Optimal Control for Quantum Optimization of Closed and Open Systems

Lorenzo Campos Venuti, Domenico D'Alessandro, and Daniel A. Lidar

Phys. Rev. Applied **16**, 054023 — Published 11 November 2021

DOI: 10.1103/PhysRevApplied.16.054023

Optimal Control for Closed and Open System Quantum Optimization

Lorenzo Campos Venuti,[1,2] Domenico D'Alessandro,[3] and Daniel A. Lidar[1,2,4,5]

[1]*Department of Physics & Astronomy, University of Southern California, Los Angeles, California 90089, USA*
[2]*Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*
[3]*Department of Mathematics, Iowa State University, Ames, Iowa 50014, USA*
[4]*Department of Electrical & Computer Engineering, University of Southern California, Los Angeles, CA 90089, USA*
[5]*Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA*

(Dated: August 13, 2021)

Optimization is one of the key applications of quantum computing where a quantum speedup has been an eagerly anticipated outcome. A promising approach to optimization using quantum dynamics is to consider a linear combination $s(t)B + (1 - s(t))C$ of two noncommuting Hamiltonians B and C, where C encodes the solution to the optimization problem in its ground state, B is a Hamiltonian whose ground state is easy to prepare, and $s(t) \in [0, 1]$ is the bounded "switching schedule" or "path", with $t \in [0, t_f]$. This approach encompasses two of the most widely studied quantum optimization algorithms: quantum annealing [QA; continuous $s(t)$] and the quantum approximate optimization algorithm [QAOA; piecewise constant $s(t)$]. While it is notoriously difficult to prove a quantum advantage for either algorithm, it is possible to compare and contrast them by finding the optimal $s(t)$. Here we provide a rigorous analysis of this quantum optimal control problem, entirely within the geometric framework of Pontryagin's Maximum Principle of optimal control theory. We extend earlier results, derived in a purely closed system setting, to open systems. This is the natural setting for experimental realizations of QA and QAOA. In the closed system setting it was shown that the optimal solution is a "bang-anneal-bang" schedule, with the bangs characterized by $s(t) = 0$ and $s(t) = 1$ in finite subintervals of $[0, t_f]$, in particular $s(0) = 0$ and $s(t_f) = 1$, in contrast to the standard prescription $s(0) = 1$ and $s(t_f) = 0$ of QA. As an example, we prove that for a single spin-$1/2$, the optimal solution in the closed system setting is the bang-bang schedule, switching midway from $s \equiv 0$ to $s \equiv 1$. For finite-dimensional environments and without any approximations we identify sufficient conditions ensuring that either the bang-anneal, anneal-bang, or bang-anneal-bang schedules are optimal, and recover the optimality of $s(0) = 0$ and $s(t_f) = 1$. However, for infinite-dimensional environments and a system described by an adiabatic Redfield master equation we do not recover the bang-type optimal solution. In fact we can only identify conditions under which $s(t_f) = 1$, and even this result is not recovered in the fully Markovian limit, suggesting that the pure anneal-type schedule is optimal. Our open system results have implications for the use of experimental quantum information processors, which are by necessity noisy, and suggest that in this practical sense the optimal schedules for quantum optimization are likely to be continuous.

I. INTRODUCTION

There is a great deal of interest in optimization algorithms that can be run on today's noisy intermediate scale quantum (NISQ) information processors [1], but so far relatively little is known in the way of guarantees of a quantum advantage. The reason is that quantum optimization algorithms are by and large heuristic, which often precludes a rigorous analysis. Two of the most promising examples are quantum annealing (QA) [2] and the quantum approximate optimization algorithm (QAOA) [3]. Both algorithms switch between two non-commuting Hamiltonians: a "driver" (or "mixer") B and a "target" (or "problem") C. The latter encodes the solution to the optimization problem as its ground state. The two algorithms can be viewed as complementary: QA switches continuously between B and C, while QAOA switches discretely; hence they are particularly well suited for analog and gate-model quantum computing devices, respectively. In addition, both algorithms are related to the quantum adiabatic algorithm [4], which is guaranteed by the venerable adiabatic theorem [5] to converge to the optimal solution in the limit of arbitrarily long evolution times [6–8]. Exactly because this guarantee is too demanding, i.e., since one would like to find paths to the ground state of C (or to a suboptimal solution that is close to it) without having to wait too long, alternative algorithms such as QA and QAOA have been put forth. QA relaxes the strict adiabaticity condition while retaining continuity of the switching "schedule", or "path" [2]. QAOA is based on a parametrized square-pulse ansatz for the path between B and C; the parameters are optimized variationally. The adiabatic algorithm becomes an instance of QAOA when the continuous evolution is replaced by pulsed segments ("Trotterized") [3, Sec. VI]. There have been numerous studies of these algorithms, including some that have compared them, with mixed results [9–12].

While rigorous statements about a quantum advantage using QA or QAOA are notoriously hard to prove, it is possible to address the question of optimality of the switching schedule, and hence to formulate rigorous results about the *relative* performance of the two algorithms. Indeed, in essence the question of which algorithm performs best boils down to an optimization of this schedule. Various results have already been established within the framework of QA, QAOA, and also the adiabatic algorithm. For example, it is well known that the latter can benefit from schedule optimization, even to an extent that can affect whether it provides a quantum speedup or not, as in the case of the Grover search problem [13, 14]. It has also been established that a variational

approach can optimize the adiabatic schedule [15]. Likewise, optimality results are known for QA [16, 17] and QAOA [18]. A natural question is whether one can jointly treat QA and QAOA under a single schedule optimization framework. The first such attempt was made by Yang *et al.* [19] using the framework of the Pontryagin Maximum Principle (PMP) of optimal control theory [20] (see also Refs. [21–23]). Yang *et al.* concluded that a strict QAOA-type discrete switching schedule is favorable to the continuous schedule of QA, and analyzed both the setting of a closed system undergoing purely unitary dynamics, and the setting of an open system subject to decoherence. However, this conclusion was later argued to be overly restrictive by Brady *et al.* [24], who showed that in general a "bang-anneal-bang" (hybrid continuous-discrete) path is optimal, in essence establishing a compromise between QA and QAOA.

The results of Brady *et al.* were obtained in the closed system setting of purely unitary dynamics. Their results would suggest that quantum computers should be built so as to simultaneously support discrete and continuous protocols. Here we reexamine these results, first, by redoing the analysis rigorously within the framework of the Pontryagin Maximum Principle, and second, by embedding the analysis within the framework of open quantum systems. The importance of the latter cannot be overstated, since it describes the realistic setting of noisy quantum computers coupled to a decohering environment. We find that while the Brady *et al.* conclusion largely holds up under scrutiny in the closed system setting (up to some important technical caveats), the situation is different in the open system setting. Namely, the finite "bang"-like initial and final segments of the "bang-anneal-bang" path shrink to points in the open system case, or disappear entirely, depending on assumptions made about the environment. The more common scenario is disappearance, i.e., "bang-anneal-bang" essentially reduces to "anneal". Thus, our work offers – in the context of optimization – a case for the development of analog quantum annealers over discrete gate-based quantum computers.

We proceed to first provide the general background for the problem, after which we outline the structure of the rest of the paper.

II. BACKGROUND

The closed system setting involves a system evolving unitarily in a d-dimensional Hilbert space $\mathcal{H}$ subject to the Schrödinger equation:

$$\frac{d}{dt}|\psi(t)\rangle = -iH(t)|\psi(t)\rangle, \qquad |\psi(0)\rangle = |\psi_0\rangle\ . \tag{1}$$

The protoypical quantum annealing problem concerns finding the optimal schedule $s(t) \in [0,1]$ for the time-dependent Hamiltonian given by [25]

$$H(t) = s(t)B + (1-s(t))C\ , \quad t \in \mathcal{I} \tag{2a}$$
$$= C + s(t)(B-C)\ . \tag{2b}$$

The *control interval* is $\mathcal{I} = [0, t_f]$. Often the Hermitian operator C is an Ising-type Hamiltonian of the form $\sum_{i=1}^n h_i\sigma_i^z + \sum_{i<j}^n J_{ij}\sigma_i^z\sigma_j^z$ (where h_i and J_{ij} are local longitudinal fields and couplings, respectively, and σ_i^z is the Pauli matrix acting on the i'th qubit), and the Hermitian operator B is a transverse field of the form $\sum_{i=1}^n \sigma_i^x$ [2]. For our purposes it only matters that $[B, C] \neq 0$.

The initial state $|\psi_0\rangle$ is assumed to be the ground state of B, and in both QA and QAOA the target state is the ground state of C. A relaxation of this, which we consider as the objective in the present work, is to minimize the expectation value of C at a given final time t_f, i.e.,

$$J := \langle\psi(t_f)|C|\psi(t_f)\rangle\ . \tag{3}$$

Minimizing J is equivalent to minimizing the energy of the C Hamiltonian, and if the global minimum is found then this corresponds to finding the ground state of C (i.e., solving the optimization problem defined by $\{h_i, J_{ij}\}$ when C is in Ising form).

It is known in quantum control theory (see, e.g., Ref. [26]) that if $\hat{\mathfrak{L}}$ is the Lie algebra generated by B and C and $e^{\hat{\mathfrak{L}}}$ the corresponding Lie group, assumed to be compact, the set of states reachable from $|\psi_0\rangle$ with free final time t_f is

$$\mathfrak{R} := \{X|\psi_0\rangle \,|\, X \in e^{\hat{\mathfrak{L}}}\}, \tag{4}$$

so that the absolute minimum of the cost J is

$$J_{\min} = \min_{|\psi\rangle\in\mathfrak{R}} \langle\psi|C|\psi\rangle\ . \tag{5}$$

If the dynamical Lie algebra $\hat{\mathfrak{L}}$ is the whole $su(d)$ then any state in the Hilbert space can be reached (starting from any other state), in particular the ground state of C, in which case the system is said to be *controllable* [27]. However, requiring full controllability may be overly restrictive, as we only need to reach a particular state. The following is a simple generalization that provides a sufficient condition for reaching the ground state.

Proposition 1. *Suppose* $[B, P_0] = [C, P_0] = 0$ *where* P_0 *is an orthogonal projector. The Hilbert space decomposes according to the block structure* $\mathcal{H} = \mathrm{Ran}(P_0) \oplus \mathrm{Ker}(P_0)$. *The Lie algebra generated by* B *and* C, $\hat{\mathfrak{L}}$, *must have the same block structure. Suppose that, according to this structure,* $\hat{\mathfrak{L}} = su(d_0) \oplus \hat{\mathfrak{L}}_1$ *with unspecified* $\hat{\mathfrak{L}}_1$ *and* $d_0 = \dim(\mathrm{Ran}(P_0))$*; then, if the initial state belongs to* $\mathrm{Ran}(P_0)$, *any state in* $\mathrm{Ran}(P_0)$ *can be reached (in finite time).*

The proof is self-evident, since the full controllability result [27] is now applicable in $\mathrm{Ran}(P_0)$. This generalization can be applied, for example, in case both B and C commute with a third operator, say M, and one knows to which sector of M the ground state of C belongs; see Appendix A for an example. In any case, it is clear that something must be assumed in order to guarantee the reachability of the ground state of C. Clearly, a necessary condition is $[B, C] \neq 0$, but even when $[B, C] \neq 0$ it is easy to come up with examples where the ground state of C cannot be reached; see Appendix

A. In the following we will tacitly assume that conditions are such that the ground state of C can be reached.

Brady *et al.* [24] used optimal control methods to prove that for the cost as defined as in Eq. (3), the optimal schedule is one where at the beginning and end of the control interval $s \equiv 0$ and $s \equiv 1$, respectively [28]. From a quantum annealing perspective this might appear as a counterintuitive result, since it means that rather than the usual "forward" formulation of quantum annealing [29, 30], where one interpolates smoothly from $H(0) = B$ to $H(t_f) = C$, the optimal protocol starts from the system being in the ground state of B but the initial Hamiltonian is C, and the final Hamiltonian is not C but rather B. The result, however, can be understood by noting that in the adiabatic approach, one interpolates so slowly from $H(0) = B$ to $H(t_f) = C$ that the system always remains in the ground state. Instead, in optimal control, we optimize over the set of possible states obtained by applying either C or B to the initial state, in a continuous fashion. In this sense, applying B at the beginning is a waste of time as it does not change the initial state. Applying C at the end, when the system is supposed to be close to the ground state of C, is similarly wasteful. This relaxation of the approach of strict adiabaticity is in line with other alternatives, such as shortcuts to adiabaticity [31, 32] and diabatic quantum annealing [33].

