

Computational complexity of quantum optimal control landscapes

Raj Chakrabarti,* Rebing Wu, and Herschel Rabitz

Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA

Kenneth Steiglitz

Department of Computer Science, Princeton University, Princeton, New Jersey 08544, USA

(Dated: 23 January 2008)

We study the Hamiltonian-independent contribution to the complexity of quantum optimal control problems. The optimization of controls that steer quantum systems to desired objectives can itself be considered a classical dynamical system that executes an analog computation. The system-independent component of the equations of motion of this dynamical system can be integrated analytically for various classes of discrete quantum control problems. For the maximization of observable expectation values from an initial pure state and the maximization of the fidelity of quantum gates, the time complexity of the corresponding computation belongs to the class continuous log (CLOG), the lowest analog complexity class, equivalent to the discrete complexity class NC. By contrast, the optimization of arbitrary nondegenerate observables from mixed initial states belongs to the class continuous P (CP), equivalent to the discrete complexity class P. Observables of different rank and degeneracy classes display qualitatively distinct polynomial scalings. These results indicate that the optimization via gradient flows of quantum observables from pure initial states is inherently less complex than that from mixed initial states, and suggest that local search algorithms will be most effective when the quantum state is pure.

PACS numbers: 03.67.Lx.,03.67.-a,02.30.Yy

I. INTRODUCTION

In recent years the methodology of optimal control theory has been applied to achieve various objectives in quantum systems. The two most common objectives are the maximization of the expectation value of an observable, and the maximization of the fidelity of a unitary transformation of the wavefunction. Whereas the maximization of observable expectation values has met with widespread success in experimental and computational incarnations [1], the achievement of high fidelity unitary transformations has proven more challenging [2], especially for large systems. However, it is not clear whether these apparent distinctions are specific to the algorithms employed, or whether they are indicative of universal and inherent features of the quantum optimization problems.

Given the computational expense of propagating the Schrödinger equation, the scaling with system size of the expense of executing a search over the space of control fields – referred to as the problem’s computational complexity – plays an essential role in determining the feasibility and efficiency of control optimizations in large quantum systems. The complexity of control optimization for quantum gates is of particular interest, as it corresponds to the classical complexity associated with the physical implementation of quantum logical operations. Recent studies have addressed the computational complexity (and related, the critical topology) of several associated problems in quantum information

science, including quantum state and process identification [3]. In contrast to these tasks, the variational problem of quantum control optimization occurs on an infinite-dimensional parameter space. However, for discrete quantum systems, both the complexity and critical topology of quantum control problems can be divided into finite- and infinite-dimensional contributions, simplifying their study.

Consider first the critical topology of quantum control variational problems. Denoting the cost functional by J and the control field by $\varepsilon(t)$, we have according to the chain rule

$$\frac{\delta J}{\delta \varepsilon(t)} = \text{Tr} \left\{ \nabla J(U(T)) \frac{\delta U^\dagger(T)}{\delta \varepsilon(t)} \right\},$$

where $U(T)$ is the dynamical propagator resulting from application of the control $\varepsilon(t)$ over the time interval $0 \leq t \leq T$. As such, the critical points of quantum control problems fall into two classes. The first type of minimizer corresponds to those control Hamiltonians that are critical points of the control objective functional, but are not critical points of the map between control fields and associated dynamical propagators (i.e., points at which $\nabla J(U(T)) = 0$, while the Frechet derivative mapping from the control variation $\delta \varepsilon(t)$ to $\delta U(T)$ at $t = T$ is surjective). These critical points are called normal extremal controls. The second type of minimizer corresponds to critical points of the latter map (i.e., points at which the mapping from $\delta \varepsilon(t)$ to $\delta U(T)$ is not locally surjective). These latter points are referred to as abnormal extremal controls. For several low-dimensional quantum control problems, it has been shown that abnormal extremals are particularly scarce

*Electronic address: rajchak@princeton.edu

[4–6], suggesting that the normal extremal controls dominate the critical point topology of these problems.

As such, attention has focused on characterizing the system-independent (or kinematic) contribution to the critical topology of quantum optimal control problems [7–9]. For a given quantum system, several different classes of quantum control problems may be envisioned, each of which has its own associated kinematic critical topology. Besides the distinction between quantum gate and observable control, observable control itself may be subdivided into several qualitatively distinct optimization problems, depending on the initial state (density matrix) of the system and the type of observable to be maximized. Mathematically, the critical topologies of observable control problems can be classified on the basis of the eigenvalue spectra of the initial density matrix ρ_0 and the observable operator Θ [8]. To date, however, the system-independent contribution to the computational complexity of these problems has not been addressed. Here, we show that this complexity can, remarkably, be analytically determined.

Our primary goal is to not to establish bounds on the absolute complexity of each of these problems, but rather to assess the comparative expense of the distinct classes (for the same system Hamiltonian), across a wide variety of systems such that the dynamical contribution to the local gradient will have no preferential orientation.

II. CONTROL OPTIMIZATION AS A DYNAMICAL SYSTEM

A universal quantum optimal control cost functional can be written as:

$$J = \Phi(U(T), T) - \text{ReTr} \left[\int_0^T \left\{ \left(\frac{\partial U(t)}{\partial t} + \frac{i}{\hbar} H(\varepsilon(t)) U(t) \right) B(t) \right\} dt \right] - \lambda \int_0^T \frac{1}{s(t)} |\varepsilon(t)|^2 dt \quad (1)$$

where $B(t)$ is a Lagrange multiplier constraining the quantum system dynamics to obey the Schrödinger equation, $\varepsilon(t)$ denotes the time-dependent control field, $s(t)$ denotes the pulse envelope, and λ weights the importance of the penalty on the total field fluence. Solutions to the optimal control problem correspond to $\frac{\delta J}{\delta \varepsilon(t)} = 0$. The problem of maximizing the expectation value of an observable corresponds to:

$$\Phi_1(\varepsilon(\cdot)) = \text{Tr}(U(T)\rho_0 U^\dagger(T)\Theta) \quad (2)$$

where ρ_0 is the initial density matrix of the system and Θ is an arbitrary observable operator.

We are interested in the convergence to the global optimum of the gradient flow trajectories of the objective functions, which are the solutions to the differential

equations

$$\frac{d\varepsilon(s, t)}{ds} = \nabla \Phi(\varepsilon(t)) = \alpha \frac{\delta \Phi_{1,2}(s, T)}{\delta \varepsilon(s, t)} \quad (3)$$

where s is a continuous variable parametrizing the algorithmic time evolution of the search trajectory and α is an arbitrary scalar that we will set to 1. The complexity of quantum control optimization is associated with the scaling with system size of the rate of convergence of the function

$$\Phi_{1,2}(s, T) - \Phi_{1,2}(\infty, T) = \alpha \int_0^s ds' \int_0^T dt \left[\frac{\Phi_{1,2}(s', T)}{\delta \varepsilon(s', t)} \right]^2$$

to zero. The gradients $\frac{\delta \Phi}{\delta \varepsilon}$ can be shown to be, respectively [10–12]:

$$\frac{\delta \Phi_1}{\delta \varepsilon(t)} = \text{Tr}([\Theta(T), \rho_0] \mu(t)), \quad (4)$$

within the electric dipole approximation, where $\mu(t) = -\frac{i}{\hbar} U^\dagger(t, 0) \mu U(t, 0)$ is the time-evolved dipole operator of the quantum system and we have adopted the shorthand notation $U \equiv U(T)$.

