Provable Speedups in Convex Optimization via Quantum Dynamics

Shouvanik Chakrabarti[1]
Global Technology Applied Research,
JPMorganChase, New York, NY, USA

Dylan Herman[1]
Global Technology Applied Research,
JPMorganChase, New York, NY, USA

Jacob Watkins[1]
Global Technology Applied Research,
JPMorganChase, New York, NY, USA

Enrico Fontana
Global Technology Applied Research,
JPMorganChase, New York, NY, USA

Brandon Augustino
Global Technology Applied Research,
JPMorganChase, New York, NY, USA

Junhyung Lyle Kim
Global Technology Applied Research,
JPMorganChase, New York, NY, USA

Marco Pistoia
Global Technology Applied Research,
JPMorganChase, New York, NY, USA

Abstract

This work investigates the possibility of quantum speedups for continuous optimization through quantum Hamiltonian simulation. We establish the first rigorous query complexity bounds for unconstrained convex optimization via digital quantum annealing, based on the non-adiabatic Quantum Hamiltonian Descent (QHD) framework. In the process, we derive rigorous resource estimates for digital quantum simulation of Schrödinger operators depending only on input simulation parameters, given black-box evaluation access to separable, Lipschitz continuous potential $b(t)f(x)$. We apply these simulation bounds to assess the complexity of high-dimensional convex optimization.

Our annealing schedule achieves *arbitrarily fast* convergence rates in the evolution time, with computational time determined solely by the cost of discretization. We show that a G-Lipschitz convex function can be optimized to an error of ϵ with $\widetilde{\mathcal{O}}(d^{1.5}G^2R^2/\epsilon^2)$ queries, given a starting point that is Euclidean distance R from optimal. Under reasonable assumptions such as the complexity of simulating Schrödinger operators, we show that $\widetilde{\Omega}(d/\epsilon^2)$ queries are necessary. This suggests QHD does not offer improvements over classical methods in the noiseless zeroth order setting.

However, we show that the QHD algorithm can tolerate $\widetilde{\mathcal{O}}(\epsilon^3/d^{1.5}G^2R^2)$ noise in function evaluation, and as a result, provides a super-quadratic query advantage over the best existing noise-tolerant classical algorithms in the high-dimensional setting. We leverage this to design a quantum algorithm for stochastic convex optimization that offers a super-quadratic speedup over all known classical and quantum algorithms in this regime. To our knowledge, these results represent the first rigorous quantum speedups for convex optimization obtained through a dynamical algorithm.

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases Convex optimization, Hamiltonian simulation, zeroth order

Digital Object Identifier 10.4230/LIPIcs.TQC.2026.4

Related Version *Full Version*: `https://arxiv.org/abs/2503.24332` [27]

Acknowledgements The authors thank Aram Harrow, Jiaqi Leng, and Xiaodi Wu for insightful discussions about Quantum Hamiltonian Descent and the simulation of real-space dynamics. They also thank their colleagues at the Global Technology Applied Research center of JPMorganChase for their support and helpful discussions.

[1] These authors contributed equally and are listed in alphabetical order.

1 Introduction

Quantum computers have the potential to solve complex optimization problems more efficiently than classical methods. The degree of quantum algorithmic speedup in optimization has been the subject of many studies spanning over two decades (for recent surveys see [1, 30]). Interest in this topic is driven both by the mathematical richness of problems in the field and by the many practical applications to machine learning, statistics, finance, and engineering.

Perhaps the most natural optimization algorithms are those that directly simulate the continuous time dynamics of some physical system. These algorithms seek to leverage the fact that such systems experience forces in the direction of minimal potential energy or related quantities. If a system can be identified where these quantities correspond to an objective function, then the simulation of that system yields an optimization algorithm. These connections have been leveraged to great success in classical optimization theory [72, 76, 47]. Their most notable success has been in explaining the so-called *acceleration phenomenon*, [62, 58, 34] wherein certain finely-tuned gradient methods have been shown to achieve asymptotically faster convergence than standard gradient descent in a variety of domains.

From the point of view of quantum algorithms, the study of optimization through simulating quantum dynamics has been ongoing since the beginning of the field. Perhaps the most famous example is the quantum adiabatic algorithm for combinatorial optimization [37], which leverages *adiabaticity* [14]. This is the property in which slowly changing Hamiltonian systems remain in their ground state (optimum) even as the Hamiltonian varies. Due to the flexibility of this framework, the adiabatic algorithm and a heuristic variant called quantum annealing [46] have been heavily studied from both theoretical and empirical perspectives [18, 19, 36, 74, 63, 4, 6, 9, 13]. For a comprehensive treatment of adiabatic quantum computing, we refer the reader to [5] and the references therein. Other optimization algorithms were originally proposed in the discrete time regime but subsequently analyzed from a dynamical/optimal control perspective, such as the Quantum Approximate Optimization Algorithm (QAOA) [35, 77, 17, 75] and the Variational Quantum Eigensolver (VQE) [61, 53, 54, 29]. However, these algorithms are primarily intended for discrete optimization.

In contrast, this work considers the continuous setting, specifically unconstrained convex optimization.

▶ **Problem 1** (Unconstrained Convex Optimization). *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a convex*[1] *G-Lipschitz continuous function with global minimum $f_\star$ and set of global minimizers $\mathcal{X}_\star$. Given an input point x_0 such that $\ell_2(x_0, \mathcal{X}_\star) \leq R$, find a point $\tilde{x} \in \mathbb{R}^d$ such that*

$$f(\tilde{x}) - f_\star \leq \epsilon.$$

Thus, R bounds the known distance from solution, and ϵ sets the acceptable error tolerance. Our focus on convex optimization stems from both practical and theoretical considerations. Practical, because unconstrained convex optimization has plentiful applications to statistics and machine learning [16]. Theoretical, because the rigorous study of the complexity of dynamical algorithms for optimization is an emerging topic, and it is instructive to focus on a domain where classical complexities are well understood [59]. Unconstrained convex optimization provides such a domain, where the optimal query complexities in the standard regimes have been characterized for decades [57]. There is also extensive research activity on variants of this setting, such as erroneous oracular access to the function or its derivatives [71, 31, 11], stochastic optimization [24, 56, 32, 69], and bandit optimization [2, 3, 21, 68, 33].

[1] Our results also hold for *star*-convex functions, all other assumptions kept the same. We only require convexity conditions to hold between a fixed global minimizer $x_\star$ and any point y in the domain.

A dynamical approach for characterizing acceleration in convex optimization was introduced by Su, Boyd, and Candès [72], and subsequently generalized by Wibisono, Wilson, and Jordan [76]. In the latter work, the authors define a class of (classical) dynamical systems in terms of a so-called Bregman Lagrangian

$$\mathcal{L}(x, v, t) = e^{\alpha_t + \gamma_t} \left(D_h(x + e^{-\alpha_t} v, x) - e^{\beta_t} f(x) \right),$$

where x denotes the position, $v = \dot{x}$ denotes the velocity, and D_h is the Bregman divergence $D_h(x, y) \coloneqq h(x) - h(y) - \langle \nabla h(x), y - x \rangle$, where h is convex and essentially smooth. Later, Leng et al. [48] quantized these dynamics in the Euclidean setting $h(x) = \|x\|^2 / 2$, which they term *Quantum Hamiltonian Descent (QHD)*. These dynamics are described by a time dependent Schrödinger equation, with Hamiltonian

$$H_{\mathrm{QHD}}(t) \coloneqq -e^{\alpha_t - \gamma_t} \frac{1}{2} \Delta + e^{\alpha_t + \beta_t + \gamma_t} f(x). \tag{QHD}$$

In both the classical and quantum frameworks, $\alpha_t, \beta_t, \gamma_t$ are differentiable real-valued functions, which we call *schedules*. Wibisono et al. [76] show that if the schedules satisfy so-called *ideal scaling* conditions, then for convex f the classical dynamics converge to a global minimum at rate $O(e^{-\beta_t})$. Leng et al. [48] show a similar result for quantum dynamics. What is remarkable about these protocols is that they offer a *non-adiabatic* pathway to annealing-based optimization, whose convergence is relatively insensitive to both initial state and spectral properties of the Hamiltonian.

Ideal scaling schedules are invariant under a reparameterization of time, and in principle, the rate of convergence can be arbitrarily fast in t. Clearly, however, the query complexity of convex optimization cannot be made arbitrarily small, as this would contradict known lower bounds in the classical [57, 59] and quantum [26, 73, 40, 39] settings. Wibisono et al. [76] offer an explanation of this phenomenon by noting that while the convergence of continuous dynamics can be arbitrarily fast, algorithms for discretizing the dynamics become unstable beyond a speed limit. They additionally argue that the maximum speed where discretization is possible corresponds to the optimal (accelerated) classical query complexities. Leng et al. [48] hint at a similar speed/discretization tradeoff for quantum algorithms, but this tradeoff has yet to be theoretically characterized.

Our main technical contribution is to give a precise theoretical characterization of the dependence of the query complexity on schedules. This leads to rigorous bounds on the query complexity of convex optimization through the simulation of QHD dynamics. We identify precise speed-discretization tradeoffs and show that our results cannot be improved under believable assumptions about initial state and the complexity of simulating this class of Schrödinger operators (e.g., Assumption 9). As a consequence, we identify regimes where QHD simulation does not offer any speedup over comparable classical algorithms, and yet others where (perhaps surprisingly) large speedups over the best known classical algorithms are possible. The next section discusses our results in detail.

2 Input Model

We study an oracular problem in which the objective function f is accessed via evaluation ("zeroth-order") queries. Our analysis is primarily concerned with the query complexity, namely the number of calls to such oracles. Quantum computing allows for several different natural access models for f; here we define the ones of relevance to this work. The number of qubits, and hence precision to which x is encoded in a finite-dimensional basis state, will be specified when necessary.

▶ **Definition 2** (ϵ_f-accurate evaluation oracle). *Let $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subseteq \mathbb{R}^d$. The unitary O_f is an ϵ_f-accurate evaluation, or binary, oracle for f if for all computational basis states $|x\rangle |y\rangle$,*

$$O_f |x\rangle |y\rangle = |x\rangle \left| y \oplus \widetilde{f}(x) \right\rangle,$$

and $\|\widetilde{f}(x) - f(x)\|_\infty < \epsilon_f$.

▶ **Definition 3** (ϵ_p-accurate phase oracle). *For $\mathcal{X} \subseteq \mathbb{R}^d$, let $f : \mathcal{X} \to \mathbb{R}$ be a bounded function with bound $\Lambda \geq \|f\|_\infty$. The unitary O_f is an ϵ_p-accurate phase oracle for f if for all computational basis states $|x\rangle$,*

$$O_f |x\rangle = e^{i\tilde{h}(x)} |x\rangle,$$

where $\|\tilde{h} - f/\Lambda\|_\infty < \epsilon_p$.

Conversions from a binary to phase oracle can be made via phase kickback, and the reverse conversion can be done with quantum phase estimation. Finally, we will make use of a "qsample" oracle for the setting of stochastic convex optimization.

