Efficient Quantum Algorithms for Quantum Optimal Control

Xiantao Li [1]   **Chunhao Wang** [2]

Abstract

In this paper, we present efficient quantum algorithms that are exponentially faster than classical algorithms for solving the quantum optimal control problem. This problem involves finding the control variable that maximizes a physical quantity at time T, where the system is governed by a time-dependent Schrödinger equation. This type of control problem also has an intricate relation with machine learning. Our algorithms are based on a time-dependent Hamiltonian simulation method and a fast gradient-estimation algorithm. We also provide a comprehensive error analysis to quantify the total error from various steps, such as the finite-dimensional representation of the control function, the discretization of the Schrödinger equation, the numerical quadrature, and optimization. Our quantum algorithms require fault-tolerant quantum computers.

1. Introduction

Optimal control is an important application of machine learning and optimization.Meanwhile, it has been demonstrated in recent works that important insights on deep learning can be obtained by interpreting the process of training a deep neural network as a discretization of an optimal control problem (E et al., 2019; He et al., 2016). In recent decades, following the emergence of quantum computers, the optimal control of quantum systems has attracted considerable attention. This is because quantum properties are largely responsible for many of the recent developments in material science and chemical engineering, and such properties are best utilized with external controls. (Geppert et al., 2004; Datta, 2005; McCreery et al., 2013). Quantum optimal control (QOC) algorithms (Brif et al., 2010; Werschnik &

[1]Department of Mathematics, Pennsylvania State University, University Park, USA [2]Department of Computer Science and Engineering, Pennsylvania State University, University Park, USA. Correspondence to: Xiantao Li <xiantao.li@psu.edu>, Chunhao Wang <cwang@psu.edu>.

Gross, 2007), which uses first principle-based computer simulations to identify the desired control variable, has been an important route toward this direction. QOC problems have also emerged in recent quantum computing and quantum information technology. For example, QOC has recently been shown to have remarkable connections and applications to machine learning techniques using quantum circuits, including variational quantum algorithms (VQA) (Choquette et al., 2021; Yang et al., 2017), quantum neural networks (Larocca et al., 2021), and quantum approximate optimization algorithm (QAOA) (Farhi et al., 2014; Cerezo et al., 2021; Leng et al., 2022). QOC is treated as a particular machine learning application in (Banchi & Crooks, 2021). One specific application of QOC to quantum machine learning is the diagnostics of quantum barren plateau (Larocca et al., 2022). In addition, solving the optimal control problems for quantum systems is a critical step for implementing quantum computers.

The problem of controlling a quantum system can be formulated based on the time-dependent Schrödinger equation (TDSE),

$$i\partial_t |\psi\rangle = \big(H_0 - u(t)\mu\big) |\psi\rangle, |\psi(0)\rangle = |\psi_0\rangle, \quad (1)$$

with an external control variable $u(t)$ that enters the Hamiltonian through an operator μ. One example comes from electron transport, where $H_0 = -\frac{\nabla^2}{2} + V(x)$ and μ is the dipole operator. To keep our discussions focused on the quantum algorithms, we consider the case with a single control. The treatment of multiple control variables and non-unitary dynamics will be presented in separate works. We assume the copies of the initial state $|\psi_0\rangle$ are given *for free*, i.e., we have access to the unitary U_{ψ_0} that prepares $|\psi_0\rangle$ as $U_{\psi_0} |0\rangle = |\psi_0\rangle$.

The goal of QOC with time duration T is to determine a control variable $u(t)$ for $t \in [0, T]$, such that a physical quantity, represented by a Hermitian operator O, is maximized in that

$$\max_{u(t)} J(u), \quad J(u) := \langle\psi(T)|O|\psi(T)\rangle - \alpha \int_0^T |u(t)|^2 \mathrm{d}t. \quad (2)$$

Here $\alpha > 0$ can be regarded as a regularization parameter imposing a soft constraint on the magnitude of the control variable $u(t)$. In the context of optimization, the

function $J(u)$ is the objective (fitness) function. Aside from the generalization to multiple control variables, another straightforward extension is to control signals with explicit parametric forms, e.g., a trigonometric expression $u(t) = \sum_j E_j \cos(\omega_j t) e^{-(t-t_j)^2/\sigma_j}$ modulated by a Gaussian envelope, which can be written in an abstract form as $u(t) = F(t, \boldsymbol{\theta})$ with θ embodying all the parameters. Then the objective function in Equation (2) can be optimized over $\boldsymbol{\theta}$ and the gradient can be trivially obtained using chain rule: $\partial J/\partial \boldsymbol{\theta} = (\partial J/\partial u) \partial F/\partial \boldsymbol{\theta}$. By restricting the control variable $u(t)$ to square-integrable functions, the objective function is clearly upper bounded. In addition, it is lower semicontinuous, which then implies that a maximum exists (Ciaramella et al., 2015). But in practice, the function is often nonconvex and possesses multiple maxima.

In addition to the wide range of applications of QOC, there is a host of mathematical analysis and computational techniques based on classical algorithms (Castro et al., 2012; Werschnik & Gross, 2007; Eitan et al., 2011; Glaser et al., 2015; Ciaramella et al., 2015; Hellgren et al., 2013; D'Alessandro, 2000; 2008; Albertini & D'Alessandro, 2002). The present paper, however, will be concerned with computer simulations using quantum algorithms, where the n-qubit Hamiltonian operator H (resp. μ) is represented as a 2^n-dimensional matrix $H \in \mathbb{C}^{2^n \times 2^n}$ (resp. $\mu \in \mathbb{C}^{2^n \times 2^n}$). In this paper, we assume the Hamiltonians H, μ, and the observable O are d-sparse: there are at most d nonzero entries in each row/column. In other words, these matrices are *efficiently row/column computable*. To access these matrices, we assume we have access to a procedure $\mathcal{P}_A$ for a matrix A that can perform the following mapping:

$$|i\rangle |0\rangle \mapsto |i\rangle |r_{i,k}\rangle, \tag{3}$$

where $r_{i,k}$ is the k-th nonzero entry of the i-th row of A. In addition, $\mathcal{P}_A$ can also perform the following mapping:

$$|i\rangle |j\rangle |0\rangle \mapsto |i\rangle |j\rangle |A(i,j)\rangle. \tag{4}$$

Note that it is easy to adapt our algorithms to other more general input models, such as the block-encoding (Low & Chuang, 2019; Chakraborty et al., 2019).

With quantum access to these matrices, we aim to determine $u(t)$ for all t such that the objective function defined in Equation (2) is maximized. However, due to the nonconvex landscape of the objective function, finding the global maximum is not realistic. Instead, we consider the notion of ϵ-second-order convergence condition (see Equation (12)) for the objective function. More specifically, we aim to solve the following problem.

Problem 1.1. Consider the n-qubit QOC in Equation (2) for time duration T. Assume[1] $\|H_0\|, \|O\|, \|\mu\| \leq 1$, and

$\alpha \geq 2/T$. Let N be the number of time intervals and define $\delta := T/N$. Suppose H_0, μ and the observable O are d-sparse, and suppose we are given access to U_{ψ_0} for preparing $|\psi_0\rangle$. Given the sparse access $\mathcal{P}_{H_0}$ to H_0, $\mathcal{P}_\mu$ to μ, and $\mathcal{P}_O$ to O, determine $\tilde{u}(t_j)$ $(t_j = j\delta)$ for all $j \in [N]$ such that $|\tilde{u}(t_j) - u(t_j)| \leq \epsilon$, where $u(t)$ is an ϵ-second-order stationary point of $J(u)$ satisfying Equation (12).

The main challenge in solving Problem 1.1 is due to the repeated solution of TDSE in Equation (1), which becomes unfavorable for classical computers because the dimension grows exponentially. In addition, many optimization algorithms require the computation of the derivatives of the objective function with respect to the control variable $u(t)$. A direct calculation of such derivatives of J in Equation (2) also requires visiting the TDSE. The most well-known classical method for optimizing Equation (2) is due to Zhu and Rabitz (Zhu & Rabitz, 1998a). The idea is to regard the TDSE in Equation (1) as a constraint, which is incorporated into the objective function using a Lagrange multiplier. For this optimization problem, algorithms can be constructed by enforcing the first-order condition. Zhu and Rabitz (Zhu & Rabitz, 1998a) established the monotone convergence property of this algorithm. This approach is later extended to further improve the convergence properties (Maday & Turinici, 2003; Eitan et al., 2011; Ciaramella et al., 2015) including formulations based on the Liouville von Neumann equation.