The gradient on $\varepsilon(t)$ is related to the gradient on $\mathcal{U}(N)$ through

$$\frac{\delta \Phi}{\delta \varepsilon(t)} = \text{Tr} \left\{ \nabla_U \Phi(U(T)) \frac{\delta U^\dagger(T)}{\delta \varepsilon(t)} \right\} \quad (5)$$

$$= \sum_{p,q} \frac{\delta U_{pq}^*}{\delta \varepsilon(t)} \frac{\partial \Phi}{\partial U_{pq}}. \quad (6)$$

Now suppose that we have the gradient flow of $\varepsilon(s, t)$ that follows (3) and let $U(s)$, the system propagator at time T driven by $\varepsilon(s, t)$, be the projected trajectory on the unitary group $\mathcal{U}(N)$. The (algorithmic) time derivative of $U(s)$ is then

$$\frac{\partial U_{ij}(s)}{\partial s} = \int_0^T \frac{\delta U_{ij}(s)}{\delta \varepsilon(s, t)} \frac{\partial \varepsilon(s, t)}{\partial s} dt \quad (7)$$

which, combined with (3) and (5), gives

$$\frac{\partial U_{ij}(s)}{\partial s} = \int_0^T \frac{\delta U_{ij}(s)}{\delta \varepsilon(s, t)} \sum_{p,q} \frac{\delta U_{pq}^*(s)}{\delta \varepsilon(s, t)} \frac{\partial \Phi}{\partial U_{pq}} dt. \quad (8)$$

It is convenient to write this equation in vector form, replacing the $N \times N$ matrix $U(s)$ with the N^2 dimensional complex vector $\mathbf{u}(s)$:

$$\frac{d\mathbf{u}(s)}{ds} = \left[\int_0^T \frac{\delta \mathbf{u}(s)}{\delta \varepsilon(s, t)} \otimes \frac{\delta \mathbf{u}^\dagger(s)}{\delta \varepsilon(s, t)} dt \right] \nabla_{\mathbf{u}} \Phi[\mathbf{u}(s)] \equiv \mathbf{G}[\varepsilon(s, \cdot)] \nabla_{\mathbf{u}} \Phi[\mathbf{u}(s)] \quad (9)$$

(where the "vectorization" \mathbf{a} of a $N \times N$ -dimensional matrix A is given by $\mathbf{a}_{i+Nj} \equiv A_{ij}$). This relation implies

that the variation of the propagator in $\mathcal{U}(N)$ caused by the natural gradient flow in the space of control fields is Hamiltonian-dependent, where the influence of the Hamiltonian is contained in the N^2 -dimensional Hermitian matrix $G[\varepsilon(s, \cdot)]$. Within the electric dipole approximation, G is given by

$$\begin{aligned} G_{ij,pq} &= \int_0^T (U(T)\mu(t))_{ij} (\mu(t)U^\dagger(T))_{pq} dt \\ &= \int_0^T (U(T,t)\mu U(t,0))_{ij} (U^\dagger(t,0)\mu U^\dagger(T,t))_{pq} dt \end{aligned}$$

G is a $N^2 \times N^2$ -dimensional positive semidefinite Hermitian matrix. If $G = I$,

$$\frac{d\mathbf{u}(s)}{ds} = \nabla\Phi[\mathbf{u}(s)]$$

It can be shown (see below) that for any constant G -matrix, the flow corresponding to the differential equation $\frac{d\mathbf{x}}{ds} = G\nabla_{\mathbf{x}}\Phi(\mathbf{x})$ is completely integrable, and corresponds to the gradient flow for an assignment problem with a modified (rotated) weight vector \mathbf{c} . As such, for any distribution of G 's, it is possible to analytically determine the corresponding distribution of convergence times for the integrated flow trajectories.

Moreover, for each of the corresponding assignment problems, there exists a corresponding U -gradient flow that also solves this assignment problem (although there is no single complex G -matrix in the U -flow differential equation that will reproduce this gradient flow). If we use an \mathbf{x} of dimension N^2 , the Hilbert space dimension of the corresponding quantum system is N . In a second approach, we numerically integrated a subset of these U -gradient flows and compared the distribution in convergence times to that analytically determined for the \mathbf{x} -flows.

As we will discuss below, it is possible to eliminate G in this differential equation, and hence to explicitly follow the $U(T)$ gradient flow, by adopting an alternative algorithmic step on $\varepsilon(t)$. Importantly, the properties of the map $\varepsilon(\cdot) \rightarrow U(T)$ for finite-dimensional quantum systems render this fairly simple to achieve across many diverse systems, albeit with an error that is Hamiltonian-dependent. For the present purpose, we can effectively ignore G because we are interested in comparing the complexity scalings of the flow trajectories for gate and observable control problems on the same quantum system. Averaged over many initial conditions, the time required for convergence of the $U(T)$ gradient flows will determine this comparative scaling.

Because the kinematic gradient flows associated with these differential equations evolve on the continuous space of a Lie group (for Φ_2) or its adjoint orbit of skew-Hermitian matrices (for Φ_1), the optimization processes can be considered to be analog rather than discrete computations. In prior work [12], we demonstrated that these flow trajectories can in fact be viewed as dynamical

systems whose equations of motion guide the algorithmic time evolution of gradient-based algorithms. Although the infinite-dimensional gradient flow equations do not admit analytical solutions, the equations of motion for the gradient flow lines on the unitary group domain, namely

$$\frac{dU}{ds} = -U[\rho_0, U^\dagger\Theta U],$$

for Φ_1 and Φ_2 , respectively, can be exactly solved and we explicitly integrate them for all distinct cases of interest. The solutions to the optimal control problem are the critical points of the flow, i.e. $\nabla\Phi(U) = 0$, or equivalently, the equilibria of the dynamical system. We use the integrated gradient flow equations to derive upper bounds on the time for convergence to the solutions of these problems and thereby assign the problems to computational complexity classes.

The possible outputs of the dynamical system executing the analog computation are the attracting fixed points of the gradient flow of the control objective functional. Just as a discrete computation can be assigned a measure of time complexity, i.e. the scaling with system size (e.g., polynomial or exponential) of the time required to solve the problem, analog computations can be assigned time complexities, which correspond to the scaling with system size of the time required for the dynamical system to converge to a vicinity of the optimal fixed point. Time complexity of an analog computation can be analytically determined for systems that are completely integrable; i.e., for those systems whose equations of motion can be exactly solved. For this purpose, we apply the recently developed theory of complexity for continuous time dynamics [13, 14].

III. ANALYTICAL INTEGRATION OF GRADIENT FLOWS

For Φ_1 , the gradient flow evolves on a polytope whose dimension varies depending on the spectrum of ρ_0 . Rewriting the gradient flow as a quadratic function on the domain of Hermitian operators $\rho_T = U(T)\rho_0U^\dagger(T)$ [12], one can solve explicitly for a unique solution $\rho_T(s)$ corresponding to $J(s)$.