More precisely, Ref. [24] showed, provided that a certain condition holds (see below), that the optimal control function starts (ends) with $s \equiv 0$ ($s \equiv 1$) in an interval of positive measure after $t = 0$ (before t_f). Elsewhere the optimal control $s(t)$ is "singular", except for possible interruptions by a sequence of "bang" controls, where $s \equiv 0$ or $s \equiv 1$. In control theory a "singular" interval or arc, is an interval of time where the PMP control Hamiltonian in Eq. (10) below does not depend on the control s. The remaining "nonsingular" arcs give rise to the "bang" controls. In the numerical simulations of Ref. [24], the control appeared to be continuous (even smooth) on such singular arcs. Hence the term "anneal" was used in lieu of "singular", with the intention of stressing the continuous (or possibly even smooth) nature of the control on the singular arcs. They suggestively called the resulting optimal control a "bang-anneal-bang" protocol. At present, a rigorous proof that the control function is continuous (let alone smooth) on singular arcs is lacking, and there is some risk of confusion in interpreting the singular arcs as always being continuous, or even differentiable as is typically assumed in QA and adiabatic quantum computing [6, 29]. Nonetheless, while keeping these caveats in mind, we shall adopt the same (numerically supported) terminology as Ref. [24], and use "continuous (or anneal) = singular" as well as "bang = nonsingular" interchangeably.

Here, we consider the *open system* version of the same optimal control problem. We reformulate the problem in terms of the density matrix ρ, whose dynamics is described by the following, rather general master equation [34]:

$$\dot{\rho} = \mathcal{L}\rho, \qquad \rho(0) = \rho_0 \,, \tag{6}$$

where the Liouvillian $\mathcal{L}$ depends linearly on the control s (and the controlled operators B, C). Note that the Liouvillian is not explicitly time-dependent (i.e. $\partial_t \mathcal{L} = 0$, $\forall t \in \mathcal{I}$) and depends on time only through the switching schedule $s(t)$. This is an important requirement that will play a crucial role in our ability to apply the Pontryagin principle in the form we need, as we discuss in more detail below. Furthermore, to be physically meaningful, $\mathcal{L}$ must preserve hermiticity, i.e., $[\mathcal{L}(X)]^\dagger = \mathcal{L}(X^\dagger)$, $\forall X$. Instead of Eq. (3), the cost J takes the form

$$J := \mathrm{Tr}\left[C\rho(t_f)\right] = \langle C, \rho(t_f)\rangle \,, \tag{7}$$

where we used the Hilbert-Schmidt scalar product $\langle X, Y\rangle := \mathrm{Tr}(X^\dagger Y)$ for operators X, Y. We shall see that a description and treatment of the optimal control problem in the setting of the density matrix is not only more general but also more elegant since the cost J is linear in the state rather than quadratic, as in Eq. (3). Moreover, we obtain the closed system result as a special case. Unlike Ref. [24], which used a mixture of the PMP and a variational (Lagrange multiplier type of) argument, we use only the PMP, which significantly simplifies the proof.

The rest of this paper is organized as follows. In Sec. III, we apply general results from optimal control theory and the necessary conditions of the PMP to the problem of minimizing J [Eq. (7)] for a given final time t_f and the general dynamical system of the form of Eq. (6). In Sec. IV we specialize to the case of closed systems, which are described by the von Neumann equation. In particular, we confirm but also sharpen the results of Ref. [24]. We also analyze in depth the optimal control problem of a single spin-1/2, and prove that the optimal schedule is of the bang-bang type. In Sec. V we consider the case of open systems. This includes both the most general case of a reduced description of quantum system obtained by tracing out the environment it is coupled to, and the case where the open quantum system is described by adiabatic master equations, both non-Markovian and Markovian. In Sec. VI we derive a "switching equation," which allows us to provide a general characterization of the switches between non-singular and singular arcs, and derive conditions for the presence or absence of singular arcs. We also give a heuristic derivation of the shortening of the length of the arcs between two switches with increasing system size. We conclude in Sec. VII. In a series of appendices we provide additional background on optimal control theory and technical details and proofs of various results from the main text.

III. STATEMENT OF THE PONTRYAGIN MAXIMUM PRINCIPLE

We state the PMP as it applies to our problem of interest (see Appendices B and C):

Theorem 1. *Assume that ρ^* and s^* are, respectively, an optimal state and control pair for the problem defined by Eqs. (6) and (7) for a fixed final time t_f [35]. Then there exists a nonzero $n \times n$ Hermitian time-dependent matrix $p = p(t)$ called the* co-state *that satisfies [36]*

$$\dot{p} = -\mathcal{L}^\dagger p \,, \tag{8}$$

with the final condition

$$p(t_f) = -C \, . \tag{9}$$

Furthermore, define the PMP control Hamiltonian *function*

$$\mathbb{H}\left(p, \rho, s\right) := \langle p, \mathcal{L}\rho \rangle \, . \tag{10}$$

We then have the maximum principle*:*

$$\mathbb{H}\left(p(t), \rho^*(t), s^*(t)\right) = \max_{v \in [0,1]} \mathbb{H}\left(p(t), \rho^*(t), v\right) \, , \tag{11}$$

and there exists a real constant λ such that

$$\mathbb{H}\left(p(t), \rho^*(t), s^*(t)\right) \equiv \lambda \, . \tag{12}$$

A few remarks are in order.

- The PMP control Hamiltonian *function* (10) is, of course, different from the Hamiltonian *operator* (2) generating the dynamics.

- Since p and ρ are Hermitian and $\mathcal{L}$ is Hermiticity-preserving [$[\mathcal{L}\left(X\right)]^\dagger = \mathcal{L}\left(X^\dagger\right) \, \forall X$], "expectation values" of the form $\langle p, \mathcal{L}\rho \rangle$ are real, and hence so is the PMP control Hamiltonian (10).

- The condition $\partial_t \mathcal{L} = 0 \, \forall t \in \mathcal{I}$ must be satisfied and is implicit in Eq. (6). In other words, $\mathcal{L}$ may not depend explicitly on time. Without this condition Eq. (12) does not hold with a constant λ.

- It is worth highlighting that at the final time the co-state becomes the (negative of the) target Hamiltonian [Eq. (9)], a fact we use repeatedly in our applications of the Theorem 1 below.

- As discussed after Theorem 8 in Appendix B, if the optimal trajectory is such that the constraint of the final time t_f is "active", i.e., a small perturbation $t_f + \delta$ allows us to decrease the cost, then $\lambda > 0$ in Eq. (12). This is an important point that will be further discussed in the next section.

- Given that the PMP is formulated in terms of real-valued quantities in the optimal control literature (see Appendix B), one must first transform the relevant equations into real-valued ones. This can easily be done since the space of $n \times n$ Hermitian matrices is isomorphic to the space of n^2 real variables via coordinatization. We discuss this in Appendix C.

- Since $p(t)$ and $\rho(t)$ are solution of differential equations, they are continuous function of time. This implies that expressions of the form $\langle p, \mathcal{K}\rho \rangle$ with the superoperator $\mathcal{K}$ independent of time (both explicitly and implicitly), are continuous functions of t, a fact which we repeatedly and implicitly use below.

IV. THE CLOSED SYSTEM CASE

We first consider the closed system case. Let us define the superoperator

$$\mathcal{K}_X := -i \left[X, \bullet\right] \, . \tag{13}$$

Note that $\mathcal{K}_X$ is linear with respect to X. For Hermitian X, $\mathcal{K}_X$ is anti-Hermitian (see Appendix D):

$$\mathcal{K}_X^\dagger = -\mathcal{K}_{X^\dagger} = -\mathcal{K}_X \, . \tag{14}$$

The von Neumann equation corresponding to Eq. (1) is

$$\dot{\rho} = \mathcal{K}_C \rho + s(t) \left(\mathcal{K}_B \rho - \mathcal{K}_C \rho\right) \, , \quad \rho(0) = \rho_0 \, , \tag{15}$$

where henceforth we denote the initial and final conditions of operators X by $X(0) := X_0$ and $X(t_f) := X_f$, respectively. I.e., one has Eq. (6) with

$$\mathcal{L} = \mathcal{K}_C + s(t) \mathcal{K}_{B-C} \, . \tag{16}$$

Since in this case $\mathcal{L}^\dagger = -\mathcal{L}$, Eq. (8) tells us that the co-state matrix p satisfies the same equation as ρ:

$$\dot{p} = \mathcal{K}_C p + s(t) \mathcal{K}_{B-C} p \, , \tag{17}$$

but with the final condition (9). The PMP control Hamiltonian reads

$$\mathbb{H} = \langle p, \mathcal{K}_C \rho \rangle + s(t) \langle p, \mathcal{K}_{B-C} \rho \rangle \, . \tag{18}$$

A. The "bang-anneal-bang" protocol is optimal

Applying Theorem 1 to the anti-Hermitian superoperator $\mathcal{L}$ of Eq. (16), we obtain the following extension of the result of Ref. [24] to the density matrix setting:

Theorem 2. *(i) Assume $s^* \in [0,1]$ is the optimal control in an interval $[0, t_f]$ minimizing the cost (7) for Eq. (15). Then there exists a nonzero Hermitian matrix solution of Eq. (17) with terminal condition (9) such that $s^* \equiv 0$ on intervals where $\langle p, \mathcal{K}_{B-C}\rho \rangle < 0$, and $s^* \equiv 1$ on intervals where $\langle p, \mathcal{K}_{B-C}\rho \rangle > 0$. On all other intervals $\langle p, \mathcal{K}_{B-C}\rho \rangle \equiv 0$ (these are called singular arcs).*

(ii) Assume furthermore that the constraint on the final time t_f is active (so that $\lambda > 0$). Then $s^*(t) = 1$ for $t \in (t_f - \epsilon, t_f]$ for some $\epsilon > 0$. Moreover, if the initial condition ρ_0 commutes with the driver Hamiltonian B, i.e., $\mathcal{K}_B \rho_0 = 0$, one also has $s^*(t) = 0$ for $t \in [0, \epsilon')$ for some $\epsilon' > 0$.

Before proving this theorem we offer a few remarks.

- Part (i) implies that the optimal control is, in general, an alternation of "bang" (nonsingular) arcs and "anneal" (singular) arcs where $\langle p, \mathcal{K}_B \rho \rangle = \langle p, \mathcal{K}_C \rho \rangle$. Using the PMP, this is an immediate consequence of the fact that the control enters linearly in the equation and it is coupled to the superoperator $\mathcal{K}_{B-C}$. *The latter is what is "special" about the quantum annealing problem.*

- Part (ii) implies that under the assumption of an active time constraint and for a particular initial condition, the optimal control starts and ends with nonsingular arcs. In particular, it starts with an arc $s \equiv 0$ and ends with an arc $s \equiv 1$.

- In practice, whether there are additional nonsingular arcs in the middle is problem dependent, and there is numerical evidence that such optimal scenarios do indeed exist [24], but such nonsingular arcs do not exist in the single qubit example discussed in Subsection IV C.

- Assuming that $\mathcal{K}_B\rho_0 = 0$ one can prove $\lambda \geq 0$ (see also Appendix B, Proposition 1), but the condition $\lambda > 0$ is more subtle and it must be assumed independently. We return to this point in the next subsection.

Proof. Part (i): Eq. (18) states that the PMP control Hamiltonian $\mathbb{H}$ depends on the control only via the term $s(t)\langle p, \mathcal{K}_{B-C}\rho\rangle$. If $\langle p, \mathcal{K}_{B-C}\rho\rangle < 0$, then to maximize this term as per Eq. (11) subject to the constraint that $s(t) \in [0,1]$, clearly we must set $s^* \equiv 0$. Likewise, if $\langle p, \mathcal{K}_{B-C}\rho\rangle > 0$, then to maximize this term subject to the same constraint requires $s^* \equiv 1$. This is the case of nonsingular arcs. Conversely, if $\langle p, \mathcal{K}_{B-C}\rho\rangle \equiv 0$ (a singular arc), then we cannot conclude anything about the control from the PMP.

Part (ii): To investigate the form of the control at the end of the control interval $[0, t_f]$, consider Eq. (12) with $\lambda > 0$. Using Eq. (9) we have $\langle p_f, \mathcal{K}_C\rho_f\rangle = \langle \mathcal{K}_C^\dagger p_f, \rho_f\rangle = \langle \mathcal{K}_C C, \rho_f\rangle = 0$. This means that $\mathbb{H}(t_f) = s(t_f)\langle p_f, \mathcal{K}_B\rho_f\rangle = \lambda > 0$, which in turn, since $s \in [0,1]$, implies that $\langle p_f, \mathcal{K}_B\rho_f\rangle > 0$. By continuity there must exist an interval $(t_f - \epsilon, t_f]$ (for some $\epsilon > 0$) such that $\langle p(t), \mathcal{K}_{B-C}\rho(t)\rangle > 0$ for $t \in (t_f - \epsilon, t_f]$, and in this interval we must have $s^*(t) = 1$ by (i).