▶ **Definition 4** (quantum stochastic oracle). *Let f be a real-valued function with domain $\mathcal{X} \subseteq \mathbb{R}^d$. For each $x \in \mathcal{X}$, let p_x be a probability density on $\mathbb{R}$ with zero mean and bounded variance. The unitary O_f is an ϵ_f-accurate stochastic oracle for f if it acts (in superposition) on a joint register $|x\rangle |y\rangle$ of domain points $x \in \mathcal{X}$ and values y as*

$$O_f |x\rangle |y\rangle = |x\rangle \int \sqrt{p_x(\xi)} |y \oplus (f(x) + \xi)\rangle d\xi. \tag{1}$$

In many ways, the quantum stochastic oracle above is a natural generalization of a stochastic classical oracle, in which $f(x) + \xi$ is returned with probability $p_x(\xi)$. For example, see [70] for a comparable setup in the first-order setting. However, we acknowledge that, in some respects, the quantum and classical oracles are less directly comparable than for more familiar deterministic evaluation oracles. For example, it's unclear how O_f could be constructed "on the fly" using a stream of classically random samples, given the need for superposition. On the other hand, if the data is loaded into quantum read-only memory (QROM), then the qsampling of this data can be done in analogous fashion to the classical case. See Remark 13 for a discussion of the consequences on query complexity, if we only have classical access to the noise.

3 Results

We present our results in two parts. First, we informally state our main theorem regarding the simulation of Schrödinger operators, given evaluation access to a potential f. Then, based on these resource estimates, we discuss results for optimization using the QHD framework.

3.1 Real space Hamiltonian simulation

Classically, the simulation of dynamical systems may be performed using various numerical integration techniques [42, 12]. In the case where (closed) quantum dynamics are simulated on a quantum computer, these integrators correspond to algorithms for *time dependent Hamiltonian simulation*, which synthesize the unitary time evolution operator as a quantum circuit. Such algorithms are well-studied in the discrete setting. A careful study of real

space simulation complexity has only been recently undertaken [28], though this analysis focuses primarily on Fourier truncation errors. In order to characterize the query complexity of optimization with QHD, we require optimized and explicit resource estimates for the complexity of simulation with black box potentials, making as few assumptions as possible and accounting for all sources of error.

We present and analyze a quantum algorithm for real-space simulation based on a pseudo-spectral method known as collocation. The main novelty of the analysis is handling the dynamical propagation of discretization error. Existing results on Schrödinger operators and partial differential equations (PDEs) are often tailored to contexts in mathematical physics, e.g. applicable only to a few spatial dimensions, and are thus not immediately useful for complexity-theoretic statements. Our work fills this gap and provides a *computational* result regarding the digital simulation of Schrödinger equations in Euclidean space, which can be applied out-of-the-box. In particular, the result applies when the dimension of the problem is taken to be an asymptotic parameter, which is of crucial importance to high-dimensional settings such as optimization.

The following informally summarizes our main simulation theorem.

▶ **Theorem 5** (Informal version of Thm 4.2 in Main Version [27]). *Consider the Schrödinger equation*

$$i\partial_t \Phi(x,t) = [-a(t)\Delta + b(t)f(x)]\Phi(x,t),$$

subject to initial data $\Phi(x,0) = \Phi_0(x)$*, where* a, b *are sufficiently smooth functions of time and* f *is a* G*-Lipschitz function with bound* $\Lambda_f \geq \|f\|_\infty$*, accessed through a* ϵ_f*-noisy binary quantum oracle* O_f*. If* $\epsilon_f = \widetilde{\mathcal{O}}(\epsilon/\|b\|_1)$*, then there is a digital quantum algorithm that outputs an* ϵ*-approximation* $|\Psi_t\rangle$ *to* $\Phi(\cdot,t)$ *(in an appropriate sense defined later) using* $\mathcal{O}\left(\Lambda\|b\|_1\right)$ *queries to* O_f*,* $\mathcal{O}\left(d^2 \ \mathrm{polylog}(1/\epsilon, \|a\|_1, \|b\|_1)\right)$ *qubits and* $\widetilde{\mathcal{O}}\left(\mathrm{poly}(d, \Lambda, \|b\|_1)\right)$ *gates.*

To the best of our knowledge, this is the first rigorous result on the quantum computational complexity of real-space quantum dynamics given solely in terms of basic simulation parameters. We prove Theorem 5 using multi-dimensional Fourier analysis and bounding various sources of error that typically occur in signal processing applications, such as spectral truncation and aliasing. These errors are connected to the so-called Sobolev norms of the functions involved, and occur because the digital quantum algorithm is unable to fully represent the infinite spectrum of the continuous problem. Our results apply to separable potentials $b(t)f(x)$ where b is differentiable and f is G-Lipschitz. These weak assumptions on smoothness are possible through a mollification argument, a common technique in signal processing and functional analysis. Further details are provided in Section 4.2 and Appendix A. As these simulation results are concerned with general classes of Schrödinger dynamics, they may be of interest beyond optimization.

3.2 Convex optimization

Our main theorem on unconstrained convex optimization via Hamiltonian simulation is as follows.

▶ **Theorem 6.** *There exists a quantum algorithm that solves Problem 1 using* $\widetilde{\mathcal{O}}\big(d^{3/2}(GR/\epsilon)^2\big)$ *queries to an evaluation oracle for* f *(Definition 2) with error* $\epsilon_f = \widetilde{\mathcal{O}}\left(\frac{\epsilon^3}{d^{3/2}G^2R^2}\right)$.

In Section 4.3, we prove this result as a simple consequence of a more general theorem. A full proof of the general result is given in the Section 5 of the Full Version [27], but here we outline a sketch of the derivation. With the simulation results of Theorem 5, we can calculate the

query complexity of optimization via QHD once we provide convergence results in continuous time. Theorem 14 provides this convergence under our modified ideal scaling conditions, which relaxes the conditions on the potential from being differentiable, as in [48], to merely Lipschitz continuous. We note that this relaxation is for *any* schedule that obeys the ideal scaling conditions. Since our goal is to completely determine the query complexity, we keep track of (possibly dimension-dependent) factors in addition to the convergence rate. Our complexity scales as $d^{3/2}$ instead of d because of a $\sqrt{d}$ factor that arises from the mismatch between the $p = \infty$ and $p = 2$ norm. We expect this may be improved with refined analysis.

Our analysis considers a modified parametrization of the QHD Hamiltonian from (QHD). In what follows we write

$$H_{\mathrm{QHD}}(t) = c_t \left(\frac{\hat{p}^2}{2m_t} + m_t \omega_t^2 f(\hat{x}) \right), \tag{2}$$

where m_t, ω_t are positive differentiable functions satisfying the ideal scaling conditions

$$\frac{\dot{m}_t}{m_t} = \lambda c_t, \quad \frac{\dot{\omega}_t}{\omega_t} \leq \frac{1}{2} \lambda c_t \tag{Ideal Scaling}$$

for some $\lambda \in \mathbb{R}_+$. Besides providing greater physical intuition, this parametrization also ensures dimensional consistency, whereas prior expressions for Ideal Scaling left this point unclear. Evidently, the parameter $c_t = e^{\alpha_t}$ merely plays the role of time parametrization. While such parametrization can affect algorithmic performance, the time dependent simulation algorithm we employ, based on the Dyson series method, will be robust to such reparametrizations.

Note that the query complexity in Theorem 6 has a dependence of $1/\epsilon^2$ on the precision, which is due to the query cost of discretization. In fact, the same complexity can be obtained from a schedule with *arbitrarily fast* convergence (and consequently, arbitrarily small simulation time). As specific examples, we now define schedules with arbitrary exponential or polynomial convergence rates. These schedules were proposed in [76] and considered in the quantum setting in [48].

▶ **Definition 7** (Exponential Schedules with Convergence Rate e^{-ct})**.** *Let $c \geq 0$ be a parameter that determines the rate of convergence, and $m_0, \omega_0 \in \mathbb{R}_+$ be arbitrary. A family of QHD Hamiltonians with exponential convergence rate e^{-ct} is defined by schedules $c_t = c$, $m_t = m_0 e^{ct}$, and $\omega_t = \omega_0 e^{ct/2}$. The resulting Hamiltonian is*

$$H_c(t) = c \left(-e^{-ct} \frac{1}{2m_0} \Delta + e^{2ct} m_0 \omega_0^2 f \right) \tag{QHD-exponential}$$

for $t \in [0, \infty)$.

▶ **Definition 8** (Polynomial Schedules with Convergence Rate t^{-k})**.** *Let $k > 0$ be a parameter that determines the rate of convergence, and $m_0, \omega_0, t_0 \in \mathbb{R}_+$ be arbitrary. A family of QHD Hamiltonians with polynomial convergence rate t^{-k} is defined by schedules $c_t = k/t$, $m_t = m_0 (t/t_0)^k$, and $\omega_t = \omega_0 (t/t_0)^{k/2}$. The resulting Hamiltonian is*

$$H_k(t) = \frac{k}{t} \left(-(t/t_0)^{-k} \frac{1}{2m_0} \Delta + (t/t_0)^{2k} m_0 \omega_0^2 f \right) \tag{QHD-polynomial}$$

for $t \in [t_0, \infty)$.

The rate of convergence for these schedules, here provided within the definitions for clarity, are simple corollaries of Theorem 14. We also show how any ideal-scaling schedule can be discretized to obtain the query complexity stated in Theorem 6. Finally, we note that, while

the choice of convergence rate is relatively arbitrary and leads to the same query complexity, the remaining parameters of the schedule such as m_0, ω_0 must be chosen carefully. This is because the convergence of QHD is shown via a Lyapunov ("nonincreasing") function $\mathcal{E}_t$ which loosely tracks the "energy" of the system. The convergence bounds are directly proportional to $\mathcal{E}_0$, which may be dimension dependent. In fact, using QHD schedules already stated in previous works [48, 49] result in a dimension dependence of $\widetilde{\mathcal{O}}(d^{2.5})$ in the query complexity, compared to the $\widetilde{\mathcal{O}}(d^{1.5})$ obtained here.

3.2.1 Schedule invariance and lower bounds on query complexity

We now consider the possibility of improvements over our upper bounds within the QHD framework. The algorithmic results we provide are independent of ideal-scaling schedule, but perhaps there are optimizations in terms of the initial parameters, the initial state, or in terms of the domain of simulation. As discussed in the previous section, we can choose schedules that converge in continuous time at arbitrarily fast rates, and thus simulate for an arbitrarily short amount of time to reach the target precision. Lower bounds on the query complexity arise from the increasing cost of simulating these arbitrarily fast schedules. We make the following assumption on the complexity of simulating Schrödinger Dynamics.