Since we are isolating the linear contribution to the control optimization problem (the part constant across arbitrary system Hamiltonians and final dynamical times), we reframe our problem in the language of linear programming (LP), where extensive work has been done on establishing analytical complexity bounds. Linear programming [15] is generally concerned with constrained linear optimization problems of the form:

$$\max \mathbf{c}^T \mathbf{x} : \mathbf{x} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^{m \times n}$ and $m \leq n$. We show that quantum observable optimization on the domain $\mathcal{U}(N)$

is a linear programming problem. Our first task is to identify A , \mathbf{b} , and \mathbf{c} . We write

$$\begin{aligned}\tilde{\rho} &= R^\dagger \rho R = \text{diag}\{\underbrace{\gamma_1, \dots, \gamma_1}_{p_1}; \dots; \underbrace{\gamma_r, \dots, \gamma_r}_{p_r}\}, \\ \tilde{\Theta} &= S^\dagger \Theta S = \text{diag}\{\underbrace{\lambda_1, \dots, \lambda_1}_{q_1}; \dots; \underbrace{\lambda_s, \dots, \lambda_s}_{q_s}\},\end{aligned}$$

where $\gamma_1 > \dots > \gamma_r$ and $\lambda_1 > \dots > \lambda_s$ are the distinct eigenvalues of ρ and Θ , respectively, and R, S are unitary transformations that diagonalize ρ, Θ respectively. Since R and S are unitary transformations, $\tilde{U} = R^\dagger U S$ is an automorphism on $\mathcal{U}(N)$. Thus we can write:

$$J(\tilde{U}) = \text{Tr}[(R^\dagger U S) \tilde{\rho} (R^\dagger U S)^\dagger \tilde{\Theta}] = \text{Tr}(\tilde{U} \tilde{\rho} \tilde{U}^\dagger \tilde{\Theta}).$$

Since F is one-to-one, there is a one-to-one mapping between the gradient flows of $J(\tilde{U})$ and $J(U)$ and in particular, they converge to the solution in the same time. Thus, without loss of generality, we will assume that ρ and Θ are diagonal, and redefine $U \rightarrow \tilde{U}$ since this simplifies the analysis.

In this case, we can express $J(U)$ as

$$J(\tilde{U}) = \sum_{i,j} |\tilde{U}_{ij}|^2 \gamma_i \lambda_j,$$

where only the (real) norms of the N^2 elements of \tilde{U} enter the expression. The parameters $|\tilde{U}_{ij}|^2$ can be assembled into an N^2 -dimensional vector, which we denote as \mathbf{x} with $\mathbf{x}_{(i-1)N+j} = |\tilde{U}_{ij}|^2$. Then, if we define the N^2 dimensional vector $\mathbf{c}_{(i-1)N+j} = \gamma_i \lambda_j$, we can write

$$J(\tilde{U}) = J(\mathbf{x}) = \mathbf{c}\mathbf{x},$$

which is in the correct form for linear programming. Note that we have $x_k > 0$ since our parameters are squared norms.

The unitarity of U imposes the following constraints on the parameters $|\tilde{U}_{ij}|^2$:

$$\begin{aligned}\sum_{j=1}^N |\tilde{U}_{ij}|^2 &= 1; \quad i = 1, \dots, N \\ \sum_{i=1}^N |\tilde{U}_{ij}|^2 &= 1; \quad j = 1, \dots, N.\end{aligned}$$

These constraints are not linearly independent, as can be seen from the fact that summing all equations in each set yields the identical constraint $|\tilde{U}_{11}|^2 + |\tilde{U}_{12}|^2 + |\tilde{U}_{13}|^2 + \dots + |\tilde{U}_{NN}|^2 = N$. There are a total $2N - 1$ independent constraints; it suffices to remove the first constraint in order to obtain a linearly independent set. The constraints can be expressed concisely in the form of the matrix equation $A\mathbf{x} = \mathbf{b}$, where $b_i = 1$, $i = 1, \dots, 2N - 1$ and

$$A_{ij} = \begin{cases} 1, & |i - j| \equiv 0 \pmod{N}, \\ 0, & \text{otherwise.} \end{cases}$$

With these definitions, we can write our complete kinematic optimization problem as a LP problem in the standard form above. Note that we have $m \leq n$, since $m = 2N - 1$ and $n = N^2$.

In the language of linear programming, this problem is isomorphic to the so-called *maximum weighted bipartite matching* (or assignment) problem, where the eigenvalues of ρ and Θ form bipartite sets and their respective products provide weights to "edges" connecting them; the goal is to find the matching (set of connecting edges) that maximizes the associated cost,

$$\max \sum_{i=1}^N \sum_{j=1}^N c_{ij} x_{ij}$$

where $\sum_{j=1}^N x_{ij} = 1$, $\sum_{i=1}^N x_{ij} = 1$, $x_{ij} \geq 0$, and $c_{ij} \in \mathbb{R}$. From this analogy, it is apparent that the constraint matrix A is the node edge incidence matrix of an undirected bipartite graph.

In linear programming problems, the constraint set $A\mathbf{x} = \mathbf{b}$ has a natural geometric interpretation in the form of a polytope, or closed polyhedron. The vertices of the polyhedron correspond to critical points, or possible solutions to the optimization problem. In general, a set of points defined by the intersection of a finite number of linear equations is termed a convex polyhedron. Specifically, the interior of the constraint set for observable maximization is more conveniently expressed in terms of $N \times N$ real matrices rather than vectors. Let B denote the $N \times N$ matrix defined by $B_{ij} = |\tilde{U}_{ij}|^2$; hence $\sum_{i=1}^N B_{ij} = 1$, $j \in 1, \dots, N$ and $\sum_{j=1}^N B_{ij} = 1$, $i \in 1, \dots, N$. Matrices of this type arise frequently in probability theory, and are called doubly stochastic or bistochastic matrices.

Definition 1 (*Doubly stochastic matrix*) *A doubly stochastic matrix is a square matrix of nonnegative real numbers, each of whose rows and columns sum to 1.*

We take A to be fixed, i.e., the polytope is fixed, while the degeneracy structures of ρ and Θ are reflected in the vector $c = (c_1, \dots, c_m; 0, \dots, 0)$. The appearance of 0 in c is from the zero eigenvalues in ρ and Θ , representing the purity of the quantum states. So actually only x_1, \dots, x_m are effective in the N^2 -dimensional LP problem. There are degeneracies in c_1, \dots, c_m that affect the behavior of the gradient flow within the polytope; the effects of these degeneracies are discussed further below. Alternatively, A could be varied with the degeneracies in ρ, Θ . In this case the constraint set is mapped to a lower dimensional polytope due to the symmetries that exist in higher dimensions, and the LP problem is formulated as an m -dimensional problem where $c = (c_1, \dots, c_m)$. In selected cases (also considered below) this formulation provides additional geometric insight.

In order to establish the geometry of the boundary of the constraint set for observable maximization, within which the gradient flow evolves, we will need to following graph-theoretic result.

