the domain $[-L, L]^d$ due to its handling of boundary conditions. To be more specific, [79, Algorithm 2] works with a *periodic* boundary condition, effectively allowing the wave function to propagate across boundaries of the domain as if the topology is $(\mathbb{R}/2L\mathbb{Z})^d$. In contrast, classical optimization methods such as Subgrad and LFMSGD do not have this feature. Arguably, this adjustment will not affect the

performance of Subgrad and LFMSGD by their design.

A.3.2 Effect of Changing L in QHD

We analyze the effect of the scaling parameter L, introduced in Section 3.3.1, on the QHD dynamics. Without loss of generality, assume that the domain of a test function $f(\boldsymbol{x})$ is $[-1,1]^d$. By rescaling the domain via $\boldsymbol{y} = L\boldsymbol{x}$, we transform $f(\boldsymbol{x})$ into $g(\boldsymbol{y}) = f(\boldsymbol{y}/L)$ with domain $[-L,L]^d$, which is then used in QHD.

The QHD evolution equation in (3.8) is

$$i\partial_t \Psi(t) = H(t)\Psi(t), \quad H(t) = \frac{1}{\lambda(t)}\left(-\frac{1}{2}\Delta_{\boldsymbol{y}}\right) + \lambda(t) g(\boldsymbol{y}), \tag{A.34}$$

where the Laplacian operator is given by $\Delta_{\boldsymbol{y}} \coloneqq \sum_{i=1}^{d} \frac{\mathrm{d}^2}{\mathrm{d}y_i^2}$. Using the chain rule, the Laplacian transforms as

$$\Delta_{\boldsymbol{y}} = L^{-2}\Delta_{\boldsymbol{x}}. \tag{A.35}$$

Substituting this into (A.34), we obtain

$$iL\partial_t \Psi(t) = \left(\frac{1}{L\lambda(t)}\left(-\frac{1}{2}\Delta_{\boldsymbol{x}}\right) + L\lambda(t) f(\boldsymbol{x})\right)\Psi(t). \tag{A.36}$$

Now, introducing the rescaled time variable $\tau \coloneqq t/L$ and defining $\Phi(\tau) = \Psi(t) = \Psi(L\tau)$, we rewrite (A.36) as

$$i\partial_\tau \Phi(\tau) = \left(\frac{1}{L\lambda(L\tau)}\left(-\frac{1}{2}\Delta_{\boldsymbol{x}}\right) + L\lambda(L\tau) f(\boldsymbol{x})\right)\Phi(\tau). \tag{A.37}$$

Therefore, the equation (A.37) represents the same QHD dynamics as in (3.8), but with the rescaled

time variable $\tau = t/L$ and the coefficient function $\mu(\tau) \coloneqq L\lambda(L\tau)$:

$$i\partial_\tau \Phi(\tau) = \left(\frac{1}{\mu(\tau)} \left(-\frac{1}{2}\Delta_{\boldsymbol{x}} \right) + \mu(\tau) f(\boldsymbol{x}) \right) \Phi(\tau). \tag{A.38}$$

A.3.3 Learning-Rate Free Momentum Stochastic Gradient Descent (LFMSGD)

LFMSGD [70] is a subgradient method that generalizes previous learning-rate free subgradient methods like DoG (Distance over Gradients) [242] and DoWG (Distance over Weighted Gradients) [243]. It starts with an initial point $\boldsymbol{x}_0$ and the update rule is

$$\eta_t = \frac{\mu_t}{\sqrt{\epsilon_0 + \sum_{i=1}^{t} \|m_i\|^2}}, \tag{A.39a}$$

$$m_{t+1} = \beta m_t + (1-\beta) g(\boldsymbol{x}_t), \tag{A.39b}$$

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \eta_t m_{t+1}, \tag{A.39c}$$

where $g(\boldsymbol{x}_t)$ is an *estimate* of $\partial f(\boldsymbol{x}_t)$, $\mu_t = \max_{i \le t}\{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\}$, $0 < \epsilon_0 \ll 1$ is a small prefixed constant that enforces the numerical stabilization and $\beta \in [0, 1)$ is the momentum parameter. The returned solution will be $\boldsymbol{x}^* \coloneqq \boldsymbol{x}_T$ where T is the total number of steps.

We choose the parameter $\beta = 0.9$ as recommended in [70]. The stochasticity in the original LFMSGD arises from the common setting in machine learning where $f = \frac{1}{N}\sum_{1 \le i \le N} f_i$ and a natural, computationally efficient approach is to choose g by cycling through the functions $\partial f_1, \partial f_2, \ldots, \partial f_N$ in sequence for each t. In our case we do not have a natrual approximation of ∂f, so we add Gaussian

noise to model g, following [52, 140]:

$$g(\boldsymbol{x}) = \partial f(\boldsymbol{x}) + \mathcal{N}(\mathbf{0}, \sigma^2). \tag{A.40}$$

A.3.4 Resolution Bottleneck of QHD Simulation

Recall that we choose spatial discretization number $N = 512$ for QHD in Section 3.3. It turns out that for some test functions, the resolution is not enough for QHD to even see a point that has smaller enough optimality gap. In Figure A.1 (cf. Figure 3.1), we highlight the minimal optimality gap that QHD can "see" after discretization with red dotted lines. It can be inferred that for functions like `ACKLEY`, QHD performance no longer improves with k increasing because of this resolution bottleneck.

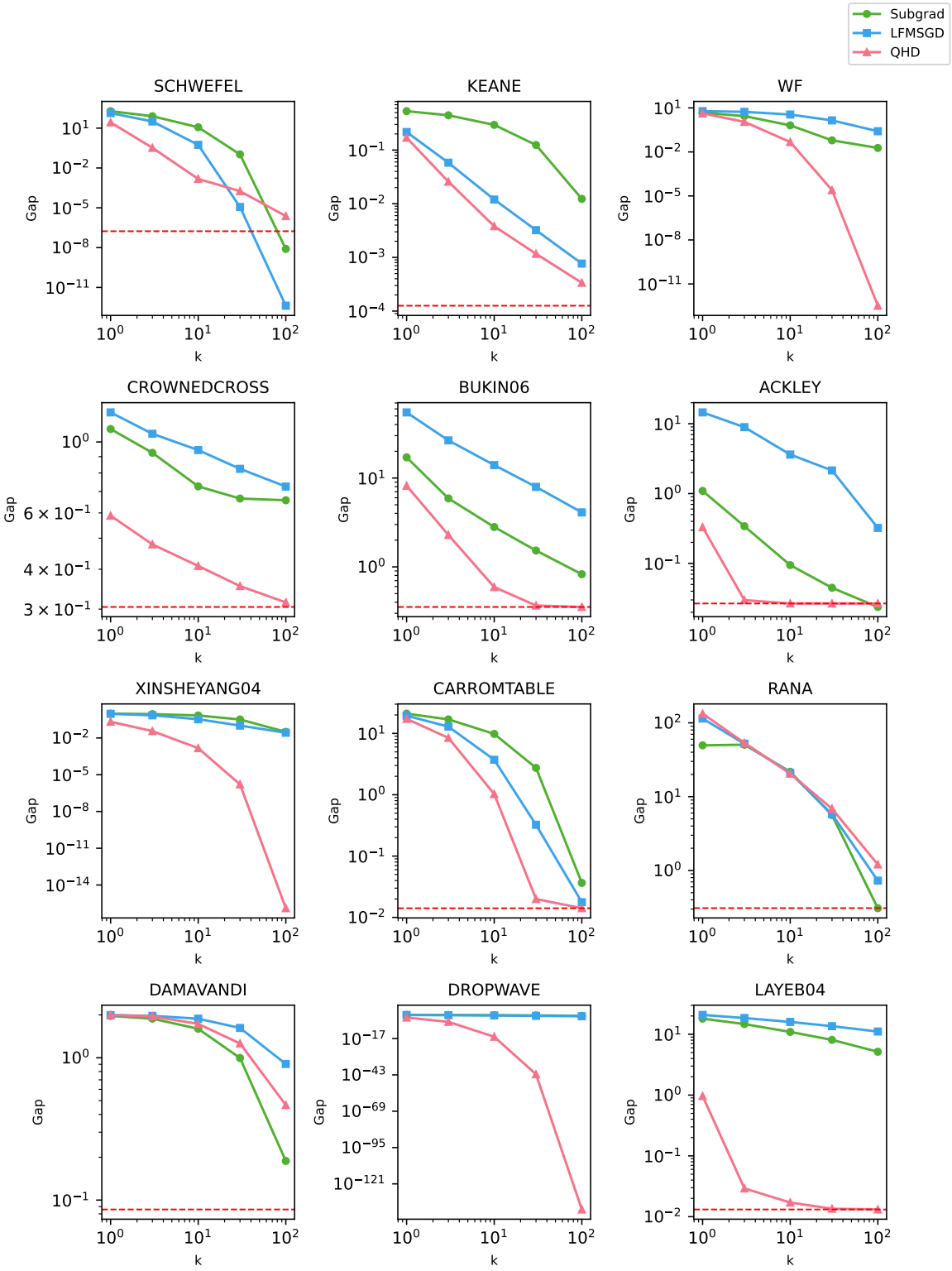


Figure A.1: Best-of-k optimality gap for 12 test functions. Minimal optimality gaps that QHD can "see" after discretization are shown as red dotted lines.

Appendix B: Appendices for Chapter 4

B.1 Quantum Simulation with Analytic Initial Data

In this section, all notations are the same as in Section 4.2.4.