▶ **Assumption 9** (No Fast-Forwarding for Time Dependent Schrödinger Operator)**.** *A potential $f : \mathbb{R}^d \to \mathbb{R}$ satisfies the "no fast-forwarding" assumption if simulating the time dependent Hamiltonian $H(t) = -a(t)\Delta + b(t)f(x)$ for $t \in [t_1, t_2]$ requires $\Omega\left(\Lambda_f \cdot \int_{t_1}^{t_2} |b(t)|\, dt\right)$ evaluation queries to f, where $\Lambda_f \geq 0$ is some functional of f, independent of t, which satisfies $\Lambda_{cf} \geq |c|\, \Lambda_f$ for any $c \in \mathbb{R}$.*

We call this assumption "no fast-forwarding" since, with constant $b(t)$, it reduces to the familiar notion for finite-dimensional, time-independent Hamiltonians, and because it is the most reasonable and natural generalization of this notion. It is possible in principle that, by restricting to convex f and for some particular choices of schedules $b(t)$, this assumption may be violated. However, we argue that it is very likely to be true for simulation schemes that work in a black box setting, for several reasons. First, the assumption is satisfied by all general algorithms for time-dependent Hamiltonian simulation of Schrödinger operators that we are aware of, including the one in this paper. Second, this dependence is inherent in the methods for time-dependent Hamiltonian simulation via the interaction picture, that these works are based on. Finally, Assumption 9 reflects intuition in the following (non-rigorous) sense: in each time slice $[t, t + dt)$ where dt is small enough that $b(t)$ is nearly constant, it follows from the no fast forwarding theorem for time independent potentials that the cost of simulating the Schrödinger operator in this time slice is at least $b(t)f dt$. Approximating the whole evolution by these time slices results in the dependence encoded in Assumption 9.

We are now ready to state our lower bounds on the query complexity of convex optimization by simulating QHD.

▶ **Theorem 10.** *Let f be a potential satisfying the no fast-forwarding assumption (Assumption 9), and the conditions of Problem 1. Then any optimization algorithm that solves Problem 1 via the simulation of QHD under ideal scaling conditions, starting from a real initial state, must make $\Omega(d\Lambda_f GR/\epsilon^2)$ queries to an evaluation oracle for f, if the continuous-time convergence is as predicted by Theorem 14.*

The additional $\sqrt{d}$ factor in our upper bounds (Theorem 6) arises from a bound on Λ_f over the hypercube. Together, our upper and lower bounds almost completely characterize the query complexity of convex optimization with QHD.

We point out that we make no assumptions beyond Lipschitz continuity of the convex functions considered in Problem 1. As such, Problem 1 gives rise to the "nonsmooth" setting in convex optimization theory [65]. It is also common to study settings with more assumptions on the function class. Concretely, it is standard to additionally assume Lipschitz continuity of gradients (this is often referred to as "smoothness" in optimization theory [59]),[2] and strong convexity. It is known that gradient methods are more efficient when we assume Lipschitz continuous gradients, or Lipschitz continuous gradients and strong convexity. In contrast, our results are equally applicable to all the settings and only require the assumptions corresponding to the nonsmooth setting.

3.2.2 Prospects for quantum speedup for convex optimization

We now compare our query complexity bounds to classical algorithms for unconstrained optimization. Our results are summarized in Table 1.

Comparison to classical algorithms for noiseless convex optimization

The first-order query complexity for the nonsmooth, smooth, and smooth and strongly convex settings is well-understood [20, 59]. When f is G-Lipschitz, one can determine an ϵ-approximate minimizer using $\mathcal{O}((GR/\epsilon)^2)$ queries. If f has L-Lipschitz continuous gradients the bound improves to $\mathcal{O}(LR^2/\epsilon)$, and when f is additionally μ-strongly convex the number of queries is at most $\mathcal{O}(\kappa \log(1/\epsilon))$, where $\kappa := L/\mu$. Note that these bounds are for *unaccelerated* methods. If one employs Nesterov's acceleration [58], these results improve to $\mathcal{O}(R\sqrt{L/\epsilon})$ for L-smooth functions, and $\mathcal{O}(\sqrt{\kappa}\log(1/\epsilon))$ for L-smooth, μ-strongly convex functions.

The *zeroth-order* complexity in these settings has only been established more recently. Most of the results along this line concern algorithms that construct gradient estimators by evaluating the function value at a fixed number of random points. The number of permitted points may vary, though the most frequent setups involve 2-point estimators [60, 33, 68] and single-point estimators [38, 3, 10, 11, 22]. Compared to the first-order query complexities, the upper bounds obtained for zeroth-order optimization suffer a poly(d) slow-down [41]. Specifically, for our setting of nonsmooth optimization, classical zeroth-order methods achieve an oracle complexity $\mathcal{O}\left(d(GR/\epsilon)^2\right)$, see, e.g., [33]. In comparison, we show query upper bounds of $\widetilde{\mathcal{O}}\left(d^{1.5}(GR/\epsilon)^2\right)$, and if the fast forwarding assumption holds and the continuous time convergence rates match those of QHD (Theorem 14), a lower bound of $\widetilde{\Omega}\left(d(GR/\epsilon)^2\right)$ which matches the classical upper bound. Thus, if the potential corresponding to the objective function satisfies Assumption 9, and if the QHD convergence bound of Theorem 14 is sharp in general, *there is no quantum speedup for zeroth order noiseless convex optimization by simulating QHD.*

Comparison to classical algorithms for noisy convex optimization

While our results for noiseless convex optimization are largely negative, we have a more positive outlook for the case of convex optimization with noisy oracles. By noisy, we assume that evaluations of f are promised to be accurate to precision ϵ_f (see for instance Definition 2). We do not assume that ϵ_f follows some probability distribution, and we will continue to use the term "noisy" for this general setting.

[2] Note that this definition of smoothness is incomparable to the typical one in analysis, where smoothness implies infinite differentiability.

Table 1 Complexity of algorithms for noisy unconstrained zero-order convex optimization. Admissible noise is an upper bound on ϵ_f, the error in the evaluation of f using a binary oracle (see Definition 2).

Classical Algorithms	Reference	Query complexity	Admissible noise
Stochastic Gradient Estimator	[64]	$\widetilde{\mathcal{O}}\left(d^4(GR/\epsilon)^6\right)$	$\max\left\{\frac{\epsilon^2}{\sqrt{d}GR}, \frac{\epsilon}{d}\right\}$
Simulated annealing	[11]	$\widetilde{\mathcal{O}}\left(d^{4.5}\right)$	ϵ/d
Quantum Algorithms	**Reference**	**Query complexity**	**Admissible noise**
Quantum simulated annealing	[51]	$\widetilde{\mathcal{O}}\left(d^3\right)$	ϵ/d
QHD simulation (this paper)	Theorem 6	$\widetilde{\mathcal{O}}\left(d^{1.5}(GR/\epsilon)^2\right)$	$\widetilde{\mathcal{O}}\left(\frac{\epsilon^3}{d^{1.5}G^2R^2}\right)$
Quantum subgradient method	[7]	$\widetilde{\mathcal{O}}\left((GR/\epsilon)^2\right)$	$\mathcal{O}\left(\frac{\epsilon^5}{d^{4.5}G^4R^4}\right)$

The best known classical algorithms with erroneous oracles are due to Belloni et al. [11] and Ristetski and Li [64]. We discuss both results, since each presents advantages over the other. [11] presents an algorithm based on simulated annealing using a hit and run walk, which requires $\widetilde{\mathcal{O}}(d^{4.5}\log(1/\varepsilon))$ evaluations of the objective, with error $\mathcal{O}(\varepsilon/d)$. On the other hand, [64] presents an algorithm based on a stochastic gradient estimator due to [38]. Their algorithm requires $\mathcal{O}(d^4/\varepsilon^6)$ evaluations of the objective with error at most $\mathcal{O}(\max(\varepsilon^2/\sqrt{d}, \varepsilon/d))$. We note that Ristetski and Li only show the query complexity of their algorithm to be polynomial scaling in d, ε but do not determine the polynomial. To make an accurate comparison, we determine this scaling in Appendix C.

In contrast, our algorithm is shown to require $\mathcal{O}(d^{1.5}/\varepsilon^2)$ quantum evaluation queries, each with error $\widetilde{\mathcal{O}}(\varepsilon^3/d^{1.5})$. We note that if $\varepsilon = \Theta(d^{-\varrho})$ for some $\varrho > 0$, our query complexities are polynomially lower than the classical algorithms. Up to polylogarithmic factors, the exponent of the speedup over [11] is $1/3 + 4\varrho/9$ and that over [64] is $(3 + 4\varrho)/(8 + 12\varrho)$. Therefore, if ϱ is small, the speedup is super-quadratic over both classical algorithms. We note that the classical algorithms are robust to higher levels of noise than ours, so it is natural to ask if they can be improved if a lower noise threshold is taken into account. We argue in Appendix D that such an improvement is not immediate from existing considerations. It is of course possible that such an improvement can be attained with more effort as there are few applicable lower bounds in general, so it is important to note that our claimed speedups are over *best known* classical algorithms, and not the best possible ones.

Comparison to quantum algorithms for noisy convex optimization

In the quantum model of computation, the best known algorithms in the noisy setting are attributed to Li and Zhang [51] and Augustino et al. [7]. The algorithm in [51] can be viewed as a quantum analogue of the simulated annealing algorithm from Belloni et al. [11] in that the classical hit and run walk is replaced with a quantum hit and run walk originally proposed in [25]. The quantum hit and run walk requires quadratically fewer samples per iteration compared to its classical counterpart, which improves the overall query complexity from $\widetilde{\mathcal{O}}(d^{4.5}\log(1/\varepsilon))$ to $\widetilde{\mathcal{O}}(d^3\log(1/\varepsilon))$. As in [11], their algorithm can tolerate error up to $\mathcal{O}(\varepsilon/d)$.

Augustino et al. [7] introduce a zero-order quantum subgradient method, wherein the subgradients are estimated via a generalization of Jordan's algorithm for gradient estimation [45] to subgradient estimation [73, 26]. The quantum subgradient method proposed in [7] achieves a query complexity of $\widetilde{\mathcal{O}}((GR/\epsilon)^2)$, which is exponentially faster than all known zero-order methods in the problem dimension. This algorithm works in both the noiseless and noisy settings, tolerating function evaluation error up to $\mathcal{O}(\epsilon^5/(d^{4.5}G^4R^4))$ in the latter.

These results and our analysis highlight a tradeoff between the query complexity and the robustness of quantum algorithms for noisy convex optimization. For the high-dimensional regime where $d \gg \text{poly}(G, R, \epsilon^{-1})$, it is clear that the quantum subgradient method is the best performing algorithm with respect to query complexity, but is the least robust in terms of the amount of noise it can tolerate in the function evaluations. Meanwhile, QHD achieves a superior query complexity compared to quantum simulated annealing [51], and is significantly more robust than the quantum subgradient method [7].

Comparison to algorithms for stochastic convex optimization

QHD's combination of the superior query complexity compared to more robust algorithms, and higher noise tolerance compared to faster algorithms, leads to an additional speedup for the problem of *unconstrained stochastic convex optimization*.

▶ **Problem 11** (Unconstrained Stochastic Convex Optimization). *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a (star) convex G-Lipschitz continuous function with global minimum $f_\star$ and set of global minimizers $\mathcal{X}_\star$. Suppose f is accessed via a random oracle (quantum or classical) which, given input x, returns $f(x) + \xi$ where ξ has zero mean and bounded variance. Given an input point x_0 such that $\ell_2(x_0, \mathcal{X}_\star) \leq R$, an algorithm solves the stochastic convex optimization problem if it returns a (random) point $\tilde{x} \in \mathbb{R}^d$ such that*

$$\mathbb{E}[f(\tilde{x})] - f_\star \leq \epsilon,$$

where $\mathbb{E}$ is taken over the randomness of the i.i.d. oracle values.