Theorem 1 (*Birkhoff-von Neumann*) [16, 17] *The convex constraint set (polytope) for the problem of observable maximization is the convex hull of all $N \times N$ permutation matrices, which is equivalent to the set of all $N \times N$ doubly stochastic matrices B over \mathbb{R} .*

Definition 2 (*Birkhoff polytope*) *The class of $N \times N$ doubly stochastic matrices is a convex polytope in \mathbb{R}^{N^2} known as the Birkhoff polytope, B_N [16]. Its dimension is $(N - 1)^2$, and it has $N!$ vertices.*

1. *Effects of rank-deficiency and eigenvalue degeneracy on the LP problem*

As discussed in the previous section, assuming either ρ or Θ is fully nondegenerate, degeneracies in the eigenvalues of the other matrix may merge critical manifolds and render the dynamics symmetric with respect to several parameters. They can alter the total number of parameters or constraints in the problem and hence alter the dimension of the space in which the constraint polytope resides, if we choose to use a variable constraint matrix A . Consider the case where $\gamma_1 = \gamma_2$. Then both the cost function and constraint equations are equivalent with respect to interchange of $|\tilde{U}_{i1}|^2$ and $|\tilde{U}_{i2}|^2$, $i = 1, \dots, N$, and the problem can be reformulated, e.g., by the reparameterization $|\tilde{U}_{i1}|^2 + |\tilde{U}_{i2}|^2 \rightarrow 2|\tilde{U}_{i1}|^2$, $i = 1, \dots, N$. As will be discussed further below, in addition to reducing the number of independent parameters and constraints, eigenvalue degeneracies can also cause the dynamics of the gradient flow to converge to an edge or face of the polyhedron and may alter the timescale for convergence to the optimal solution (if they correspond to the optimal vertices). If degeneracies exist in both matrices, this does not result in any further reduction in the number of independent parameters beyond that caused by the individual degeneracies. For example, if $\lambda_1 = \lambda_2$ and $\gamma_1 = \gamma_2$, the cost function is invariant under interchange of $|\tilde{U}_{12}|^2$ and $|\tilde{U}_{21}|^2$, but the constraint set is not invariant under such an interchange.

The number of parameters and corresponding number of constraints is also reduced if one of the matrices ρ , Θ is rank-deficient. When either ρ or Θ possess null eigenvalues, the dimension of the vector \mathbf{x} decreases from its maximal value of N^2 . If any λ_i , $\gamma_j = 0$, the number of parameters in the problem decreases by N and the number of constraints decreases by 1 since the parameters in the corresponding constraint equation disappear. For example, consider the case where $\lambda_1 = 0$ and all other eigenvalues are nonzero. Then the parameters $|\tilde{U}_{1j}|^2$, $j = 1, \dots, N$ are eliminated from the LP problem, along with the corresponding constraint $\sum_{j=1}^N |\tilde{U}_{1j}|^2 = 1$. Note also that the equality constraints $\sum_{i=1}^N |\tilde{U}_{ij}|^2 = 1$ are now replaced by the corresponding inequality constraints $\sum_{i=2}^N |\tilde{U}_{ij}|^2 \leq 1$. If both ρ and Θ have null eigenvalues, all equality constraints in the original LP problem are replaced by inequality constraints

in the reduced dimensional problem.

Although it is possible to hold the constraint matrix A and hence the polytope fixed in the presence of degeneracies or rank deficiencies, greater geometric insight can be obtained by projecting the polytope onto a lower dimensional space based on the associated symmetries. This reduced dimensional polytope resides in \mathbb{R}^{n^2} , where n is the number of independent parameters, is of dimension $\mathbb{R}^{(n-m)^2}$, where m is the number of constraints, and has a number of vertices equal to the number of ways of matching the eigenvalues of ρ , Θ where matchings involving null eigenvalues are treated as equivalent. For example, if ρ is fully nondegenerate and Θ has rank $N - k$, the number of vertices is $\frac{N!}{k!}$. For $k = N - 1$, the polytope is a simplex. Note that for the reduced dimensional polytope corresponding to rank deficient or degenerate ρ , Θ , the interior of the polytope will no longer be populated with doubly stochastic matrices. Can we further characterize the interior of the polytope?

These expressions for the total number of parameters and constraints as functions of eigenvalue degeneracies and the number of null eigenvalues will be required in the complexity calculations below.

2. *Basic and nonbasic sets*

By an appropriate change of basis, any vertex of the polytope can be written in terms of its basic and nonbasic sets: $x = (x_N, x_B)$. A basic solution to the LP problem is one where (under an appropriate change of basis), $n - m$ of the variables $x_i = 0$. Basic solutions correspond to vertices of the LP polytope. The basic solution, i.e., the vertex, that maximizes the objective function is called the optimal solution.

We define the $m \times (n - m)$ matrix A_N as the matrix whose column vectors are the basis vectors corresponding to the x_i that are equal to zero at a given vertex, and the corresponding $m \times m$ matrix A_B as the matrix whose columns correspond to the remaining x_i . A_N consists of eigenvectors $\mu_i = e_i + \sum_{j \in A_B} \alpha_{ji} e_j$, $i \in N$. Then, in this basis, the constraint matrix can be partitioned $A = (A_N, A_B)$, and $\mathbf{x}_B = A_B^{-1} \mathbf{b}$. Note it is possible that some of the elements of x_B can also be zero; this phenomenon, referred to as degeneracy, is discussed further below. For the observable maximization problem, for example, there will always be $|U_{ij}|^2$'s that are zero (zeros in the permutation matrix vertices) in the m -dimensional solution basis A_B .

3. *Analytical solution to the gradient flow equations*

The gradient flows corresponding to least squares (maximum weighted bipartite) matching problems were shown to be integrable by Brockett and Bloch [18], who also established the equivalence of these flows to the notion of a Lax pair from the theory of completely inte-

grable dynamical systems [19]. Specifically, it can be shown that the gradient flow of cost function ϕ_1 (However,) the formal analytical solution was established by Faybusovich, who studied the explicit expression for the gradient flow in terms of A , \mathbf{b} and \mathbf{c} from the LP formulation of the problem [20, 21]. The gradient flow of the cost function evolves on the interior of the polytope defined by the constraint set. The vector field introduced by Faybusovich is a projection of the gradient of linear cost function onto the constraint set, relative to a Riemannian metric that enforces the positivity constraints $x \geq 0$:

$$\text{grad } J(x) = [D(x) - D(x)A^T(AD(x)A^T)^{-1}AD(x)]c$$

where $D(\mathbf{x}) = \text{diag}\{x_1, \dots, x_n\} \in \mathbb{R}^{n \times n}$. Faybusovich gave the following formal expression for the integrated gradient flow:

$$x_i(s) = x_i(0) \exp \left(-\Delta_i s + \sum_{j \in A_B} \alpha_{ji} \log \frac{x_j(s)}{x_j(0)} \right)$$

where the index $i \in A_N$, A_B is the basic set for the optimal solution (with corresponding nonbasic set A_N), $\Delta_i = -c_i + \sum_{j \in A_B} \alpha_{ji} c_j$ and $\alpha_{ji} = |(A_B^{-1}A_N)_{ji}|$. This solution describes the evolution of the $n - m$ independent variables \mathbf{x}_N as they converge to zero, in terms of the m constrained variables \mathbf{x}_B . It can be shown that Δ_i is the inner product $\langle c_i, \mu_i \rangle$, where μ_i , $i \in A_N$ is a set of vectors perpendicular to the rows of A , and parallel to faces of polyhedron. An important feature of the Faybusovich solution is that it only applies to LP problems formulated on nondegenerate constraint polyhedra, i.e. polyhedra whose vertices have only nonzero components x_i . Although general observable maximization gradient flows do not satisfy this condition, a simple perturbation, described in the next section, enables the analytical solution to be applied.