The argument for the initial time is similar but instead of Eq. (9) it uses the extra assumption $\mathcal{K}_B\rho_0 = 0$. Let us evaluate the control Hamiltonian at $t = 0$. Because of the assumption $\mathcal{K}_B\rho_0 = 0$ we have $\mathbb{H}(t=0) = (1 - s(0))\langle p_0, \mathcal{K}_C\rho_0\rangle = \lambda > 0$. Since $s \in [0,1]$ this implies that $\langle p_0, \mathcal{K}_C\rho_0\rangle > 0$ and $\langle p_0, \mathcal{K}_{-C}\rho_0\rangle < 0$. By continuity there must exist an $\epsilon' > 0$ such that, for $t \in [0, \epsilon')$, $\langle p(t), \mathcal{K}_{B-C}\rho(t)\rangle < 0$ and in this interval we must have $s^*(t) = 0$ by (i). □

B. The active constraint assumption and a sharpening of the results of Ref. [24]

The condition $\lambda > 0$ (that is, an active constraint on the final time t_f) requires some extra discussion. It is a known fact in the geometric theory of quantum control systems, and it follows as an application of general results on control systems on Lie groups (see, e.g., Ref. [27, Th. 7.2]), that there exists a *critical time* t_c such that, the set of states reachable at time t, coincides for every $t \geq t_c$. In other words, the reachable set does not grow past a certain time t_c. Therefore, for every $t_f \geq t_c$ the time constraint is never active. The minimum time $t_{\min}$ to reach the ground state of C is $\leq t_c$. If the final time t_f is greater than or equal to $t_{\min}$, then again the time constraint can never be active. In order to avoid this situation, it was claimed in Ref. [24] that having $t_f < t_{\min}$, is sufficient for having $\lambda > 0$ in Eq. (12). Their argument only uses $[\rho_0, B] = 0$. However, in Sec. IV C below we give an example satisfying this assumption for which $\lambda = 0$ for arbitrarily small t_f. Thus, the assumption $t_f < t_{\min}$ is certainly necessary for $\lambda > 0$ but is in fact not sufficient. Rather, $\lambda > 0$ is a feature of the optimal trajectory rather than of the problem itself. This can be explained more easily geometrically, as we now do.

The optimal cost at time t_f is the minimum of a continuous function on the reachable set of states $\mathfrak{R}_{t_f}$ (see Appendix B). It is also known, under conditions that apply in our case, that the reachable set $\mathfrak{R}_{t_f}$ varies continuously with t_f [37]. We can map the space of Hermitian matrices ρ diffeomorphically to $\mathbb{R}^{n^2}$ (see Appendix C) and consider its reachable set there. Since the cost function (7) is linear on this set, the minimum occurs on the boundary. Therefore, the optimal trajectory is a curve starting from the initial condition ρ_0 and ending on the boundary of $\mathfrak{R}_{t_f}$. At the endpoint, the trajectory will have a tangent vector which indicates its future direction. Now $\lambda > 0$ if, going (infinitesimally) in that direction combined with an increase in the size of the reachable set $\mathfrak{R}_{t_f+\epsilon}$ for some small ϵ, will result in a reduced cost, and this is what we mean by the time constraint being active. If t_f is such that the reachable set does not increase at t_f, for instance if $t_f \geq t_c$, then clearly this is not possible and we must have $\lambda = 0$. However, it is also possible that the reachable set increases but not in a way to (strictly) decrease the cost, in particular the portion of the boundary where we landed might not move at all, or it might move but not in a direction that decreases the cost. This geometric discussion is illustrated with figures in Appendix E.

The phenomenon that the optimal cost does not decrease with an increasing final time t_f may occur even though t_f is arbitrarily small. Let us denote by $J_{\min}(t_f)$ the minimum cost (7) as a function of t_f. The example we provide below (Sec. IV C) shows that, even assuming $[\rho_0, B] = 0$, we can have $J_{\min}(t_f) = J_{\min}(0)$ for $t_f \in [0, \epsilon)$ and some $\epsilon > 0$, that is, the cost cannot be lowered for some time, independently of the control. However, under the additional assumption that ρ_0 is the *nondegenerate ground state of* B this does not happen, and we have the following theorem which we prove in Appendix F:

Theorem 3. *Assume that ρ_0 in Eq. (6) is the nondegenerate ground state of B. Then there exists an $\epsilon > 0$ such that, for every $t_f \in (0, \epsilon)$, $J_{\min}(t_f) < J_{\min}(0) = \mathrm{Tr}(C\rho_0)$.*

In other words, if we start from the nondegenerate ground state of B we can always decrease the cost for sufficiently small t_f. Note that this, however, does not prove that $\lambda > 0$. As we have explained, the condition $\lambda > 0$ is a condition about the optimal trajectory, and it is an open problem to find sufficient conditions such that every optimal trajectory satisfies the $\lambda > 0$ requirement for sufficiently small t_f.

C. Example: optimal control of a spin-$1/2$ particle

We now give an example showing that without the assumption that ρ_0 is the nondegenerate ground state of B, the cost (7) cannot be lowered even for arbitrarily small t_f's.

Consider a spin-$1/2$ particle (qubit) in a magnetic field. The model is given by Eq. (15) with $C = \frac{1}{2}\sigma^z$ and $B = \frac{1}{2}\sigma^x$. As an orthonormal, Hermitian operator basis we choose $F_i = \frac{1}{\sqrt{2}}\sigma_i$, where we denote the standard Pauli matrices $\sigma^x \equiv \sigma_1$ etc., i.e.:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{19}$$

They satisfy the su(2) commutation relations

$$[\sigma_1, \sigma_2] = 2i\sigma_3, \quad [\sigma_3, \sigma_1] = 2i\sigma_2, \quad [\sigma_2, \sigma_3] = 2i\sigma_1 . \tag{20}$$

We parametrize the density matrix as $\rho = \frac{1}{2}(\mathbb{1} + \boldsymbol{v}\cdot\boldsymbol{\sigma})$, where $\boldsymbol{v} \in \mathbb{R}^3$ is the Bloch vector ($\|\boldsymbol{v}\| \leq 1$) and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)^T$. The Bloch vector satisfies Eq. (15) where, using $\mathcal{K}_{ij} = \mathrm{Tr}[F_i \mathcal{K}(F_j)]$ (see Appendix C), we have

$$\mathcal{K}_B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathcal{K}_C = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{21}$$

Equivalently, the dynamics are given by the Bloch equation

$$\dot{\boldsymbol{v}} = \boldsymbol{\mathcal{M}}(s)\boldsymbol{v} \tag{22a}$$

$$\boldsymbol{\mathcal{M}}(s) = \begin{pmatrix} 0 & -(1-s) & 0 \\ 1-s & 0 & -s \\ 0 & s & 0 \end{pmatrix}. \tag{22b}$$

Geometrically, $\mathcal{K}_C$ is the infinitesimal generator of a counterclockwise rotation about the v_3 axis, while $\mathcal{K}_B$ is the infinitesimal generator of a counterclockwise rotation about the v_1 axis. For $s \in (0,1)$, $\boldsymbol{\mathcal{M}}(s) = (1-s)\mathcal{K}_C + s\mathcal{K}_B$ generates a counterclockwise rotation about an intermediate axis in the (v_1, v_3) plane. The cost (7) in this case becomes $J = \frac{1}{2}v_3$, i.e., it corresponds to the minimization of the v_3 component. Furthermore, let us assume for simplicity that the initial state ρ_0 is pure ($\|\boldsymbol{v}\| = 1$). There are only two such states compatible with the condition $[B, \rho_0] = 0$ (equivalently: $\mathcal{K}_B \boldsymbol{v}_0 = \boldsymbol{0}$): the σ_1 eigenstates, i.e., $\boldsymbol{v}_0 = (\pm 1, 0, 0)^T$.

Now, if the initial state is $\boldsymbol{v}_0 = (1,0,0)^T$, i.e., the excited state of B, then for sufficiently small t we have $v_3(t) \geq 0$ independently of the control $s \in [0,1]$ (see Appendix D). Therefore, an optimal control in $[0, t_f]$ for t_f small will be $s \equiv 0$ (which will keep the value of v_3 at zero). The value of the minimum cost $J_{\min}(t_f)$ is equal to $J(0)$ for any arbitrarily small t_f. The constraint on the final time is not active here, *even for arbitrarily small* t_f. As a consequence, in this case we cannot draw the conclusions of Theorem 2 following from the assumption $\lambda > 0$. On the other hand, for $\boldsymbol{v}_0 = (-1,0,0)^T$, which corresponds to the (nondegenerate) *ground state*, with sufficiently small t_f we can lower the cost according to Theorem 3. We prove in Appendix H that the optimal control in this case is a simple bang-bang protocol:

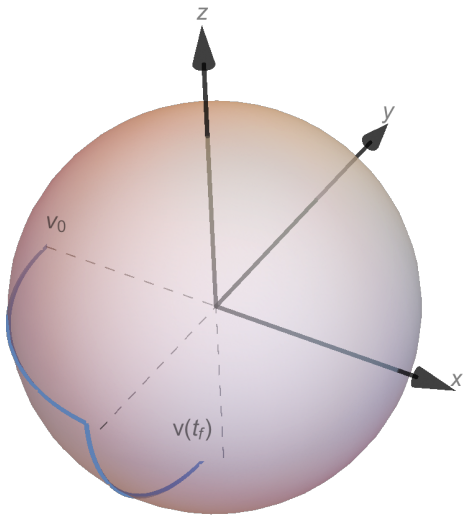

FIG. 1. (Color online) Single qubit case. Optimal trajectory on the Bloch sphere with initial condition $\boldsymbol{v}_0 = (-1,0,0)^T$. Here $t_f = 0.95\pi$ so the ground state of C, corresponding to the point $(0,0,-1)$, is not reached exactly.

Theorem 4. *The optimal control for the system of one spin-$1/2$ particle considered above, starting from the ground state and minimizing the cost* $\mathrm{Tr}(\sigma_3\rho)$ *in time* $t_f < \pi$*, is the sequence* $s^* \equiv 0$ *for time* $\frac{t_f}{2}$ *followed by* $s^* \equiv 1$ *for time* $\frac{t_f}{2}$ *(see Fig. 1).*

Here $t_c = \pi$, i.e., if $t_f \geq \pi$ one trivially finds the ground state exactly (by a $\pi/2$ rotation from the -1 eigenstate of σ^x to the -1 eigenstate of σ^y, followed by another $\pi/2$ rotation to the -1 eigenstate of σ^z) and one cannot do better by increasing t_f. This optimal bang-bang schedule result for a single spin-$1/2$ joins previous such results for systems as diverse as pairs of one-dimensional quasicondensates [38] or "gmon" qubits [39], as well as braiding of Majorana zero modes [40].

V. THE OPEN SYSTEM CASE

In this section we generalize the results for closed systems to the open system setting. We consider two different approaches: an approximation-free treatment of a system + environment where both are finite-dimensional, and a master equation approach subject to a Markovian approximation, which applies for infinite-dimensional environments [41–45]. We show that under a number of additional assumptions, we can (partially) recover the results from the closed system setting, but that the bangs characterizing the latter are not a particularly robust feature in the open system setting.

A. Optimal control for the Liouville-von Neumann equation

One approach for extending the closed system results of the previous section to open systems is to consider the *full dynamics* of a jointly evolving system + environment. In this case ρ in Eq. (6) is the density matrix of the system and the environment, with an initial condition which is usually taken to be of the factorized form $\rho(0) := \rho_0 \otimes \rho_E$, where now ρ_0 refers to the initial state of the system only. The Liouville-von Neumann equation is [extending Eq. (15)]:

$$\dot{\rho} = \mathcal{K}_{H_{\rm tot}} \rho \,, \quad \rho(0) = \rho_0 \otimes \rho_E \,, \tag{23}$$

where $\mathcal{K}$ is defined in Eq. (13), with the total Hamiltonian

$$H_{\rm tot} = H_S \otimes \mathbb{1}_E + H_I + \mathbb{1}_S \otimes H_E \,. \tag{24}$$

Here H_I is the interaction between the system and environment, H_E generates the dynamics of the environment, while the system Hamiltonian, as before, contains the controllable part:

$$H_S(t) = C + s(t)\,(B - C) \,. \tag{25}$$

Finally the cost is given by

$$J = {\rm Tr}[(C \otimes \mathbb{1}_E)\,\rho] \,. \tag{26}$$

Theorem 1 holds with $\mathcal{L} = \mathcal{K}_{H_{\rm tot}}$ and the PMP control Hamiltonian has the form

$$\mathbb{H} = \langle p, \mathcal{K}_{H_S \otimes \mathbb{1}_E} \rho\rangle + \langle p, \mathcal{K}_{H_I} \rho\rangle + \langle p, \mathcal{K}_{\mathbb{1}_S \otimes H_{\rm E}} \rho\rangle \,. \tag{27}$$

The treatment of Sec. IV applies, *mutatis mutandis*. In particular, condition (9) is replaced by

$$p(t_f) = -C \otimes \mathbb{1}_E \,. \tag{28}$$

Remarkably, no additional modifications of the statement of the PMP Theorem 1 are needed. Moreover, it is clear from Eqs. (25) and (28) that once again the control enters $\mathbb{H}$ only via the term $s(t)\langle p, \mathcal{K}_{(B-C)\otimes \mathbb{1}_E}\rho\rangle$, so that the proof of Part (i) of Theorem 2 applies without any change. This shows that:

Corollary 1. *In the general open system setting of Eq. (23) with the cost (26), the optimal control $s(t)$ is an alternation of bang arcs where $s \equiv 0$ (when $\langle p, \mathcal{K}_{(B-C)\otimes \mathbb{1}}\rho\rangle < 0$) or $s \equiv 1$ (when $\langle p, \mathcal{K}_{(B-C)\otimes \mathbb{1}}\rho\rangle > 0$), and singular arcs where $\langle p, \mathcal{K}_{C\otimes \mathbb{1}}\rho\rangle \equiv \langle p, \mathcal{K}_{B\otimes \mathbb{1}}\rho\rangle$.*

Let us consider the generalization of Part (ii) of Theorem 2, which addresses the characterization of the control function at the beginning and at the end. We first consider the final arc. We have the following:

Theorem 5. *Assume $s^* \in [0,1]$ is the optimal control in an interval $[0, t_f]$ minimizing the cost (26) for Eq. (23). Assume furthermore that $[H_I, C \otimes \mathbb{1}_E] = 0$ and that the final time constraint is active, i.e., $\lambda > 0$. Then $s^*(t) = 1$ for $t \in (t_f - \epsilon, t_f]$ for some $\epsilon > 0$.*

Note that the assumption $[H_I, C \otimes \mathbb{1}_E] = 0$ implies that the choice of control $s \equiv 0$ leaves the cost $J = \langle C \otimes \mathbb{1}, \rho\rangle$ unchanged since in this case $H_{\rm tot}$ commutes with $C \otimes \mathbb{1}_E$.