The following result is an immediate consequence of combining the result in Theorem 6 with quantum mean estimation [55].

▶ **Theorem 12.** *There exists a quantum algorithm that solves Problem 11 using $\widetilde{\mathcal{O}}\left(d^3(GR/\epsilon)^5\right)$ queries to a stochastic evaluation oracle for f in Definition 4.*

A proof is provided in Section 6 of the full version [27], and can be informally summarized as follows. Our algorithm for convex optimization evaluates f up to error $\widetilde{\mathcal{O}}\left(\frac{\epsilon^3}{d^{3/2}G^2R^2}\right)$. On the other hand, Montanaro [55] proves that if one has access to an oracle O_f for evaluating f up to error ϵ_f, then $\widetilde{\mathcal{O}}(1/\epsilon_f)$ applications of O_f suffice to estimate $\mathbb{E}[f(x)]$.

Accordingly, the performance of QHD is favorable over the simulated annealing algorithm of Belloni et al. [11] and the quantum algorithms [51, 7] in the zero-order stochastic setting, as highlighted in Table 2. Indeed, QHD achieves a super-quadratic speedup in the dimension over the algorithm in [11], and a sub-quadratic speedup in d over quantum simulated annealing [51]. The query complexity of QHD in this setting also outperforms that of the quantum subgradient method [7] in every parameter. We further remark that the dependence on dimension is close to optimal, due to an $\Omega(d^2(GR/\epsilon)^2)$ lower bound from Shamir [67].

Table 2 Complexity of algorithms for unconstrained zero-order stochastic convex optimization. See Problem 11 for a statement of the problem .

Classical Algorithms	Reference	Query complexity
Simulated annealing	[11]	$\widetilde{\mathcal{O}}\left(d^{7.5}(GR/\epsilon)^2\right)$
Quantum Algorithms	**Reference**	**Query complexity**
Quantum simulated annealing	[51]	$\widetilde{\mathcal{O}}\left(d^5 GR/\epsilon\right)$
QHD simulation (this paper)	Theorem 12	$\widetilde{\mathcal{O}}\left(d^3(GR/\epsilon)^5\right)$
Quantum subgradient method	[7]	$\widetilde{\mathcal{O}}\left(d^{4.5}(GR/\epsilon)^7\right)$

▶ Remark 13. As discussed earlier, the assumption that qsamples are available for the noise ξ may not be justified in all settings. In case we have only a classical sampler for ξ, we must resort to classical mean estimation to produce a noisy oracle. In this case, QHD simulation results in a query complexity of $\widetilde{\mathcal{O}}\left(d^{4.5}\left(\frac{GR}{\epsilon}\right)^8\right)$ whereas the quantum subgradient method incurs a query complexity of $\widetilde{\mathcal{O}}\left(d^9\left(\frac{GR}{\epsilon}\right)^{12}\right)$. QHD simulation still provides the fastest stochastic convex optimization in this regime, but this regime is much more limited than under qsample access for the noise. The quantum subgradient method no longer provides any speedup.

Related work

Several recent works also consider the use of Hamiltonian dynamics for continuous optimization without adiabaticity. For example, [23] considers several approaches for quantum dynamics-based optimization, including a similar time-dependent Hamiltonian scheme (see Sec. IV.B and V.B therein). While similar, our work has slightly different emphasis and scope. In [23], the authors analyze local convergence in nonconvex optimization by analyzing quadratic approximations to local minima. Although the assumption of convexity is dropped, this analysis tacitly assumes f is twice differentiable and strongly convex in a neighborhood of local optima. In contrast, our work only applies in the convex setting, but makes global convergence claims (via a Lyapunov argument) and only requires f be Lipschitz continuous. Overall, we believe our work and that of [23] both shed light on the potential for quantum-based optimization beyond the computation of gradients of f.

4 Overview of Methods

4.1 Lyapunov convergence

Our first result proves the continuous time convergence of QHD under ideal scaling conditions. The proof is a slight strengthening of [48, Theorem 1]. In addition to the generalization of ideal scaling, we remark on two additional points: firstly, we retain the pre-factor of the error dependence in addition to the convergence rate. This is crucial in order to determine the dimension dependence of the final query complexity (as we do in the next section). Secondly, we show that the convergence requires only the Lipschitz continuity of the potential, not necessarily differentiability.

▶ **Theorem 14** (QHD convergence for Lipschitz Potentials). *Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex and Lipschitz continuous, with global minimum $f_\star$ and minimizer $x_\star \in \mathcal{X}_\star$. Let $H_{\mathrm{QHD}}(t)$ be the QHD Hamiltonian with schedules c_t, m_t, ω_t obeying the ideal scaling conditions (Ideal Scaling). Then, for any $\Phi_t \in \mathrm{Dom}(-i\nabla, x, f(x))$ solving the Schrödinger equation with Hamiltonian H_{QHD} and initial condition Φ_0,*

$$\mathbb{E}_{|\Phi_t|^2} f(x) - f_\star \leq \mathcal{E}_0 \omega_t^{-2},$$

where

$$\mathcal{E}_0 \coloneqq \frac{1}{2}\langle(-im_0^{-1}\nabla + \lambda(x - x_\star))^2\rangle_{\Phi_0} + \omega_0^2\langle f(x) - f_\star\rangle_{\Phi_0}.$$

See Appendix B for a proof. Theorem 14 implies that QHD for nonsmooth convex optimization converges at an arbitrarily fast rate in t. To see this, suppose that $r(t)$ is a monotonically increasing function such that $e^{-r(t)}$ is the desired convergence rate. Simply set $\omega_t^2 = e^{r(t)}$ and c_t, m_t according to ideal scaling with equality.

4.2 Hamiltonian simulation

Implementing QHD can be abstracted as the task of simulating $\Phi(x,t) \in \mathcal{H} = L_2([-\frac{1}{2}, \frac{1}{2}]^d)$ on a Euclidean space, evolved under a kinetic-plus-potential type Hamiltonian

$$i\partial_t \Phi(x,t) = [-a(t)\Delta + V(x,t)]\Phi(x,t) = H(t)\Phi(x,t) \qquad \text{(Real Space Schrödinger)}$$

from some initial state $\Phi(x,0) = \Phi_0(x)$ under periodic boundary conditions (PBC). Particularly, the QHD case only considers separable potentials of the form $V(x,t) = b(t)g(x)$. Here g is a natural restriction of some Lipschitz function f to the hypercube $S \coloneqq [-\frac{1}{2}, \frac{1}{2}]^d$, defined by $g(x) \coloneqq f(x_0 + 2Rx)$. For the case of QHD, f is the function we want to optimize.

The first step is to approximately map the spatially-continuous evolution onto a qubit Hamiltonian, which is done via a *spectral method*. Spectral methods begin by imposing some truncation scheme on the solution, such as

$$\Psi(x,t) = \sum_{\mathbf{n} \in \mathcal{N}} \widehat{\Psi}_{\mathbf{n}}(t)\chi_{\mathbf{n}}(x) \tag{3}$$

and stipulating some conditions on the modified coefficients $\widehat{\Psi}_{\mathbf{n}}$ so that Ψ approximates Φ for the observable of interest. Here $\mathcal{N} \coloneqq \{-N, -N+1, \ldots, N-2, N-1\}^d$, $\chi_{\mathbf{n}}(x) = e^{i2\pi\mathbf{n}\cdot x}$ with $\mathbf{n} \in \mathbb{Z}^d$. We let $\mathcal{H}_N \coloneqq \mathrm{span}\{\chi_{\mathbf{n}}\}_{\mathbf{n}\in\mathcal{N}}$.

One popular and practical pseudo-spectral method is the so-called *collocation* method, which seeks to replicate the Schrödinger equation exactly on a predefined set of grid points $\mathcal{G} \subset S$. The collocation method seeks a $\Psi \in \mathcal{H}_N$ such that the residual is zero on $\mathcal{G}$:

$$[(i\partial_t - H(t))\Psi](x,t) = 0, \forall x_{\mathbf{j}} \in \mathcal{G}. \tag{4}$$

We take a uniform rectangular grid $\mathcal{G} = \{x_{\mathbf{j}}\}$ such that $x_{\mathbf{j}} = \mathbf{j}/2N,\ \mathbf{j} \in \mathcal{N}$. Then one can show [27, Lemma 4.1] that the solution to (4) can be obtained by simulating the following $d\log(N)$-qubit quantum Hamiltonian:

$$\widehat{H}(t) = a(t)(2\pi)^2 \mathsf{QFT}^{\otimes d}\hat{\mathbf{n}}^2\mathsf{QFT}^{\dagger\otimes d} + b(t)G = K(t) + V(t), \tag{5}$$

where $G = \sum_{\mathbf{j}\in\mathcal{N}} g(x_{\mathbf{j}}) |\mathbf{j}\rangle \langle\mathbf{j}|$ is the discretized potential. The second component is the time-dependent digital simulation of (5), which is performed utilizing the interaction picture (IP) and rescaled Dyson series approach of [52].

Let U_K stand for the unitary propagators under $K(t)$. Given a digital Schrödinger wavefunction $|\Psi_t\rangle$ generated by $\widehat{H}(t)$, we define the $\left|\Psi_t^K\right\rangle$ in the IP as

$$\left|\Psi_t^K\right\rangle \coloneqq U_K^\dagger(t,0)\left|\Psi_t\right\rangle. \tag{6}$$

From here on we drop the 0 argument to U_K. It can be checked that

$$i\partial_t\left|\Psi_t^K\right\rangle = H_I(t)\left|\Psi_t^K\right\rangle, \quad H_I(t) \coloneqq U_K^\dagger(t)V(t)U_K(t). \tag{7}$$

By solving (7), the original solution of interest $|\Psi_t\rangle$ can be obtained by applying $U_K(t)$, which is generally presumed efficient, and is indeed so in our case. In fact,

$$U_K(t) = \mathsf{QFT}^{\otimes d}\exp\left\{-i(2\pi\hat{\mathbf{n}})^2\int_0^t a(\tau)d\tau\right\}\mathsf{QFT}^{\dagger\otimes d} \tag{8}$$

which can be implemented using d QFTs and inverses, along with standard techniques for diagonal phase operations. The rescaled approach also includes a time-reparameterization $t \mapsto \xi(t) = \int_0^t \Lambda b(\tau)d\tau$ applied to $H_I(t)$, which results in

$$\widetilde{H}_I(t) = U_K^\dagger(t)(G/\Lambda)U_K(t), \tag{9}$$

where $\Lambda \geq \|g\|_\infty$. This acts as a preconditioning step. Following [52], a truncated Dyson series (Trotterization plus linear combination of unitaries) approach is used to simulate the evolution under $\widetilde{H}_I(t)$, which can be done with $\widetilde{O}(\Lambda T)$ queries to a Ham-$T(\widetilde{H}_I)$ oracle.