We consider the special case where ρ represents a pure state (i.e., only one λ_i is nonzero) and Θ has an arbitrary spectrum. In this case, we have only one constraint ($m = 1$), and N variables ($n = N$) which we write as x_1, \dots, x_N . The constraint polyhedron in this case is nondegenerate, permitting direct application of the Faybusovich flow.

In the case that ρ_0 has only one nonzero eigenvalue, corresponding to an initial pure state, it then follows that under the change of variables $\rho_T(s) = |\psi(s)\rangle\langle\psi(s)|$, $|\psi(s)\rangle = (c_1(s), \dots, c_N(s))$, $x(s) \equiv (|c_1(s)|^2, \dots, |c_N(s)|^2)$, the gradient flow can be explicitly integrated to give [12]:

$$x(s) = \frac{(e^{2s\lambda_1}|c_1(0)|^2, \dots, e^{2s\lambda_N}|c_N(0)|^2)}{\sum_{i=1}^N |c_i(0)|^2 e^{2s\lambda_i}} \quad (10)$$

where $\lambda_1, \dots, \lambda_N$ denote the eigenvalues of Θ .

IV. ANALOG COMPLEXITY OF QUANTUM CONTROL OPTIMIZATION

In the case of objective functional Φ_1 , the input for the computation may be viewed as being the eigenvalues of the observable operator, while the initial density matrix determines the initial condition. For the purpose of assigning a complexity class to the problem, the latter is generally taken as fixed. The time complexity is then expressed as a function of the eigenvalues of Θ . However, as discussed below, the complexity of the optimal control problem can vary sharply as a function of the initial state as well. For objective functional Φ_2 , we consider the input for the computation to be the target unitary transformation W . The time complexity is expressed in terms of the eigenvalues of W . Again, time complexity can be modulated by varying the initial condition of the search away from the identity matrix I .

It is possible to assess the convergence of the optimal control search in terms of the difference in objective function values $\Phi(s) - \Phi(\infty)$, but because of the degeneracy of solutions corresponding to a given value of Φ , a more precise assessment of complexity can be achieved in terms of the convergence of the distance between the current guess and the global solution to the problem. As mentioned, the distance $\|E(s) - E(\infty)\|^2 = \int_0^T E(s, t)E(\infty, t)dt$ is not an appropriate measure because it is highly system-specific (precluding uniform behavior as a function of system size) and obscures the underlying geometry of convergence since Φ is not explicitly a function of $E(t)$. The most appropriate choice is a distance on the space of solutions that displays the lowest degeneracy for a given value of Φ , while remaining independent of the system Hamiltonian. For Φ_1 , this distance is $\|U(s) - U(\infty)\|_F^2$, whereas for Φ_2 , it is the Euclidean distance on the polytope wherein the flow evolves.

The attracting region R of an attracting fixed point is the subset of phase space wherein the distance to the point is monotonically decreasing with time. We consider the quantum control optimization problem to be solved to a desired precision, and the computation halted, if the gradient flow trajectory enters within an ϵ_p -radius of the global optimum and is also within its attracting region. This definition ensures that for inputs on which the ϵ_p -vicinity of the attractor is larger than the attracting region, entrance into the attracting region is required for halting of the computation. The computational complexity of the control optimization algorithm is then defined as the scaling with system size of the convergence time $t_c(H) = \max[t_c(\epsilon), t_c(R)]$ to the intersection of the attracting region and the ball of radius ϵ_p , starting from a given initial condition.

A. Quantum observable maximization

From the Faybusovich flow, it is apparent that the computation time is determined primarily by the system size dependence of Δ_i and α_{ij} . Let $\beta_i = \sum_{j \in A_B} \left(\alpha_{ji} \log \frac{x_{j+n-m}(s)}{x_{j+n-m}(0)} \right)$. Then the computation time is roughly given by $T = \max_i \left(\frac{\beta_i}{\Delta_i} + \frac{|\log \epsilon|}{\Delta_i} \right)$. In order to obtain bounds on the complexity scaling of observable maximization problems, we therefore need to obtain bounds on β_i and $\min(\Delta_i)$. We first derive two results that enable analytical calculation of these bounds.

1. Unimodularity of the constraint matrix

In principle, it is possible for arbitrary LP problems to have exponential complexity. Analytical derivation of the complexity of observable maximization problems is simplified by a property displayed by the constraint matrix A called total unimodularity. An integer matrix A is totally unimodular if every square nonsingular submatrix B of A satisfies $\det B = \pm 1, 0$.

Theorem 2 *A sufficient condition for a totally unimodular (TUM) constraint matrix A is satisfaction of the following three conditions:*

- 1) Every column of A consists of two nonzero entries.
- 2) Every entry in A is 0, 1 or -1 .
- 3) If two nonzero entries in a column have same sign, the row of one is in B , and the row of other is in C , where B, C are two disjoint sets.

Moreover, for a TUM constraint matrix with only 0, 1 entries, all basic and nonbasic sets A_B, A_N corresponding to such a constraint matrix will have only 0, 1 entries.

Proof. Perform induction on size of (square) matrices. For single matrix entries, true by hypothesis. Let B be any submatrix of size k .

If B has a zero column, $\det B = 0$.

If B has no zero columns and has a column with one 1-entry, then we can expand its determinant along that column and the result follows from the induction hypothesis.

If B has only columns with two 1-entries, from hypothesis we have that

$$\sum_{i \in I_1} a_{ij} = \sum_{i \in I_2} a_{ij} = 1, \forall j$$

and therefore, we can find a linear combination of the rows that is equal to zero. Hence, $\det(C) = 0$. QED.

Moreover, A will remain unimodular under any change of basis. As such, the unimodularity of the constraint matrix A implies that the basic and nonbasic sets A_B, A_N at each of the vertices of the LP polytope will only have elements equal to zero or one.

2. Perturbation of degenerate Birkhoff polyhedra

Since the vertices of the Birkhoff polyhedron are permutation matrices, and since these matrices contain only n nonzero elements, it is apparent that each vertex of the polyhedron is associated with more than one A_B . A degenerate polyhedron is one with more than one A_B that defines a vertex. Hence, the Birkhoff polyhedron is degenerate. The Faybusovich vector field can only be used to integrate interior point flows for nondegenerate polyhedra. However, it is possible to infinitesimally perturb a degenerate polyhedron so that it becomes nondegenerate, while preserving the original solutions.

Finding extreme points of the convex set of solutions of $\sum_i x_i A_i = \mathbf{b}$ is equivalent to finding intersections of the ray $\alpha \mathbf{b} \mid \alpha \geq 0$ with hypertetrahedra formed on linearly independent sets of A_j from (A_i, \dots, A_n) . Geometrically, a degenerate situation is one in which the vector \mathbf{b} lies on a bounding hyperplane or edge of the convex cone determined by its vectors.

We need to perturb \mathbf{b} so that it lies on the interior of this convex cone. We achieve this by adding a vector function $\delta \mathbf{b}(\epsilon)$ of a real parameter ϵ to \mathbf{b} . This vector can be expanded on the basis of columns of the constraint matrix A . Each coefficient in this expansion must act on an entirely different scale from all the other coefficients and the data for the problem. Then, there will only be a finite number of values of ϵ for which the new ray hits boundaries of the tetrahedra. Thus for ϵ sufficiently small, degeneracy is avoided. The following standard procedure achieves this effect.