Lemma B.1.1. *Let $\Psi(x,t)$ denote the exact solution of (4.20) and $\widetilde{\Psi}(x,t)$ denote the approximated solution by the Fourier spectral method (truncated up to frequency n).*[1] *We assume that the initial data $\Psi_0(x)$ is periodic and analytic in $x \in \Omega$. Then, for any integer $n \geq 1$, the error from the Fourier spectral method satisfies*

$$\max_{x,t} |\Psi(x,t) - \widetilde{\Psi}(x,t)| \leq 2r^{n/2+1}, \tag{B.1}$$

where $0 < r < \min(\frac{1}{2}, \frac{1}{At})$, A is an absolute constant that only depends on $\Psi_0(x)$.

Proof. We give the proof in one dimension, as the same argument is readily generalized to arbitrary finite dimensions. We assume the initial data $\Psi_0(x)$ is periodic over $[0, 2\pi]$. The analyticity implies that, for any $x \in [0, 2\pi]$, there is a constant C such that

$$\left|\Psi_0^{(k)}(x)\right| \leq C^{k+1}(k!). \tag{B.2}$$

[1] More details on the Fourier spectral method can be found in Section 2.2 (in particular Lemma 1) in [89].

Therefore, the function $\Psi_0(x)$ admits an analytic continuation in the strip

$$\Gamma_0 = \left\{ z \in \mathbb{C} : |z - x| < \frac{1}{C}, x \in [0, 2\pi] \right\}.$$

Due to [244, Proposition 1], there is an absolute constant A such that for any $t \in [0, T]$,

$$\left\| \Psi^{(k)}(\cdot, t) \right\| \leq At \left\| \Psi_0^{(k)} \right\| \leq AC^{k+1} t(k!), \tag{B.3}$$

which implies that the wave function $\Psi(x, t)$ is periodic and analytic for any finite t. The strip on which $\Psi(x, t)$ admits an analytic continuation is

$$\Gamma_t = \left\{ z \in \mathbb{C} : |z - x| < \frac{1}{ACt}, x \in [0, 2\pi] \right\}.$$

Suppose that the function $\Psi(x, t)$ allows an exact, infinite trigonometric polynomial representation (see [89, Lemma 16]),

$$\Psi(x, t) = \sum_{k=0}^{\infty} c_k(t) e^{ikx}. \tag{B.4}$$

Let $\tilde{\Psi}(t, x)$ be the truncated Fourier series up to $k = n/2$, then the error from the Fourier spectral method satisfies

$$|\Psi(x, t) - \tilde{\Psi}(x, t)| \leq \sum_{k=n/2+1}^{\infty} |c_k(t)|. \tag{B.5}$$

Meanwhile, the function $\Psi(x, t)$ has an analytic continuation defined by $\Psi(x, t) = \sum_{k=0}^{\infty} c_k(t) z^k$ for

$z \in \Gamma_t$. By Cauchy's integral formula, for any simply connected curve γ on the strip Γ_t, we have that

$$c_k(t) = \frac{1}{k!} \int_\gamma \Psi^{(k)}(z,t) \mathrm{d}z. \tag{B.6}$$

Together with (B.3), it turns out that $|c_k(t)| \leq r^k$ for some $0 < r < \frac{1}{At}$. Therefore, the error from the Fourier spectral method satisfies

$$|\Psi(x,t) - \tilde{\Psi}(x,t)| \leq \sum_{k=n/2+1}^{\infty} |c_k(t)| \leq \frac{r^{n/2+1}}{1-r}. \tag{B.7}$$

Moreover, if we force $r < 1/2$, the error is bounded by $2r^{n/2+1}$. □

To ensure that the simulation error is bounded by η, we may choose the truncation number

$$n = \left\lceil 2 \left(\frac{\log(4/\eta)}{\log(1/r)} - 1 \right) \right\rceil \leq \mathcal{O}\left(\log(1/\eta)\right).$$

Next, by plugging this truncation number in Equation (113) in [89], we prove Theorem 4.2.4.

B.2 Sub-Gaussian Ground States

B.2.1 Probability Toolbox

Lemma B.2.1. *Let $X \in \mathbb{R}$ be a random variable such that $\mathbb{E}[e^{X^2/K}] \leq 2$, where $K > 0$ is a real number. Then, for any $s \in \mathbb{R}$, we have*

$$\mathbb{E}[e^{sX}] \leq e^{2s^2K}. \tag{B.8}$$

Proof. This is a standard result of sub-Gaussian random variables. Here, we give the proof for the completeness of this paper. By Markov's inequality, we have

$$\Pr[|X| > t] = \Pr\left[e^{X^2/K} > e^{t^2/K}\right] \leq \frac{\mathbb{E}[e^{X^2/K}]}{e^{t^2/K}} \leq 2e^{-t^2/K}.$$

With the above inequality, we can estimate the k-th moment of X. For any positive integer $k \geq 1$,

$$\begin{aligned}\mathbb{E}[|X|^k] &= \int_0^\infty \Pr\big[|X|^k > t\big]\,\mathrm{d}t = \int_0^\infty \Pr\big[|X| > t^{1/k}\big]\,\mathrm{d}t \\ &\leq 2\int_0^\infty e^{-t^{2/k}/K}\,\mathrm{d}t = K^{k/2}k\Gamma(k/2),\end{aligned}$$

where $\Gamma(z) \coloneqq \int_0^\infty t^{z-1}e^{-t}\,\mathrm{d}t$ is the Gamma function.

For any $s \in \mathbb{R}$, we use the Taylor expansion of the exponential function e^{sX} and apply the dominated convergence theorem,

$$\begin{aligned}\mathbb{E}[e^{sX}] &\leq 1 + \sum_{k=1}^\infty \frac{s^k\mathbb{E}[|X|^k]}{k!} \leq 1 + \sum_k \frac{(s^2K)^{k/2}k\Gamma(k/2)}{k!} \\ &= 1 + \sum_{k=1}^\infty \frac{(s^2K)^k 2k\Gamma(k)}{(2k)!} + \sum_{k=1}^\infty \frac{(s^2K)^{k+1/2}(2k+1)\Gamma(k+1/2)}{(2k+1)!} \\ &\leq 1 + \left(2 + \sqrt{s^2K}\right)\sum_{k=1}^\infty \frac{(s^2K)^k k!}{(2k)!} \leq 1 + \left(1 + \sqrt{\frac{s^2K}{4}}\right)\sum_{k=1}^\infty \frac{(s^2K)^k}{k!} \leq e^{2s^2K}.\end{aligned}$$

Note that we use the inequality $2(k!)^2 \leq (2k)!$ in the second-to-last step. □

Lemma B.2.2. *Let $X \in \mathbb{R}$ be a random variable such that $\mathbb{E}[e^{X^2/K}] \leq \alpha$ with $\alpha > 1$. Then, we have*

$$\mathbb{E}[e^{X^2/(\alpha-1)K}] \leq 2. \tag{B.9}$$

Proof. Let $r > 1$ be a real-valued parameter. Using the Fubini theorem,

$$\mathbb{E}[e^{X^2/rK}] = 1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[(X^2/rK)^k]}{k!} \leq 1 + \frac{1}{r} \sum_{k=1}^{\infty} \frac{\mathbb{E}[(X^2/K)^k]}{k!} \leq 1 + \frac{\alpha - 1}{r}.$$

Therefore, if we choose $r = \alpha - 1$, we end up with $\mathbb{E}[e^{X^2/(\alpha-1)K}] \leq 2$. □

Theorem B.2.3 (Theorem 1, [245]). *Let $A \in \mathbb{R}^{m \times n}$ be a matrix, and let $\Sigma = A^T A$. Suppose that $x = (x_1, \ldots, x_n)$ is a random vector such that, for some $\sigma \geq 0$,*

$$\mathbb{E}\left[e^{r^T x}\right] \leq e^{\|r\|^2 \sigma^2/2} \tag{B.10}$$

for all $r \in \mathbb{R}^d$. Then, for all $c > 0$,

$$\Pr\left[\|Ax\|^2 > \sigma^2 \left(\mathrm{Tr}(\Sigma) + 2\sqrt{\mathrm{Tr}(\Sigma^2)c} + 2\|\Sigma\|c\right)\right] \leq e^{-c}. \tag{B.11}$$

B.2.2 Proof of Lemma 4.4.2

Proof. Note that the ground state of the Hamiltonian $\hat{H}(\lambda)$ is exactly the same as

$$2\lambda\hat{H}(\lambda) = -\nabla^2 + 2\lambda^2 w(x),$$

where the effective potential function is $\lambda^2 w(x)$. By [56, Theorem 3.4], there is a constant $\alpha > 0$ such that

$$\int_{\mathbb{R}} e^{d_0(x)} |\phi_\lambda(x)|^2 \, \mathrm{d}x \leq \alpha, \tag{B.12}$$

where $d_0(x)$ is the Agmon distance between x and 0 (see [56, Definition 3.2]). It is well known that the Agmon distance scales as the square root of the potential function, i.e.,

$$d_0(x) \sim (\lambda^2 w(x))^{1/2} \text{ as } |x| \to \infty. \tag{B.13}$$

By Definition 4.3.2, the function $w(x)$ grows as fast as a quartic function, which means there exists a positive constant c' such that $w(x) \geq c'(x - x^*)^4$. It turns out that

$$d_0(x) \geq c''\lambda(x - x^*)^2,$$

where c'' is an absolute constant. It follows from (B.12) that

$$\int_{\mathbb{R}} e^{c''\lambda(x-x^*)^2} |\phi_\lambda(x)|^2 \, \mathrm{d}x \leq \alpha. \tag{B.14}$$