Nevertheless, the complexity associated with one iteration of these classical algorithms has the scaling of $\mathcal{O}(\text{poly}(2^n)T)$, where n is the number of system qubits and we did not include ϵ in the complexity. The exponential dependence on n comes from the fact that at each time step, the wave functions have to be computed from previous time steps using an approximation of the unitary operator. Depending on the specific methods, this may involve matrix multiplications or solutions of a linear system of equations (Castro et al., 2004). The exponential scaling significantly limits the scope of the application, since 2^n is usually proportional to the domain size and the number of electrons.

Main Contributions. In this paper, we present *efficient* quantum algorithms for solving the quantum optimal control problem defined in Equation (2) that scale polynomially in n — our quantum algorithms have *exponential speedups* compared with best known classical algorithms. One key contribution is a complexity analysis that takes into account various error contributions, including (**i**) the finite-dimensional approximation of $u(t)$, (**ii**) the discretization of the objective function in Equation (2), (**iii**) the approximation of the wave function, and (**iv**) the optimization error. Such comprehensive overall estimates are extremely valuable to assess the resources needed for a specific QOC problem. We start

---

[1]More generally, if $\|H_0\|, \|\mu\| \leq \Lambda$, the overall complexity of our algorithm will pick up an extra $\mathcal{O}(\Lambda)$ factor because of Equation (26) and the fact that $\|A\|_{\max} \leq \|A\|$ for a matrix A.

with the Hamiltonian simulation method for time-dependent problems (Berry et al., 2020) to access the objective function, thus enabling a quantum gradient estimation (Gilyén et al., 2019a) approach, so that a gradient-based optimization method can be applied. Although our approach is formulated for the TDSE, it can be readily generalized to other control problems where the dynamics is not Hamiltonian, e.g., ODEs implemented with quantum spectral method (Childs & Liu, 2019).

The following theorem summarizes our main result, where the $\widetilde{\mathcal{O}}$ notation has omitted logarithmic factors.

Theorem 1.2 (Summarizing Theorem 3.9 and Theorem 4.1). *Assume that $u(t) \in C^2([0, T])$. There exists a quantum algorithm that, with probability at least $2/3^2$, solves Problem 1.1 using $\widetilde{\mathcal{O}}\left(\frac{dLT^2}{\epsilon^3}\right)$ queries to $\mathcal{P}_{H_0}$ and $\mathcal{P}_\mu$, and $\widetilde{\mathcal{O}}\left(\frac{dnLT^2}{\epsilon^3} + \frac{LT^{3/2}}{\epsilon^{5/2}}\right)$ additional 1- and 2-qubit gates.*

Further, when the control variable $u(t)$ is smooth, the gate complexity becomes $\widetilde{\mathcal{O}}\left(\frac{dnLT^2}{\epsilon^3} + \frac{LT}{\epsilon^2}\right)$. The Lipschitz constant L can be bounded by Equation (14)

Although the smooth control variables does not improve the gate complexity asymptotically, it can reduce the number of gates in practical implementations. We also use Table 1 for a clear comparison of different smoothness assumptions on the control variable $u(t)$.

Smoothness	Queries	Additional gates
Without smoothness	$\widetilde{\mathcal{O}}\left(\frac{dLT^2}{\epsilon^3}\right)$	$\widetilde{\mathcal{O}}\left(\frac{dnLT^2}{\epsilon^3} + \frac{LT^{3/2}}{\epsilon^{5/2}}\right)$
Smooth control	$\widetilde{\mathcal{O}}(\frac{dLT^2}{\epsilon^3})$	$\widetilde{\mathcal{O}}\left(\frac{dnLT^2}{\epsilon^3} + \frac{LT}{\epsilon^2}\right)$

Table 1. The query and gate complexities for solving Problem 1.1.

It is important to note that our algorithms require fault-tolerant quantum computers, and hence they are not suitable for near-term quantum devices.

Related work. To the authors' knowledge, the first attempt to solve the quantum control problem using quantum algorithms has been proposed by (Li et al., 2017). This approach requires a classical device to update the control variable $u(t)$ using a gradient-based method. More recently, Brif et al. (Brif et al., 2010) considered a specific application, where a molecular wave function is evolved under a laser pulse. The algorithm is also hybrid in nature, utilizing a classical device to update the control variable. The approach outputs the objective function, rather than the gradient, for the Nelder-Mead optimization method. None of the above-mentioned works uses the full power of quantum computation, includ-

ing the state-of-the-art quantum algorithms for Hamiltonian simulation algorithms and gradient estimation. More importantly, precise query complexity that takes into account all levels of approximations is not provided, making it difficult to assess the applicability of the algorithms to specific quantum control problems in practice.

The QOC problems show great similarity with variational quantum algorithms (VQA) in general (Cerezo et al., 2021), except that in most VQAs, the unitary operations are associated with parameterized gates, rather than time-dependent Hamiltonians in QOC, where the smoothness of the control function also plays a role in the error. Several other algorithms have recently been proposed to help with the computation of the gradient in VQA (Crooks, 2019; Wierichs et al., 2022; Kyriienko & Elfving, 2021).

Very recently, Leng, Peng, Qiao, Lin, and Wu (Leng et al., 2022) proposed a differentiable programming framework for quantum computers to evaluate gradients of time-evolved states. One of their application is QOC. In their model, the control function $u(t)$ is fixed while leaving freedoms to the parameters of u to be tuned. In contrast, we work on a more general control model and deal with the control function in real time directly. In addition, our work focuses on quantum algorithms on fault-tolerant quantum computers, and we use state-of-the-art techniques to improve the complexity.

Notation. We provide a brief introduction to quantum computation in Section 2.1. Here, we summarize the notations used in this paper. First of all, we should distinguish quantum states from a vector in the context of classical computation, although a quantum state is mathematically represented as a vector. In the context of quantum dynamics, we use the Dirac notation $|\cdot\rangle$ to represent quantum states. For vectors in the context of classical computation, e.g., the gradient vector and the control variable at multiple time steps, we use the bold font, e.g., $\boldsymbol{u}$. For a positive integer k, we use $[k]$ to denote the set $\{1, \ldots, k\}$. For a vector $\boldsymbol{u}$, we use the subscript with a normal font to denote its entries, i.e., $u_1, \ldots, u_d$ are the entries of $\boldsymbol{u}$. For a vector $\boldsymbol{u}$, we use $\|\boldsymbol{u}\|$ to denote its Euclidean norm, i.e., $\|\boldsymbol{u}\| = \sqrt{\sum_{j=1}^{d} |u_j|^2}$. We use $\|\boldsymbol{u}\|_\infty$ to denote its maximum norm, i.e., $\|\boldsymbol{u}\|_\infty = \max_j |u_j|$.

In addition, we denote by $C^2[0, T]$ the class of twice continuously differential functions defined in $[0, T]$. For a function f, we use $\|f\|_2$, or $\|f(t)\|_2$ to denote its L_2 norm. Namely, $\|f(t)\|_2 = \|f\|_2 = \int_0^T f(t)^2 \, dt$.

2. Preliminaries

To solve Problem 1.1, we need to deal with several levels of approximations, including a finite-dimensional representation of the function $u(t)$, the time discretization of the TDSE in Equation (1), the approximation of the gradient

---

[2]Using standard techniques, the success probability can be boosted to a constant arbitrarily close to 1 while only introducing a logarithmic factor in the complexity.

of $\langle \psi(T) | O | \psi(T) \rangle$, and the optimization error. In this section, we first give a simplistic yet sufficient introduction to quantum computation. Then we analyze the error caused by the discretization of $u(t)$, the discretization of the TDSE, and optimization in subsequent subsections.