Proof. Let us compute the PMP Hamiltonian at $t = t_f$. Note first that it follows from Eq. (28) that

$$\begin{aligned} \langle p_f, \mathcal{K}_{C\otimes \mathbb{1}_E} \rho_f\rangle &= \langle \mathcal{K}^\dagger_{C\otimes \mathbb{1}_E} p_f, \rho_f\rangle \\ &= \langle \mathcal{K}_{C\otimes \mathbb{1}_E} C \otimes \mathbb{1}_E, \rho_f\rangle = 0 \,. \end{aligned} \tag{29}$$

The other two terms of the PMP control Hamiltonian Eq. (27) also vanish at $t = t_f$: using the same calculation as in Eq. (29) the second term vanishes because of the assumption $[H_I, C \otimes \mathbb{1}_E] = 0$, and the third term does as well because $[\mathbb{1}_S \otimes H_E, C \otimes \mathbb{1}_E] = 0$. So we obtain $\mathbb{H} = s(t_f)\langle p_f, \mathcal{K}_{B\otimes \mathbb{1}_E}\rho_f\rangle = \lambda > 0$. This implies that $\langle p_f, \mathcal{K}_{B\otimes \mathbb{1}_E}\rho_f\rangle > 0$ and by continuity there must exist an interval $(t_f - \epsilon, t_f]$ for some $\epsilon > 0$ such that $\langle p(t), \mathcal{K}_{(B-C)\otimes \mathbb{1}_E}\rho(t)\rangle > 0$ for $t \in (t_f - \epsilon, t_f]$. Finally we must have $s^* \equiv 1$ in this interval by Corollary 1. □

Regarding the arc at the beginning we have instead:

Theorem 6. *Assume $s \in [0,1]$ is the optimal control in an interval $[0, t_f]$ minimizing the cost Eq. (26) for Eq. (23). Assume that $[\rho_E, H_E] = 0$ and that $[H_I, \rho_0 \otimes \rho_E] = 0$. Assume furthermore that $[B, \rho_0] = 0$ and that the final time constraint is active $\lambda > 0$. Then the control satisfies $s^*(t) = 0$ for $t \in [0, \epsilon)$ for some $\epsilon > 0$.*

Note that the assumption $[H_I, \rho_0 \otimes \rho_E] = 0$ implies that the interaction alone does not modify the initial state of the system.

Proof. We abbreviate the proof since it is very similar to the ones we presented above in more detail. Using the assumption $[H_I, \rho_0 \otimes \rho_E] = 0$ and $[B, \rho_0] = 0$, evaluating $\mathbb{H}(t = 0)$ we obtain $\mathbb{H}(0) = (1 - s(0))\,\langle p(0), \mathcal{K}_{C\otimes \mathbb{1}_E}\rho(0)\rangle = \lambda > 0$. This implies that $\langle p(0), \mathcal{K}_{C\otimes \mathbb{1}_E}\rho(0)\rangle > 0$ or equivalently that $\langle p(0), \mathcal{K}_{-C\otimes \mathbb{1}_E}\rho(0)\rangle < 0$. By continuity this in turn implies that there exist an interval $[0, \epsilon)$ for some $\epsilon > 0$ such that $\langle p(t), \mathcal{K}_{(B-C)\otimes \mathbb{1}_E}\rho(t)\rangle < 0$ for $t \in [0, \epsilon)$. Finally we must have $s^* \equiv 0$ in this interval by Corollary 1. □

We comment on the implications of the additional assumptions used in these theorems in Sec. VII.

B. Optimal control for quantum master equation dynamics

The treatment of the open system case in the previous subsection did not involve any approximations. On the other hand, we tacitly assumed that the environment is *finite dimensional*. This was helpful since all the results on optimal control which we have elaborated upon in Sec. IV and used so far, are classically stated and proved for finite dimensional systems. Extending such results, in particular concerning the PMP and the topology and continuity of the reachable sets for infinite dimensional systems, is possible and is a current area of research in control theory (see, e.g., Refs. [20, 46]),

although the results in this area become considerably more technical. An alternative we discuss in this subsection is to replace the Liouville-von Neumann equation (23) with an approximate quantum master equation. This can be viewed as an investigation of the result of Sec. V A when the environment dimension is sent to infinity.

Without loss of generality we write $H_I = \sum_\alpha S_\alpha \otimes E_\alpha$, where $S_\alpha = S_\alpha^\dagger$ and $E_\alpha = E_\alpha^\dagger$ $\forall \alpha$. The goal is now to find a master equation for the dynamics of the system density matrix in the case of a time-dependent system Hamiltonian. After the Born approximation and tracing out the environment, one arrives at a time dependent Redfield master equation (see, e.g., the Schrödinger picture Redfield master equation (SPRME) derived in Ref. [47]). From this point there are multiple ways to proceed, e.g., by introducing an additional adiabatic approximation or an additional Markovian approximation, or both. These different paths, and exactly how they are taken, lead to a plethora of different master equations [47–56]. We next focus on two representative cases of master equations derived from first principles.

1. *Adiabatic Redfield Master Equation*

The Adiabatic Redfield Master Equation (ARME) is derived in Ref. [47]. It results from assuming that $t_f \gg \tau_B$, where τ_B is the environment time scale, and the adiabatic approximation $\mathcal{T} \exp\left[-i \int_{t-r}^{t} H_S(t')dt'\right] \approx e^{-irH_S(t)}$, dropping a correction of $O\left((r/t_f)^2\right)$. The ARME has the form of Eq. (6) with a time-dependent Redfield generator $\mathcal{L}$ given by

$$\mathcal{L} = \mathcal{K}_{H_S} + \mathcal{D} \tag{30a}$$

$$\mathcal{D}\rho = \sum_{\alpha\beta} \int_0^{t_{\max}} dr\, G_{\alpha\beta}(r)\, [S_\beta(-r)\rho, S_\alpha] + \text{h.c.}, \tag{30b}$$

where the system Hamiltonian is as in Eq. (25). $G_{\alpha\beta}(t)$ is the environment correlation function

$$G_{\alpha\beta}(t) = \langle E_\alpha(t) E_\beta(0)\rangle = \text{Tr}[E_\alpha(t)E_\beta(0)\rho_E]\,, \tag{31}$$

where $\langle X\rangle$ denotes the environmental thermal average of X. When $G_{\alpha\beta}(r)$ decays exponentially, the relative error of the resulting dynamics due to the adiabatic approximation above is $O\left((\tau_B/t_f)^2\right)$. Finally,

$$S_\beta(-r) = e^{-irH_S(s)} S_\beta e^{irH_S(s)}\,. \tag{32}$$

The parameter $t_{\max}$ can either be set to t_f or infinity on account of the fact that the environment correlation function decays very rapidly. The ARME is not in Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) form [57–59], hence does not generate a completely positive map. However, it generates non-Markovian dynamics, hence has a wider range of applicability than Markovian master equations, within its range of applicability [51].

Crucially, the generator in Eq. (30) with Eqs. (25), (31) and (32) *depends on time only through the control function* $s(t)$. This implies that the PMP control Hamiltonian $\mathbb{H} = \langle p, \mathcal{L}\rho\rangle$ is constant and hence the PMP in the form of Theorem 1 is directly applicable [60]. However, the control now enters nonlinearly in $\mathcal{D}$, in particular in an exponential through Eq. (32). As a consequence it is not possible to derive the form of the control on the nonsingular arcs, or even to determine simple equations for the appearance of singular arcs. One can ask, however, what remains of the results of the previous subsection. We do not have an analog of Theorem 6 for the initial arc. However, if we again make the assumption of Theorem 5 that $[S_\alpha, C] = 0$ $\forall\alpha$, then the analog of this theorem for the final arc holds, even when the environment is infinite-dimensional, and under the approximations used to derive Eq. (30). However, instead of an arc, we obtain a bang only at a point:

Theorem 7. *Assume $s^* \in [0,1]$ is the optimal control in an interval $[0, t_f]$ minimizing the cost Eq. (7) for Eq. (6) with $\mathcal{L}$ given by Eq. (30). Assume further that $[S_\alpha, C] = 0$ $\forall\alpha$ and that the final time constraint is active ($\lambda > 0$). Then the optimal control satisfies $s^*(t_f) = 1$.*

Proof. It is convenient to write the Redfield dissipator as

$$\mathcal{D}\rho = \sum_{\alpha\beta} \left(\left[W_{\alpha\beta}\rho, S_\alpha\right] + \left[S_\alpha, \rho W_{\alpha\beta}^\dagger\right] \right) \tag{33a}$$

$$W_{\alpha\beta}(t) = \int_0^{t_{\max}} dr\, G_{\alpha\beta}(r)\, S_\beta(-r)\,. \tag{33b}$$

Using Eq. (33a) one obtains, for the adjoint of $\mathcal{D}$:

$$\mathcal{D}^\dagger X = \sum_{\alpha\beta} \left(W_{\alpha\beta}^\dagger\, [X, S_\alpha] + [S_\alpha, X]\, W_{\alpha\beta} \right) \tag{34}$$

(see Appendix D). From the above expression and the assumption $[S_\alpha, C] = 0$, $\forall\alpha$ we obtain $\mathcal{D}^\dagger C = 0$. Using $p_f = -C$ we have $\langle p_f, \mathcal{D}\rho_f\rangle = -\langle \mathcal{D}^\dagger C, \rho_f\rangle = 0$. Evaluating the PMP control Hamiltonian at the final time we obtain $\mathbb{H}(t_f) = s(t_f)\langle p_f, \mathcal{K}_B\rho_f\rangle = \lambda > 0$. This implies that $\langle p_f, \mathcal{K}_B\rho_f\rangle > 0$, and so $s(t_f) = 1$ from the maximum principle. □

Since, as argued in Sec. VI B, the size of the bang intervals is generically expected to shrink when the total system size increases, this result can be seen as a generalization of Theorem 5 when the environment dimension is sent to infinity and the bang interval at the end shrinks to a point.

The implications of the additional assumptions used in Theorem 7 are discussed in Sec. VII.

2. *Markovian, completely positive master equations*

The most significant drawback of the ARME is the violation of complete positivity, which means that the density matrix can develop unphysical, negative eigenvalues. Hence we also consider Markovian, completely positive master equations. There are a variety of such master equations derived from first principles under different assumptions. However, in most cases the generator $\mathcal{L}$ is explicitly time-dependent (e.g.,

the coarse-grained master equation (CGME) [51, Eq. (22)], the master equation of Ref. [52, Eq. (21)], the non-adiabatic master equation (NAME) [53, Eq. (16)], and the universal Lindblad equation (ULE) [54, Eq. (27)]) and hence we cannot apply Theorem 1.

In this subsection we give an example of a Markovian master equation derived from first principles where, like in the ARME case, the generator $\mathcal{L}$ depends on time only through the schedule $s(t)$. In this case the PMP can be applied in the simplified form described in Theorem 1.

Consider the "geometric-arithmetic master equation" (GAME) [55, Eq. (46)], which is claimed there to have a higher degree of accuracy than all the previous Markovian master equations. In the adiabatic limit it has the Schrödinger picture form

$$\mathcal{L} = \mathcal{K}_{H_S} + \mathcal{D} \tag{35a}$$

$$\mathcal{D}\rho = \sum_j \left([L_\alpha(s)\rho, L_\alpha^\dagger(s)] + [L_\alpha(s), \rho L_\alpha^\dagger(s)] \right) , \tag{35b}$$

where $L_\alpha(s) = S_\alpha \circ \sqrt{\gamma(s)}$, the circle denotes the Hadamard (element-wise) product, γ is the spectral density matrix [Fourier transform of the environment correlation function (31)] whose elements $\gamma_{nm} := \gamma(\omega_{nm}(s))$ depend on the instantaneous Bohr frequencies $\omega_{nm}(s) = E_n(s) - E_m(s)$, where $H_S(s)|n(s)\rangle = E_n(s)|n(s)\rangle$, and the dependence on time is only through the schedule $s(t)$ [61]. The adjoint dissipator is now:

$$\mathcal{D}^\dagger X = \sum_\alpha \left((S_\alpha \circ \sqrt{\gamma})^\dagger [X, S_\alpha \circ \sqrt{\gamma}] + [(S_\alpha \circ \sqrt{\gamma})^\dagger, X] S_\alpha \circ \sqrt{\gamma} \right) . \tag{36}$$

Unfortunately, since $[S_\alpha, C] = 0 \not\Rightarrow [S_\alpha \circ \sqrt{\gamma}, C] = 0$, this means that even if $[S_\alpha, C] = 0$, we do *not* obtain $\mathcal{D}^\dagger C = 0$ as in the ARME case, and hence the proof of Theorem 7 does not carry through [62].