Let $X \sim |\phi_\lambda(x)|^2$ and $\xi \coloneqq X - x^*$, we rewrite (B.14) as $\mathbb{E}[e^{c''\lambda\xi^2}] \leq \alpha$. Without loss of generality, we assume $\alpha > 1$. Then, Lemma B.2.2 implies that $\mathbb{E}[e^{c''\lambda\xi^2/(\alpha-1)}] \leq 2$. Now, we invoke Lemma B.2.1 to obtain that

$$\mathbb{E}[e^{s\xi}] \leq e^{s^2\sigma^2/2}, \tag{B.15}$$

where $\sigma^2 = \beta/\lambda$ with $\beta = 4(\alpha - 1)/c''$. Note that β only depends on our choice of w. □

B.3 Auxiliary Lemmas

Lemma B.3.1. *Suppose that ψ_1, ψ_2 are two unit vectors in $\mathcal{H}$ such that*

$$\|\psi_1 - \psi_2\| \leq \delta. \tag{B.16}$$

Let O be a bounded operator on $\mathcal{H}$ such that $\|O\| \leq 1$. Then, we have

$$|\langle\psi_1| O |\psi_1\rangle - \langle\psi_2| O |\psi_2\rangle| \leq 2\delta. \tag{B.17}$$

Proof. By the triangle inequality,

$$|\langle\psi_1| O |\psi_1\rangle - \langle\psi_2| O |\psi_2\rangle| \leq |(\langle\psi_1| - \langle\psi_2|)O |\psi_1\rangle| + |\langle\psi_2| O(|\psi_1\rangle - |\psi_2\rangle)| \leq 2\delta.$$

□

The following is the proof of Lemma 4.4.3.

Proof. To simulate the QHD dynamics as given in (4.49), we consider the change of variable,

$$\xi = \lambda_f + \frac{1}{\lambda_f} - \frac{1}{\lambda(t)} = \lambda_f \left(1 - e^{-2t}\right) + \frac{1}{\lambda_f}. \tag{B.18}$$

We see that $\xi(t)$ is an increasing function in t such that $\xi(0) = \frac{1}{\lambda_f}$, $\xi(t_f) = \lambda_f$. Moreover, we can

express t in terms of ξ:

$$t(\xi) = -\frac{1}{2}\log\left(-\frac{\xi}{\lambda_f} + 1 + \frac{1}{\lambda_f^2}\right). \tag{B.19}$$

Using the change of variable $t \mapsto \xi(t)$ and we define $|\Psi^\epsilon(\xi)\rangle = |\Psi^\epsilon(\xi(t))\rangle$, the Schrödinger equation (4.49) becomes

$$i\epsilon\frac{\mathrm{d}}{\mathrm{d}\xi}|\Psi^\epsilon(\xi)\rangle = \frac{1}{2}\left[-\frac{1}{2}\nabla^2 + \varphi(\xi)f\right]|\Psi^\epsilon(\xi)\rangle, \tag{B.20}$$

where

$$\varphi(\xi) \coloneqq \frac{1}{(-\xi + \lambda_f + 1/\lambda_f)^2} \in [1/\lambda_f^2, \lambda_f^2]. \tag{B.21}$$

We can also absorb the parameter ϵ into the Hamiltonian by dilating the time scale $s \coloneqq \xi/\epsilon \in J \coloneqq [\frac{1}{\epsilon\lambda_f}, \frac{\lambda_f}{\epsilon}]$. Let $\varphi(s) \coloneqq \varphi(\epsilon s)$, we end up with the effective dynamics described by (4.52). □

Appendix C: Appendices for Chapter 6

C.1 Adiabatic Quantum Computing

In Theorem 4.4.1 (ii), we used the following lemmas:

Lemma C.1.1. *Let $H(t)$ be an explicitly s-sparse Hamiltonian path for $t \in [0,1]$. Suppose that $\|H(t)\|$, $\|\dot{H}(t)\|$, and $\|\ddot{H}(t)\|$ are all upper bounded by M and that the spectral gap of $H(t)$ is at least δ for all t. Furthermore, assume that the ground state of $H(0)$ is η-close to an efficiently preparable state $|\phi\rangle$. Then, there is a quantum algorithm, given oracle access to $H(t)$, that simulates the Schrödinger dynamics*

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = TH(t)|\psi(t)\rangle\,, \quad |\psi(0)\rangle = |\phi\rangle\,, \quad T = \Theta\left(\frac{M^2}{\delta^3\eta}\right), \tag{C.1}$$

producing a final state $|\psi(1)\rangle$ that is 2η-close to the ground state of $H(1)$ with query and gate complexity $(nsTM)^{1+o(1)}$.

Lemma C.1.2. *Let $H(t) = -\Delta + V(t,x)$ be a Hamiltonian path for $t \in [0, t_{\mathrm{f}}]$. Suppose that $|V(t,x)|$, $\|\dot{H}(t)\| \leq \max_x |\dot{V}(t,x)|$, and $\|\ddot{H}(t)\| \leq \max_x |\ddot{V}(t,x)|$ are all upper bounded by M and that the spectral gap of $H(t)$ is at least δ for all t. Furthermore, assume that the ground state of $H(0)$ is η-close to an efficiently preparable state $|\phi\rangle$. Then, there is a quantum algorithm, given oracle access to $V(t,x)$, that simulates the Schrödinger dynamics*

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = TH(t)|\psi(t)\rangle\,, \quad |\psi(0)\rangle = |\phi\rangle\,, \quad T = \Theta\left(\frac{M^2}{\delta^3\eta}\right), \tag{C.2}$$

producing a final state $|\psi(t_{\mathrm{f}})\rangle$ *that is* 2η*-close to the ground state of* $H(t_{\mathrm{f}})$ *with query and gate complexity* $(nTM)^{1+o(1)}$.

The correctness of Lemma C.1.1 directly follows from combining the *quantum adiabatic theorem* (Theorem C.1.3) and the rescaled Dyson-series algorithm [223] that simulates (C.1) with query and gate complexity $(nsTM)^{1+o(1)}$.

Similarly, The correctness of Lemma C.1.2 follows from Theorem C.1.3 and an algorithm [89, Theorem 3] that simulates (C.2) with query and gate complexity $(nTM)^{1+o(1)}$.

C.1.1 Quantum Adiabatic Theorem

Given a time-dependent Hamiltonian $H(t)$ for $t \in [0,1]$, denote by $|\phi(t)\rangle$ the ground state of $H(t)$. The *quantum adiabatic theorems* provide conditions under which the evolution remains close to $|\phi(t)\rangle$. Specifically, consider the following Schrödinger equation:

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = TH(t)|\psi(t)\rangle, \quad t \in [0,1], \tag{C.3}$$

with an initial state $|\psi(0)\rangle := |\phi(0)\rangle$ and a sufficiently large evolution time T. This dynamics is considered adiabatic because it can be rewritten via the change of variable $s := tT$, often referred to as a *time dilation* argument:

$$i\frac{\mathrm{d}}{\mathrm{d}s}|\psi(s/T)\rangle = \widetilde{H}(s)|\psi(s/T)\rangle, \quad s \in [0,T], \tag{C.4}$$

where $\widetilde{H}(s) := H(s/T)$. In this formulation, $\widetilde{H}(s)$ varies infinitesimally slowly as $T \to \infty$, ensuring adiabatic evolution. Below, we introduce a version of the theorem that establishes a sufficient bound

on T based on the operator norm of $\dot{H}(t)$ and $\ddot{H}(t)$ and the spectral gap of $H(t)$, following a standard operator-theoretic approach. For an elementary proof with slightly weaker bounds but offering more intuition, we refer the reader to [246].

Theorem C.1.3 ([57, 59, 247]). *Suppose $H(t)$ is piecewise twice differentiable and has a nondegenerate ground state for all $t \in [0,1]$. Furthermore, $\|\dot{H}(t)\|$ and $\|\ddot{H}(t)\|$ are upper bounded by M and the spectral gap of $H(t)$ is at least δ for all t. Then (C.3) produces a final state $|\psi(1)\rangle$ satisfying*

$$\| \, |\psi(1)\rangle - |\phi(1)\rangle \, \| \leq \varepsilon, \tag{C.5}$$

provided that

$$T \geq \frac{1}{\varepsilon} \cdot O\left(\frac{M^2}{\delta^3}\right). \tag{C.6}$$

Remark C.1.4. Note that $H(t)$ itself can be unbounded in Theorem C.1.3.

C.2 Explicit Technical Building Blocks

We present clean statements of the main technical building blocks employed implicitly in the proofs of Theorem 4.4.1 and Theorem 6.10.2. These building blocks have been modified to meet the specific requirements of our approach and are therefore (re)stated here for clarity and potential reuse.

The following proposition reduces any stoquastic sparse Hamiltonian to a TFD Hamiltonian. Its proof is implicit in those of Lemma 6.4.3 and 6.4.4.

Proposition C.2.1. *Let n, s, M be sufficiently large, and let $\eta, \varepsilon > 0$ be sufficiently small. Then, there exist some $a > 0$ and a mapping that transforms any n-qubit s-sparse stoquastic Hamiltonian H*

satisfying $\|H\| \leq M$ into a $(6n+4s)$-qubit stoquastic TFD Hamiltonian

$$K = -a \sum_{i \in [6n+4s]} X_i + D \tag{C.7}$$

with D diagonal, such that the following holds.

(i) $(K, \mathcal{E})$ simulates H with error (η, ε), where $\mathcal{E}\colon |x\rangle \mapsto |x, x, 0^s, 0^s, 0^n, x, x, 0^s, 0^s, 0^n\rangle$ is the encoding isometry.