2.1. A brief introduction to quantum computation

The most daunting obstacle for learning quantum computing is probably its notations. Once one is getting familiar with the Dirac notation, quantum computing can be understood as matrix-vector multiplications. For an n-qubit state, we use a 2^n-dimensional complex (column) vector with unit Euclidean norm to mathematically describe it, and this vector is denoted, in the Dirac notation, by $|\psi\rangle$. We use $\langle\psi|$ to denote the conjugate transpose of $|\psi\rangle$, and then the inner product between two states can be conveniently written as $\langle\psi|\phi\rangle$. When the symbol in the Dirac notation is a natural number, i.e., $|0\rangle, \ldots, |2^n - 1\rangle$, we use such states to specifically denote the set of the *standard basis* of a 2^n-dimensional vector space, same as the basis vectors $\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{2^n-1}$ that we usually see in linear algebra textbooks. With respect to the standard basis, let $\alpha_0, \ldots, \alpha_{2^n-1}$ be the entries of $|\psi\rangle$, i.e., $|\psi\rangle = \begin{bmatrix} \alpha_0, \ldots, \alpha_{2^n-1} \end{bmatrix}^T$. It can be conveniently written as a *superposition*, $|\psi\rangle = \alpha_0 |0\rangle + \cdots + \alpha_{2^n-1} |2^n - 1\rangle$. When $|\psi\rangle$ is *measured*, with probability $|\alpha_j|^2$, it will *collapse* to $|j\rangle$. Note that a measurement is an irreversible process: a state cannot be restored once it collapses to some basis state.

The building blocks of quantum algorithms are 1- and 2-qubit gates: they are modeled as 2×2 (1-qubit), or 4×4 (2-qubit) unitary matrices. A network of quantum gates form a quantum circuit, which, acting on n qubits, is mathematically represented as a $2^n \times 2^n$ unitary matrix. In essence, a quantum algorithm starts from an n-qubit initial state $|0\rangle$ that is easy to prepare, and applies a quantum circuit, represented by U, yielding the state $|\psi\rangle = U |0\rangle$. Then a measurement is performed on $|\psi\rangle$ to obtain the desired outcomes. Readers may refer to (Nielsen & Chuang, 2010) for a thorough introduction to quantum computation.

2.2. Approximation Error of the Objective Function

We first represent the control variable $u(t)$ at finitely many time steps with the length of a time interval denoted by δ: Let $t_j = j\delta$ be discrete time steps, and $N = T/dt$ be the number of time steps. Then we represent $u(t)$ via piece-wise linear interpolation, denoted by $u_I(t)$, given by,

$$u_I(t) = \sum_{j=0}^{N} u_j b_j(t), \tag{5}$$

with $b_j(t)$ being the nodal basis functions (i.e., the hat functions) centered at time steps t_j. The interpolating function can also be integrated directly, which leads to a quadrature

approximation for the integral in Equation (2). The error from these approximations follows the standard error bound, e.g., see (Kincaid et al., 2009).

Theorem 2.1 (Perturbation under C^2 assumption.). *Assume that $u(t) \in \mathrm{C}^2([0, T])$, i.e., it is twice continuously differentiable, then the error from the piecewise-linear interpolation is bounded as follows,*

$$\|u(t) - u_I(t)\|_2 \leq C \|u''(t)\|_2 \delta^2. \tag{6}$$

In addition, the integral of $u(t)^2$ can be approximated by a composite trapezoid rule, with weights denoted by w_j,

$$\int_0^T u(t)^2 \, \mathrm{d}t = \delta \sum_{j=0}^{N} w_j u(t_j)^2 + \mathcal{O}\left(T\delta^2\right). \tag{7}$$

The quadrature formula ($w_0 = w_N = 1/2$ and $w_1 = w_2 = \cdots = w_{N-1} = 1$) can be derived from the linear interpolation of $u(t)$ in each time interval by integrating the nodal functions directly. In addition, the error carries a prefactor that is proportional to the second-order derivatives of the integrand, which is bounded due to the assumption that $u(t) \in \mathrm{C}^2([0, T])$.

With this representation of the control function, the problem is reduced to a finite-dimensional problem. We will use $\boldsymbol{u}$ to represent the nodal values of the interpolation: $\boldsymbol{u} = \{u_j\}_{j=1}^{N}$, with u_j being the nodal value of $u(t)$ at t_j. We will then write the corresponding objective function as $\widetilde{J}(\boldsymbol{u})$, a function with N variables. Namely,

$$\widetilde{J}(\boldsymbol{u}) = \langle \psi_N | O | \psi_N \rangle - \alpha\delta \sum_{j=0}^{N} w_j u_j^2. \tag{8}$$

Here we have also included the approximation of $|\psi(T)\rangle$, denoted by $|\psi_N\rangle$.

Corollary 2.2. *Assume that the wave function is approximated by $|\psi_N\rangle$, with an error comparable to the quadrature error in Equation (7), i.e.,*

$$\||\psi_N\rangle - |\psi(T)\rangle\| = \mathcal{O}(T\delta^2). \tag{9}$$

Further assume that the maximum of $J[u]$ is achieved at some $u^*(t) \in \mathrm{C}^2([0, T])$, and the maximum is not degenerate. Then for sufficiently small ϵ, by choosing,

$$N = \mathcal{O}\left(\frac{T^{3/2}}{\epsilon^{1/2}}\right), \tag{10}$$

the approximate objective function $\widetilde{J}(\boldsymbol{u})$ has a maximum $\boldsymbol{u}^*$, for which the corresponding interpolant satisfies,

$$\|u(t) - u_I(t)\|_2 \leq \epsilon.$$

Proof. This can be proved by using the continuous dependence of the wave function $|\psi(t)\rangle$ on the control variable $u(t)$, e.g., see Proposition 4 of (Ciaramella et al., 2015). $\square$

2.3. Optimization method

A gradient-base algorithm updates the control variables using the gradient (Nesterov, 2018) to generate an approximating sequence $\{\boldsymbol{u}^{(k)}\}_{k\geq 0}$,

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \eta\big(\nabla J(\boldsymbol{u}^{(k)}) + \boldsymbol{\zeta}_k\big). \tag{11}$$

Notice that the QOC in Equation (2) is customarily formulated as a maximization problem. Therefore, the algorithm updates the control variables along the gradient *ascent* direction. In addition, we considered the perturbed gradient method (Jin et al., 2021), which introduced a Gaussian noise $\boldsymbol{\zeta}_k$ with mean zero and $\mathbb{E}[\boldsymbol{\zeta}_k^2] = r^2$ for some $r > 0$ to help the solution with fulfilling an approximation second-order condition.

The convergence to the global maximum of a gradient-based algorithm requires certain convexity assumptions on J, e.g., the star-convexity (Zhou et al., 2018), which, however, may not be easy to verify in practice. Alternatively, one can seek approximate optimality conditions. Jin et al. (Jin et al., 2021) introduced ϵ-first and second order conditions,

$$\|\nabla J(\boldsymbol{u})\| < \epsilon, \text{ and } \quad \nabla^2 J(\boldsymbol{u}) \leq \sqrt{\rho\epsilon}I. \tag{12}$$

Here ρ is the Lipschitz constant of the Hessian. The following result, adapted from Theorem 4.1 of (Jin et al., 2021) (with $\sigma = \mathcal{O}(\epsilon^2)$) summarizes the complexity associated with optimization.

Theorem 2.3 (Adapted from Theorem 4.1 in (Jin et al., 2021))**.** *Assume that the gradient ∇J is $L-$Lipschitz and it can be estimated with an unbiased estimation of variance $\sigma^2 = \mathcal{O}(\epsilon^2)$ and that the learning rate is $\eta < \Theta\left(\frac{1}{L}\right)$. With probability at least $1-\delta$, the optimization method defined in Equation (11) will visit an approximate stationary point that fulfills the first and second order conditions in Equation (12), at least once within the following number of iterations,*

$$k = \widetilde{\mathcal{O}}\left(\log\frac{1}{\delta}L\frac{J(\boldsymbol{u}^*) - J(\boldsymbol{u}^{(0)})}{\epsilon^2}\right). \tag{13}$$

For the objective function in Equation (8), the Lipschitz constant can be estimated using the derivative bounds from Lemma 3.5,

$$L \leq T\delta\|\mu\|^2 + 2\alpha\delta. \tag{14}$$

Since we do not assume an upper bound on α, we will treat L as a separate constant.

3. Quantum algorithm and complexity analysis

In order to access the objective function in Equation (2), a time discretization is needed to approximate the integral and the wave function $\psi(T)$ by numerical integrators, as alluded

to in the previous section. To begin with, we write $\widetilde{J}$ as two terms: $\widetilde{J}(\boldsymbol{u}) = \widetilde{J}_1(\boldsymbol{u}) + \widetilde{J}_2(\boldsymbol{u})$, where

$$\widetilde{J}_1(\boldsymbol{u}) := \langle\psi_N|\, O\, |\psi_N\rangle \quad \text{and} \tag{15}$$

$$\widetilde{J}_2(\boldsymbol{u}) := \alpha\delta\sum_{j=1}^{N} w_j u(t_j)^2. \tag{16}$$

Note the $\widetilde{J}_2$ is easy to compute on a classical computer. Thus we only focus on $\widetilde{J}_1$, and use a quantum algorithm to speed up its calculation.