We rigorously show (see proof of [27, Theorem 4.2]) that under the resources specified in Theorem 5, the digital state outputted by our algorithm, $|\Psi_T\rangle$, satisfies:

$$|\langle\Psi_T|G|\Psi_T\rangle - \langle\Phi(\cdot,T)|g|\Phi(\cdot,T)\rangle| \leq \|f\|_\infty\epsilon, \tag{10}$$

where $\Phi(\cdot,T)$ is the continuum solution to (Real Space Schrödinger).

4.3 Combining for full algorithm

We now prove Theorem 6 by combining the Lyapunov convergence results of Section 4.1 with the Hamiltonian simulation results of Section 4.2. These results together lead to a complete resource analysis of discretized QHD. We will actually prove the following stronger theorem.

▶ **Theorem 15.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a G-Lipschitz convex function with $\mathcal{X}_\star$ its (non-empty, convex) set of global minima. Let there further be a known point $x_0 \in \mathbb{R}^d$ and parameters $R, R_\infty, \Lambda, \Lambda_\infty \in \mathbb{R}_+$ such that for some minimum $x_\star \in \mathcal{X}_\star$ it holds that $\|x_0 - x_\star\| \leq R$, $\|x_0 - x_\star\|_\infty \leq R_\infty$, $f(x_0) - f(x_\star) \leq \Lambda$, and $\max_{x\in B_\infty(x_0,4R_\infty)} f(x) \leq \Lambda_\infty$.[3] There exists then a quantum algorithm (Algorithm 1) that solves Problem 1 using $\widetilde{\mathcal{O}}\left(\frac{dGR\Lambda_\infty}{\epsilon^2}\right)$ queries to a $\widetilde{\mathcal{O}}\left(\frac{\epsilon^3}{dGR\Lambda_\infty}\right)$-accurate evaluation oracle, or $\widetilde{\mathcal{O}}\left(\left[\frac{dGR\Lambda_\infty}{\epsilon^2}\right]^2\frac{1}{\epsilon}\right)$ queries to a $\widetilde{\mathcal{O}}\left(\frac{\epsilon^3}{dGR\Lambda_\infty^2}\right)$-accurate phase oracle.*

Let us first prove Theorem 6 as a consequence of the above.

[3] Not all of $R, R_\infty, \Lambda, \Lambda_\infty$ must be supplied in advance, as bounds on these parameters can be obtained from others and Lipschitz continuity. We state the result in terms of the parameters to accommodate cases where stronger bounds than these are available.

Proof of Thm 6. We note that f satisfies all the conditions of Theorem 15. By the conditions of Problem 1 we are given a point x_0 such that $\|x_0 - x_\star\| \leq R$, hence, $\|x_0 - x_\star\|_\infty \leq R$. Furthermore, since f is G-Lipschitz, we have that $f(x_0) - f(x_\star) \leq GR$ and $\max_{x \in B_\infty(x_0, 4R_\infty)} f(x) \leq GR\sqrt{d}$. We can therefore apply Theorem 15 with $R_\infty = R$, $\Lambda = GR$, and $\Lambda_\infty = GR\sqrt{d}$. The result follows. ◀

Algorithm 1 gives pseudocode for our protocol.

Algorithm 1 Convex Optimization via Simulation of Quantum Hamiltonian Descent.

Input: Evaluation oracle O_f for f satisfying conditions of Theorem 15 with parameters $R, R_\infty, \Lambda, \Lambda_\infty$; Schedule c_t, m_t, ω_t satisfying Ideal Scaling with equality, with initial conditions $\omega_0 = 1$, $m_0 = R^{-1}\sqrt{d/(GR + \Lambda)}$ and $\lambda = G^{-1}R^{-2}(\Lambda + GR)^{3/2}$ (See Definitions 7,8 for examples); Target error parameter ϵ.

Output: With probability at least 2/3, a point $\tilde{x}$ such that $f(\tilde{x}) - f_\star \leq \epsilon, \forall x_\star \in \mathcal{X}_\star$.

Procedure:

1: Compute minimum time T_ϵ such that $\omega_{T_\epsilon}^{-2} \leq \epsilon/24$.

2: Prepare initial state $|\Psi_0\rangle$ as an appropriately discretized symmetric Gaussian wavefunction

$$\Phi_0(x) = \frac{1}{(2\pi R^2)^{d/4}} e^{-(x-x_0)^2/4R^2}.$$

3: Simulate, using Theorem 5, centered at x_0 with simulation accuracy $\epsilon/6\Lambda_\infty$ and $R = 4R_\infty$, the evolution of $|\Psi_0\rangle$ under H_{QHD} as in (QHD), for time T_ϵ. Obtain state $|\Psi_{T_\epsilon}\rangle$.

4: Measure $|\Psi_{T_\epsilon}\rangle$ in the computational basis to obtain $\tilde{x}$; return $\tilde{x}$.

We must refer to the Main Version [27] of this paper for a full proof of Theorem 15. Nonetheless, we outline the requisite steps. First, in Lemma 5.1 of [27], we show that for an initial state taken as gaussian wavepacket

$$\Phi_0(x) = (2\pi\sigma^2)^{-d/4} e^{-(x-x_0)^2/4\sigma^2}$$

the Lyapunov constant is bounded as

$$\mathcal{E}_0 \leq \frac{d}{(2m_0\sigma)^2} + 2\lambda^2(d\sigma^2 + R^2) + \omega_0^2(GR + \Lambda).$$

Next, we bound the simulation cost, showing that for ideal scaling parameters, our interaction picture algorithm has complexity

$$\int_0^T c_t m_t \omega_t^2 dt = \frac{m_0\omega_0^2}{2\lambda}(\omega_T^4/\omega_0^4 - 1). \tag{11}$$

The result then follows from judicious choice of the initial parameters m_0, ω_0, c_0.

4.4 Conditional lower bound

In this section, we investigate the dependence of the query complexity derived in previous sections on choice of dynamics within the QHD framework. In particular, is there room to reduce the complexity, by careful optimization of the schedule functions c_t, m_t, ω_t? We answer this question in the negative, under some reasonable assumptions on the query complexity of Hamiltonian simulation and the initial state, in the above Theorem 10. These lower

bounds, however, are highly conditional on the convergence rate matching the one from the Lyapunov argument. With faster convergence rates than ω_T^{-2}, lower complexities may indeed be achievable.

We use the following helper lemma, which is a specific version of the Heisenberg uncertainty principle in d dimensions.

▶ **Lemma 16.** *Let $\phi \in \mathrm{Dom}(x_j p_k) \cap \mathrm{Dom}(p_j x_k) \subseteq L_2(\mathbb{R}^d)$ for all $j, k \in [d]$, with $p = -i\nabla$. Suppose $\langle x \rangle_\phi = x_0$ for some $x_0 \in \mathbb{R}$ and suppose $\langle p \rangle_\phi = 0$. Let $\sigma^2 = \langle (x - x_0)^2 \rangle_\phi$ be the variance of ϕ. Then the variance of the momentum can be bounded as $\langle p^2 \rangle_\phi = \Omega(d^2/\sigma^2)$.*

See Appendix for proof. We are now able to give a proof of the conditional lower bound.

Proof of Thm 10. Assume, without loss of generality, that Φ_0 has expected value x_0 in position. Using Eq.(11) and Assumption 9, the total query complexity of simulation is given by

$$\Lambda_f \int_0^T c_t m_t \omega_t^2 dt = \Lambda_f \frac{m_0 \omega_0^2}{2\lambda} \left((\omega_T/\omega_0)^4 - 1 \right).$$

Since the convergence in continuous-time follows from Theorem 14, in order to ensure error $\leq \epsilon$, the schedule and simulation time must be concurrently chosen to ensure that $\mathcal{E}_0 \omega_T^{-2} \leq \epsilon$. Note that throughout the argument, $\mathcal{E}_0$ has a possible dependence on problem parameters including d, ϵ. As a consequence, the total query complexity can be lower bounded by

$$\frac{\Lambda_f m_0 \omega_0^2}{2\lambda} \left(\frac{\mathcal{E}_0^2}{\epsilon^2 \omega_0^4} - 1 \right). \tag{12}$$

We recall that for a state satisfying our conditions, and with initial variance $\sigma^2 = O(R^2)$ (all expectations with respect to Φ_0):

$$\begin{aligned}
\mathcal{E}_0 &\coloneqq \frac{1}{2} \langle (-i m_0^{-1} \nabla + \lambda(x - x_\star))^2 \rangle + \omega_0^2 \langle f(x) - f_\star \rangle \\
&= \frac{1}{2} m_0^{-2} \langle p^2 \rangle + \frac{1}{2} \lambda^2 \langle (x - x_0)^2 \rangle + \omega_0^2 \langle f(x) - f_\star \rangle \\
&= \Omega \left(\max \left(m_0^{-2} \langle p^2 \rangle, \lambda^2 \langle (x - x_\star)^2 \rangle, \ \omega_0^2 \langle f(x) - f_\star \rangle \right) \right) \\
&= \Omega \left(\max \left(d/(m_0 \sigma)^2, \lambda^2 \langle (x - x_\star)^2 \rangle, \ \omega_0^2 \langle f(x) - f_\star \rangle \right) \right).
\end{aligned}$$

Let $l(d, \epsilon)$ be a desired lower bound on $\mathcal{E}_0$. In order for this bound to be satisfied for a general f satisfying only the conditions of Problem 1, we must ensure that:

- $d/(m_0 \sigma)^2 = \mathcal{O}(l(d, \epsilon))$, therefore $m_0 = \Omega(\sigma \sqrt{d/l(d, \epsilon)})$.
- $\lambda^2 \langle (x - x_\star)^2 \rangle = \mathcal{O}(l(d, \epsilon))$, therefore $\lambda^{-1} = \Omega(\sqrt{(R^2 + \sigma^2)/l(d, \epsilon)})$.
- $\omega_0^2 \langle f(x) - f_\star \rangle = \mathcal{O}(l(d, \epsilon))$, therefore $\omega_0^2 = \mathcal{O}(l(d, \epsilon)/GR)$.

Plugging these requirements into Equation 12 we obtain the lower bound

$$\Omega \left(\Lambda_f \cdot \frac{1}{\sigma} \sqrt{\frac{d}{l(d, \epsilon)}} \cdot \frac{\sqrt{R^2 + \sigma^2}}{\sqrt{l(d, \epsilon)}} \cdot \frac{GR}{l(d, \epsilon)} \cdot \left(\frac{l(d, \epsilon)^2}{\epsilon^2} - \frac{l(d, \epsilon)^2}{G^2 R^2} \right) \right).$$

We observe that the second term in the parentheses is dominated when $\epsilon \to 0$, and the remaining term is minimized when $\sigma = \Theta(R)$. Thus simplifying, we obtain a query complexity lower bound of $\Omega(\sqrt{d} GR \Lambda_f / \epsilon^2)$. ◀

Although the assumption of $\Psi_0(x)$ having real amplitudes appears natural, the reader might wonder whether other choices can avert the lower bound. A hint in this direction is that, domain questions briefly aside, the operator $(p + x)^2$ in the Lyapunov operator $\widehat{\mathcal{E}}_0$ has eigenfunction $e^{ix^2/2}$ with eigenvalue 0, independent of d. This suggests there may be some proper, normalized wavefunction which minimizes the Lyapunov constant. This possibility may be worth exploring in future research; however, we argue that such an approach may amount to shifting complexity from time evolution to state preparation. The state in question appears to be highly oscillatory, whereas a real, log-concave initial state such as a Gaussian admits a number of well-studied methods for preparing.