(ii) $\|K\| \leq \mathrm{poly}(1/\eta, 1/\varepsilon, n, s, M)$.

(iii) An oracle for K can be efficiently implemented: a query to this oracle can be simulated using $O(ns + s^2)$ queries to an oracle for H, with gate complexity $\mathrm{poly}(n, s)$.

The following proposition linearizes a Hamiltonian $K(t)$ to a Hamiltonian $H(t)$ that is linear in t. Note that the reduction *preserves stoquasticity*. Its proof is implicit in those of Lemma 6.4.5, 6.4.6, and 6.4.7.

Proposition C.2.2. *Let $K(t)$ be a n-qubit Hamiltonian path that is explicitly s-sparse. Assume that for any $t \in [0,1]$, $K(t)$ has norm $\|K(t)\| \leq M$ and spectral gap at least δ. Assume in addition that $K(t)$ is L-Lipschitz in t.*

Let $\lambda \leq O(\delta/M)$ be arbitrary. For some $\ell = \Theta(L/\delta)$ and $\Delta \geq \mathrm{poly}(\ell, 1/\lambda)$, define

$$H(t) = \frac{\Delta}{\lambda}\left(\sum_{i=0}^{\ell} i(i+1-2t)\,|i\rangle\langle i| - \lambda \sum_{i=0}^{\ell-1} \left(|i\rangle\langle i+1| + |i+1\rangle\langle i|\right)\right) + \sum_{i=0}^{\ell} K(i/\ell) \otimes |i\rangle\langle i|\,. \tag{C.8}$$

Then the following holds.

(i) For $t \in [0,1]$, $H(t)$ has spectral gap $\delta/4$, norm $\mathrm{poly}(1/\lambda, \ell)$, and sparsity at most $s+2$.

(ii) Let $|\psi\rangle$ be a ground state of $K(0)$. Then $|\psi\rangle \otimes |0\rangle$ is $O(\lambda)$-close to a ground state of $H(0)$.

(iii) Let $|\psi'\rangle$ be a ground state of $K(1)$. Then $|\psi'\rangle \otimes |\ell\rangle$ is $O(\lambda)$-close to a ground state of $H(1)$.

(iv) An oracle for $H(t)$ can be efficiently implemented: a query to this oracle can be simulated using $O(1)$ queries to an oracle for $K(t)$, with gate complexity $\mathrm{poly}(n, s)$.

Remark C.2.3. We note that the statement can be strengthened such that $H(t)$ simulates $K(\tau)$ for some monotonically increasing function $\tau = \tau(t)$ satisfying $\tau(0) = 0$ and $\tau(1) = 1$, although certain parameter dependencies need to be adjusted.

The following proposition transforms any hard instances exhibiting quantum advantage via adiabatic quantum optimization to hard instances exhibiting quantum advantage via Quantum Hamiltonian Descent (with the help of a quick-vanishing term $\nu(t)g(x)$). Its proof is implicit in that of Theorem 6.10.2.

Proposition C.2.4. *Let n, M, q be sufficiently large and $\varepsilon, \delta > 0$ be sufficiently small. Let H_{TFD} be an n-qubit TFD Hamiltonian such that $\|H_{\mathrm{TFD}}\| \leq M$, and that all coefficients of X_i in H_{TFD} are no more than -1. Let D be an n-qubit diagonal Hamiltonian such that its eigenvalues lie in $[-1, 0)$. Suppose that the Hamiltonian path $H(t)$ of the form*

$$H_{\mathrm{TFD}} \longrightarrow D \tag{C.9}$$

has quantum advantage given oracle access to D:

(i) Any classical algorithm requires at least q queries to find $u \in \{0, 1\}^n$ that minimizes $\langle u|D|u\rangle$, with success probability greater than $1/q$.

(ii) $H(t)$ *has spectral gap* δ*, norm* M*, and the ground state of* $H(0) = H_{\mathrm{TFD}}$ *is* ε*-close to a product state.*

Then, there exist functions $g(x)$ *and* $\nu(t)$*, and a family of functions* $f\colon \mathcal{X} \to \mathbb{R}$ *with a fixed box domain* $\mathcal{X} \subset \mathbb{R}^n$ *where the diameter of* $\mathcal{X}$ *is at most* $\mathrm{poly}(M)$*, such that the Hamiltonian path* $\widehat{H}(t)$ *of the form*

$$\widehat{H}(t) \coloneqq -\Delta + V(t,x), \quad V(t,x) \coloneqq \nu(t)g(x) + tf(x), \quad \forall t \geq 0, x \in \mathcal{X} \tag{C.10}$$

has quantum advantage given oracle access to $f(\cdot)$*:*

(i) Any classical algorithm requires at least q *queries to find an* x *such that* $f(x) - f(x^*) \leq \delta$*, with success probability greater than* $1/q$*.*

(ii) $\widehat{H}(t)$ *has spectral gap at least* $\frac{1}{\mathrm{poly}(n,1/\varepsilon,1/\delta)}$*, and the ground state of* $\widehat{H}(0)$ *is* 2ε*-close to a product state. Moreover, for any* $\eta > 0$*, there exists* $t_{\mathrm{f}} = \mathrm{poly}(M, 1/\eta)$ *such that* $\langle\phi(t_{\mathrm{f}})|f|\phi(t_{\mathrm{f}})\rangle - f(x^*) \leq \eta$*, where* $|\phi(t)\rangle$ *is the ground state of* $\widehat{H}(t)$*.*

(iii) $f(x)$*,* $g(x)$*, and* $\nu(t)$ *satisfy the following:*

- $f(x)$ *is bounded within* $\pm\mathrm{poly}(n, M)$ *and it is* $\mathrm{poly}(n, M)$*-Lipschitz. Moreover,* $f(x)$ *has* 2^n *local minima, and the function values at these minima correspond to the diagonal entries of the diagonal matrix* D*.*
- $g(x)$ *is bounded within* $\pm\mathrm{poly}(n, M, \log\frac{1}{\varepsilon}, \log\frac{1}{\delta})$ *and it is* $\mathrm{poly}(n, M, \log\frac{1}{\varepsilon}, \log\frac{1}{\delta})$*-Lipschitz.*
- $\dfrac{\mathrm{d}^k\nu}{\mathrm{d}t^k} = e^{-\Theta(\sqrt{t})}$ *for any* k*.*

Remark C.2.5. The functions $f(x)$ and $g(x)$ can be made smooth (i.e., infinitely differentiable) by a more sophisticated construction of $\widehat{D}$ in Definition 6.10.4, for example by convolving a mollifier

with a more carefully designed base function. We chose not to present this approach formally, as the benefits are minimal and proofs of certain claims, e.g., Fact 6.10.5 (i), would be more complicated.

Bibliography

[1] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 4th edition, 2022.

[2] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2011.

[3] Mauro E. S. Morales, Lirandë Pira, Philipp Schleich, Kelvin Koor, Pedro C. S. Costa, Dong An, Alán Aspuru-Guzik, Lin Lin, Patrick Rebentrost, and Dominic W. Berry. Quantum linear system solvers: A survey of algorithms and applications, 2025. arXiv:2411.02522.

[4] Aaron J. Ostrander. *Quantum Algorithms for Differential Equations*. PhD thesis, University of Maryland, College Park, 2019.

[5] Amira Abbas, Andris Ambainis, Brandon Augustino, Andreas Bärtschi, Harry Buhrman, Carleton Coffrin, Giorgio Cortiana, Vedran Dunjko, Daniel J. Egger, Bruce G. Elmegreen, Nicola Franco, Filippo Fratini, Bryce Fuller, Julien Gacon, Constantin Gonciulea, Sander Gribling, Swati Gupta, Stuart Hadfield, Raoul Heese, Gerhard Kircher, Thomas Kleinert, Thorsten Koch, Georgios Korpas, Steve Lenk, Jakub Marecek, Vanio Markov, Guglielmo Mazzola, Stefano Mensa, Naeimeh Mohseni, Giacomo Nannicini, Corey O'Meara, Elena Peña Tapia, Sebastian Pokutta, Manuel Proissl, Patrick Rebentrost, Emre Sahin, Benjamin C. B. Symons, Sabine Tornow, Víctor Valls, Stefan Woerner, Mira L. Wolf-Bauwens, Jon Yard, Sheir Yarkoni, Dirk Zechiel, Sergiy Zhuk, and Christa Zoufal. Challenges and opportunities in quantum optimization. *Nature Reviews Physics*, 6(12):718–735, 12 2024.

[6] I. M. Georgescu, S. Ashhab, and Franco Nori. Quantum simulation. *Rev. Mod. Phys.*, 86:153–185, Mar 2014.

[7] Jingwen Suo, Licheng Wang, Sijia Yang, Wenjie Zheng, and Jiankang Zhang. Quantum algorithms for typical hard problems: a perspective of cryptanalysis. *Quantum Information Processing*, 19(6):178, Apr 2020.

[8] Mark Horowitz and Emily Grumbling. *Quantum computing: progress and prospects*. National Academies Press, 2019.

[9] Ryan Babbush, Jarrod R. McClean, Michael Newman, Craig Gidney, Sergio Boixo, and Hartmut Neven. Focus beyond quadratic speedups for error-corrected quantum advantage. *PRX Quantum*, 2(1), March 2021.

[10] Jens Eisert and John Preskill. Mind the gaps: The fraught road to quantum advantage, 2025. arXiv:2510.19928.

[11] Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *35th Annual Symposium on Foundations of Computer Science, Santa Fe, New Mexico, USA, November 20-22, 1994*, pages 124–134. IEEE Computer Society, 1994.