3.1. Obtaining the final state

To have quantum access to $\widetilde{J}_1$, we need to efficiently prepare of $|\psi_N\rangle$ via a time-marching procedure, i.e.,

$$|\psi_N\rangle = V(t_N, t_{N-1})V(t_{N-1}, t_{N-2})\cdots V(t_1, t_0)|\psi_0\rangle. \tag{17}$$

Note that $t_N = T$. At each step, the dynamics is evolved via an operator $V(t_j, t_{j-1})$. For instance, in the Dyson series approach (Low & Wiebe, 2018; Kieferová et al., 2019), the operator V can be constructed as follows,

$$V(t_{j+1}, t_j) = \sum_{k=0}^{K}\frac{(-i\delta)^k}{M^k k!}\sum_{j_1, j_2, \cdots, j_k = 0}^{M-1}\mathcal{T}H(s_{j_k})\cdots H(s_{j_1}). \tag{18}$$

Here the symbol $\mathcal{T}$ indicates that only those points where $s_{j_1} \leq s_{j_2} \leq \cdots \leq s_{j_k}$ are collected. The points are selected to approximate the integrals in the Dyson series expansion (Berry et al., 2017). The complexity estimate was presented in Theorem 9 of (Low & Wiebe, 2018). We use a more efficient simulation algorithm by Berry, Childs, Su, Wang, and Wiebe (Berry et al., 2020) where the dependence on $T\max_t\|H(t)\|_{\max}$ has later been relaxed to an averaged L_1 norm in time.

Lemma 3.1 ((Berry et al., 2020))**.** *Assume that n-bit Hamiltonian $H(t)$ is d-sparse for all $t \in [0, T]$. The TDSE in Equation (1) can be simulation until time T within error ϵ and failure probability $\mathcal{O}(\epsilon)$ using*

$$\mathcal{O}\left(d\int_0^T \mathrm{d}\tau\|H(\tau)\|_{\max}\frac{\log(dH_{\max}T/\epsilon)}{\log\log(dH_{\max}T/\epsilon)}\right), \tag{19}$$

queries to the oracle of $H(t)$ and $\widetilde{\mathcal{O}}(dn\int_0^T \mathrm{d}\tau\|H(\tau)\|_{\max})$ additional 1- and 2-qubit gates.

Once $|\psi_N\rangle$ from Equation (17) is prepared by Lemma 3.1, we can use it to estimate the gradient of $\widetilde{J}$ defined in Equation (8). We take advantage of quantum gradient estimation algorithms to estimate $\nabla\widetilde{J}$. These algorithms require certain quantum access to the function. In particular, we define the two commonly used oracles, namely, the *probability oracle* and the *phase oracle* as follows.

Definition 3.2 (Probability oracle). Consider a function $f : \mathbb{R}^d \to [0, 1]$. The *probability oracle* for f, denoted by U_f, is a unitary defined as

$$U_f |\boldsymbol{x}\rangle |\boldsymbol{0}\rangle = \tag{20}$$
$$|\boldsymbol{x}\rangle \left(\sqrt{f(\boldsymbol{x})} |1\rangle |\phi_1(\boldsymbol{x})\rangle + \sqrt{1 - f(\boldsymbol{x})} |0\rangle |\phi_0(\boldsymbol{x})\rangle \right),$$

where $|\phi_1(\boldsymbol{x})\rangle$ and $|\phi_0(\boldsymbol{x})\rangle$ are arbitrary states.

Definition 3.3 (Phase oracle). Consider a function $f : \mathbb{R}^d \to \mathbb{R}$. The *phase oracle* for f, denoted by O_f, is a unitary defined as

$$O_f |\boldsymbol{x}\rangle |\boldsymbol{0}\rangle = e^{if(\boldsymbol{x})} |\boldsymbol{x}\rangle |\boldsymbol{0}\rangle . \tag{21}$$

The following result from Theorem 14 of (Gilyén et al., 2019a) shows an efficient conversion from the probability oracle to the phase oracle.

Lemma 3.4. *Consider a function $f : \mathbb{R}^d \to [0, 1]$. Let U_f be the probability oracle for f. Then, for any $\epsilon \in (0, 1/3)$, we can implement an ϵ-approximate of the phase oracle O_f for f, denoted by $\widetilde{O}_f$, such that $\left\| \widetilde{O}_f |\psi\rangle |\boldsymbol{x}\rangle - O_f |\psi\rangle |\boldsymbol{x}\rangle \right\| \leq \epsilon$, for all state $|\psi\rangle$. This implementation uses $\mathcal{O}(\log(1/\epsilon))$ invocations to U_f and $U_f^\dagger$, and $\mathcal{O}(\log\log(1/\epsilon))$ additional qubits.*

We will show how to construct the required oracle for $\widetilde{J}_1$ given the circuit that prepares $|\psi_N\rangle$ in the proof of Theorem 3.9. Before that, let us focus on higher-order derivatives of $\widetilde{J}_1$, which are crucial in quantum algorithms for estimating the gradient of $\widetilde{J}_1$.

3.2. High-order derivatives of the objective function

In this subsection, we bound the magnitude of higher-order derivatives of $\widetilde{J}_1$. In particular, we show the following lemma. The proof can be found in Appendix A.

Lemma 3.5. *Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k) \in [N+1]^k$ be an index sequence[3]. The derivatives of the control function $\widetilde{J}_1$ with respect to the control variables satisfy:*

$$\left\| \frac{\partial^{\boldsymbol{\alpha}} \widetilde{J}_1}{\partial u_{\alpha_1} u_{\alpha_2} \cdots u_{\alpha_k}} \right\| \leq (k+1)! \left(\delta \|\mu\| \right)^k . \tag{22}$$

3.3. Quantum gradient estimation

In this subsection, we show how to efficiently estimate the gradient of $\widetilde{J}_1$. In the following, we state a quantum algorithm for estimating gradients due to Gilyén, Arunachalam, and Wiebe (Gilyén et al., 2019a), which is an extension Jordan's gradient estimation algorithm (Jordan, 2005). Note

---

[3]for a precise definition of an index sequence, see Definition 4 of (Gilyén et al., 2019a)

Algorithm 1 Quantum algorithm for solving QOC

Given $T > 0$ and $\epsilon > 0$, set $N = \frac{T^{3/2}}{\epsilon^{1/2}}$, and $k_{\max} = \mathcal{O}(\frac{L}{\epsilon^2})$ where L is bounded by Equation (14); set $u(t) = 0$
for $k = 1 : k_{\max}$ **do**
    Use Lemmas 3.1 and 3.4 to construct the phase oracle for $\widetilde{J}_1(\boldsymbol{u})$
    Use Lemma 3.7 to estimate $\boldsymbol{g}^{(k)} \approx \nabla J(\boldsymbol{u}^{(k)})$
    Update control variable: $\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \eta \boldsymbol{g}^{(k)} + \eta \boldsymbol{\zeta}_k$
end for

---

that there exists a nearly-optimal gradient estimation, i.e., Theorem 25 of (Gilyén et al., 2019a); however, this result does not directly apply to our setting since the derivatives, as characterized in Lemma 3.5, do not satisfy the bound $c^k k^{k/2}$. To give a more efficient estimation algorithm than just relying on the bound on the Hessian, we use the following result based on a higher-order finite difference method.

Lemma 3.6 (Rephrased from Theorem 23 of (Gilyén et al., 2019a)). *Suppose the access to $f : [-1, 1]^N \to \mathbb{R}$ is given via a phase oracle O_f. If f is $(2m+1)$-times differentiable and for all $\boldsymbol{x} \in [-1, 1]^N$ it holds that*

$$|\partial_{\boldsymbol{r}}^{2m+1} f(\boldsymbol{x})| \leq B \quad for \; \boldsymbol{r} = \boldsymbol{x}/\|\boldsymbol{x}\|, \tag{23}$$

then there exists a quantum algorithm that output an approximate gradient $\boldsymbol{g}$ such that $\|\boldsymbol{g} - \nabla f(\boldsymbol{0})\|_\infty \leq \epsilon$ with probability at least $1 - \rho$ using

$$\widetilde{\mathcal{O}} \left(\frac{N^{1/2} B^{1/(2m)} N^{1/(4m)} \log(N/\rho)}{\epsilon^{1+1/(2m)}} \right) \tag{24}$$

queries to O_f, and $\widetilde{\mathcal{O}}(N)$ additional 1- and 2-qubit gates.