Theorem 3 [22] *Let A_j denote the column vectors of the constraint matrix A . Define x_{ij} by the relations $A_j = \sum_{i=1}^m x_{ij} A_i$, and let $z_j = \sum_{i=1}^m x_{ij} c_j$, $j = 1, \dots, n$. The constraint equations in the original LP problem can then be expressed $x_1 A_1, \dots, x_n A_n = \mathbf{b}$. Perturb \mathbf{b} by replacing it with $\mathbf{b} + \sum_{j=1}^n \epsilon^j A_j$. We then obtain a perturbed LP problem with constraint set*

$$x_1 A_1, \dots, x_n A_n = \mathbf{b} + \epsilon A_1, \dots, \epsilon^n A_n = \mathbf{b}(\epsilon),$$

and cost function $f_\epsilon(\mathbf{x}) \equiv \sum_i c_i x_i + \sum_{j=1}^n \epsilon^j z_j$.

Then every extreme point in the original problem has a counterpart in the new ϵ problem, and in the limit of small ϵ , these counterparts can be brought arbitrarily close to the original solutions while avoiding LP degeneracies.

See Supporting Information for proof.

3. Derivation of the complexity bound

For observable maximization, we first derive the complexity bound for control of pure states before considering the more general case of mixed states. For optimization of Φ_1 starting from an initial pure state, we can establish a bound on $t_c(\epsilon)$ as follows. We assume

without loss of generality that the eigenvalues of Θ are arranged such that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_N$. For nondegenerate Θ , the global optimum is then the unit vector $e_{i^*} = e_1$ and the distance of the current point $x(s)$ on the search trajectory to the global optimum can be written:

$$\begin{aligned} \|x(s) - x(\infty)\|^2 &= \\ \|x(s)\|^2 - 2 \frac{(e^{2s\lambda_1} |c_1(0)|^2, \dots, e^{2s\lambda_N} |c_N(0)|^2) \cdot e_1}{\sum_{i=1}^N |c_i(0)|^2 e^{2s\lambda_i}} + 1 \\ &\leq 2 - 2 \frac{e^{2s\lambda_1} |c_1(0)|^2}{\sum_{i=1}^N e^{2s\lambda_i} |c_i(0)|^2} \end{aligned} \quad (11)$$

since $\sum_{i=1}^N x_i = \sum_{i=1}^N c_i^2 = 1$ and therefore $\sum_{i=1}^N x_i^2 = \sum_{i=1}^N c_i^4 \leq 1$. Defining $\mu \equiv \lambda_1 - \lambda_2$, we then have

$$\sum_{i=1}^N e^{2s(\lambda_i - \lambda_1)} |c_i(0)|^2 \leq e^{-2\mu s} + |c_1(0)|^2$$

such that the bound on the distance to the solution becomes

$$\|x(s) - x(\infty)\|^2 \leq 2 - 2(1 + e^{-2\mu s} |c_1(0)|^{-2})^{-1}.$$

For simplicity, we choose the identity vector $\frac{1}{N}(1, \dots, 1)$ as the initial state. More generally, $x_{i^*}(0)$ will scale inversely with N for randomly chosen initial conditions on the interior of the simplex. We then obtain the following bound on the convergence time to the global optimum:

$$t_c(\epsilon) \leq \left\lceil \frac{\ln\left(\frac{\epsilon^2 |c_1(0)|^2}{2}\right)}{2\mu} \right\rceil = \frac{1}{2\mu} \ln\left(\frac{2N}{\epsilon^2}\right)$$

for ϵ small. Note that the time scale for convergence to the global optimum (the linearization of the coefficient in the exponential) is given by μ . This is equal to the lowest eigenvalue of the Hessian of the objective function, $\mathcal{H}_A(\rho(\infty)) = -\sum_{j \neq 1} (\alpha_{j1}^2 + \beta_{j1}^2)(\lambda_j - \lambda_1)$, where α_{jk}, β_{jk} are the real, complex parts of an arbitrary Hermitian matrix A , obtained in [8]¹. If the maximum eigenvalue of Θ is degenerate with multiplicity k , such that $\lambda_1 = \lambda_2 = \dots = \lambda_k$, and $\lambda_k > \lambda_j$, $j = k+1, \dots, N$, then the dynamics converges to the point $x_\infty = \frac{1}{k}(1, \dots, 1, 0, \dots, 0)$ [12]. In this case, the distance to the global optimum becomes

$$\|x(s) - x_\infty\|^2 \leq 2 - 2 \frac{e^{2s\lambda_1} \sum_{i=1}^k |c_i(0)|^2}{\sum_{j=1}^N e^{2s\lambda_j} |c_j(0)|^2}, \quad (12)$$

corresponding to

$$t_c(\epsilon) \leq \frac{1}{2\mu} \ln\left(\frac{2Nk}{\epsilon^2}\right).$$

A bound on the convergence time to the attracting region of the solution, $t_c(R)$, can be derived as follows for the general case of a k -fold degenerate maximal eigenvalue of Θ . Again, without loss of generality, we assume $\lambda_1, \dots, \lambda_k \equiv \lambda_{(1)} > \lambda_{k+1} > \dots > \lambda_N$. In the present case, $t_c(R) = t_c$ such that for all $s > t_c$,

$$\begin{aligned} \frac{d}{ds} \|x(s) - x(\infty)\|^2 &= \\ e^{2s\lambda_1} \sum_{i=1}^k |c_i(0)|^2 \sum_{j=1}^N e^{2s\lambda_j} |c_j(0)|^2 (\lambda_{(1)} - \lambda_j) &< 0. \end{aligned} \quad (13)$$

This corresponds to the condition $\dot{x}_i < 0$ for $i > k$ [12]. This condition holds iff $\sum_{j=1}^N \lambda_j x_j > \lambda_i$, $i = k+1, \dots, N$. Insertion of the analytical solution above gives

$$\sum_{j=1}^N \lambda_j |c_j(0)|^2 e^{\lambda_j s} > \lambda_{k+1} \sum_{j=1}^N |c_j(0)|^2 e^{\lambda_j s}.$$

For the purposes of obtaining a bound on t_c , we can rewrite this

$$\begin{aligned} \sum_{i=1}^k (\lambda_{(1)} |c_i(0)|^2 - \lambda_{k+1} |c_{k+1}(0)|^2) e^{\lambda_{(1)} t_c} &> \\ (N - k - 2) \lambda_{k+1} |c_{k+1}(0)|^2 e^{\lambda_{k+1} t_c}, \end{aligned} \quad (14)$$

which can be solved for $t_c(R)$ to give:

$$\begin{aligned} t_{c,1}(R) &\leq \frac{1}{\lambda_{(1)} - \lambda_{k+1}} \\ \ln \left\{ \frac{(N - k - 2) \lambda_{k+1} |c_{k+1}(0)|^2}{\sum_{i=1}^k [\lambda_{(1)} |c_i(0)|^2 - \lambda_{k+1} |c_{k+1}(0)|^2]} \right\}. \end{aligned} \quad (15)$$

Therefore, the attracting region of the solution $x(\infty) = \frac{1}{k}(1, \dots, 1, 0, \dots, 0)$ does not cover the entire domain. Again taking the initial condition $\frac{1}{N}(1, \dots, 1)$, this bound becomes $t_{c,1}(R) \leq \frac{1}{\mu} \ln \frac{(N-k-2)\lambda_{k+1}}{k(\lambda_{(1)} - \lambda_{k+1})}$. The upper bound on the computation time of the problem is then given by

$$\begin{aligned} t_{c,1}(H) &= \max [t_{c,1}(\epsilon), t_{c,1}(R)] \leq \\ \frac{1}{2\mu} \left[\ln\left(\frac{2Nk}{\epsilon^2}\right) + 2 \ln \frac{(N-k-2)\lambda_{k+1}}{k(\lambda_{(1)} - \lambda_{k+1})} \right]. \end{aligned} \quad (16)$$

Now let us consider the general case of observable maximization for mixed states.