[12] Stephen P. Jordan, Noah Shutty, Mary Wootters, Adam Zalcman, Alexander Schmidhuber, Robbie King, Sergei V. Isakov, Tanuj Khattar, and Ryan Babbush. Optimization by decoded quantum interferometry. *Nature*, 646(8086):831–836, Oct 2025.

[13] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, page 212–219, New York, NY, USA, 1996. Association for Computing Machinery.

[14] Gilles Brassard, Peter Høyer, and Alain Tapp. *Quantum cryptanalysis of hash and claw-free functions: Invited paper*, page 163–169. Springer Berlin Heidelberg, 1998.

[15] Dorit Aharonov, Vaughan Jones, and Zeph Landau. A polynomial quantum algorithm for approximating the Jones polynomial. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, page 427–436, New York, NY, USA, 2006. Association for Computing Machinery.

[16] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, STOC '11, page 333–342, New York, NY, USA, 2011. Association for Computing Machinery.

[17] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, Oct 2019.

[18] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15), October 2009.

[19] Francois Le Gall. Improved quantum algorithm for triangle finding via combinatorial arguments. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, page 216–225. IEEE, October 2014.

[20] Harry Buhrman and Robert Spalek. Quantum verification of matrix products, 2005. arXiv:quant-ph/0409035.

[21] Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM Journal on Computing*, 37(1):210–239, 2007.

[22] Christoph Durr and Peter Hoyer. A quantum algorithm for finding the minimum, 1996. arXiv:quant-ph/9607014.

[23] David Bulger, William P Baritompa, and Graham R Wood. Implementing pure adaptive search with Grover's quantum algorithm. *Journal of optimization theory and applications*, 116:517–529, 2003.

[24] William P Baritompa, David W Bulger, and Graham R Wood. Grover's quantum algorithm applied to global optimization. *SIAM Journal on Optimization*, 15(4):1170–1184, 2005.

[25] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H Booth, et al. The variational quantum eigensolver: a review of methods and best practices. *Physics Reports*, 986:1–128, 2022.

[26] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.

[27] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016.

[28] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

[29] Catherine C McGeoch. *Adiabatic quantum computation and quantum annealing: Theory and practice*. Morgan & Claypool Publishers, 2014.

[30] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse Ising model. *Physical Review E*, 58(5):5355, 1998.

[31] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution, 2000. arXiv:quant-ph/0001106.

[32] Troels F Rønnow, Zhihui Wang, Joshua Job, Sergio Boixo, Sergei V Isakov, David Wecker, John M Martinis, Daniel A Lidar, and Matthias Troyer. Defining and detecting quantum speedup. *science*, 345(6195):420–424, 2014.

[33] Kostas Blekos, Dean Brand, Andrea Ceschini, Chiao-Hui Chou, Rui-Hao Li, Komal Pandya, and Alessandro Summer. A review on Quantum Approximate Optimization Algorithm and its variants. *Physics Reports*, 1068:1–66, 2024.

[34] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D Lukin. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Physical Review X*, 10(2):021067, 2020.

[35] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm, 2014. arXiv:1411.4028.

[36] Matthew B Hastings. A short path quantum algorithm for exact optimization. *Quantum*, 2:78, 2018.

[37] Patrick Rebentrost, Maria Schuld, Leonard Wossnig, Francesco Petruccione, and Seth Lloyd. Quantum gradient descent and Newton's method for constrained polynomial optimization. *New Journal of Physics*, 21(7):073023, 2019.

[38] Iordanis Kerenidis and Anupam Prakash. Quantum gradient descent for linear systems and least squares. *Physical Review A*, 101(2):022316, 2020.

[39] Shouvanik Chakrabarti, Andrew M Childs, Tongyang Li, and Xiaodi Wu. Quantum algorithms and lower bounds for convex optimization. *Quantum*, 4:221, 2020.

[40] Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Convex optimization using quantum oracles. *Quantum*, 4:220, 2020.

[41] Fernando G.S.L. Brandao and Krysta M. Svore. Quantum speed-ups for solving semidefinite programs. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 415–426, 2017.

[42] Fernando G. S. L. Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu. Quantum SDP solvers: large speed-ups, optimality, and applications to quantum learning. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 27:1–27:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[43] Joran van Apeldoorn and András Gilyén. Improvements in quantum SDP-solving with applications. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 99:1–99:15, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[44] Joran van Apeldoorn and András Gilyén. Quantum algorithms for zero-sum games, 2019. arXiv:1904.0318.

[45] Iordanis Kerenidis and Anupam Prakash. A quantum interior point method for LPs and SDPs. *ACM Transactions on Quantum Computing*, 1(1), October 2020.

[46] Mohammadhossein Mohammadisiahroudi, Ramin Fakhimi, and Tamás Terlaky. Efficient use of quantum linear system algorithms in interior point methods for linear optimization, 2023. arXiv:2205.01220.

[47] Brandon Augustino, Giacomo Nannicini, Tamás Terlaky, and Luis F. Zuluaga. Quantum interior point methods for semidefinite optimization. *Quantum*, 7:1110, September 2023.

[48] Zeguan Wu, Mohammadhossein Mohammadisiahroudi, Brandon Augustino, Xiu Yang, and Tamás Terlaky. An inexact feasible quantum interior point method for linearly constrained quadratic ooptimization. *Entropy*, 25(2), 2023.

[49] Simon Apers and Sander Gribling. Quantum speedups for linear programming via interior point methods, 2026. arXiv:2311.03215.

[50] Aaron Sidford and Chenyi Zhang. Quantum speedups for stochastic optimization. *Advances in Neural Information Processing Systems*, 36, 2024.

[51] Chenyi Zhang, Jiaqi Leng, and Tongyang Li. Quantum algorithms for escaping from saddle points. *Quantum*, 5:529, August 2021.

[52] Jiaqi Leng, Ethan Hickman, Joseph Li, and Xiaodi Wu. Quantum Hamiltonian descent, 2023. arXiv:2303.01471.

[53] Zherui Chen, Yuchen Lu, Hao Wang, Yizhou Liu, and Tongyang Li. Quantum Langevin dynamics for optimization. *Communications in Mathematical Physics*, 406(3):52, Feb 2025.

[54] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. Schuyler Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papageorge, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, Colm A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, Robert S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, Blake R. Johnson, M. Reagor, M. P. da Silva, and C. Rigetti. Unsupervised machine learning on a hybrid quantum computer, 2017. arXiv:1712.05771.

[55] S. Ebadi, A. Keesling, M. Cain, T. T. Wang, H. Levine, D. Bluvstein, G. Semeghini, A. Omran, J.-G. Liu, R. Samajdar, X.-Z. Luo, B. Nash, X. Gao, B. Barak, E. Farhi, S. Sachdev, N. Gemelke, L. Zhou, S. Choi, H. Pichler, S.-T. Wang, M. Greiner, V. Vuletić, and M. D. Lukin. Quantum optimization of maximum independent set using Rydberg atom arrays. *Science*, 376(6598):1209–1215, June 2022.

[56] P. D. Hislop and I. M. Sigal. *Introduction to Spectral Theory: With Applications to Schrödinger Operators*, volume 113 of *Applied Mathematical Sciences*. Springer Science & Business Media, New York, NY, United States, 2012.

[57] Stefan Teufel. *Adiabatic perturbation theory in quantum dynamics*. Springer Berlin Heidelberg, 2003.

[58] Sevag Gharibian, Yichen Huang, Zeph Landau, and Seung Woo Shin. Quantum Hamiltonian complexity. *Foundations and Trends® in Theoretical Computer Science*, 10(3):159–282, 2015.

[59] Sabine Jansen, Mary-Beth Ruskai, and Ruedi Seiler. Bounds for the adiabatic approximation with applications to quantum computation. *Journal of Mathematical Physics*, 48(10):102111, 10 2007.

[60] Jérémie Roland and Nicolas J. Cerf. Quantum search by local adiabatic evolution. *Phys. Rev. A*, 65:042308, 3 2002.

[61] Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Journal on Computing*, 37(1):166–194, 2007.

[62] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 90:015002, 1 2018.

[63] Jiaqi Leng, Joseph Li, Yuxiang Peng, and Xiaodi Wu. Expanding hardware-efficiently manipulable hilbert space via Hamiltonian embedding. *Quantum*, 9:1857, September 2025.

[64] D-Wave Systems Inc. D-wave documentation, 2024. Accessed: 2026-03-07.

[65] Dylan Herman, Guneykan Ozgul, Anuj Apte, Junhyung Lyle Kim, Anupam Prakash, Jiayu Shen, and Shouvanik Chakrabarti. Mechanisms for quantum advantage in global optimization of nonconvex functions, 2025. arXiv:2510.03385.

[66] Aris Daniilidis and Dmitriy Drusvyatskiy. Pathological subgradient dynamics. *SIAM Journal on Optimization*, 30(2):1327–1338, 2020.

[67] Rodolfo Ríos-Zertuche. Examples of pathological dynamics of the subgradient method for Lipschitz path-differentiable functions. *Mathematics of Operations Research*, 47(4):3184–3206, November 2022.

[68] Damek Davis, Dmitriy Drusvyatskiy, Sham Kakade, and Jason D. Lee. Stochastic subgradient method converges on tame functions. *Foundations of Computational Mathematics*, 20(1):119–154, Feb 2020.

[69] Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

[70] Xiaoyin Hu, Nachuan Xiao, Xin Liu, and Kim-Chuan Toh. Learning-rate-free momentum SGD with reshuffling converges in nonsmooth nonconvex optimization. *Journal of Scientific Computing*, 102(3):85, Feb 2025.