Based on Lemma 3.6, we have the following result, where the proof is postponed in Appendix B.

Lemma 3.7. *Let $\widetilde{J}_1$ be defined as in Equation (15). Suppose we are given access to the phase oracle $O_{\widetilde{J}_1}$ for $\widetilde{J}_1$. Then, there exists a quantum algorithm that outputs an approximate gradient $\boldsymbol{g}$ such that $\left\| \boldsymbol{g} - \nabla \widetilde{J}_1 \right\| \leq \epsilon$ with probability at least $1 - \rho$ using $\widetilde{\mathcal{O}}\left(T \log(N/\rho)/\epsilon\right)$ queries to $O_{\widetilde{J}_1}$, and $\widetilde{\mathcal{O}}(N)$ additional 1- and 2-qubit gates.*

3.4. Proof of the main theorem

We outline the combined algorithm in Algorithm 1. The bound on $k_{\max}$ follows from Equation (13).

Before proving the performance of Algorithm 1, we need to bound the L_1 norm $\int_0^T \mathrm{d}\tau \|H(t)\|_{\max}$, which depends on the L_1 norm of control function. Although in general, no explicit formula is available for $u(t)$, we can deduce some a priori bounds on $u(t)$.

Lemma 3.8. *Suppose that* $\|O\| \leq 1$, *and* $\alpha > \frac{2}{T}$. *Let* $u^{(1)}, \ldots, u^{(k)}$ *be a sequence of control variables obtained from each iteration of Algorithm 1. Let* $u^{(1)}(t) = 0$ *be the initial guess. Then for all* $j \in [k]$, *the* L_1 *norms of the control functions satisfy* $\int_0^T |u^{(1)}(t)| dt \leq T$.

Proof. First of all, note that $J(u^{(1)}) \geq -1$. For all $j = 2, \ldots, k$, we have $J(u^{(j)}) \geq J(u^{(1)}) \geq -1$. Again using the fact that $\|O\| \leq 1$, we find $\int_0^T |u^{(j)}(t)|^2 dt \leq 2/\alpha \leq T$. Using Cauchy–Schwarz inequality, we have,

$$\int_0^T \left|u^{(j)}(t)\right| dt \leq \left[\int_0^T \left|u^{(j)}(t)\right|^2 dt\right]^{\frac{1}{2}} T^{\frac{1}{2}} \leq T. \quad (25)$$

$\square$

The inequality $J(u^{(j)}) \geq J(u^{(1)})$ follows from a function ascent property in Equation (31), which as we will illustrate in the proof of Theorem 3.9, occurs with high probability. Because $\|H_0\|, \|\mu\| \leq 1$, Lemma 3.8 implies that

$$\int_0^T d\tau \|H(t)\|_{\max} = \mathcal{O}(T). \quad (26)$$

Now, we are ready to state the main result of this section.

Theorem 3.9. *Assume that* $u(t) \in C^2([0, T])$. *There exists a quantum algorithm that, with probability at least* $2/3$, *solves Problem 1.1 using* $\widetilde{\mathcal{O}}\left(\frac{dLT^2}{\epsilon^3}\right)$ *queries to* $\mathcal{P}_{H_0}$ *and* $\mathcal{P}_\mu$, *and* $\widetilde{\mathcal{O}}\left(\frac{dnLT^2}{\epsilon^3} + \frac{LT^{3/2}}{\epsilon^{5/2}}\right)$ *additional 1- and 2-qubit gates[4]. The Lipschitz constant* L *can be bounded by Equation (14).*

Proof. Recall that by using Lemma 3.1, we can obtain $|\psi_N\rangle$ given $|\psi_0\rangle$. Now, we show how to construct the oracle to estimate the gradient of $\widetilde{J}$. We first construct the probability oracle $U_{\widetilde{J}_1}$ for $\widetilde{J}_1$ by a Hadamard test circuit. Note that this oracle is well-defined because we have assumed that $\|O\| \leq 1$. In addition, we need access to a block-encoding U_O of O, which is a unitary of the form

$$U_O = \begin{bmatrix} O & \cdot \\ \cdot & \cdot \end{bmatrix}. \quad (27)$$

Hence, $\langle 0| \langle\psi_N| U_O |0\rangle |\psi_N\rangle = \langle\psi_N| O |\psi_N\rangle$. Let c-$U_O$ denote the controlled U_O. The Hadamard test circuit $(H \otimes I)(\text{c-}U_O)(H \otimes I)$ acting on $|0\rangle |\psi_N\rangle$ produces,

$$\sqrt{f(\boldsymbol{u})} |1\rangle |\phi_1(\boldsymbol{u})\rangle + \sqrt{1 - f(\boldsymbol{u})} |0\rangle |\phi_0(\boldsymbol{u})\rangle, \quad (28)$$

[4] In the gate complexity, the second term is dominated by the first; however, we keep both to demonstrate how smooth control can improve the gate complexity non-asymptotically. (See Section 4.)

for $f(\boldsymbol{u}) := -\frac{1}{2} \langle\psi_N| O |\psi_N\rangle + \frac{1}{2}$. Note that we actually constructed the probability oracle for $f(\boldsymbol{u})$ instead of $\langle\psi_N| O |\psi_N\rangle$. However, this is fine because the gradient of the latter is the gradient of the former multiplied by a factor $-1/2$. In addition, we note that the block-encoding U_O can be efficiently constructed using the sparse-access oracles for O (see Lemma 48 of (Gilyén et al., 2019b)). It is important to note that the implementation of the time-dependent Hamiltonian simulation by using Lemma 3.1 uses some control variable $\boldsymbol{u}$ as part of the input. Such Hamiltonian simulation circuits can be made coherent, i.e., it can be used when the register containing the control variable $\boldsymbol{u}$ is in superposition. In addition, we use U_{ψ_0} to prepare the initial state $|\psi_0\rangle$ from $|0\rangle$ for Hamiltonian simulation. Due to the bound on the L_1 norm in Equation (26), we can implement the probability oracle $U_{\widetilde{J}_1}$ as defined in Definition 3.2 using $\widetilde{\mathcal{O}}(dT)$ queries to $\mathcal{P}_{H_0}$ and $\mathcal{P}_\mu$. We highlight that the oracle for the time-dependent Hamiltonian can be efficiently implemented. The values of the control function away from the nodal points can be easily computed by linear interpolation. Thus, we store $\mathcal{O}(N)$ parameters in QRAM to construct the input oracle. The gate complexity of the addressing scheme for QRAM is $\widetilde{\mathcal{O}}(N)$, which is polynomial in T and ϵ due to Equation (10). The circuit depth of the addressing scheme is $\mathcal{O}(\log N)$. As a result, considering the gate- and time-complexities of implementing QRAM, the Hamiltonian simulation is still efficient. Further, in special cases of our control models where the form of $u(t)$ is known, while the parameters in $u(t)$ need to be determined, we do not need QRAM because $u(t)$ can be computed by an efficient subroutine.

Next, we use Lemma 3.4 to construct an ϵ-approximate phase oracle $\widetilde{O}_f$ defined in Definition 3.3 using $\mathcal{O}(1)$ queries to $U_{\widetilde{J}_1}$ and hence $\widetilde{\mathcal{O}}(dT)$ queries to $\mathcal{P}_{H_0}$ and $\mathcal{P}_\mu$.