While these arguments are not a proof that in general Markovian dynamics do not admit $s(t_f) = 1$ as an optimal control solution, we conjecture that in fact, they do not. It thus appears that the "counterintuitive" appearance of the driver Hamiltonian at the end of the control interval is not a feature of the optimal schedule in the Markovian limit of open quantum systems. We revisit this point in Sec. VII.

VI. SWITCHING OPERATOR AND ANALYSIS OF THE OPTIMAL CONTROL

In order to study the qualitative behavior of the optimal control law, in particular its switching properties and the existence and nature of the singular arcs, it is convenient to introduce one more operator, besides the state ρ and the co-state p, which we call the *switching operator*. The switching operator determines the behavior of the optimal control, i.e., the points where there is a switch between $s = 0$ and $s = 1$, and where there is a singular arc. For clarity we focus on the closed system case of Sec. IV but our definitions and treatment naturally extend with a change of notation to the open system case of Sec. V A.

A. Switching equation

The switching operator S is the Hermitian operator defined as

$$S := i\,[p, \rho] \ . \tag{37}$$

In the closed system case the Liouvillian has the form $\mathcal{L} = \mathcal{K}_H$, with the system Hamiltonian of Eq. (2). One has the following property:

$$\mathcal{K}_H\left([X, Y]\right) = [\mathcal{K}_H(X), Y] + [X, \mathcal{K}_H(Y)] \tag{38}$$

valid for any operators X, Y, H. Now, differentiating Eq. (37), we obtain

$$\dot{S} = i\,[\dot{p}, \rho] + i\,[p, \dot{\rho}] \tag{39a}$$

$$= i\,[\mathcal{K}_H p, \rho] + i\,[p, \mathcal{K}_H \rho] \tag{39b}$$

$$= i\mathcal{K}_H\left([p, \rho]\right) = \mathcal{K}_H S\,, \tag{39c}$$

where in the third equality we used Eq. (38). Thus, S satisfies the same equation as ρ and p. To understand why S determines the optimal control switching times, let us define

$$x_C := \langle p, \mathcal{K}_C \rho\rangle = -i\,\mathrm{Tr}\left(p\,[C, \rho]\right) \tag{40a}$$

$$= -i\,\mathrm{Tr}\left(C\,[\rho, p]\right) = \mathrm{Tr}\left(CS\right) = \langle C, S\rangle\,, \tag{40b}$$

and similarly

$$x_B := \langle p, \mathcal{K}_B \rho\rangle = \langle B, S\rangle\,, \tag{41}$$

so that the PMP control Hamiltonian Eq. (18) can be written in a form closely resembling the Hamiltonian (2b):

$$\mathbb{H} = x_C + s\,(x_B - x_C) \ . \tag{42}$$

The quantity $x_B - x_C = \langle B - C, S\rangle$, that is, the orthogonal component of S along $B - C$, regulates the switches of the candidate optimal control. By the same PMP argument we have used repeatedly in our proofs, when $x_B - x_C < 0$ we have $s \equiv 0$, and when $x_B - x_C > 0$ we have $s \equiv 1$. The switch occurs when $x_B - x_C = 0$, while a singular arc occurs when $x_B - x_C \equiv 0$ for an interval of positive measure.

The initial condition S_0 of the switching operator determines the optimal control candidate s uniquely. This initial condition is not completely arbitrary. In particular, under the assumptions of Theorem 2 the following holds. Since $s(0) = 0$, we have $x_C(0) = \langle C, S_0\rangle = \lambda$. Furthermore, since $[B, \rho_0] = 0$, we have

$$x_B(0) = \langle B, S_0\rangle = i\,\mathrm{Tr}(B[p_0, \rho_0]) = i\,\mathrm{Tr}\left(p_0[\rho_0, B]\right) = 0\,. \tag{43}$$

At the final time t_f, since $s(t_f) = 1$, we have from Eq. (42) $x_B(t_f) = \lambda$, while using Eq. (9) we have

$$x_C(t_f) = \langle C, S_f\rangle = i\,\mathrm{Tr}\left(C[p_f, \rho_f]\right) = i\,\mathrm{Tr}\left(\rho_f[C, p_f]\right) = 0\,. \tag{44}$$

Thus, at $t = 0$ we have $x_B(0) = 0$ and in the initial arc $s \equiv 0$ and $x_C \equiv \lambda$ by Eq. (42). The next arc can be either nonsingular with $s \equiv 1$ ($x_B > x_C$) or a singular arc with $x_B \equiv x_C$. Either way, at the switching point we must have $x_B = \lambda$, hence we reach the point $(x_C, x_B) = (\lambda, \lambda)$ at the end of the first arc. When $(x_C, x_B) = (\lambda, \lambda)$ there is either a switch to an arc with $s \equiv 1$, or a return to $s \equiv 0$, or a singular arc where $(x_C, x_B) \equiv (\lambda, \lambda)$. On this arc s is unspecified, but nonetheless certain equations need to be satisfied and they can be used to obtain information on the dynamics on such singular arcs (see Appendix G). Note that every switching event, whether from a bang arc to a singular arc or *v.v.*, or from a bang arc to another bang arc, happens at $(x_C, x_B) = (\lambda, \lambda)$. When $s \equiv 1$, from Eq. (42) we have that x_B is constant, while x_C is allowed to change. Therefore the optimal control can be described schematically as in Fig. 2 in the (x_C, x_B) plane.

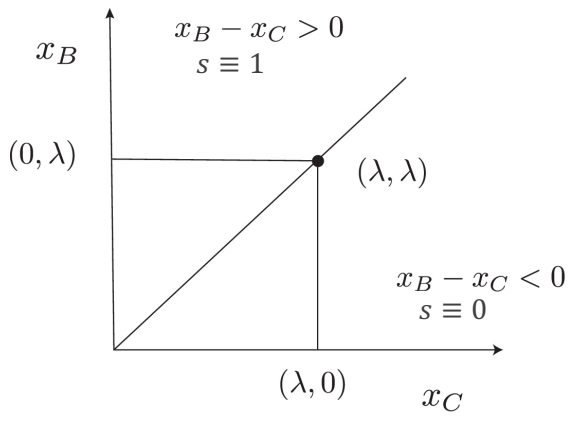

FIG. 2. Switching diagram for optimal control candidates. Starting from the point $(\lambda, 0)$ the dynamics stay on the line $x_C = \lambda$. In principle even negative values of x_B can be attained, but eventually the point (λ, λ) must be reached, followed by an alternation of vertical and horizontal lines (where in principle one can also extend to negative values of x_C) through (λ, λ) with time intervals where the dynamics do not move from (λ, λ), representing the singular arcs. In the last interval, the dynamics follow a horizontal line (with $s \equiv 1$) reaching the point $(x_C, x_B) = (0, \lambda)$.

In the case where the final time is not active, i.e., when we do not have the guarantee $\lambda > 0$ in Eq. (42), the above reasoning can still be applied to conclude $x_B(0) = 0$ and $x_C(t_f) = 0$. Furthermore, if $\lambda = 0$, from Eq. (42), we obtain $(1 - s(0))x_C(0) = 0$ and $s(t_f)x_B(t_f) = 0$, in addition to the maximization condition Eq. (11).

B. Shortening of nonsingular arcs

1. *Dependence on n*

Here we give a heuristic argument that explains why the terminal arcs ($s \equiv 0$ and $s \equiv 1$) for the optimal control become shorter as the number n of spins (or qubits) increases. In fact the heuristic holds also for intermediate arcs taking place anywhere along the optimal trajectory.

Consider a bang arc where $s(t) = 0$ for $t \in [t_0, t_1]$. Equation (39) for the switching operator in this region is $\dot{S} = \mathcal{K}_C S$ with the initial condition $S(t_0) = S_0$ for some S_0. The solution in this interval is $S(t) = e^{-i(t-t_0)C} S_0 e^{i(t-t_0)C}$. The coordinate x_C equals λ in the interval: $x_C = \lambda = \langle C, S(t)\rangle = \langle C, S_0\rangle$. The switching happens when $x_B = \lambda$, i.e., at the first solution t_1 of $\langle B, S(t_1)\rangle = \langle C, S_0\rangle$. More explicitly, the interval of the bang arc $\Delta t := t_1 - t_0$ is given by the first solution of

$$x_B(\Delta t) := \mathrm{Tr}\left(B e^{-i\Delta t C} S_0 e^{i\Delta t C}\right) = \mathrm{Tr}\left(C S_0\right) . \quad (45)$$

Using the spectral resolution $C = \sum_k E_k \Pi_k$ (with eigenvalues E_k and eigenprojectors Π_k), the left-hand hand side of Eq. (45) can be written as

$$x_B(\Delta t) = \sum_{kl} M_{kl} e^{-i\Delta t \omega_{kl}} , \quad (46)$$

with amplitudes $M_{kl} := \mathrm{Tr}\left(B\Pi_k S_0 \Pi_l\right)$ and Bohr frequencies $\omega_{kl} = E_k - E_l$. The function $x_B(\Delta t)$ is a real trigonometric polynomial with $O(d^2)$ terms (where d is the Hilbert space dimension) starting from $\mathrm{Tr}(BS_0)$ at $\Delta t = 0$, and reaching $\lambda > 0$ at Δt. Now, as the number of qubits n increases, both $d = 2^n$ and the frequencies increase. Both of these facts contribute to making $x_B(t)$ oscillate faster. As a consequence, the solution Δt of Eq. (45) tends to decrease with n. The same considerations hold for the case of an $s \equiv 1$ arc with B and C interchanged. See Appendix I for additional comments. Note that this heuristic applies both to the initial and final bang arcs, as well as to possible intermediate bang arcs if they are present.

2. *Dependence on t_f*

It was concluded in Ref. [24, Sec. S3] that "these bangs should become smaller and smaller as t_f is increased. Eventually in the true $t_f \to \infty$ adiabatic limit, these bangs disappear recovering the standard form expected for quantum adiabatic computing." However, there is in fact no guarantee that the optimal control coincides with the adiabatic path even in this limit, since the adiabatic theorem provides a sufficient, but not a necessary condition for convergence to the minimum of the cost function. Indeed, it is easy to construct a counterexample, as we now do. First note that, as we have seen, when $t_f < t_c$, the optimal schedule always starts with a bang $s \equiv 0$ and ends with a bang $s \equiv 1$ (provided the initial state commutes with B). We have not addressed the question of uniqueness of this optimal schedule, which we leave for future work. However, when $t_f > t_c$, the optimal schedule is certainly not unique, as one has the possibility of "wasting time" by adding a bang $s \equiv 0$ at the end (thus applying C there), or by adding a bang $s \equiv 1$ at the beginning (thus applying B there), or both. The resulting schedules do not resemble the smooth adiabatic schedule interpolating slowly from $s(0) = 1$ to $s(t_f) = 0$.

VII. SUMMARY AND DISCUSSION

The quest to discover the optimal schedule for quantum optimization algorithms such as QA and QAOA naturally leads to the use of optimal control theory via Pontryagin's principle. Previous work concluded that QAOA is optimal [19], but a more careful analysis showed that in fact a hybrid bang-anneal-bang protocol is generally optimal for closed systems when not enough time is allowed for the desired state to be reached perfectly [24]. Here we confirmed this result using a density matrix approach, which both generalizes the analysis to mixed states and simplifies it since it makes the cost-function linear in the state. We also showed that the assumption that t_f is smaller than the critical time t_c needed to reach the ground state of C exactly, is necessary but not sufficient for the result of Ref. [24], by giving a counterexample to the latter.

We introduced a switching operator and found its equation of motion, which characterizes the points at which the optimal schedule switches between the two different types of bang arcs and the anneal arc. In Theorem 4 we gave the explicit optimal schedule for the example of a single spin-$1/2$ particle, which consists of two bangs of equal duration.

Using the density matrix formulation we extended the theory to the open system setting, both for the exact reduced system dynamics in the case of a finite-dimensional environment, and under the approximation of dynamics governed by a master equation due to coupling to an infinite-dimensional environment. We proved that in the first setting, depending on additional assumptions concerning the initial states of the system and the environment and their interaction, either an anneal-bang (Theorem 5) or bang-anneal (Theorem 6) schedule is optimal.