[71] Yang Xiang, DY Sun, W Fan, and XG Gong. Generalized simulated annealing algorithm and its application to the thomson model. *Physics Letters A*, 233(3):216–220, 1997.

[72] David J Wales and Jonathan PK Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997.

[73] Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106:25–57, 2006.

[74] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, New York, 2006.

[75] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. Gurobi Optimization, LLC, 13.0 edition, 2026. Accessed: 2026-02-22.

[76] Yizhou Liu, Weijie J. Su, and Tongyang Li. On quantum speedups for nonconvex optimization via quantum tunneling walks. *Quantum*, 7:1030, June 2023.

[77] Jiaqi Leng, Yufan Zheng, and Xiaodi Wu. A quantum-classical performance separation in nonconvex optimization, 2023. arXiv:2311.00811.

[78] Ahmet Burak Catli, Sophia Simon, and Nathan Wiebe. Exponentially better bounds for quantum optimization via dynamical simulation, 2025. arXiv:2502.04285.

[79] Jiaqi Leng, Yufan Zheng, Zhiyuan Jia, Lei Fan, Chaoyue Zhao, Yuxiang Peng, and Xiaodi Wu. Quantum hamiltonian descent for non-smooth optimization, 2025. arXiv:2503.15878.

[80] Shouvanik Chakrabarti, Dylan Herman, Jacob Watkins, Enrico Fontana, Brandon Augustino, Junhyung Lyle Kim, and Marco Pistoia. On speedups for convex optimization via quantum dynamics, 2025. arXiv:2503.24332.

[81] Jiaqi Leng, Zhiyan Ding, Zherui Chen, and Lin Lin. Operator-level quantum acceleration of non-logconcave sampling, 2025. arXiv:2505.05301.

[82] András Gilyén, Matthew B. Hastings, and Umesh Vazirani. (sub)exponential advantage of adiabatic quantum computation with no sign problem. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 1357–1369, New York, NY, USA, 2021. Association for Computing Machinery.

[83] Yufan Zheng, Jiaqi Leng, Yizhou Liu, and Xiaodi Wu. On the computational complexity of Schrödinger operators, 2024. arXiv:2411.05120.

[84] Jiaqi Leng, Kewen Wu, Xiaodi Wu, and Yufan Zheng. (Sub)exponential quantum speedup for optimization, 2025. arXiv:2504.14841.

[85] Weijie Su, Stephen Boyd, and Emmanuel J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016. `http://jmlr.org/papers/v17/15-084.html`.

[86] Andre Wibisono, Ashia C. Wilson, and Michael I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.

[87] Yurii Nesterov. A method for solving the convex programming problem with convergence rate O(1/k2). In *Dokl akad nauk Sssr*, volume 269, page 543, 1983.

[88] Leon Armenovich Takhtadzhian. *Quantum mechanics for mathematicians*, volume 95. American Mathematical Soc., 2008.

[89] Andrew M. Childs, Jiaqi Leng, Tongyang Li, Jin-Peng Liu, and Chenyi Zhang. Quantum simulation of real-space dynamics. *Quantum*, 6:860, November 2022.

[90] Samuel Kushnir, Jiaqi Leng, Yuxiang Peng, Lei Fan, and Xiaodi Wu. QHDOPT: A software for nonlinear optimization with quantum hamiltonian descent. *INFORMS Journal on Computing*, 37(1):107–124, 2025.

[91] Annabella Astorino, Antonio Fuduli, and Enrico Gorgone. Non-smoothness in classification problems. *Optimisation Methods & Software*, 23(5):675–688, 2008.

[92] Ohad Shamir and Tong Zhang. Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. In *International conference on machine learning*, pages 71–79. PMLR, 2013.

[93] Victor A Kovtunenko, Hiromichi Itou, Alexander M Khludnev, and Evgeny M Rudoy. Non-smooth variational problems and applications, 2022.

[94] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[95] Stephen Boyd, Lin Xiao, and Almir Mutapcic. Subgradient methods. *lecture notes of EE392o, Stanford University, Autumn Quarter*, 2004(01), 2003.

[96] Angelia Nedic and Dimitri P Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization*, 12(1):109–138, 2001.

[97] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[98] Brian Bullins and Richard Peng. Higher-order accelerated methods for faster non-smooth optimization, 2019. arXiv:1906.01621.

[99] Huan Li and Zhouchen Lin. Accelerated proximal gradient methods for nonconvex programming. *Advances in Neural Information Processing Systems*, 28, 2015.

[100] D. R. Sahu, J. C. Yao, M. Verma, and K. K. Shukla. Convergence rate analysis of proximal gradient methods with applications to composite minimization problems. *Optimization*, 70(1):75–100, 2021.

[101] Lorenzo Stella. *Proximal Envelopes: Smooth Optimization Algorithms for Nonsmooth Problems*. Phd thesis, IMT School for Advanced Studies Lucca, Lucca, Italy, 2017. PhD in Computer, Decision and Systems Science, XXVIII Cycle.

[102] Yurii Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103:127–152, 2005.

[103] Joseph Kreimer and Reuven Y Rubinstein. Nondifferentiable optimization via smooth approximation: General analytical approach. *Annals of Operations Research*, 39:97–119, 1992.

[104] Marjo Haarala, Kaisa Miettinen, and Marko M Mäkelä. New limited memory bundle method for large-scale nonsmooth optimization. *Optimization Methods and Software*, 19(6):673–692, 2004.

[105] A Yu Kitaev. Quantum measurements and the Abelian stabilizer problem, 1995. arXiv:quant-ph/9511026.

[106] Dominic W Berry. High-order quantum algorithm for solving linear differential equations. *Journal of Physics A: Mathematical and Theoretical*, 47(10):105301, 2014.

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

[108] Ankit Garg, Robin Kothari, Praneeth Netrapalli, and Suhail Sherif. No quantum speedup over gradient descent for non-smooth convex optimization. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2021.

[109] Chengchang Liu, Chaowen Guan, Jianhao He, and John Lui. Quantum algorithms for non-smooth non-convex optimization, 2024. arXiv:2410.16189.

[110] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192, 2007.

[111] Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. In *29th International Conference on Machine Learning, ICML 2012*, pages 449–456, 2012.

[112] Amir Beck. *First-order methods in optimization*. SIAM, 2017.

[113] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[114] Arkadii Semenovich Nemirovski and David Borisovich Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley & Sons, New York, 1983.

[115] Victor Ivrii. 100 years of Weyl's law. *Bulletin of Mathematical Sciences*, 6(3):379–452, 2016.

[116] Grigorii Vladimirovich Rozenblum. Distribution of the discrete spectrum of singular differential operators. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, 1976(1):75–86, 1976.

[117] Bin Shi, Simon S Du, Michael I Jordan, and Weijie J Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*, pages 1–70, 2022.

[118] Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, United States, 2nd edition, 2010.

[119] Moslem Zamani and François Glineur. Exact convergence rate of the last iterate in subgradient methods, 2023. arXiv:2307.11134.

[120] GitHub Contributors. QHD-experiment - Github Repository, 2025. `https://github.com/lwins-lights/QHD-experiment`.

[121] Aaron Defazio, Ashok Cutkosky, Harsh Mehta, and Konstantin Mishchenko. Optimal linear decay learning rate schedules and further refinements, 2023. arXiv:2310.07831.

[122] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25, 2012.

[123] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.

[124] M Van den Berg. On condensation in the free-boson gas and the spectrum of the Laplacian. *Journal of Statistical Physics*, 31:623–637, 1983.

[125] Mark S Ashbaugh and Rafael Benguria. Optimal lower bound for the gap between the first two eigenvalues of one-dimensional Schrödinger operators with symmetric single-well potentials. *Proceedings of the American Mathematical Society*, 105(2):419–424, 1989.

[126] Shing-Tung Yau. Nonlinear analysis in geometry, monographies de lenseignement mathématique, vol. 33. *L'Enseignement Mathématique, Geneva*, page 54, 1986.

[127] Ben Andrews and Julie Clutterbuck. Proof of the fundamental gap conjecture. *Journal of the American Mathematical Society*, 24(3):899–916, 2011.

[128] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2018.

[129] Shing-Tung Yau. Gap of the first two eigenvalues of the Schrödinger operator with nonconvex potential, 2009. arXiv:0902.2253.

[130] Evans M. Harrell. Double wells. *Communications in Mathematical Physics*, 75(3):239–261, 1980.

[131] Barry Simon. Semiclassical analysis of low lying eigenvalues. i. non-degenerate minima: Asymptotic expansions. *Annales de l'Institut Henri Poincaré, Section A: Physique Théorique*, 38(3):295–308, 1983.

[132] Tosio Kato. On the adiabatic theorem of quantum mechanics. *Journal of the Physical Society of Japan*, 5(6):435–439, 1950.

[133] Evgeny Mozgunov and Daniel A Lidar. Quantum adiabatic theorem for unbounded Hamiltonians with a cutoff and its application to superconducting circuits. *Philosophical Transactions of the Royal Society A*, 381(2241):20210407, 2023.

[134] Stephen Wiesner. Simulations of many-body quantum systems by a quantum computer, 1996. arXiv:quant-ph/9603028.

[135] Christof Zalka. Efficient simulation of quantum systems by quantum computers. *Fortschritte der Physik: Progress of Physics*, 46(6-8):877–879, 1998.

[136] Dong An, Di Fang, and Lin Lin. Time-dependent unbounded Hamiltonian simulation with vector norm scaling. *Quantum*, 5:459, 2021.