Using Lemma 3.7, we can obtain a vector $\boldsymbol{g}$ such that $\left\|\boldsymbol{g} - \nabla\widetilde{J}_1/2\right\| \leq \epsilon$ with probability at least $1 - \rho$ using $\widetilde{\mathcal{O}}(T \log(N/\rho)/\epsilon)$ queries to $O_{\widetilde{J}_1}$. Hence, the number of queries to $\mathcal{P}_{H_0}$ and $\mathcal{P}_\mu$ is

$$\widetilde{\mathcal{O}}\left(\frac{dT^2 \log(N/\rho)}{\epsilon}\right). \quad (29)$$

Now we use Theorem 2.3, and analyze the error caused by the estimation error in the spectral norm. We consider the gradient iteration method with the noise $\zeta_k = 0$,

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \eta(\boldsymbol{g}_k + \boldsymbol{\zeta}_k). \quad (30)$$

The main departure from the optimization algorithm in Equation (11) is that an approximation $\boldsymbol{g}_k$ of the gradient of a loss function $\nabla J(\boldsymbol{u}^{(k)})$ is used. We denote the error by

$$\boldsymbol{e}_k := \boldsymbol{g}_k - \nabla J(\boldsymbol{u}^{(k)}).$$

The proof of Theorem 2.3 in (Jin et al., 2021) relies on a descent property: Lemma B.1 in (Jin et al., 2021). Following this approach, we first show that the following inequality holds with probability at least $1 - \nu$,

$$J(\boldsymbol{u}^{(k)}) \geq J(\boldsymbol{u}^{(0)}) + \frac{\eta}{2} \sum_{j=0}^{k-1} \left\| \nabla J(\boldsymbol{u}^{(j)}) \right\|^2 - \frac{k\eta\epsilon^2}{4} \quad (31)$$

where the algorithm in Lemma 3.7 is implemented with $\rho = \nu/k$. We leave the derivation of this inequality to Appendix C.

In light of Equation (8), the fact that $w_j = \mathcal{O}(1)$, and the derivative bounds in Lemma 3.5, we see that the Lipschitz constant of ∇J is $L = \mathcal{O}((1+\alpha)\sqrt{N}\delta)$. Now if we choose the number of iterations,

$$k = 4L \frac{J^* - J(\boldsymbol{u}^{(0)})}{\epsilon^2}, \quad (32)$$

and suppose that $\left\| \nabla J(\boldsymbol{u}^{(j)}) \right\| > \epsilon, \forall j \in [k]$. Namely, the iterations never reached a critical point. Then the descent property implies that, $J(\boldsymbol{u}^{(k)}) - J(\boldsymbol{u}^{(0)}) \geq \frac{k\eta\epsilon^2}{4}$, with probability $1 - \nu$, which would lead to a contradiction $J(\boldsymbol{u}^{(k)}) - J(\boldsymbol{u}^{(0)}) > J^* - J(\boldsymbol{u}^{(0)})$. Consequently, the iterations in Equation (30) will fulfill the first order condition. The rest of the proof can be completed by following the proof in (Jin et al., 2021).

We use ν instead of ρ in Equation (29). The complexity only depends on ν logarithmically. As a result, we have the claimed overall query complexity. The gate complexity for each iteration is dominated by the gradient estimation and time-dependent Hamiltonian simulation , i.e., $\widetilde{\mathcal{O}}(dnLT^2/\epsilon + N) = \widetilde{\mathcal{O}}(dnLT^2/\epsilon + T^{3/2}/\epsilon^{1/2})$. $\qquad \square$

4. Quantum dynamics with smooth control

The estimates in the previous sections have been obtained based on the mild assumption that the control function $u(t)$ is only C^2. Meanwhile, there are various scenarios where the control function is parameterized using smooth functions (Song et al., 2022; Machnes et al., 2018). For example, rather than controlling point-wise values of $u(t)$, one can express $u(t)$ as Fourier series and then the control parameters are reduced to Fourier coefficients. One important implication is that the corresponding solution of the TDSE is smooth as well, in which case, the algorithms presented in Section 3 can be greatly improved. In particular, both the wave function $\psi(T)$ and the integral of the control function within a time interval can be approximated with arbitrary

order p. Namely,

$$\begin{aligned} |\psi(\delta)\rangle - |\psi_1\rangle &= \widetilde{\mathcal{O}} \left(\frac{\delta^{p+1}}{(p+1)!} \right), \\ \int_0^\delta u(t)^2 dt - \delta \sum_j w_j u_j &= \widetilde{\mathcal{O}} \left(\frac{\delta^{p+1}}{(p+1)!} \right). \end{aligned} \quad (33)$$

for arbitrary order p. Without loss of generality, we have expressed the approximations here for the first time step.

For the approximation of the integral, Gaussian quadrature can be used to obtain maximum accuracy. For the approximation of the wave function, such accuracy has been obtained in the Dyson series approach (Kieferová et al., 2019) even without this smoothness assumption.

To ensure that the error from the approximation of $|\psi(T)\rangle$ is within precision ϵ, we force the one-step error to be below ϵ/N. Since p can be chosen arbitrarily, the step size $\delta = \mathcal{O}(1)$. In light of Equation (33), this error bound can be achieved if we choose p as follows,

$$p = \Theta \left(\frac{\log N/\epsilon}{\log \log N/\epsilon} \right) = \Theta \left(\frac{\log T/\epsilon}{\log \log T/\epsilon} \right). \quad (34)$$

Consequently, the number of time steps $N = T/\delta$ is improved from Equation (10) to

$$N = \Theta \left(T \frac{\log T/\epsilon}{\log \log N/\epsilon} \right) = \widetilde{\mathcal{O}} (T). \quad (35)$$

Overall, the smoothness of the control variable $u(t)$ allows us to use high-order approximations and it reduces the extra gates used in Lemma 3.6 by a factor of $T^{1/2}/\epsilon^{1/2}$. However, the smoothness does not improve the query complexity and the overall gate complexity because the number of queries depend on $T = N\delta$, no matter how small N is.

To summarize this section, we have the following theorem.

Theorem 4.1. *Assume that $u(t)$ is smooth. There exists a quantum algorithm that, with probability at least $2/3$, solves Problem 1.1 using $\widetilde{\mathcal{O}} \left(\frac{dLT^2}{\epsilon^3} \right)$ queries to $\mathcal{P}_{H_0}$ and $\mathcal{P}_\mu$, and $\widetilde{\mathcal{O}} \left(\frac{dnLT^2}{\epsilon^3} + \frac{LT}{\epsilon^2} \right)$ additional 1- and 2-qubit gates. The Lipschitz constant L can be bounded by Equation (14).*

Proof. The proof is similar to that of Theorem 3.9. We use the number of steps N as in Equation (35) to obtain this improved gate complexity. $\qquad \square$

5. Numerical Example

Here we consider a one-dimensional model similar to (Zhu & Rabitz, 1998b). The Hamiltonian H_0 is a tri-diagonal

matrix corresponding to a three-point discrete Laplacian operator. The operator μ and O are diagonal,

$$\mu = \sum_r r e^{-r/r_0} |r\rangle\langle r|, \quad O = \sum_r \frac{\gamma_0}{\pi} e^{-\gamma_0^2 r^2} |r\rangle\langle r|.$$

The step size is set to $\delta = 0.02$.

In the optimization, we choose the learning rate to be 0.04. Using $u(t) = 0$ as the initial guess, we apply Equation (11) for 2000 iterations. Figure 1 shows that the objective function has reached a plateau. As comparison, we run the algorithms with exactly computed gradient, and perturb gradients with noise. We observe that even with noise perturbations, the gradient-based algorithm still converges to the same maximum. In addition, as shown in Figure 2, the resulting control variable exhibits some statistical fluctuations; however, it still remains close to the optimal control variable, indicating its resilience to noisy perturbations.
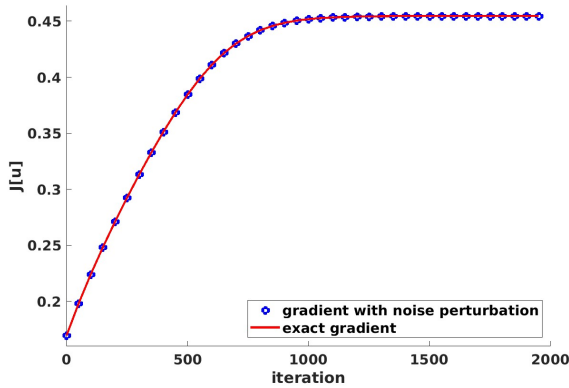
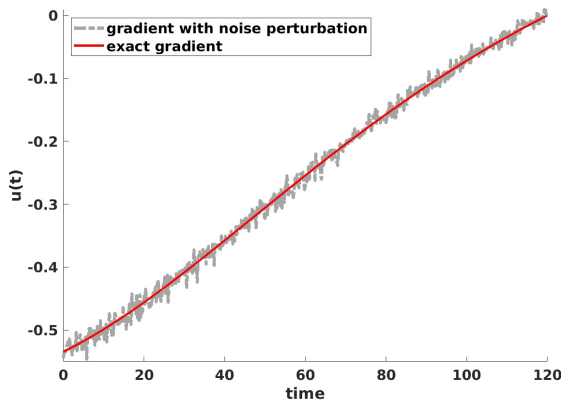*Figure 1.* The loss function during 2000 iterations.