In the second setting (infinite-dimensional environment) we considered both an adiabatic Redfield equation accounting for non-Markovian dynamics but without a complete positivity guarantee, and a completely positive Markovian master equation. In the former (Redfield) case we could only prove that the optimal schedule terminates with the driver Hamiltonian, i.e., $s(t_f) = 1$ (Theorem 7). One could interpret this result as a manifestation of the phenomenon of the shortening of the nonsingular arcs as the *total* system (i.e., the subsystem plus its environment) size increases. Indeed, in the Redfield case the environment Hilbert space dimension is infinite, which is consistent with the bang arc having shrunk down to a point. In the fully Markovian case, even this last remnant of the bang-arc was not recovered, as we found no evidence of natural conditions under which $s(t_f) = 1$ holds.

Let us now comment on the differences between these theorems and their closed system counterpart, Part (ii) of Theorem 2. Regarding Theorem 5 concerning the final arc, the main change is the addition of the assumption that the interaction Hamiltonian commutes with the cost function, i.e., $[H_I, C \otimes \mathbb{1}_E] = 0$. Writing the interaction in the general form $H_I = \sum_\alpha S_\alpha \otimes E_\alpha$, where S_α and E_α are system and environment operators, respectively, the assumption is equivalent to $[S_\alpha, C] = 0\ \forall \alpha$, which is the same assumption as in Theorem 7. Thus C must belong to the commutant of the algebra generated by the set $\{S_\alpha\}$, i.e., $C \in \mathrm{Alg}\{S_\alpha\}'$ [63]. For example, if $S_\alpha = \sum_{i=1}^n \sigma_i^\alpha$ for a system of n qubits, where $\alpha \in \{x, y, z\}$, i.e., the collective decoherence case [64, 65], then, if C is at most a two-body interaction, it follows that it must be of the Heisenberg interaction form: $C = \sum_{ij} J_{ij} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j$, where J_{ij} are constants [66]. Or, for a classical target Hamiltonian arising in optimization such as the Ising-type Hamiltonian mentioned in Sec. II, this means that the interaction must be of the pure-dephasing type, i.e., $S_\alpha \propto \sigma^z$ (or products of σ^z over different qubits). This is a realistic model, e.g., for superconducting qubits undergoing flux noise [67].

Regarding Theorem 6 concerning the initial arc, the main change is the addition of the two assumptions that (i) the environment Hamiltonian commutes with the environment's initial state ($[H_E, \rho_E] = 0$), and (ii) the interaction Hamiltonian commutes with the joint system-environment initial state ($[H_I, \rho_0 \otimes \rho_E] = 0$). The first of these is natural and is known as the stationary environment assumption [43]. It is satisfied, e.g., if the environment is in thermal equilibrium, i.e., in the Gibbs state: $\rho_E \propto e^{-\beta H_E}$, where β is the inverse temperature. The second assumption means that $\rho_0 \in \mathrm{Alg}\{S_\alpha\}'$ and $\rho_E \in \mathrm{Alg}\{E_\alpha\}'$. This assumption is the least natural of the ones we have encountered so far. E.g., it is clearly violated in the standard quantum annealing setting where ρ_0 is the ground state of a transverse field $-\sum_i \sigma_i^x$ and $S_\alpha \propto \sigma^z$. Even the condition $\rho_E \in \mathrm{Alg}\{E_\alpha\}'$ is not very natural. For example, for a bosonic environment one typically has E_α as the position operator of an oscillator, while H_E might be the number operator, in which case the Gibbs state ρ_E would not commute with E_α.

We note that while the conditions given in Theorem 6 are sufficient, we do not know if they are necessary, which we thus leave as an open problem. We conjecture that the initial bang does not appear as a feature of optimal schedules for open systems coupled to an infinite-dimensional environment. This state of affairs would be reminiscent of the existence of an arrow of time for open systems, which breaks the symmetry between the initial and final times (see, e.g., Ref. [47] for a similar effect in the pure QA setting). On the other hand, given the naturalness of the sufficient conditions under which a final bang arc (Theorem 5) or a schedule terminating with the driver Hamiltonian (Theorem 7) are optimal, such schedules may find utility in the design of quantum algorithms for optimization problems in the setting of open quantum systems. This is true in particular for systems that are well described by the adiabatic Redfield master equation, e.g., superconducting flux qubits used for quantum annealing [68–71].

However, the conditions under which the adiabatic Redfield master equations hold need not apply in general, e.g., for Hamiltonians that arise naturally in systems such as Rydberg atoms or transmons, which have been used to demonstrate QAOA [10, 72]. Especially for quantum optical systems such as Rydberg atoms, the Markovian limit may be more appropriate, and we have not found evidence of the optimality of an initial or final bang arc in this limit, or even the optimality of $s(t_f) = 1$.

While our analysis does not strictly rule out bang-type schedules for open systems coupled to infinite-dimensional

environments, we conjecture that they are indeed not a feature of optimal schedules in this case, primarily due to the shortening of arcs in the open system setting. If this could be confirmed, it would mean that after all, continuous annealing-type schedules are optimal for optimization purposes when using open quantum systems, which would have implications for all NISQ-era optimization algorithms.

Finally, we remark that throughout this work we have used a simplified form of the PMP as described in Theorem 1. A more general PMP for time-dependent dynamics is described in Ref. [73]. The main difference is that $\mathbb{H} \equiv \lambda$ [Eq. (12)] is no longer valid in the given form. This equation is, in fact, a special case of another one which contains the derivative of the dynamics with respect to t, and this term vanishes when the dynamics are not explicitly time-dependent, as in our case, where we considered adiabatic time-dependent master equations. The time-dependence of the dynamics in typical quantum master equations [47–56] is, however, different from the one of models usually encountered in classical control theory, and therefore further study is required before the PMP can be applied to a broader class of quantum master equations.

VIII. CONCLUSIONS

Optimization is one of the key areas where a quantum speedup has long been an eagerly anticipated outcome. Several promising heuristic approaches have been proposed toward this end, including the analog quantum annealing (QA) algorithm and the digital quantum approximate optimization algorithm (QAOA). While a result identifying the general conditions for a quantum speedup under either QA or QAOA appears to be out of reach, in this work we have undertaken a rigorous analysis that unifies these algorithms under the framework of quantum optimal control theory, and found conditions describing the optimal path from the initial to the target Hamiltonian. Previously it was shown that the optimal path is of the "bang-anneal-bang" type, meaning that the path should start and end with QAOA-like segments, but in between it should resemble QA. This analysis was limited to the setting of closed quantum systems, which is an idealization describing quantum computers that are not subjected to any noise or decoherence. Our analysis extends to open quantum systems, i.e., a realistic description of quantum computers that interact with a noisy environment. We demonstrated that the conclusions derived for closed quantum systems are modified in such a way that the initial and final "bang" segments either shrink to a point or disappear entirely. This means that the optimal solution in the open system setting more closely resembles QA. Our result has implications for the use of actual devices, which are by necessity noisy in the NISQ era and likely beyond, and suggest that in this practical sense the optimal schedules for quantum optimization are likely to be continuous.

ACKNOWLEDGMENTS

LCV's and DL's research is based upon work (partially) supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) and the Defense Advanced Research Projects Agency (DARPA), via the U.S. Army Research Office contract W911NF-17-C-0050. DD's research was supported by the NSF under Grant ECCS 1710558. DL's research was also sponsored by the Army Research Office and was accomplished under Grant Number W911NF-20-1-0075. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, DARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Appendix A: Examples of reachability/unreachability of the ground state of C

Here we provide two examples, one where Prop. 1 guarantees reachability of the ground state of C despite the algebra generated by B and C being smaller than the full $su(d)$, another illustrating that $[B, C] \neq 0$ is not a sufficient condition.

1. First example

Let $B = \sigma^z \otimes \mathbb{1}$, $C = \sigma^x \otimes \sigma^x + \sigma^y \otimes \sigma^y + \sigma^z \otimes \sigma^z$. Both B and C commute with $M = \sigma^z \otimes \mathbb{1} + \mathbb{1} \otimes \sigma^z$. We know that the ground state of C is a singlet $\frac{1}{\sqrt{2}}(|{\downarrow\uparrow}\rangle - |{\uparrow\downarrow}\rangle)$ with total spin zero and hence belongs to the sector $M = 0$. This statement follows from a theorem of Marshall [74, 75] and holds for general antiferromagnetic Heisenberg models defined on a bipartite lattice. In particular, it does not require knowledge of the ground state. The projector P_0 is given by

$$P_0 = |{\downarrow\uparrow}\rangle\langle{\downarrow\uparrow}| + |{\uparrow\downarrow}\rangle\langle{\uparrow\downarrow}| \,. \tag{A1}$$

In $\mathrm{Ran}(P_0)$ one has:

$$P_0 B P_0 = \sigma^z \tag{A2a}$$
$$P_0 C P_0 = -\mathbb{1} + 2\sigma^x \,. \tag{A2b}$$

Therefore in $\mathrm{Ran}(P_0)$, B and C generate the full $su(2)$ algebra and any state in $\mathrm{Ran}(P_0)$ can be reached starting from any state in $\mathrm{Ran}(P_0)$, in particular the ground state of C.

2. Second example

It is straightforward to find examples where the ground state of C cannot be reached even when $[B, C] \neq 0$. Consider, e.g., $B = \sigma^x \otimes \mathbb{1}$ and $C = \sigma^z \otimes \mathbb{1} + \mathbb{1} \otimes \sigma^z$. In this case the

Lie algebra generated by B and C is $su(2)\otimes u(1)\neq su(4)$ and only the first qubit can be fully steered anywhere on the Bloch sphere. As a consequence, the ground state $|\downarrow\downarrow\rangle$ of C cannot be reached unless one starts with a state of the form $\rho_0=\tilde{\rho}\otimes|\downarrow\rangle\langle\downarrow|$, with $\tilde{\rho}$ being an arbitrary single qubit state.

Appendix B: General results on optimal control; the Pontryagin Maximum Principle

We review here standard results in optimal control theory emphasizing a geometric viewpoint and the results needed for the applications in the main body of the paper.

1. Setup

In optimal control theory (see, e.g., Ref. [73]) one considers a general control system

$$\dot{x}=f\left(x,s\right),\qquad x(0)=x_0, \tag{B1}$$

with $x\in\mathbb{R}^N$ [76], the control s with values from a compact subset $S\subseteq\mathbb{R}^M$, f a smooth map $\mathbb{R}^N\times\mathbb{R}^M\to\mathbb{R}^N$ that does not depend explicitly on time. The terminal cost (of *Mayer type* [77])

$$J:=\phi\left(x(t_f)\right)\ , \tag{B2}$$

is to be minimized at the terminal (final) time t_f, where ϕ is a smooth function. In particular, and this is the case that interests us, one can fix t_f so that J only depends on the final state $x(t_f)$.

The geometric approach to the necessary condition of optimal control (see, e.g., Ref. [78]) is based on the concept of a reachable set (or attainable set) for Eq. (B1) with values of the control in S, which we denote by $\mathfrak{R}_t$. The set $\mathfrak{R}_t$ is the set of values for the state x that can be reached (from x_0) at time *exactly* t for control functions with values in the set S. With this definition, the minimum of the cost J in Eq. (B2) is the minimum of the function ϕ over $\mathfrak{R}_{t_f}$. One also defines the reachable set $\mathfrak{R}_{\leq t_f}:=\cup_{0\leq t\leq t_f}\mathfrak{R}_t$, and if $\mathfrak{R}_t$ is non decreasing with t, $\mathfrak{R}_{t_f}=\mathfrak{R}_{\leq t_f}$. This is the case for Eq. (15) if one assumes, as we do, that ρ_0 commutes with B (in this case the system can remain in the state ρ_0 for an arbitrary length of time with the choice $s\equiv 1$). If the set of admissible controls $\tilde{S}$ is compact (as assumed in the main text) and under general conditions on the map f in Eq. (B1), which are also satisfied in our cases, Filippov's theorem (see, e.g., [78, Th. 10.1]) states that the reachable sets are *compact* and this implies the existence of the minimum of the function ϕ and therefore of the optimal control [79]. The introduction of the concept of a reachable set effectively reduces the optimal control problem to a *static* optimization problem for the function ϕ, where the set of possible dynamics is described by the reachable set $\mathfrak{R}_t$, that roughly separates the minimization problem from the analysis of the dynamics.

2. The Pontryagin Maximum Principle (PMP)

The basic necessary conditions of optimality are given by the Pontryagin Maximum Principle (PMP), which we restate below in a more general formulation than in the main text, but in a context relevant for the problem of interest to us, i.e., a fixed final time and a free final state. Assume that $s^*=s^*(t)$ is the optimal control function and $x^*=x^*(t)$ the optimal trajectory. We shall refer to (x^*,s^*) as an *optimal pair*. We have the following.