[137] Dong An, Di Fang, and Lin Lin. Time-dependent Hamiltonian simulation of highly oscillatory dynamics and superconvergence for Schrödinger equation. *Quantum*, 6:690, 2022.

[138] Division of Information Technology of University of Maryland. Zaratan HPC cluster. `https://hpcc.umd.edu/hpcc/zaratan.html`. Accessed: 2023-08-16.

[139] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[140] Bin Shi, Weijie Su, and Michael I. Jordan. On learning rates and Schrödinger operators. *Journal of Machine Learning Research*, 24(379):1–53, 2023.

[141] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 6 1982.

[142] Bela Bauer, Sergey Bravyi, Mario Motta, and Garnet Kin-Lic Chan. Quantum algorithms for quantum chemistry and quantum materials science. *Chemical Reviews*, 120(22):12685–12717, Nov 2020.

[143] Steven M. Girvin and Kun Yang. *Modern Condensed Matter Physics*. Cambridge University Press, Cambridge, 2019.

[144] Errol G. Lewars. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Springer Science+Business Media B.V., Dordrecht, The Netherlands, 2nd edition, 2011.

[145] Ira N Levine, Daryle H Busch, and Harrison Shull. *Quantum Chemistry*, volume 6. Pearson Prentice Hall, Upper Saddle River, NJ, 2009.

[146] Yuan Su, Dominic W. Berry, Nathan Wiebe, Nicholas Rubin, and Ryan Babbush. Fault-tolerant quantum simulations of chemistry in first quantization. *PRX Quantum*, 2:040332, Nov 2021.

[147] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum algorithms for quantum field theories. *Science*, 336(6085):1130–1133, 2012.

[148] John Preskill. Simulating quantum field theory with a quantum computer. In *Proceedings of The 36th Annual International Symposium on Lattice Field Theory — PoS(LATTICE2018)*, volume 334, page 024, 2019.

[149] Samuel L. Braunstein and Peter van Loock. Quantum information with continuous variables. *Rev. Mod. Phys.*, 77:513–577, Jun 2005.

[150] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, and Seth Lloyd. Gaussian quantum information. *Rev. Mod. Phys.*, 84:621–669, May 2012.

[151] Izrail Moiseevich Gel'fand and Boris Moiseevich Levitan. On the determination of a differential equation from its spectral function. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 15(4):309–360, 1951.

[152] Michael Reed and Barry Simon. *Methods of modern mathematical physics*, volume 1. Academic Press, New York, NY, United States, 1972.

[153] J. Avron, I. Herbst, and B. Simon. Schrödinger operators with magnetic fields. I. general interactions. *Duke Mathematical Journal*, 45(4):847 – 883, 1978.

[154] Barry Simon. Schrödinger semigroups. *Bulletin of the American Mathematical Society*, 7(3):447–526, 1982.

[155] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer Science & Business Media, Berlin, Heidelberg, New York, 2013.

[156] Scott Aaronson. Why quantum chemistry is hard. *Nature Physics*, 5(10):707–708, Oct 2009.

[157] Scott Aaronson, DeVon Ingram, and William Kretschmer. The acrobatics of BQP. In Shachar Lovett, editor, *37th Computational Complexity Conference (CCC 2022)*, volume 234 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 20:1–20:17, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[158] Barbara M. Terhal and David P. DiVincenzo. Classical simulation of noninteracting-fermion quantum circuits. *Phys. Rev. A*, 65:032325, Mar 2002.

[159] E. Knill. Fermionic linear optics and matchgates, 2001.

[160] Leslie G. Valiant. Quantum computers that can be simulated classically in polynomial time. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, STOC '01, page 114–123, New York, NY, USA, 2001. Association for Computing Machinery.

[161] Daniel Gottesman. The Heisenberg representation of quantum computers. In *22nd International Colloquium on Group Theoretical Methods in Physics*, pages 32–43, 7 1998.

[162] Sergey Bravyi. Monte Carlo simulation of stoquastic Hamiltonians. *Quantum Information and Computation*, 2015.

[163] Stephen P. Jordan, Hari Krovi, Keith S. M. Lee, and John Preskill. BQP-completeness of scattering in scalar quantum field theory. *Quantum*, 2:44, January 2018.

[164] Norbert Schuch and Frank Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nature Physics*, 5(10):732–735, Oct 2009.

[165] Andrew M. Childs, David Gosset, and Zak Webb. The Bose-Hubbard model is QMA-complete. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming*, pages 308–319, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[166] Leo Zhou and Dorit Aharonov. Strongly universal Hamiltonian simulators, 2021. arXiv:2102.02991.

[167] Ulysse Chabaud, Michael Joseph, Saeed Mehraban, and Arsalan Motamedi. Bosonic quantum computational complexity, 2024. arXiv:2410.04274.

[168] Varun Upreti and Ulysse Chabaud. Bounding the computational power of bosonic systems, 2025. arXiv:2503.03600.

[169] Alexei Yu Kitaev, Alexander H. Shen, and Mikhail N. Vyalyi. *Classical and Quantum Computation*. Number 47 in Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, United States, 2002.

[170] Adam D Bookatz. QMA-complete problems. *Quantum Information and Computation*, 14(5&6):361–383, 2014.

[171] Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local Hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006.

[172] Roberto Oliveira and Barbara Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quantum Information and Computation*, 8(10):900–924, 2008.

[173] Toby S. Cubitt, Ashley Montanaro, and Stephen Piddock. Universal quantum Hamiltonians. *Proceedings of the National Academy of Sciences*, 115(38):9497–9502, 2018.

[174] Toby Cubitt and Ashley Montanaro. Complexity classification of Local Hamiltonian Problems. *SIAM Journal on Computing*, 45(2):268–316, 2016.

[175] Stephen Piddock and Ashley Montanaro. The complexity of antiferromagnetic interactions and 2D lattices. *Quantum Information & Computation*, 17(7-8):636–672, 2017.

[176] Dorit Aharonov, Daniel Gottesman, Sandy Irani, and Julia Kempe. The power of quantum systems on a line. *Communications in Mathematical Physics*, 287(1):41–65, Apr 2009.

[177] Sean Hallgren, Daniel Nagaj, and Sandeep Narayanaswami. The local Hamiltonian problem on a line with eight states is QMA-complete. *Quantum Information & Computation*, 13(9-10):721–750, 2013.

[178] Sergey Bravyi, David P. DiVincenzo, Roberto Oliveira, and Barbara M. Terhal. The complexity of stoquastic local Hamiltonian problems. *Quantum Inf. Comput.*, 8(5):361–385, 2008.

[179] Sergey Bravyi, Arvid J. Bessen, and Barbara M. Terhal. Merlin-Arthur games and stoquastic complexity, 2006. arXiv:quant-ph/0611021.

[180] Sergey Bravyi and Barbara Terhal. Complexity of stoquastic frustration-free Hamiltonians. *SIAM Journal on Computing*, 39(4):1462–1485, 2010.

[181] Jiaqing Jiang. Local hamiltonian problem with succinct ground state is MA-complete. *PRX Quantum*, 6:020312, Apr 2025.

[182] Sevag Gharibian and François Le Gall. Dequantizing the Quantum Singular Value Transformation: Hardness and applications to quantum chemistry and the quantum PCP conjecture. *SIAM Journal on Computing*, 52(4):1009–1038, 2023.

[183] Jordi Weggemans, Marten Folkertsma, and Chris Cade. Guidable Local Hamiltonian Problems with Implications to Heuristic Ansatz State Preparation and the Quantum PCP Conjecture. In Frédéric Magniez and Alex Bredariol Grilo, editors, *19th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2024)*, volume 310 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:24, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[184] Chris Cade, Marten Folkertsma, Sevag Gharibian, Ryu Hayakawa, François Le Gall, Tomoyuki Morimae, and Jordi Weggemans. Improved Hardness Results for the Guided Local Hamiltonian Problem. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 32:1–32:19, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[185] Andrew D. King, Alberto Nocera, Marek M. Rams, Jacek Dziarmaga, Roeland Wiersema, William Bernoudy, Jack Raymond, Nitin Kaushal, Niclas Heinsdorf, Richard Harris, Kelly Boothby, Fabio Altomare, Mohsen Asad, Andrew J. Berkley, Martin Boschnak, Kevin Chern, Holly Christiani, Samantha Cibere, Jake Connor, Martin H. Dehn, Rahul Deshpande, Sara Ejtemaee, Pau Farre, Kelsey Hamer, Emile Hoskinson, Shuiyuan Huang, Mark W. Johnson, Samuel Kortas, Eric Ladizinsky, Trevor Lanting, Tony Lai, Ryan Li, Allison J. R. MacDonald, Gaelen Marsden, Catherine C. McGeoch, Reza Molavi, Travis Oh, Richard Neufeld, Mana Norouzpour, Joel Pasvolsky, Patrick Poitras, Gabriel Poulin-Lamarre, Thomas Prescott, Mauricio Reis, Chris Rich, Mohammad Samani, Benjamin Sheldan, Anatoly Smirnov, Edward Sterpka, Berta Trullas Clavera, Nicholas Tsai, Mark Volkmann, Alexander M. Whiticar, Jed D. Whittaker, Warren Wilkinson, Jason Yao, T. J. Yi, Anders W. Sandvik, Gonzalo Alvarez, Roger G. Melko, Juan Carrasquilla, Marcel Franz, and Mohammad H. Amin. Beyond-classical computation in quantum simulation. *Science*, 388(6743):199–204, 2025.

[186] A. Ciani and B. M. Terhal. Stoquasticity in circuit QED. *Phys. Rev. A*, 103:042401, Apr 2021.