Figure 2. The control function after 2000 iterations.

References

Albertini, F. and D'Alessandro, D. The Lie algebra structure and controllability of spin systems. *Linear Algebra and its Applications*, 350(1-3):213–235, July 2002. ISSN 00243795. doi: 10.1016/S0024-3795(02) 00290-2. URL https://linkinghub.elsevier. com/retrieve/pii/S0024379502002902.

Banchi, L. and Crooks, G. E. Measuring analytic gradients of general quantum evolution with the stochastic parameter shift rule. *Quantum*, 5:386, 2021.

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

Berry, D. W., Childs, A. M., Su, Y., Wang, X., and Wiebe, N. Time-dependent Hamiltonian simulation with l^1-norm scaling. *Quantum*, 4:254, 2020.

Brif, C., Chakrabarti, R., and Rabitz, H. Control of quantum phenomena: past, present and future. *New Journal of Physics*, 12(7):075008, July 2010. ISSN 1367-2630. doi: 10.1088/1367-2630/12/7/075008. URL http://stacks.iop.org/1367-2630/ 12/i=7/a=075008?key=crossref. 77a344443577b6cc05c060a399fa38e1.

Castro, A., Marques, M. A., and Rubio, A. Propagators for the time-dependent kohn–sham equations. *The Journal of chemical physics*, 121(8):3425–3433, 2004.

Castro, A., Werschnik, J., and Gross, E. K. U. Controlling the Dynamics of Many-Electron Systems from First Principles: A Combination of Optimal Control and Time-Dependent Density-Functional Theory. *Physical Review Letters*, 109(15), October 2012. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.109. 153603. URL https://link.aps.org/doi/10. 1103/PhysRevLett.109.153603.

Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S. C., Endo, S., Fujii, K., McClean, J. R., Mitarai, K., Yuan, X., Cincio, L., et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

Chakraborty, S., Gilyén, A., and Jeffery, S. The power of block-encoded matrix powers: Improved regression techniques via faster Hamiltonian simulation. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, 2019.

Childs, A. M. and Liu, J.-P. Quantum spectral methods for differential equations. *arXiv preprint arXiv:1901.00961*, 2019.

Choquette, A., Di Paolo, A., Barkoutsos, P. K., Sénéchal, D., Tavernelli, I., and Blais, A. Quantum-optimal-control-inspired ansatz for variational quantum algorithms. *Physical Review Research*, 3(2):023092, 2021.

Ciaramella, G., Borzì, A., Dirr, G., and Wachsmuth, D. Newton Methods for the Optimal Control of Closed Quantum Spin Systems. *SIAM Journal on Scientific Computing*, 37 (1):A319–A346, January 2015. ISSN 1064-8275, 1095-7197. doi: 10.1137/140966988. URL http://epubs. siam.org/doi/10.1137/140966988.

Crooks, G. E. Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition. *arXiv preprint arXiv:1905.13311*, 2019.

D'Alessandro, D. Topological properties of reachable sets and the control of quantum bits. *Systems & Control Letters*, 41(3):213–221, October 2000. ISSN 01676911. doi: 10.1016/S0167-6911(00) 00063-3. URL https://linkinghub.elsevier. com/retrieve/pii/S0167691100000633.

D'Alessandro, D. *Introduction to quantum control and dynamics*. Chapman & Hall/CRC applied mathematics and nonlinear science series. Chapman & Hall/CRC, Boca Raton, 2008. ISBN 978-1-58488-884-0. OCLC: ocn123539646.

Datta, S. *Quantum transport: Atom to transistor*. Cambridge University Press, 2005.

Deuflhard, P. and Bornemann, F. *Scientific computing with ordinary differential equations*, volume 42. Springer Science & Business Media, 2002.

E, W., Han, J., and Li, Q. A mean-field optimal control formulation of deep learning. *Res. Math. Sci.*, 6:1–41, 2019.

Eitan, R., Mundt, M., and Tannor, D. J. Optimal control with accelerated convergence: Combining the Krotov and quasi-Newton methods. *Physical Review A*, 83(5), May 2011. ISSN 1050-2947, 1094-1622. doi: 10.1103/ PhysRevA.83.053426. URL https://link.aps. org/doi/10.1103/PhysRevA.83.053426.

Farhi, E., Goldstone, J., and Gutmann, S. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

Geppert, D., Seyfarth, L., and de Vivie-Riedle, R. Laser control schemes for molecular switches. *Appl. Phys. B*, 79(8):987–992, 2004.

Gilyén, A., Arunachalam, S., and Wiebe, N. Optimizing quantum optimization algorithms via faster quantum gradient computation. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1425–1444. SIAM, 2019a.

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

Glaser, S. J., Boscain, U., Calarco, T., Koch, C. P., Köckenberger, W., Kosloff, R., Kuprov, I., Luy, B., Schirmer, S., Schulte-Herbrüggen, T., Sugny, D., and Wilhelm, F. K. Training Schrödinger's cat: quantum optimal control: Strategic report on current status, visions and goals for research in Europe. *The European Physical Journal D*, 69(12), December 2015. ISSN 1434-6060, 1434-6079. doi: 10.1140/epjd/ e2015-60464-1. URL http://link.springer. com/10.1140/epjd/e2015-60464-1.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, 2016.

Hellgren, M., Räsänen, E., and Gross, E. K. U. Optimal control of strong-field ionization with time-dependent density-functional theory. *Physical Review A*, 88(1), July 2013. ISSN 1050-2947, 1094-1622. doi: 10.1103/ PhysRevA.88.013414. URL https://link.aps. org/doi/10.1103/PhysRevA.88.013414.

Jin, C., Netrapalli, P., Ge, R., Kakade, S. M., and Jordan, M. I. A short note on concentration inequalities for random vectors with subgaussian norm. *arXiv preprint arXiv:1902.03736*, 2019.

Jin, C., Netrapalli, P., Ge, R., Kakade, S. M., and Jordan, M. I. On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points. *Journal of the ACM (JACM)*, 68(2):1–29, 2021.

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

Kieferová, M., Scherer, A., and Berry, D. W. Simulating the dynamics of time-dependent hamiltonians with a truncated dyson series. *Physical Review A*, 99(4):042314, 2019.

Kincaid, D., Kincaid, D. R., and Cheney, E. W. *Numerical analysis: mathematics of scientific computing*, volume 2. American Mathematical Soc., 2009.

Kyriienko, O. and Elfving, V. E. Generalized quantum circuit differentiation rules. *Physical Review A*, 104(5): 052417, 2021.

Larocca, M., Ju, N., García-Martín, D., Coles, P. J., and Cerezo, M. Theory of overparametrization in quantum neural networks. *arXiv preprint arXiv:2109.11676*, 2021.

Larocca, M., Czarnik, P., Sharma, K., Muraleedharan, G., Coles, P. J., and Cerezo, M. Diagnosing barren plateaus with tools from quantum optimal control. *Quantum*, 6: 824, 2022.

Leng, J., Peng, Y., Qiao, Y.-L., Lin, M., and Wu, X. Differentiable analog quantum computing for optimization and control. In *Advances in Neural Information Processing Systems*, 2022.

Li, J., Yang, X., Peng, X., and Sun, C.-P. Hybrid quantum-classical approach to quantum optimal control. *Physical review letters*, 118(15):150503, 2017.

Low, G. H. and Chuang, I. L. Hamiltonian simulation by qubitization. *Quantum*, 3:163, 2019.

Low, G. H. and Wiebe, N. Hamiltonian simulation in the interaction picture. *arXiv preprint arXiv:1805.00675*, 2018.

Machnes, S., Assémat, E., Tannor, D., and Wilhelm, F. K. Tunable, flexible, and efficient optimization of control pulses for practical qubits. *Physical review letters*, 120 (15):150401, 2018.

Maday, Y. and Turinici, G. New formulations of monotonically convergent quantum control algorithms. *The Journal of Chemical Physics*, 118(18):8191–8196, May 2003. ISSN 0021-9606, 1089-7690. doi: 10.1063/ 1.1564043. URL http://aip.scitation.org/ doi/10.1063/1.1564043.

McCreery, R. L., Yan, H., and Bergren, A. J. A critical perspective on molecular electronic junctions: there is plenty of room in the middle. *Phys. Chem. Chem. Phys.*, 15(4):1065–1081, 2013.