Disclaimer

References

1 Amira Abbas, Andris Ambainis, Brandon Augustino, Andreas Bärtschi, Harry Buhrman, Carleton Coffrin, Giorgio Cortiana, Vedran Dunjko, Daniel J. Egger, Bruce G. Elmegreen, et al. Challenges and opportunities in quantum optimization. *Nature Reviews Physics*, pages 1–18, 2024.

2 Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *COLT*, pages 28–40. Citeseer, 2010. URL: `http://colt2010.haifa.il.ibm.com/papers/COLT2010proceedings.pdf#page=36`.

3 Alekh Agarwal, Dean P. Foster, Daniel J. Hsu, Sham M. Kakade, and Alexander Rakhlin. Stochastic convex optimization with bandit feedback. *Advances in Neural Information Processing Systems*, 24, 2011.

4 Dorit Aharonov, Wim Van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Review*, 50(4):755–787, 2008. `doi:10.1137/080734479`.

5 Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Reviews of Modern Physics*, 90(1):015002, 2018.

6 Boris Altshuler, Hari Krovi, and Jérémie Roland. Anderson localization makes adiabatic quantum optimization fail. *Proceedings of the National Academy of Sciences*, 107(28):12446–12450, 2010.

7 Brandon Augustino, Dylan Herman, Enrico Fontana, Junhyung Lyle Kim, Jacob Watkins, Shouvanik Chakrabarti, and Marco Pistoia. Fast convex optimization with quantum gradient methods. *arXiv preprint*, 2025. `arXiv:2503.17356`.

8 Brandon Augustino, Jiaqi Leng, Giacomo Nannicini, Tamás Terlaky, and Xiaodi Wu. A quantum central path algorithm for linear optimization. *arXiv preprint arXiv:2311.03977*, 2023. `doi:10.48550/arXiv.2311.03977`.

9 Ryan Babbush, Peter J. Love, and Alán Aspuru-Guzik. Adiabatic quantum simulation of quantum chemistry. *Scientific Reports*, 4(1):6603, 2014.

10 Francis Bach and Vianney Perchet. Highly-smooth zero-th order online optimization. In *Conference on Learning Theory*, pages 257–283. PMLR, 2016. URL: `http://proceedings.mlr.press/v49/bach16.html`.

11 Alexandre Belloni, Tengyuan Liang, Hariharan Narayanan, and Alexander Rakhlin. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In *Conference on Learning Theory*, pages 240–265. PMLR, 2015. URL: `http://proceedings.mlr.press/v40/Belloni15.html`.

12 Michael Betancourt, Michael I. Jordan, and Ashia C. Wilson. On symplectic optimization. *arXiv preprint*, 2018. `arXiv:1802.03653`.

13 Sergio Boixo, Troels F. Rønnow, Sergei V. Isakov, Zhihui Wang, David Wecker, Daniel A. Lidar, John M. Martinis, and Matthias Troyer. Evidence for quantum annealing with more than one hundred qubits. *Nature Physics*, 10(3):218–224, 2014.

14 Max Born and Vladimir Fock. Beweis des adiabatensatzes. *Zeitschrift für Physik*, 51(3):165–180, 1928.

15 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.

16 Stephen P. Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

17 Lucas T. Brady, Christopher L. Baldwin, Aniruddha Bapat, Yaroslav Kharkov, and Alexey V. Gorshkov. Optimal protocols in quantum annealing and quantum approximate optimization algorithm problems. *Physical Review Letters*, 126(7):070505, 2021.

18 J. Brooke, David Bitko, Rosenbaum, and Gabriel Aeppli. Quantum annealing of a disordered magnet. *Science*, 284(5415):779–781, 1999.

19 J. Brooke, T.F. Rosenbaum, and G. Aeppli. Tunable quantum tunnelling of magnetic domain walls. *Nature*, 413(6856):610–613, 2001.

20 Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015. `doi:10.1561/2200000050`.

21 Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012. `doi:10.1561/2200000024`.

22 Sébastien Bubeck, Ronen Eldan, and Yin Tat Lee. Kernel-based methods for bandit convex optimization. *Journal of the ACM (JACM)*, 68(4):1–35, 2021. `doi:10.1145/3453721`.

23 Ahmet Burak Catli, Sophia Simon, and Nathan Wiebe. Efficient quantum optimization via dynamical simulation. *arXiv preprint*, 2026. URL: `https://arxiv.org/abs/2502.04285`, `arXiv:2502.04285`.

24 Nicolo Cesa-Bianchi, Alex Conconi, and Claudio Gentile. On the generalization ability of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50(9):2050–2057, 2004. `doi:10.1109/TIT.2004.833339`.

25 Shouvanik Chakrabarti, Andrew M Childs, Shih-Han Hung, Tongyang Li, Chunhao Wang, and Xiaodi Wu. Quantum algorithm for estimating volumes of convex bodies. *ACM Transactions on Quantum Computing*, 4(3):1–60, 2023. `doi:10.1145/3588579`.

26 Shouvanik Chakrabarti, Andrew M. Childs, Tongyang Li, and Xiaodi Wu. Quantum algorithms and lower bounds for convex optimization. *Quantum*, 4:221, 2020. `doi:10.22331/Q-2020-01-13-221`.

27 Shouvanik Chakrabarti, Dylan Herman, Jacob Watkins, Enrico Fontana, Brandon Augustino, Junhyung Lyle Kim, and Marco Pistoia. On speedups for convex optimization via quantum dynamics. *arXiv preprint arXiv:2503.24332*, 2025. `doi:10.48550/arXiv.2503.24332`.

28 Andrew M. Childs, Jiaqi Leng, Tongyang Li, Jin-Peng Liu, and Chenyi Zhang. Quantum simulation of real-space dynamics. *Quantum*, 6:860, 2022. `doi:10.22331/Q-2022-11-17-860`.

29 Alexandre Choquette, Agustin Di Paolo, Panagiotis Kl Barkoutsos, David Sénéchal, Ivano Tavernelli, and Alexandre Blais. Quantum-optimal-control-inspired ansatz for variational quantum algorithms. *Physical Review Research*, 3(2):023092, 2021.

30 Alexander M. Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T. Hann, Michael J. Kastoryano, Emil T. Khabiboulline, Aleksander Kubica, et al. Quantum algorithms: A survey of applications and end-to-end complexities. *arXiv preprint*, 2023. arXiv:2310.03011.

31 Olivier Devolder, François Glineur, and Yurii Nesterov. First-order methods of smooth convex optimization with inexact oracle. *Mathematical Programming*, 146(1):37–75, 2014. doi:10.1007/S10107-013-0677-5.

32 John C. Duchi. Introductory lectures on stochastic optimization. *The mathematics of data*, 25:99–186, 2018.

33 John C. Duchi, Michael I. Jordan, Martin J. Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015. doi:10.1109/TIT.2015.2409256.

34 Alexandre d'Aspremont, Damien Scieur, and Adrien Taylor. Acceleration methods. *Foundations and Trends in Optimization*, 5(1-2):1–245, 2021. doi:10.1561/2400000036.

35 Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm, 2014. arXiv:1411.4028.

36 Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science*, 292(5516):472–475, 2001.

37 Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution, 2000. arXiv:quant-ph/0001106.

38 Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 385–394, USA, 2005. Society for Industrial and Applied Mathematics. URL: http://dl.acm.org/citation.cfm?id=1070432.1070486.

39 Ankit Garg, Robin Kothari, Praneeth Netrapalli, and Suhail Sherif. Near-optimal lower bounds for convex optimization for all orders of smoothness. *Advances in Neural Information Processing Systems*, 34:29874–29884, 2021. URL: https://proceedings.neurips.cc/paper/2021/hash/fa6c94460e902005a0b660266190c8ba-Abstract.html.

40 Ankit Garg, Robin Kothari, Praneeth Netrapalli, and Suhail Sherif. No Quantum Speedup over Gradient Descent for Non-Smooth Convex Optimization. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 53:1–53:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2021.53.

41 Alexander V. Gasnikov, Darina Dvinskikh, Pavel Dvurechensky, Eduard Gorbunov, Aleksandr Beznosikov, and Alexander Lobanov. Randomized gradient-free methods in convex optimization. In *Encyclopedia of Optimization*, pages 1–15. Springer, 2023.

42 Ernst Hairer, Christian Lubich, and Gerhard Wanner. Structure-preserving algorithms for ordinary differential equations. *Geometric numerical integration*, 31, 2006.

43 Brian C. Hall. *Quantum theory for mathematicians*. Springer, 2013.

44 Samuel B Hopkins. Mean estimation with sub-gaussian rates in polynomial time. *The Annals of Statistics*, 48(2):1193–1213, 2020.

45 Stephen P. Jordan. Fast quantum algorithm for numerical gradient estimation. *Physical Review Letters*, 95(5):050501, 2005.

46 Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse Ising model. *Physical Review E*, 58(5):5355, 1998.

47 Walid Krichene, Alexandre Bayen, and Peter L. Bartlett. Accelerated mirror descent in continuous and discrete time. *Advances in Neural Information Processing Systems*, 28, 2015.

48 Jiaqi Leng, Ethan Hickman, Joseph Li, and Xiaodi Wu. Quantum Hamiltonian Descent, 2023. arXiv:2303.01471. doi:10.48550/arXiv.2303.01471.

49 Jiaqi Leng, Yufan Zheng, Zhiyuan Jia, Chaoyue Zhao, Yuxiang Peng, and Xiaodi Wu. Quantum Hamiltonian Descent for Non-smooth Optimization", 2025. arXiv:2503.15878.

50 Jiaqi Leng, Yufan Zheng, and Xiaodi Wu. A quantum-classical performance separation in nonconvex optimization, 2023. arXiv:2311.00811. doi:10.48550/arXiv.2311.00811.

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

52 Guang Hao Low and Nathan Wiebe. Hamiltonian simulation in the interaction picture. *arXiv preprint*, 2018. arXiv:1805.00675.

53 Alicia B. Magann, Christian Arenz, Matthew D. Grace, Tak-San Ho, Robert L. Kosut, Jarrod R. McClean, Herschel A. Rabitz, and Mohan Sarovar. From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms. *PRX Quantum*, 2(1):010101, 2021.

54 Oinam Romesh Meitei, Bryan T Gard, George S Barron, David P Pappas, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. Gate-free state preparation for fast variational quantum eigensolver simulations. *npj Quantum Information*, 7(1):155, 2021.

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

56 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. doi:10.1137/070704277.

57 Arkadi Semenovič Nemirovskii and David Borisovich Yudin. *Problem complexity and method efficiency in optimization*. Wiley-Interscience, 1983.