Theorem 4 *The CT complexity of observable maximization with n, m given for various rank, degeneracy classes of ρ , Θ in section III 1 is to be completed...*

Proof.

Bound on $|\alpha_{ij}|$. $\alpha_{ij} = (A_B^{-1} A_N)_{ij}$ corresponds to the elements of a product of two matrices whose components are 1, 0, according to theorem 2. Therefore, if

¹ It can be shown that the expression derived in [8] for $\mathcal{H}_A(\mathcal{U}(\infty))$ is equivalent to that for $\mathcal{H}_A(\rho(\infty))$.

each column of A_B, A_N has at most k nonzero entries, $|\alpha_{ji}| \leq k \forall i \in A_B, j \in A_N$; here, $k = 2$, so we have established a bound on $|\alpha_{ji}| \leq 2$.

Bound on Δ_i . Next consider the bound on $\Delta_{i,min}$. Since $\Delta_i = \langle c_i, \mu_i \rangle = -c_i + \sum_{j \in A_B} \alpha_{ji} c_j$ is the dot product of a real-valued vector with an integer valued vector, it can be arbitrarily small in general, unlike $|\alpha_{ij}|$. Note, however, that since for the purpose of the complexity calculation we are assuming A_B is the optimal vertex, the integrality of α_{ij} implies that $\Delta_{i,min} \geq 0$. (If it is equal to zero, the flow converges to an edge or face of the polyhedron rather than a vertex.) {From the above expression, it is apparent that $\Delta_{i,min}$ will generally scale as the difference of m times the mean value of the products of the nonzero eigenvalues of ρ , Θ and the largest c_i , $i \in A_N$.} Because these eigenvalues are generally randomly distributed, their scaling cannot be considered in the complexity bound.

Bound on $\sum_{j=1}^m |\log x_j(\infty)|$.

Here again, ∞ denotes the parameter vector corresponding to the optimal vertex. Recall that we need to break the degeneracy in the polytope vertices. We therefore write $\mathbf{b}' = \mathbf{b} + \delta\mathbf{b}$, where from theorem 3, the perturbation that breaks the degeneracy of the polytope is $\delta\mathbf{b} = (\delta, \delta^2, \dots, \delta^m)$ (renaming $\epsilon \rightarrow \delta$ to avoid redundancy).

We have $\mathbf{x}' = \mathbf{x} + \delta\mathbf{x} = A_B^{-1}\mathbf{b} + A_B^{-1}\delta\mathbf{b}$. Since the optimal solution is a permutation matrix, the corresponding \mathbf{x} is a m -dimensional vector with exactly N elements equal to 1 and all others null. The rows of A_B can be written $\mathbf{a}^T \in \{0, 1\}^m$. To calculate a bound we assume the worst case for A_B^{-1} , i.e. that each row contains m elements equal to 1. Then we obtain

$$\delta\mathbf{x}_j = \sum_{i=1}^m \delta^i, \quad j = 1, \dots, m,$$

and \mathbf{x}' is the sum of this with a permutation vector. Thus, $\sum_{j=1}^m |\log x_j(\infty)| \leq \log N + m |\log(\sum_{i=1}^m \delta^i)| \leq \log N + m \log(m\delta^m)$.

Total bound. We can now apply the above bounds on $|\alpha_{ij}|$, Δ_i , and $\sum_{j=1}^m \log x_j(\infty)$ to calculate the complexity of observable maximization gradient flows. In order to avoid divergences near the boundary of the polytope, we choose the center of the polytope as the initial condition. In the center of the Birkhoff polytope is the van der Waerden matrix, all of whose entries are equal to $1/N$.

The distance between the current parameter vector at algorithmic time s and the optimal vertex $x(\infty)$ can be written:

$$\begin{aligned} \|x(s) - x(\infty)\|^2 &= \|x(s)\|^2 + m + 2\langle x(s), x(\infty) \rangle \\ &\leq 2m + 2\langle x(s), x(\infty) \rangle. \end{aligned} \quad (17)$$

Inserting the integrated flow into this equation (ignoring the contribution of the initial condition in the logarithm), and applying the various bounds estimated above, we have

$$\begin{aligned} x_i(s) &= \frac{1}{N} \exp \left(-\Delta_i s + \sum_{j \in A_B} \alpha_{ji} \log \frac{x_j(s)}{x_j(0)} \right) \\ &\leq \frac{1}{N} \exp \left(-\Delta_i s + 2(\log N + m \log m + m^2 \log \delta) \right) \end{aligned}$$

Assuming Δ_i does not scale, it can then be shown that $\|x(s) - x(\infty)\| < \epsilon$ if

TO BE COMPLETED; estimates including degeneracy-breaking suggest polynomial scaling.

So the complexity class is polynomial for any problem where the rank deficiency or number of degeneracies does not scale with the problem size. Only for those where the number of constraints m is constant with problem (system) size does the scaling become logarithmic.

4. Summary of rank, degeneracy effects on complexity

We give results for a few cases of special interest for illustrative purposes: (give explicit scaling with N and several choices of $n_1, \dots, n_r; m_1, \dots, m_s$.)

Degeneracies or rank deficiencies decrease the number of vertices and faces of the polyhedron; number of vertices change very quickly, but complexity is not sensitive to that rapid change

Note that the scaling is not determined by the number of free parameters or natural coordinates, but rather the number of constrained parameters.

B. Assignment to complexity classes

A continuous time problem is said to be in the complexity class CLOG (continuous log) if it has a polynomial number of variables (here, the system dimension) and a logarithmic time complexity. Kinematic optimization of objective functions Φ_1 , where ρ is a pure state, and Φ_2 therefore belong to the complexity class CLOG. CLOG is the analog counterpart of the classical complexity class NC₁, the class of problems that can be efficiently solved on a discrete parallel computer, meaning problems that are decidable in polylogarithmic time on a parallel computer with a polynomial number of processors. It is the lowest time complexity class, lying immediately below P (polynomial time complexity). For real control optimization algorithms evolving in discrete time, the kinematic component of these problems will have NC₁ complexity. By contrast, kinematic optimizations of objective function Φ_1 where ρ is a mixed state

belong to the complexity class CP, equivalent to the discrete complexity class P. The order of the polynomial scaling depends on the rank and degeneracy structure of ρ and Θ .