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

Nielsen, M. A. and Chuang, I. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

Song, Y., Li, J., Hai, Y.-J., Guo, Q., and Deng, X.-H. Optimizing quantum control pulses with complex constraints and few variables through autodifferentiation. *Physical Review A*, 105(1):012616, 2022.

Werschnik, J. and Gross, E. K. U. Quantum Optimal Control Theory. *arXiv:0707.1883 [quant-ph]*, July 2007. URL http://arxiv.org/abs/0707. 1883. arXiv: 0707.1883.

Wierichs, D., Izaac, J., Wang, C., and Lin, C. Y.-Y. General parameter-shift rules for quantum gradients. *Quantum*, 6: 677, 2022.

Yang, Z.-C., Rahmani, A., Shabani, A., Neven, H., and Chamon, C. Optimizing variational quantum algorithms using pontryagin's minimum principle. *Physical Review X*, 7(2):021027, 2017.

Zhou, Y., Yang, J., Zhang, H., Liang, Y., and Tarokh, V. Sgd converges to global minimum in deep learning via star-convex path. In *International Conference on Learning Representations*, 2018.

Zhu, W. and Rabitz, H. A rapid monotonically convergent iteration algorithm for quantum optimal control over the expectation value of a positive definite operator. *J.Chem. Phys.*, 109(2):385–391, 1998a.

Zhu, W. and Rabitz, H. A rapid monotonically convergent iteration algorithm for quantum optimal control over the expectation value of a positive definite operator. *The Journal of Chemical Physics*, 109(2):385–391, July 1998b. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.476575. URL http://aip.scitation. org/doi/10.1063/1.476575.

A. The proof of Lemma 3.5

Proof. To understand how the objective function J depends on the control variables, we first explicitly write the piecewise-linear interpolating function using shape functions $b_j(t)$,

$$u_I(t) = \sum_j u_j b_j(t),$$

where $b_j(t)$'s are standard hat functions. We first examine the derivatives of $|\psi(T)\rangle$ with respect to the control variables. Following the TDSE in Equation (1), we can write the derivative of $\psi(t)$ with respect to u_j, $\phi_1(t) := \partial_{u_j} \psi(t)$, as follows,

$$i\partial_t \phi_1 = H(t)\phi_1 - \mu b_j(t)\psi(t). \tag{36}$$

Known as a variational equation (Deuflhard & Bornemann, 2002), such an equation characterizes the dependence of the solution on the parameters in an evolution equation. By using variation-of-constant formula, we find

$$\phi_1(t) = i \int_0^t U(t,t')\mu\psi(t')b_j(t')\, \mathrm{d}t',$$

where $U(t,t')$ is the unitary operator generated by $H(t)$. A straightforward bound thus follows,

$$\|\phi(t)\| \le \|\mu\|\delta,\ \forall 0 \le t \le T.$$

Here we used the fact that

$$\int_0^T |b_j(t)|\, \mathrm{d}t \le \delta.$$

We can continue with this calculation by letting $\phi_2(t) := \partial_{u_j}\phi_1(t)$ for a second order derivative:

$$i\partial_t \phi_2 = H(t)\phi_2 - 2\mu b_j(t)\phi_1(t).$$

Therefore, $\|\phi_2(t)\| \le 2(\delta\|\mu\|)^2$. Mixed derivatives can be treated similarly and they follow a similar bound. By defining $\phi_3(t) := \partial_{u_j}\phi_2(t)$, we obtain,

$$i\partial_t \phi_3 = H_0\phi_3 - 3\mu b_j(t)\phi_2(t).$$

By repeating this argument for higher-order derivatives, we arrive at the inequality.

$$\left\| \frac{\partial^{\boldsymbol{\alpha}}}{\partial u_{\alpha_1} u_{\alpha_2} \cdots u_{\alpha_k}} \psi(T) \right\| \le k!\, (\delta\|\mu\|)^k. \tag{37}$$

Consequently, the function $\widetilde{J}_1$ has derivative bounds,

$$\left\| \frac{\partial^{\boldsymbol{\alpha}} \widetilde{J}_1}{\partial u_{\alpha_1} u_{\alpha_2} \cdots u_{\alpha_k}} \right\| \le \sum_{j=0}^k \binom{k}{j}(k-j)!j!\,(\delta\|\mu\|)^k \le (k+1)!\,(\delta\|\mu\|)^k. \tag{38}$$

$\square$

B. Proof of Lemma 3.7

Proof. First observe that $\widetilde{J}_1(\boldsymbol{u})$ depends on $\boldsymbol{u}$ smoothly. By Lemma 3.5, we have

$$B = \left| \nabla_{\boldsymbol{r}}^{2m+1} J \right| = \mathcal{O}((2m+1)!\delta^{2m+1}). \tag{39}$$

Therefore, according to Lemma 3.6, the query complexity to achieve the error bound in the spectral norm is

$$\widetilde{\mathcal{O}}\left(\frac{N}{\epsilon} \frac{B^{\frac{1}{2m}} N^{\frac{1}{4m}}}{\epsilon^{\frac{1}{2m}}} \right). \tag{40}$$

We notice that by Stirling's approximation,

$$(2m+1)! \approx \left(\frac{2m+1}{e}\right)^{2m+1}. \tag{41}$$

Therefore, it suffices to choose

$$m = \mathcal{O}\left(\frac{\log\left(T^{1/2}/\epsilon^{3/4}\right)}{\log\log\left(T^{1/2}/\epsilon^{3/4}\right)}\right), \tag{42}$$

so that both m and $(T^{1/2}/\epsilon^{3/4})^{1/(2m)}$ are poly-logarithmic factors. As a result, we have the claimed query complexity. □

C. The function ascent property Equation (31)

We start by noticing that,

$$
\begin{aligned}
J(\boldsymbol{u}^{(k+1)}) \geq{}& J(\boldsymbol{u}^{(k)}) + \left(\nabla J(\boldsymbol{u}^{(k)}), \boldsymbol{u}^{(k+1)} - \boldsymbol{u}^{(k)}\right) - \frac{L}{2}\left\|\boldsymbol{u}^{(k+1)} - \boldsymbol{u}^{(k)}\right\|^2 \\
\geq{}& J(\boldsymbol{u}^{(k)}) + \eta\left\|\nabla J(\boldsymbol{u}^{(k)})\right\|^2 + \eta\left(\nabla J(\boldsymbol{u}^{(k)}), \boldsymbol{\zeta}_k + \boldsymbol{e}_k\right) - \frac{\eta}{2}\left\|\nabla J(\boldsymbol{u}^{(k)}) + \boldsymbol{e}_k + \boldsymbol{\zeta}_k\right\|^2 \\
\geq{}& J(\boldsymbol{u}^{(k)}) + \frac{\eta}{2}\left\|\nabla J(\boldsymbol{u}^{(k)})\right\|^2 - \eta\|\boldsymbol{\zeta_k}\|^2 - \eta\|\boldsymbol{e}_k\|^2.
\end{aligned}
$$

The first line was obtained from the Lipschitz condition on the gradient. Equation (30) is then used to arrive at the second line.

By a telescoping sum, we arrive at the desired function ascent property,

$$
\begin{aligned}
J(\boldsymbol{u}^{(k)}) \geq{}& J(\boldsymbol{u}^{(0)}) + \frac{\eta}{2}\sum_{j=0}^{k-1}\left\|\nabla J(\boldsymbol{u}^{(j)})\right\|^2 \\
& - \eta\sum_{j=0}^{k-1}\|\boldsymbol{\zeta_j}\|^2 - \eta\sum_{j=0}^{k-1}\|\boldsymbol{e_j}\|^2.
\end{aligned}
\tag{43}
$$

The noise term $\sum_{j=0}^{k-1}\|\boldsymbol{\zeta_j}\|^2$ has been treated probabilistically in (Jin et al., 2021) using a concentration inequality (Jin et al., 2019), which provides a bound of $\frac{1}{8}k\epsilon^2$ with high probability. Meanwhile, we set a failure probability $\rho = \nu/k$ in the gradient estimation such that,

$$\mathbb{P}\left(\|\boldsymbol{e}_j\|^2 < \epsilon^2/8\right) \geq (1-\rho),$$

implying that

$$\mathbb{P}\left(\|\boldsymbol{e}_j\|^2 < \epsilon^2/8, \ \forall 1 \leq j \leq k\right) \geq (1-\rho)^k \geq 1 - \nu. \tag{44}$$