58 Yurii Nesterov. A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. In *Dokl akad nauk Sssr*, volume 269, page 543, 1983.

59 Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

60 Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017. doi:10.1007/S10208-015-9296-2.

61 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.

62 Boris T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

63 Ben W. Reichardt. The quantum adiabatic optimization algorithm and local minima. In *Proceedings of the thirty-sixth annual ACM Symposium on Theory of Computing*, pages 502–510, 2004. doi:10.1145/1007352.1007428.

64 Andrej Risteski and Yuanzhi Li. Algorithms and matching lower bounds for approximately-convex optimization. *Advances in Neural Information Processing Systems*, 29, 2016.

65 Terry Rockafellar. Nonsmooth optimization, 1994.

66 Walter Rudin. *Real and complex analysis*. McGraw-Hill, Inc., 1987.

67 Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on learning theory*, pages 3–24. PMLR, 2013. URL: http://proceedings.mlr.press/v30/Shamir13.html.

68 Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017. URL: https://jmlr.org/papers/v18/16-632.html.

69 Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

70 Aaron Sidford and Chenyi Zhang. Quantum speedups for stochastic optimization. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 35300–35330. Curran Associates, Inc., 2023. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/6ed9931d6e1fb6a85efa1b2c014a47e1-Paper-Conference.pdf.

71 Yaron Singer and Jan Vondrák. Information-theoretic lower bounds for convex optimization with erroneous oracles. *Advances in Neural Information Processing Systems*, 28, 2015.

72 Weijie Su, Stephen Boyd, and Emmanuel J. Candes. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016. URL: `https://jmlr.org/papers/v17/15-084.html`.

73 Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Convex optimization using quantum oracles. *Quantum*, 4:220, 2020. `doi:10.22331/Q-2020-01-13-220`.

74 Wim Van Dam, Michele Mosca, and Umesh Vazirani. How powerful is adiabatic quantum computation? In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 279–287. IEEE, 2001. `doi:10.1109/SFCS.2001.959902`.

75 Lorenzo Campos Venuti, Domenico D'Alessandro, and Daniel A. Lidar. Optimal control for quantum optimization of closed and open systems. *Physical Review Applied*, 16(5):054023, 2021.

76 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.

77 Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamon. Optimizing variational quantum algorithms using Pontryagin's minimum principle. *Physical Review X*, 7(2):021027, 2017.

78 Chenyi Zhang, Jiaqi Leng, and Tongyang Li. Quantum algorithms for escaping from saddle points. *Quantum*, 5:529, August 2021. `doi:10.22331/q-2021-08-20-529`.

A Additional Details for Hamiltonian Simulation

Simulation by truncated Dyson series (previously mentioned in Section 4.2) requires a certain block encoding of the interaction picture Hamiltonian $\widetilde{H}_I$, which we now define.

▶ **Definition 17** (Ham-T Oracle [52]). *Let $H(t)$ be a time-dependent Hamiltonian such that for some $\Lambda \in \mathbb{R}_+$, $\|H(t)\| \leq \Lambda$ for all $t \in [0,T]$. Let $\{t_j\}_{j=1}^M \subset [0,T]$ be a discrete set of points. The associated* Ham-T *oracle is a unitary block encoding such that*

$$(\langle 0|_a \otimes I)\mathsf{Ham}\text{-}T(H)(|0\rangle_a \otimes I) = C(H)/\Lambda,$$

where

$$C(H) \coloneqq \sum_{j=0}^{M} |j\rangle\langle j| \otimes H(t_j)$$

is the time-controlled Hamiltonian and a is an auxiliary register.

Our Hamiltonian (9) admits such a block encoding almost immediately.

▶ **Lemma 18** (Ham-T Construction, Main Version [27] Lemma 4.2). *Let $\widetilde{H}_I$ be the interaction picture Hamiltonian of equation (9) with $G|\mathbf{j}\rangle \coloneqq g(x_\mathbf{j})|\mathbf{j}\rangle$, and let $\Lambda \geq \|g\|_\infty$ be a promised upper bound. Then there exists an ϵ_H-accurate block encoding U_g of G, such that*

$$\left\| \mathsf{Ham}\text{-}T(\widetilde{H}_I) - C(U_K)^\dagger U_g C(U_K) \right\| \leq \epsilon_H, \tag{13}$$

with probability of failure at most δ, and where Ham-$T(\widetilde{H}_I)$ *is given in Definition 17. The total costs are as follows.*

- ***Queries**: Q_f queries to a quantum oracle O_f, where*

$$Q_f = \begin{cases} 2 & \text{if } O_f \text{ is a } (\Lambda\epsilon_H)\text{-accurate binary oracle (Definition 2),} \\ \mathcal{O}(\log(1/\delta)/\epsilon_H) & \text{if } O_f \text{ is a } \epsilon_H\text{-accurate controlled phase oracle (Definition 3).} \end{cases} \tag{14}$$

- ***Qubits**: $\mathcal{O}(\log(1/\epsilon_H) + \log(M) + \log(1/\delta))$ auxiliary qubits.*
- ***Gates**: $\mathcal{O}(\log(1/\delta)\log^2(1/\epsilon_H) + d\log(N)^2 + d\log(N)\log(M))$ additional elementary quantum gates.*

The overall resource count for simulating $\widetilde{H}_I$ (Main Version [27, Theorem 4.3]) then follows from plugging in the various parameters of our setting into the results of [52] and accounting for the cost to implement $\mathsf{Ham}\text{-}T(\widetilde{H}_I)$. This is roughly $\widetilde{\mathcal{O}}(T\Lambda Q_f)$ queries to the function quantum oracle O_f, as long as $\epsilon_f^{-1} = \widetilde{\mathcal{O}}\left(\frac{\|b\|_1}{\epsilon}\right)$ (Main Version [27, Theorem 4.4]) under overall desired error ϵ.

We also sketch the procedure for analyzing the discretization error, i.e. how does the digital state produced by simulating $\widehat{H}(t)$ compare to Real Space Schrödinger. In Main Version [27, Corollary 4.1], we show that digital output of our algorithm takes the form

$$|\Psi_t\rangle = \frac{1}{\sqrt{(2N)^d}} \sum_{\mathbf{j}\in\mathcal{N}} \Psi_{\mathbf{j}}(t)\,|\mathbf{j}\rangle\,.$$

We can use this to define a continuous wave function over S, using the notion of *Fourier interpolant*:

$$\Psi(x,t) = \sum_{\mathbf{n}\in\mathcal{N}} \left[\frac{1}{(2N)^d} \sum_{\mathbf{j}\in\mathcal{N}} \overline{\chi}_{\mathbf{n}}(x_{\mathbf{j}})\Psi_{\mathbf{j}}(t)\right] \chi_{\mathbf{n}}(x) \in \mathcal{H}_N. \tag{15}$$

In Main Version [27, Theorem 4.6], we show that the d-dimensional pseudospectral method applied to (Real Space Schrödinger) with N^d grid points satisfies:

$$\|\Phi(T) - \Psi(T)\| \lesssim \|a\|_1 \cdot \max_{t\in[0,T]} \left[\left(\frac{\pi}{4}\right)^{d/4} \frac{2|\Phi(t)|_{\mathcal{H}^{m+2}} + |\Phi(t)|_{\mathcal{H}^m}}{\sqrt{(m-d/2)\Gamma(d/2)N^m}} + \frac{2|\Phi(t)|_{\mathcal{H}^3} + |\Phi(t)|_{\mathcal{H}^1}}{N}\right],$$

where $|\cdot|_{\mathcal{H}^s}$ is the s-th *Sobolev seminorm* defined as $|\phi|^2_{\mathcal{H}^s} := \sum_{\alpha\in\mathbb{N}^n:\|\alpha\|_1=s}\langle\partial^\alpha\phi, \partial^\alpha\phi\rangle$. This result is obtained by bounding both the d-dimensional Fourier truncation and aliasing errors encountered by mapping to a finite grid. As done in Main Version [27, Theorem 4.9] this error bound can be used to show (10). Extending the results of [15], we derive (Main Version [27, Lemma C.9]) a bound on the growth of the s-th Sobolev seminorm over t in terms of those of the potential g:

$$|\Phi(T)|_{\mathcal{H}^s} \le 2|\Phi(0)|_{\mathcal{H}^s} + \left(\frac{2\sum_{k=1}^s \binom{s}{k}|g|_{\mathcal{H}^k}\|b\|_1}{s}\right)^s.$$

Under the assumption that f is G-Lipschitz, convex, and satisfies PBC, we (show Main Version [27, Lemma C.11]) that there exists a convex function very close to g, called g_σ, whose Sobolev seminorms are well-behaved:

$$\|g - g_\sigma\|_\infty = \mathcal{O}\left(\sqrt{d}GR\sigma\right), \quad |g_\sigma|_{\mathcal{H}^m} = \mathcal{O}\left(\|g\|_\infty \left(\frac{Cd}{\sigma}\right)^{3d}\right), m \le d + \mathcal{O}(1),$$

for some constant $C > 0$. This uses the notion of the mollification of a function. If g does not satisfy PBC, then we can use an additional, yet mild, assumption (Main Version [27, Definition 4.5]) to make this work. Due to the linearity of quantum simulation, the $\|g - g_\sigma\|_\infty$ term only causes a proportional blow-up in the overall error (Main Version [27, Lemma C.1]). With an appropriately chosen σ, this all suffices to ensure that the total number of qubits $d\log(N)$ remains $\mathcal{O}\left(d^2 \ \mathrm{polylog}(1/\epsilon, \|a\|_1, \|b\|_1)\right)$ for (10) to hold.

Our work is not the first to analyze quantum algorithms for real space simulation. Indeed, the pseudo-spectral framework of Section 4.2 was previously considered in [28], which made a significant step towards a rigorous characterization of real-space quantum Hamiltonian simulation. The framework and analysis has been subsequently used for various applications [78, 48, 50, 49, 8]. However, we argue that existing literature leaves out a critical piece of analysis that is essential to obtain a complete end-to-end analysis of quantum real-space algorithms is absent.

The output of the pseudo-spectral method is *neither a truncated Fourier series of the true solution nor a trigonometric interpolation.* However, [28] provides only a proof of the required grid spacing to ensure low Fourier truncation error, which does not by itself rigorously imply good pseudo-spectral method performance. For the truncation error bound, the results of [28] are given in terms of a quantity g', which depends on properties of the solution wavefunction and are left unbounded. Note that no such quantity appears in our main simulation result (Theorem 5). As mentioned above, this is essential, since we demand a simulation theorem that is applicable to black-box potentials. The Fourier truncation result of [28] was only proven for one-dimension, which we generalized to multiple dimensions.

B Select Proofs of Results

Here we provide some selected proofs for results stated in the main body of this paper, chosen according to importance and simplicity. As usual, see Main Version [27] for all results and proofs.