Quantum optimal control problems aimed at maximizing the expectation value of any observable operator starting from a pure state of the system all belong to the same analog time complexity class. However, their characteristic time scale for exponential convergence, when the optimization algorithm follows the gradient flow trajectories as faithfully as possible, differs in a predictable fashion. Specifically, the characteristic time scale for exponential convergence is given by $\min_{j \neq 1} |\lambda_j - \lambda_1|$, i.e., the magnitude of the difference between the largest and second largest eigenvalue of the observable operator. This corresponds to the magnitude of the smallest Hessian eigenvalue of the objective functional near the solution. As such, the rate of convergence for the problem of driving a pure initial state to a pure final state - which corresponds to $\lambda_1 = 1, \lambda_i = 0 \quad \forall i > 1$ above - is the greatest, such that this problem has the lowest computational complexity of all observable maximization problems.

Although the problems of pure state observable maximization and gate control belong to the same complexity class with respect to the scaling of the number of iterations required as a function of system size, they display distinct behavior in other regards. In particular, it can be shown [12] that for both problems, there exists a set of initial conditions from which the flow does not converge to the solution. For Φ_1 , initial conditions where $c_{i^*} = 0$ (i.e., where the initial state resides on the so-called basin boundary of the simplex) do not converge. The scaling of the convergence time as a function of the distance to this pathological initial condition differs for the two problems [12]. In these cases, the initial guess U_0 should be modulated to facilitate the convergence of gradient algorithms.

As such, a central issue concerns the accessibility of these lower bounds on scaling through the design of gradient-based algorithms in both the laboratory and in simulations.

V. DISCUSSION

The complexity results derived herein are based on analysis of the kinematic contribution to optimal control gradient flows, and have neglected the dynamical contribution of the $\varepsilon(\cdot) \mapsto U(T)$ map to the scaling of search effort. For an arbitrary Hamiltonian, if one averages over many initial conditions and target gates/states, the relative convergence times of gate and state control optimization should correspond to those of the kinematic gradient flows [12]. However, since the Hamiltonian, and hence the properties of the $\varepsilon(\cdot) \mapsto U(T)$ map must change with system dimension, the absolute scaling for

either problem may in principle deviate from that predicted here.

From a practical perspective, since the kinematic complexity of these problems is either logarithmic or polynomial, the scaling of control search effort should not prohibit the efficient application of gradient-based algorithms to high-dimensional systems. For state-to-state coherence transfer problems, this prediction is borne out by both simulations and experimental evidence [23]. However, the reported scaling of gate optimization simulations [2] is worse than that predicted according to the present theory, suggesting that a reexamination of current quantum gate control algorithms is warranted. For control problems whose kinematic complexity class exceeds CLOG, it may be advantageous to apply matrix tracking algorithms (below) rather than gradient-based algorithms in order to achieve the above lower bounds. Evidence suggests that observable maximization problems starting from an initial nondegenerate mixed state belong to this category [24].

For general ρ_0 and Θ , the gradient flow for objective function Φ_1 is an "isospectral" flow [20, 25, 26]. In N dimensions, this flow has N integrals of the motion that are in involution, which is the classical definition of complete integrability for a dynamical system. From the point of view of the modern theory of integrable systems, the double bracket flow can be shown to represent a type of Lax pair, a general form that can be adopted by all completely integrable dynamical systems [19]. For this reason [12], the system-independent complexity of all discrete quantum observable maximization problems can in principle be analytically determined, as shown above. In contrast, the kinematic gradient flows of the gate and observable control problems in classical mechanics [4, 27] are not integrable, and hence cannot be assigned to analytic complexity classes.

In the present work, we have focused attention on the complexity of optimizations that employ local gradient search algorithms. The favorable scaling of the convergence times for kinematic gradient flows, and the scarcity of abnormal extremals in discrete quantum control problems [4] motivate a rigorous definition of the universal complexity of quantum control problems in terms of the scaling of the expense of tracking such kinematic paths through elimination of the matrix G in the algorithmic step (9). These global algorithms require an additional computational overhead of N^4 in order to invert the matrix G , as well as overhead for statistical estimation of the states or dynamical propagators based on experimental observations [4]. Assignment of quantum control problems to complexity classes based on the convergence of such global algorithms must account for deviations from the desired track induced by properties of $\varepsilon(\cdot) \mapsto U(T)$ that vary across homologous families of Hamiltonians. Such assignment, which requires extensive sampling over Hamiltonian space, is the subject of a separate work.

-
- [1] M. Shapiro and P. Brumer, *Phys. Rep.* **425**, 195 (2006).
- [2] Palao and R. Kosloff, *Phys. Rev. Lett.* **89**, 1883011 (2002).
- [3] M. Mohseni, A. Reza khani, and D. Lidar (2007), [quant-ph/0702131v1](#).
- [4] R. Chakrabarti and H. Rabitz, *Int. Rev. Phys. Chem.* **26** (2007).
- [5] R. Wu and H. Rabitz, in preparation (2007).
- [6] D. D'Alessandro and M. Dahleh, *IEEE Trans. Autom. Control* **46**, 866 (2001).
- [7] H. Rabitz, M. Hsieh, and C. Rosenthal, *Science* **303**, 1998 (2004).
- [8] H. Rabitz, M. Hsieh, and C. Rosenthal, *J. Chem. Phys.* **124**, 204107 (2006).
- [9] H. Rabitz, M. Hsieh, and C. Rosenthal, *Phys. Rev. A* **72**, 052337 (2005).
- [10] T.-S. Ho and H. Rabitz, *J. Photochem. Photobiol. A* **180**, 226 (2006).
- [11] T. Ho, In preparation (2007).
- [12] R. Chakrabarti, R. Wu, and H. Rabitz, In preparation (2007).
- [13] H. Siegelmann, A. Ben-Hur, and S. Fishman, *Phys. Rev. Lett.* **83**, 1463 (1999).
- [14] A. Ben-Hur, H. Siegelmann, and S. Fishman, *J. Complexity* **18**, 51 (2002).
- [15] R. Saigal, *Linear Programming* (Kluwer Academic, Boston, 1995).
- [16] G. Birkhoff, *Univ. Nac. Tucuman Rev. A* (1946).
- [17] J. von Neumann, in *Contributions to the Theory of Games*, edited by H. W. Kuhn and A. W. Tucker (Princeton University Press, Princeton, NJ, 1956), pp. 247–254.
- [18] A. Bloch, R. Brockett, and T. Ratiu, *Commun. Math. Phys.* **147**, 57 (1992).
- [19] O. Babelon, D. Bernard, and M. Talon, *Introduction to classical integrable systems*, vol. 60 of *Cambridge Monographs on Mathematical Physics* (Cambridge, Cambridge, 2003).
- [20] L. Faybusovich, *Physica D* **23**, 309 (1991).
- [21] L. Faybusovich, *IMA J. Math. Control Inform.* **8**, 135 (1991).
- [22] A. Charnes, *Econometrica* **20**, 135 (1952).
- [23] K. Moore, M. Hsieh, and H. Rabitz, To be submitted (2007).
- [24] G. Riviello, M. Hsieh, and H. Rabitz (2007).
- [25] A. Bloch, *Contemp. Math.* **114**, 77 (1990).
- [26] A. Bloch, *Hamiltonian and Gradient Flows, Algorithms and Control*, vol. 3 of *Fields Institute Communications* (Oxford University Press, Oxford, 1995).
- [27] R. Wu, R. Chakrabarti, and H. Rabitz (2007), [quant-ph/0708.2118](#).