Theorem 8. *Assume that (x^*,s^*) is an optimal pair. Then there exists a nonzero vector of functions $p=p(t)\in\mathbb{R}^n$ called the* co-state *that satisfies the terminal problem*

$$\dot{p}^T=-p^T\frac{\partial f}{\partial x}\left(x^*(t),s^*(t)\right),\quad p^T\left(t_f\right)=-\frac{\partial\phi}{\partial x}\left(x^*(t_f)\right), \tag{B3}$$

with f defined in Eq. (B1) and ϕ defined in Eq. (B2). Furthermore, define the Hamiltonian function

$$\mathbb{H}\left(p,x,s\right):=p^Tf\left(x,s\right)\ . \tag{B4}$$

Then we have (maximum principle):

$$\mathbb{H}\left(p,x^*,s^*\right)=\max_{v\in\tilde{S}}\mathbb{H}\left(p,x^*,v\right)\ , \tag{B5}$$

(where $\tilde{S}$ is the set of admissible controls) and

$$\mathbb{H}\left(p(t),x^*(t),s^*(t)\right)\equiv\lambda, \tag{B6}$$

for a constant λ.

The constant λ describes the dependence of the optimal cost on the terminal time t_f. To see this, given the optimal control s^* defined in $[0,t_f]$, calculate the variation of the cost with this control at $t=t_f$,

$$\begin{aligned}\left.\frac{d}{dt}\right|_{t=t_f}\phi(x(t))&=\frac{\partial\phi}{\partial x}(x(t_f))f\left(x(t_f),s(t_f)\right)\\&=-p^T\left(t_f\right)f\left(x(t_f),s(t_f)\right)\\&=-\lambda\ ,\end{aligned} \tag{B7}$$

where we used Eqs. (B1), (B3) and (B6). In particular, $\lambda>0$ indicates that it is possible to lower the cost by increasing the time or, in other words, the constraint $t=t_f$ is *active*. If $\lambda=0$ the constraint on the final time is not active. If $\lambda<0$, the above calculation shows that the cost is actually *increasing* with t at $t=t_f$.

If there exists a value s_0 in the admissible control set $\tilde{S}$ such that $f(x_0,s_0)=0$ in Eq. (B1), we can show that we must have $\lambda\geq 0$. The argument is as follows. Assume $\lambda<0$ in Eq. (B7). Since $\phi(x(t))$ is increasing in t, there exists an $\epsilon>0$ such that $\phi(x(t_f-\epsilon))<\phi(x(t_f))=J$. Now construct the following control function

$$s_1(t)=\begin{cases}s_0 & t\in[0,\epsilon]\\ s^*(t-\epsilon) & t\in(\epsilon,t_f]\ ,\end{cases}, \tag{B8}$$

i.e., s_1 leaves the cost unchanged in the first interval of time $[0,\epsilon]$ and then follows the optimal schedule shifted by ϵ. Let us denote by $\phi_1(t)$ the cost function at time t obtained with control s_1. Then, by construction, $\phi_1(t_f)=\phi(x(t_f-\epsilon))<\phi(x(t_f))$, which contradicts the fact that s^* is optimal. Therefore λ cannot be negative. Summarizing, we have:

Proposition 1. *Assume there exists a value $s_0\in\tilde{S}$ such that $f(x_0,s_0)=0$ in Eq. (B1). Then $\lambda\geq 0$ in Eq. (B6), and $\lambda>0$ if and only if the constraint on the final time t_f is active.*

For the problem of interest in the main body of the paper the above assumption on the existence of the value $s_0\in\tilde{S}$ is valid in Theorems 2 and 6, since we assume that the initial condition commutes with the Hamiltonian B (i.e., we can choose $s_0=1$). Furthermore, we note, concerning reachable sets, that (i) because of the existence of such a value $s_0\in\tilde{S}$, we have $\mathcal{R}_{\leq t}=\mathcal{R}_t$; (ii) from standard result in control theory, (e.g., Ref. [37, Th. 1] and references therein) we know that for the bilinear class of models [such as Eq. (15)], the sets $\mathcal{R}_{\leq t}=\mathcal{R}_t$ are compact and continuous with t with respect to the Hausdorff metric.

Since the optimal cost J is the minimum of the continuous function $\phi=\phi(x)$ on $\mathfrak{R}_{t_f}$, it depends continuously on t_f, and since $\mathfrak{R}_{t_f}$ is nondecreasing with t_f, the optimal cost is *nonincreasing* with t_f. That the constraint on the final time is active means that the cost is actually strictly *decreasing*. Notice in particular that the function $\phi=\mathrm{Tr}(C\rho)$ giving the cost in Eq. (7) is *linear* in the state (ρ); hence the minimum is necessarily achieved on the *boundary* of the reachable set [80]. Therefore $\lambda>0$ in Eq. (B6) implies that at the optimal final point, the boundary of the reachable set $\mathfrak{R}_{t_f}$ "moves" in such a way so as to make the cost decrease.

Appendix C: Coordinatization in terms of an orthonormal real matrix basis

The application to quantum systems of the PMP, which is typically formulated over real vector spaces as in Appendix B, has to account for the fact that in the quantum case the equations are complex-valued. To show how this can be done we start with some basic preliminaries.

We denote the Hilbert-Schmidt scalar product between operators A and B acting on an n-dimensional Hilbert space $\mathcal{H}$ by $\langle A,B\rangle:=\mathrm{Tr}(A^\dagger B)$. For superoperators $\mathcal{L}$, we denote the Hilbert-Schmidt adjoint of $\mathcal{L}$ by $\mathcal{L}^\dagger$, which is defined via

$$\langle\mathcal{L}^\dagger(A),B\rangle:=\langle A,\mathcal{L}(B)\rangle\ \ \forall A,B\,. \tag{C1}$$

We now choose an orthonormal basis $\{F_j\}$ for the *real* vector space of Hermitian $n\times n$ matrices with Hilbert-Schmidt inner product $\langle A,B\rangle=\mathrm{Tr}(AB)$. Let us "coordinatize" $X=\sum_j \boldsymbol{X}_j F_j$ in this basis, where henceforth we use the notation $\boldsymbol{X}=\{\boldsymbol{X}_j\}$ for the vector of real-valued coordinates of the operator X. Then:

$$\begin{aligned}\langle A,B\rangle&=\sum_{jk}\boldsymbol{A}_j\boldsymbol{B}_k\,\mathrm{Tr}(F_jF_k)=\sum_{jk}\boldsymbol{A}_j\boldsymbol{B}_k\delta_{jk}\\&=\boldsymbol{A}^T\boldsymbol{B}\,,\end{aligned} \tag{C2}$$

so that in these coordinates the inner product $\langle A,B\rangle$ corresponds to the standard inner product in $\mathbb{R}^{n^2}$. In particular, the Hermitian density operator $\rho=\sum_i\boldsymbol{\rho}_iF_i$, is now represented by a real, n^2-dimensional vector $\boldsymbol{\rho}$ [81]. Since the Liouvillian is Hermitian preserving, i.e., $[\mathcal{L}(X)]^\dagger=\mathcal{L}\left(X^\dagger\right)\forall X$, after coordinatization the operator $\mathcal{L}$ can be seen as an operator $\mathbb{R}^{n^2}\mapsto\mathbb{R}^{n^2}$. Indeed, denoting the corresponding matrix by $\boldsymbol{\mathcal{L}}$ in the chosen basis, i.e., $\boldsymbol{\mathcal{L}}_{ij}=\mathrm{Tr}\left(F_i\mathcal{L}\left(F_j\right)\right)$, one has

$$\begin{aligned}\boldsymbol{\mathcal{L}}^*_{ij}&=\mathrm{Tr}\left[\left(F_i\mathcal{L}\left(F_j\right)\right)^\dagger\right]=\mathrm{Tr}\left(\left[\mathcal{L}\left(F_j\right)\right]^\dagger F_i^\dagger\right)\\&=\mathrm{Tr}\left(\mathcal{L}\left(F_j\right)F_i\right)=\boldsymbol{\mathcal{L}}_{ij}\,.\end{aligned} \tag{C3}$$

i.e., the matrix $\boldsymbol{\mathcal{L}}$ is real and defines an operator $\mathbb{R}^{n^2}\mapsto\mathbb{R}^{n^2}$. Accordingly, Eq. (6) is transformed into a real-valued equation:

$$\dot{\boldsymbol{\rho}}=\boldsymbol{\mathcal{L}}\boldsymbol{\rho},\qquad\boldsymbol{\rho}(0)=\boldsymbol{\rho}_0\,. \tag{C4}$$

The cost (7) takes the form

$$J=\boldsymbol{C}^T\boldsymbol{\rho}(t_f)\,. \tag{C5}$$

We are now ready to state the PMP in the standard setting of real-valued functions, in the form needed for our purposes. Namely:

Theorem 9. *Assume $(\boldsymbol{\rho}^*,s^*)$ is an optimal pair for the problem defined by Eqs. (C4) and (C5) for a fixed final time t_f [82]. Then there exists a co-state vector $\boldsymbol{p}$ that satisfies [83]*

$$\dot{\boldsymbol{p}}^T=-\boldsymbol{p}^T\boldsymbol{\mathcal{L}},\quad\boldsymbol{p}(t_f)=-\boldsymbol{C}\,. \tag{C6}$$

Furthermore, define the PMP control Hamiltonian function

$$\mathbb{H}\left(\boldsymbol{p},\boldsymbol{\rho},s\right)=\boldsymbol{p}^T\boldsymbol{\mathcal{L}}\boldsymbol{\rho}\,. \tag{C7}$$

We then have the maximum principle*:*

$$\mathbb{H}\left(\boldsymbol{p}(t),\boldsymbol{\rho}^*(t),s^*(t)\right)=\max_{v\in[0,1]}\mathbb{H}\left(\boldsymbol{p}(t),\boldsymbol{\rho}^*(t),v\right)\,, \tag{C8}$$

and there exists a real non-negative constant λ such that

$$\mathbb{H}\left(\boldsymbol{p}(t),\boldsymbol{\rho}^*(t),s^*(t)\right)=\lambda\,. \tag{C9}$$

Theorem 1 is readily obtained applying coordinatization in reverse. To see that Eq. (C6) corresponds to Eq. (8), let us write Eq. (C6) explicitly as $\dot{\boldsymbol{p}}_i=-\sum_j\boldsymbol{\mathcal{L}}_{ji}\boldsymbol{p}_j$. Next, since $\boldsymbol{\mathcal{L}}$ is real we have $\boldsymbol{\mathcal{L}}_{ji}=\langle F_j,\mathcal{L}F_i\rangle=\langle F_j,\mathcal{L}F_i\rangle^*=\langle\mathcal{L}F_i,F_j\rangle=\langle F_i,\mathcal{L}^\dagger F_j\rangle$. Thus

$$\dot{p}=\sum_i\dot{\boldsymbol{p}}_iF_i=-\sum_{ij}\boldsymbol{\mathcal{L}}_{ji}\boldsymbol{p}_jF_i \tag{C10a}$$

$$=-\sum_{ij}\langle F_i,\mathcal{L}^\dagger F_j\rangle\boldsymbol{p}_jF_i \tag{C10b}$$

$$=-\sum_i\langle F_i,\mathcal{L}^\dagger p\rangle F_i=-\mathcal{L}^\dagger p\,. \tag{C10c}$$

Finally, the correspondence between Eqs. (C7) and (10) is a direct consequence of Eq. (C2): $\langle p,\mathcal{L}\rho\rangle=\boldsymbol{p}^T\boldsymbol{\mathcal{L}}\boldsymbol{\rho}$.

Appendix D: Proof of various formulas

1. Proof of Eq. (14)

The proof is, for for arbitrary operators A, B, X:

$$\begin{aligned}
\langle A, \mathcal{K}_X^\dagger(B)\rangle &= \langle \mathcal{K}_X(A), B\rangle && \text{(D1a)}\\
&= \langle -i[X,A], B\rangle = i\,\mathrm{Tr}([X,A]^\dagger B) && \text{(D1b)}\\
&= i\,\mathrm{Tr}(A^\dagger X^\dagger B - A^\dagger B X^\dagger) && \text{(D1c)}\\
&= i\,\mathrm{Tr}(A^\dagger [X^\dagger, B]) = i\langle A, [X^\dagger, B]\rangle && \text{(D1d)}\\
&= \langle A, -\mathcal{K}_{X^\dagger}(B)\rangle\ , && \text{(D1e)}
\end{aligned}$$

where we used $\langle X, Y\rangle = \mathrm{Tr}[X^\dagger Y]$.

2. Proof of Eq. (34)

Let $\mathcal{D}\rho := [W\rho, V^\dagger] + [W, \rho V^\dagger]$. Then:

$$\begin{aligned}
\langle X, \mathcal{D}\rho\rangle &= \mathrm{Tr}\left(X^\dagger [W\rho, V_\alpha^\dagger] + X^\dagger [V, \rho W^\dagger]\right)\\
&= \mathrm{Tr}\Big[\big(V^\dagger X^\dagger W - X^\dagger V^\dagger W\\
&\qquad + W^\dagger X^\dagger V - W^\dagger V X^\dagger\big)\rho\Big] && \text{(D2a)}\\
&= \mathrm{Tr}\Big[\big(W^\dagger X V - W^\dagger V X\\
&\qquad + V^\dagger X W - X V^\dagger W\big)^\dagger \rho\Big] && \text{(D2b)}\\
&= \mathrm{Tr}\Big[\big(W^\dagger [X,V] + [V^\dagger, X] W\big)^\dagger \rho\Big] && \text{(D2c)}\\
&= \langle \mathcal{D}^\dagger X, \rho\rangle\ , && \text{(D2d)}
\end{aligned}$$

which yields Eq. (34) when we replace W by $W_{\alpha\beta}$, V by $S_\alpha = S_\alpha^\dagger$, and sum over α, β.