Proof of Thm 14. As in [48], we proceed with a quantum Lyapunov argument. Let $p \coloneqq -i\nabla$ and

$$\widehat{\mathcal{E}}_t \coloneqq \frac{1}{2}(p/m_t + \lambda(x - x_\star))^2 + \omega_t^2(f(x) - f_\star). \tag{16}$$

Observe that $\widehat{\mathcal{E}}_t$ is the sum of two positive semi-definite operators (on the appropriate subdomain), hence positive. Define $\mathcal{E}_t \coloneqq \langle\widehat{\mathcal{E}}_t\rangle_t$, where $\langle\cdot\rangle_t$ is shorthand for $\langle\cdot\rangle_{\Phi_t}$ hereon. Then $\mathcal{E}_t \in C^1$ is a non-negative function. We wish to prove that $\mathcal{E}_t$ is a *Lyapunov function*, meaning $\partial_t\mathcal{E}_t \leq 0$. This would imply

$$\omega_t^2\langle f(x) - f_\star\rangle_t \leq \mathcal{E}_T \leq \mathcal{E}_0$$

which amounts to the theorem statement.

Toward this end, we note the well-known fact that

$$\partial_t\langle\widehat{\mathcal{E}}_t\rangle_t = i\langle[H_{\mathrm{QHD}}(t), \widehat{\mathcal{E}}_t]\rangle_t + \langle\partial_t\widehat{\mathcal{E}}_t\rangle_t.$$

For the first term on the right, we evaluate the commutator as

$$i[H_{\mathrm{QHD}}(t), \widehat{\mathcal{E}}_t] = \frac{i}{4}\frac{c_t}{m_t}[p^2, P_t^2] + \frac{i}{2}\frac{c_t\omega_t^2}{m_t}[p^2, f] + \frac{i}{2}c_t m_t\omega_t^2[f, P_t^2]$$

where $P_t \coloneqq p/m_t + \lambda(x - x_\star)$. We handle each term on the right separately, making frequent use of the derivation property of the commutator: $[AB, C] = [A, C]B + A[B, C]$. We also use

$$[p, f] = -i\nabla f$$

where ∇f is understood as the gradient of f where it exists, and 0 elsewhere. By Rademacher's theorem, the "elsewhere" set is of measure zero [66]. The first term is

$$\frac{i}{4}\frac{c_t}{m_t}[p^2, P_t^2] = \lambda\frac{c_t}{m_t}\left(\frac{p^2}{m_t} + \lambda\Re(p \cdot \Delta x)\right)$$

where $\Re(A)$ denotes the Hermitian part and $\Delta x = x - x_\star$. Next,

$$\frac{i}{2}\frac{c_t\omega_t^2}{m_t}[p^2, f] = \frac{1}{2}\frac{c_t\omega_t^2}{m_t}(p \cdot \nabla f + \nabla f \cdot p).$$

Finally,

$$\frac{i}{2}c_t m_t \omega_t^2[f, P_t^2] = -\frac{1}{2}\frac{c_t\omega_t^2}{m_t}(p \cdot \nabla f + \nabla f \cdot p) - \lambda c_t\omega_t^2 \Delta x \cdot \nabla f.$$

The last two are easiest to combine, as there is a cancellation of one of the terms. Adding all the terms together one obtains

$$i\langle[H(t), \widehat{\mathcal{E}}_t]\rangle_t = \lambda c_t\frac{\langle p^2\rangle_t}{m_t^2} + \lambda^2 c_t\frac{\langle\Re(\Delta x \cdot p)\rangle_t}{m_t} - \lambda c_t\omega_t^2\langle\Delta x \cdot \nabla f\rangle_t.$$

Meanwhile, the partial derivative term is given by

$$\langle\partial_t\widehat{\mathcal{E}}_t\rangle_t = -\frac{\dot{m}_t}{m_t}\frac{\langle p^2\rangle_t}{m_t^2} - \lambda\frac{\dot{m}_t}{m_t}\frac{\langle\Re(\Delta x \cdot p)\rangle_t}{m_t} + 2\dot{\omega}_t\omega_t\langle f - f_\star\rangle_t.$$

Adding the previous two lines term by term, one can see that the $\langle p^2\rangle_t$ and $\langle\Re(\Delta x \cdot p)\rangle_t$ terms cancel if c_t and m_t are chosen such that

$$\lambda c_t - \frac{\dot{m}_t}{m_t} = 0$$

which amounts to the first ideal scaling condition. Taking this from here on, we have

$$\begin{aligned}\partial_t\mathcal{E}_t &= 2\dot{\omega}_t\omega_t\langle f - f_\star\rangle_t - \lambda c_t\omega_t^2\langle\Delta x \cdot \nabla f\rangle_t \\ &= \omega_t^2\left(2\frac{\dot{\omega}_t}{\omega_t}\mathbb{E}_{x\sim|\Phi_t|^2}[f - f_\star] - \lambda c_t\mathbb{E}_{x\sim|\Phi_t|^2}[\Delta x \cdot \nabla f]\right).\end{aligned} \tag{17}$$

Next, we make use of the (star-)convexity of the function f, which implies that

$$0 \le f(x) - f_\star \le \Delta x \cdot \nabla f(x) \tag{18}$$

almost everywhere. Hence the expression holds unconditionally under the expectation value, leaving

$$\partial_t\mathcal{E}_t \le \omega_t^2\mathbb{E}_{x\sim|\Phi_t|^2}[\Delta x \cdot \nabla f]\left(2\frac{\dot{\omega}_t}{\omega_t} - \lambda c_t\right). \tag{19}$$

By (18), the expectation value is nonnegative, and thus $\partial_t\mathcal{E}$ is non-positive provided that the second ideal scaling condition holds. ◀

Proof. Observe that $\langle (x-x_0)^2 \rangle_\phi = \sum_{i=1}^d \langle (x-x_0)_i^2 \rangle_\phi$ and $\langle p^2 \rangle_\phi = \sum_{i=1}^d \langle p_i^2 \rangle_\phi$. Using the Robertson form of the Heisenberg Uncertainty Principle in one dimension [43], we notice that $\langle p_i^2 \rangle_\phi \geq \varsigma / \langle (x-x_0)_i^2 \rangle_\phi$ where ς is an absolute constant.

We then note that,

$$\begin{aligned}\langle p^2 \rangle_\phi \cdot \langle (x-x_0)^2 \rangle_\phi &= \left(\sum_{i=1}^d \langle p_i^2 \rangle_\phi \right) \cdot \left(\sum_{i=1}^d \langle (x-x_0)_i^2 \rangle_\phi \right), \\ &\geq \varsigma \left(\sum_{i=1}^d \langle (x-x_0)_i^{-2} \rangle_\phi \right) \cdot \left(\sum_{i=1}^d \langle (x-x_0)_i^2 \rangle_\phi \right) \geq \varsigma d^2,\end{aligned}$$

where the last inequality arises from the well known AM-HM inequality. The result follows. ◀

C Query Complexity of the Li-Risteski Algorithm

As a comparison point, let's consider the classical algorithm from Li and Risteski [64] for approximately convex optimization. To our knowledge, this is the most competitive classical algorithm in our setting of noisy zeroth-order convex optimization. Although Li and Risteski only demonstrates polynomial scaling, here we supply the detailed scaling to obtain the first row of Table 1. See [64, Algorithm 1] for a description of the algorithm. Since we are interested in the case of unconstrained optimization, we will make this simplification while analyzing the query complexity. We will scale the input and output space such that $G = 1$ and $R = 1$, as a consequence we must perform the optimization up to error $\varepsilon \coloneqq \epsilon / GR$

The algorithm executes $T = \Theta(d^2/\varepsilon^4)$ iterations. The primary computational bottleneck in each step is estimating the expectation

$$f_t = \mathbb{E}_{w \in S^d} \left[\frac{d}{r} f(x_t + rw) w \right] = \mathbb{E}_{w \in S^d} \left[\frac{d}{r} (f(x_t + rw) - f(x_t)) w \right],$$

within accuracy $\Theta(\varepsilon)$ in ℓ_2 norm. Here x_t is a point in $B_2(x_0, R)$, $r = \Theta(\varepsilon)$ and w is uniformly sampled from the d-dimensional hypersphere S_d. In the worst case,

$$\Sigma = \mathrm{Var}_{w \in S^d} \left[\frac{d}{r} (f(x_t + rw) - f(x_t)) w \right] \leq d^2 \cdot \mathrm{Var}_{w \in S^d} \left[\|w\| w \right] = d^2 \mathrm{Var}_{w \in S^d} [w] = d \mathbb{I}_d,$$

where the inequality from symmetry and the Lipschitz continuity of f. Using results from Hopkins [44] on optimal mean estimation in Euclidean norm, we observe that given some fixed desired success probability, the ℓ_2 norm accuracy of the estimated mean decays with the number of samples N as

$$\mathcal{O}\left(\sqrt{\mathrm{Tr}(\Sigma)/N} + \sqrt{\|\Sigma\|/N} \right) = \mathcal{O}(d/\sqrt{N}).$$

In order to obtain $\mathcal{O}(\varepsilon)$ error in ℓ_2 norm, the number of samples taken per round must be $\sim d^2/\varepsilon^2$. Each sample requires a noisy evaluation of the function f, resulting in a final query complexity of $\mathcal{O}(d^4/\varepsilon^6)$.

Because their algorithm tolerates high evaluation noise, it is natural to wonder if refinements of the complexity are possible assuming lower noise levels. We essentially rule out this possibility in Appendix D. In summary, our algorithm appears to offer genuine speedups over known classical algorithms for stochastic convex optimization.

D Prospects for Reducing Classical Complexity

Here we discuss whether leading classical algorithms for noisy zeroth-order convex optimization can be improved if the noise tolerance is allowed to be lowered to that of the present quantum algorithm. It is not apparent to us that such an improvement is possible.

First, we note that the query complexity of the algorithm of Belloni et al [11] is based on the analysis of a classical hit-and-run walk. The bottleneck is the classical complexity of the algorithm and is unchanged even if the queries are chosen to be noiseless. The robustness of the algorithm arises from the natural robustness of walk based algorithms, and the lack of any gradient estimation. Thus reducing the noise tolerance does not lead to any improvements in oracle complexity.

Next, consider the algorithm of Ristetski and Li [64, Algorithm 1], which is based on a gradient estimator. It is evident by inspection of the algorithm that the query complexity is determined by a parameter r that is the radius of smoothing. As usual, we rescale the problem by GR, letting $\varepsilon = \epsilon/GR$ and $\varepsilon_f = \epsilon_f/GR$. A calculation borrowed from [38] shows that the choice for r must satisfy

$$f(x) - f_\star - 2r \geq \frac{\varepsilon_f \sqrt{d}}{r} \text{ whenever } f(x) - f_\star \geq \varepsilon.$$

Plugging in the 2nd inequality to the first, it suffices that r lies in the range

$$r = \frac{\varepsilon}{4}\left(1 \pm \sqrt{1 - 8\varepsilon_f\sqrt{d}/\varepsilon^2}\right). \tag{20}$$

Assuming a noise $\varepsilon_f = \mathcal{O}(\varepsilon^3/d)$, we choose the largest r consistent with these bounds, which minimizes sample complexity (see proof of Thm 3.2 [64]). This is $r = \Theta(\varepsilon)$, the same choice of r as the original work. Thus, following through the rest of the analysis (which is independent of the noise parameter) leads to no improvements.