[187] Dorit Aharonov, Alex B. Grilo, and Yupan Liu. StoqMA vs. MA: the power of error reduction, 2021. arXiv:2010.02835.

[188] Adam R. Klivans and Dieter van Melkebeek. Graph nonisomorphism has subexponential size proofs unless the polynomial-time hierarchy collapses. *SIAM Journal on Computing*, 31(5):1501–1526, 2002.

[189] Peter Bro Miltersen and N. V. Vinodchandran. Derandomizing Arthur–Merlin games using hitting sets. *Computational Complexity*, 14(3):256–279, Dec 2005.

[190] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. *J. ACM*, 69(4), August 2022.

[191] Elliott H. Lieb and Robert Seiringer. *The Stability of Matter in Quantum Mechanics*. Cambridge University Press, Cambridge, 2009.

[192] Cristina E. González-Espinoza, Paul W. Ayers, Jacek Karwowski, and Andreas Savin. Smooth models for the coulomb potential. *Theoretical Chemistry Accounts*, 135(12):256, Oct 2016.

[193] Sergey Bravyi, David P. DiVincenzo, and Daniel Loss. Schrieffer–Wolff transformation for quantum many-body systems. *Annals of Physics*, 326(10):2793–2826, 2011.

[194] Daniel Gottesman, Alexei Kitaev, and John Preskill. Encoding a qubit in an oscillator. *Phys. Rev. A*, 64:012310, Jun 2001.

[195] Sergey Bravyi and Matthew Hastings. On complexity of the quantum Ising model. *Communications in Mathematical Physics*, 349(1):1–45, 1 2017.

[196] Andrew Jackson, Theodoros Kapourniotis, and Animesh Datta. Accreditation of analogue quantum simulators. *Proceedings of the National Academy of Sciences*, 121(6):e2309627121, 2024.

[197] Stephen P. Jordan, David Gosset, and Peter J. Love. Quantum-Merlin-Arthur–complete problems for stoquastic Hamiltonians and Markov matrices. *Phys. Rev. A*, 81:032331, Mar 2010.

[198] Dominik Janzing and Pawel Wocjan. BQP-complete problems concerning mixing properties of classical random walks on sparse graphs, 2006. arXiv:quant-ph/0610235.

[199] Shu Nakamura and Yukihide Tadano. On a continuum limit of discrete Schrödinger operators on square lattice. *Journal of Spectral Theory*, 11(1):355–367, March 2021.

[200] Lov K. Grover. Synthesis of quantum superpositions by quantum computation. *Phys. Rev. Lett.*, 85:1334–1337, Aug 2000.

[201] M. Dimassi and J. Sjostrand. *Spectral Asymptotics in the Semi-Classical Limit*. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 1999.

[202] Bernard Helffer. *Semi-classical Analysis for the Schrödinger Operator and Applications*, volume 1336 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, Heidelberg, 1988.

[203] Shu Nakamura. A remark on eigenvalue splittings for one-dimensional double-well Hamiltonians. *Letters in Mathematical Physics*, 11(4):337–340, May 1986.

[204] 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.

[205] Daniel R. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997.

[206] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM Journal on Computing*, 47(3):982–1038, 2018.

[207] Rolando D. Somma, Daniel Nagaj, and Mária Kieferová. Quantum speedup by quantum annealing. *Phys. Rev. Lett.*, 109:050501, 7 2012.

[208] Andrew M. Childs, Richard Cleve, Enrico Deotto, Edward Farhi, Sam Gutmann, and Daniel A. Spielman. Exponential algorithmic speedup by a quantum walk. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 59–68, New York, NY, USA, 2003. Association for Computing Machinery.

[209] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 04 2014.

[210] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[211] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, August 1996.

[212] Takashi Yamakawa and Mark Zhandry. Verifiable quantum advantage without structure. *J. ACM*, 71(3), 6 2024.

[213] Scott Aaronson. How much structure is needed for huge quantum speedups?, 2022. arXiv:2209.06930.

[214] Matthew P. Harrigan, Kevin J. Sung, Matthew Neeley, Kevin J. Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Daniel Eppens, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Alan Ho, Sabrina Hong, Trent Huang, L. B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Martin Leib, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Charles Neill, Florian Neukart, Murphy Yuezhen Niu, Thomas E. O'Brien, Bryan O'Gorman, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Andrea Skolik, Vadim Smelyanskiy, Doug Strain, Michael Streif, Marco Szalay, Amit Vainsencher, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Leo Zhou, Hartmut Neven, Dave Bacon, Erik Lucero, Edward Farhi, and Ryan Babbush. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, 3 2021.

[215] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem, 2015. arXiv:1412.6062.

[216] Boaz Barak, Ankur Moitra, Ryan O'Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright. Beating the random assignment on constraint satisfaction problems of bounded degree. In Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 110–123, Dagstuhl, Germany, 2015. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[217] Joao Basso, David Gamarnik, Song Mei, and Leo Zhou. Performance and limitations of the QAOA at constant levels on large sparse hypergraphs and spin glass models. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 335–343, 2022.

[218] Edward Farhi, Sam Gutmann, Daniel Ranard, and Benjamin Villalonga. Lower bounding the MaxCut of high girth 3-regular graphs using the QAOA, 2025. arXiv:2503.12789.

[219] Stephen P. Jordan and Edward Farhi. Perturbative gadgets at arbitrary orders. *Phys. Rev. A*, 77:062329, 6 2008.

[220] Jacob D. Biamonte and Peter J. Love. Realizable Hamiltonians for universal adiabatic quantum computers. *Phys. Rev. A*, 78:012352, 7 2008.

[221] Dorit Aharonov and Amnon Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 20–29, New York, NY, USA, 2003. Association for Computing Machinery.

[222] Dominic W. Berry, Graeme Ahokas, Richard Cleve, and Barry C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Communications in Mathematical Physics*, 270(2):359–371, 3 2007.

[223] Dominic W. Berry, Andrew M. Childs, Yuan Su, Xin Wang, and Nathan Wiebe. Time-dependent Hamiltonian simulation with L^1-norm scaling. *Quantum*, 4:254, 4 2020.

[224] Sergey Bravyi and Matthew Hastings. On complexity of the quantum ising model, 2014. arXiv:1410.0703.

[225] Andrew M. Childs, Dmitri Maslov, Yunseong Nam, Neil J. Ross, and Yuan Su. Toward the first quantum simulation with quantum speedup. *Proceedings of the National Academy of Sciences*, 115(38):9456–9461, 2018.

[226] Wikipedia contributors. Weyl's inequality, 2025. [Online; accessed 25-January-2025].

[227] Joseph Frederick Elliott. *The characteristic roots of certain real symmetric matrices*. PhD thesis, University of Tennessee, Knoxville, 1953. `https://trace.tennessee.edu/utk_gradthes/2384/`.

[228] R.A. Usmani. Inversion of Jacobi's tridiagonal matrix. *Computers & Mathematics with Applications*, 27(8):59–66, 1994.

[229] Konrad Schmüdgen. *Unbounded self-adjoint operators on Hilbert space*, volume 265. Springer Science & Business Media, 2012.

[230] Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[231] Lin Lin. Dissipative preparation of many-body quantum states: Toward practical quantum advantage. *APL Computational Physics*, 1(1):010901, 2025.

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

[233] Elizabeth Crosson and Aram W. Harrow. Simulated quantum annealing can be exponentially faster than classical simulated annealing. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 714–723, 2016.

[234] Vasil S Denchev, Sergio Boixo, Sergei V Isakov, Nan Ding, Ryan Babbush, Vadim Smelyanskiy, John Martinis, and Hartmut Neven. What is the computational value of finite-range tunneling? *Physical Review X*, 6(3):031015, 2016.

[235] Frank H Clarke. *Optimization and nonsmooth analysis*. SIAM, 1990.

[236] H Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer Science & Business Media, 2011.

[237] E. Polak, J.E. Higgins, and D.Q Mayne. A barrier function method for minimax problems. `https://doi.org/10.1007/BF01586049`, 1992.

[238] Andrea Gavana. Global optimization benchmarks and AMPGO. `https://infinity77.net/global_optimization/index.html`.

[239] Luca Baronti and Marco Castellani. A Python benchmark functions framework for numerical optimisation problems, 2024. arXiv:2406.16195.

[240] Kayvan Nalaie, Toktam Saghafi, and Mohammad-R. Akbarzadeh-T. A report on numeric benchmark functions. `https://sciip2.um.ac.ir/images/260/content/Numeric%20Benchmark%20Functions.pdf`, 2015.

[241] Abdesslem Layeb. New hard benchmark functions for global optimization, 2022. arXiv:2202.04606.

[242] Maor Ivgi, Oliver Hinder, and Yair Carmon. DoG is SGD's best friend: A parameter-free dynamic step size schedule. In *International Conference on Machine Learning*, pages 14465–14499. PMLR, 2023.

[243] Ahmed Khaled, Konstantin Mishchenko, and Chi Jin. Dowg unleashed: An efficient universal parameter-free gradient descent method. *Advances in Neural Information Processing Systems*, 36:6748–6769, 2023.

[244] Jean Bourgain. On growth of Sobolev norms in linear schrödinger equations with smooth time dependent potential. *Journal d'Analyse Mathématique*, 77(1):315–348, 1999.

[245] Daniel Hsu, Sham Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(none):1 – 6, 2012.

[246] Andris Ambainis and Oded Regev. An elementary proof of the quantum adiabatic theorem, 2006. arXiv:quant-ph/0411152.

[247] Andrew M. Childs. Lecture notes on quantum algorithms, 2025. https://www.cs.umd.edu/~amchilds/qa/.