3. Proof of $v_3(t) \geq 0$ for sufficiently small t

Let us compute the Dyson series solution of Eq. (22a) to second order:

$$\begin{aligned}
\boldsymbol{v}(t) = \Big[\mathbb{1} + \int_0^t dt_1\, \boldsymbol{\mathcal{M}}(s(t_1)) + \\
\int_0^t dt_1 \int_0^{t_1} dt_2\, \boldsymbol{\mathcal{M}}(s(t_1))\boldsymbol{\mathcal{M}}(s(t_2)) + O(t^3)\Big]\boldsymbol{v}_0\ .
\end{aligned} \tag{D3}$$

Using Eq. (22b), for the initial condition $\boldsymbol{v}_0 = (1,0,0)^T$ there is no contribution from the first (and in fact also the third) order, while the second order contributes via $[\boldsymbol{\mathcal{M}}(s(t_1))\boldsymbol{\mathcal{M}}(s(t_2))]_{31} = s(t_1)(1-s(t_2)) \geq 0$. Hence, using $s(t) \in [0,1]$, $v_3(t) = \int_0^t dt_1 \int_0^{t_1} dt_2\, s(t_1)(1-s(t_2)) + O(t^4) \geq 0$ for sufficiently small t. In contrast, for the initial condition $\boldsymbol{v}_0 = (-1,0,0)^T$ we have $v_3(t) \leq 0$ by the same argument.

Appendix E: Optimal control and the geometry of the reachable set

The cost considered in this work is linear in the state ρ, which after coordinatization we have identified with a point $\boldsymbol{\rho}$ in $\mathbb{R}^{n^2}$, that is, $J := \mathbf{C}^T\boldsymbol{\rho}$. In $\mathbb{R}^{n^2}$, we also consider the reachable sets $\mathfrak{R}_t$ at various times t. It is of interest to consider the level lines (hyperplanes) $\mathbf{C}^T\boldsymbol{\rho} = k$ for various k's. If k is the minimum cost at the final time t_f, the level line $\mathbf{C}^T\boldsymbol{\rho} = k$ intersects the boundary of the reachable set $\mathfrak{R}_{t_f}$ at the point $\boldsymbol{\rho}(t_f)$. In our case, the reachable sets $\mathfrak{R}_t$ are always nondecreasing with t. Figure 3(a) describes the regular situation of an active time constraint ($\lambda > 0$ in the main text). The intersection occurs at a point where the reachable set is increasing with time. Therefore an increase (decrease) of the final time t_f results in a decrease (increase) of the optimal cost. However, in principle, a different situation may occur which is described in Fig. 3(b). In this case, the reachable set $\mathfrak{R}_t$ increases with t but not at the point where the optimum occurs. In this case the final time constraint is not active ($\lambda = 0$). Notice that by continuity of the reachable set $\mathfrak{R}_t$ with t, the point $\boldsymbol{\rho}(t_f)$ where the optimum was achieved with the final time t_f will give the optimal for $t_f + \epsilon$ for sufficiently small $\epsilon > 0$. The corresponding control will be a zero control which keeps the state at the initial value for time ϵ followed by the same control applied to reach $\boldsymbol{\rho}(t_f)$.

Since we have not claimed uniqueness of the optimal control, the two situations may occur simultaneously for two different optimal trajectories. For one of them the final time constraint is active, while for the other one it is not. Given the additional structure of our problem, it should be possible to say more about the geometry of the reachable sets for the systems of interest here beside what is known from, for instance, Ref. [37]. However, this is beyond the scope of this work.

Appendix F: Proof of Theorem 3

Proof. Let us denote by $\tilde{J}(t_f)$ the cost function obtained with control $\tilde{s} := \tilde{s}(t)$. By definition, the optimal control satisfies $J_{\min}(t_f) \leq \tilde{J}(t_f)$. Hence, to prove $J_{\min}(t_f) < J_{\min}(0) = \langle C, \rho_0\rangle$, it suffices to show that there exists a control $\tilde{s}$ for Eq. (6) such that the corresponding cost $\tilde{J}(t_f)$ satisfies $\tilde{J}(t_f) < J_{\min}(0)$. This is the proof strategy we employ here.

Specifically, we consider a bang-bang switching schedule $\tilde{s}$ in the interval $[0, 2t]$ and denote the corresponding cost starting from ρ_0 by $\tilde{J}(2t)$. We will show that $\tilde{J}(2t) < \tilde{J}(0)$ for sufficiently small t, which gives

$$J_{\min}(2t) \leq \tilde{J}(2t) < \tilde{J}(0) = J_{\min}(0) = \mathrm{Tr}(C\rho_0)\ , \tag{F1}$$

and this proves the theorem with $t_f = 2t$.

The class of controls we consider is $\tilde{s} \equiv 1$ [corresponding to $\mathcal{K}_C$ in Eq. (15)] for an interval of length t, followed by $\tilde{s} \equiv 0$ [corresponding to $\mathcal{K}_B$ in Eq. (15)] for a second interval of length t. This gives for $J(t)$ [Eq. (7)]:

$$\tilde{J}(2t) = \mathrm{Tr}\left(Ce^{-iBt}e^{-iCt}\rho_0 e^{iCt}e^{iBt}\right)\ . \tag{F2}$$

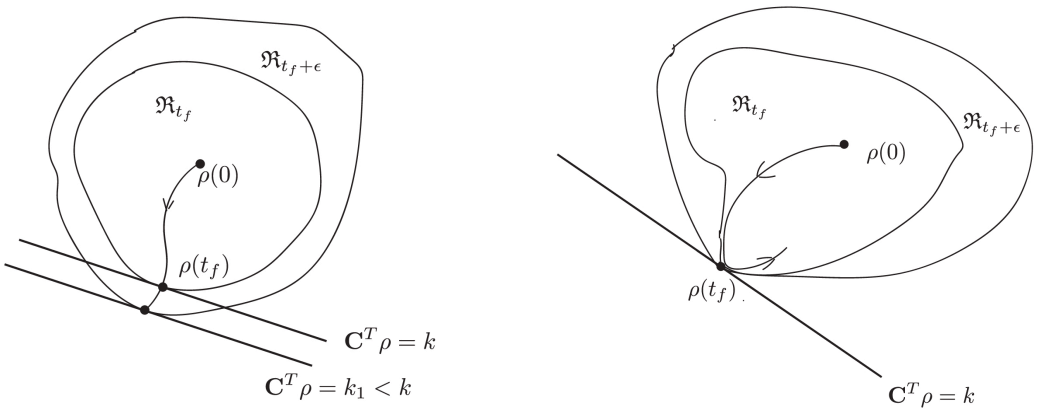

FIG. 3. (a) Behavior of the optimal cost in the regular case where the reachable set increases with the final time at the optimal point $\boldsymbol{\rho}(t_f)$. In this case we expect $\lambda > 0$. (b) Behavior of the optimal cost in the case where the reachable set increases with the final time but not at the optimal final point $\boldsymbol{\rho}(t_f)$. In this case we expect $\lambda = 0$.

We work in a basis where B is diagonal with eigenvalues in decreasing order: $B = \mathtt{diag}(\lambda_n, \lambda_{n-1}, \ldots, \lambda_1)$ and $\lambda_1 < \lambda_j$, for each $j = 2, 3, \ldots, n$ (nondegeneracy). In this basis $\rho_0 = \mathtt{diag}(0, \ldots, 0, 1)$.

In Eq. (F2), the Baker-Campbell-Hausdorff (BCH) formula yields:

$$\tilde{J}(2t) = \mathrm{Tr}\left[e^{iBt}Ce^{-iBt}\times \left(\rho_0 - i[C,\rho_0]t - [C,[C,\rho_0]]\frac{t^2}{2} + O(t^3)\right)\right] \quad \text{(F3)}$$

Applying the BCH formula again, this time to $e^{iBt}Ce^{-iBt}$, we obtain:

$$\tilde{J}(2t) = \mathrm{Tr}\left[\left(C + i[B,C]t - [B,[B,C]]\frac{t^2}{2}\right)\times \left(\rho_0 - i[C,\rho_0]t - [C,[C,\rho_0]]\frac{t^2}{2}\right)\right] + O(t^3). \quad \text{(F4)}$$

Expanding, we obtain:

$$\begin{aligned}\tilde{J}(2t) = {}& \mathrm{Tr}(C\rho_0) + i\,\mathrm{Tr}([B,C]\rho_0)t - \mathrm{Tr}([B,[B,C]]\rho_0)\frac{t^2}{2} \\ & - i\,\mathrm{Tr}(C[C,\rho_0])t + \mathrm{Tr}\left([B,C][C,\rho_0]\right)t^2 \\ & - \mathrm{Tr}\left(C[C,[C,\rho_0]]\right)\frac{t^2}{2} + O(t^3). \end{aligned} \quad \text{(F5)}$$

Several of the terms in the above equation vanish. In particular,

$$\mathrm{Tr}\left([B,C]\rho_0\right) = \mathrm{Tr}\left([\rho_0,B]C\right) = 0\,, \quad \text{(F6)}$$

since B and ρ_0 commute. $\mathrm{Tr}([B,[B,C]]\rho_0) = 0$ for the same reason. $-i\,\mathrm{Tr}(C[C,\rho_0]) = -i\,\mathrm{Tr}(\rho_0[C,C]) = 0$, and $\mathrm{Tr}\left(C[C,[C,\rho_0]]\right) = \mathrm{Tr}\left([C,\rho_0][C,C]\right) = 0$. Therefore we have

$$\tilde{J}(2t) - \mathrm{Tr}(C\rho_0) = \mathrm{Tr}\left([B,C][C,\rho_0]\right)t^2 + O(t^3)\,. \quad \text{(F7)}$$

Write

$$B = \begin{pmatrix}\Lambda & 0\\ 0 & \lambda_1\end{pmatrix}, \qquad C = \begin{pmatrix}C_1 & a\\ a^\dagger & c\end{pmatrix}, \quad \text{(F8)}$$

with $\Lambda = \mathtt{diag}(\lambda_n, \ldots, \lambda_2)$, C_1 an $(n-1)\times(n-1)$ Hermitian matrix, c a real number and a an $(n-1)$-th dimensional complex vector. With these notations, we have

$$[B,C] = \begin{pmatrix}[\Lambda, C_1] & (\Lambda - \lambda_1\mathbb{1})a\\ a^\dagger(\lambda_1\mathbb{1} - \Lambda) & 0\end{pmatrix} \quad \text{(F9)}$$

$$[C,\rho_0] = \begin{pmatrix}0 & a\\ -a^\dagger & 0\end{pmatrix}. \quad \text{(F10)}$$

From this we obtain:

$$\mathrm{Tr}\left([B,C][C,\rho_0]\right) = 2a^\dagger(\lambda_1\mathbb{1} - \Lambda)a\,, \quad \text{(F11)}$$

so that we have, from Eq. (F7):

$$\tilde{J}(2t) - \mathrm{Tr}(C\rho_0) = -t^2\left(\sum_{j=2}^{n}(\lambda_j - \lambda_1)|a_j|^2\right) + O(t^3)\,, \quad \text{(F12)}$$

where a_j are the components of a. Since $\lambda_j > \lambda_1$ for each j, we have for sufficiently small t:

$$\tilde{J}(2t) - \mathrm{Tr}(C\rho_0) = \tilde{J}(t) - J_{\min}(0) < 0\,, \quad \text{(F13)}$$

as required. We have assumed here that at least one of the components of a is nonzero. If that were not the case then $[\rho_0, C] = 0$ and ρ_0 would be fixed not just under B but also under C. There would then be no dynamics, which is a case that is naturally excluded. □

Appendix G: Singular arcs

Along singular arcs we have $x_C \equiv x_B$, i.e.,

$$\langle B, S\rangle \equiv \langle C, S\rangle \,. \tag{G1}$$

Differentiating Eq. (G1), using Eq. (39) we find $\langle B, \mathcal{K}_H S\rangle = \langle C, \mathcal{K}_H S\rangle$. Using $H = sB + (1-s)C$ and the antihermiticity of $\mathcal{K}_H$ we thus obtain:

$$\langle \mathcal{K}_H^\dagger(B), S\rangle = \langle \mathcal{K}_H^\dagger(C), S\rangle \implies \tag{G2a}$$
$$\langle (1-s)[C,B]^\dagger, S\rangle = \langle s[B,C]^\dagger, S\rangle \implies \tag{G2b}$$
$$-(1-s)\langle [C,B], S\rangle = s\langle [C,B], S\rangle \implies \tag{G2c}$$
$$\langle [C,B], S\rangle = 0 \,. \tag{G2d}$$

Analogously, differentiating Eq. (G2) and using the antihermiticity of $\mathcal{K}_H$ again, setting $D := [C,B]$, we have:

$$0 = \langle D, \dot{S}\rangle = \langle D, \mathcal{K}_H(S)\rangle = \langle \mathcal{K}_H^\dagger(D), S\rangle \tag{G3a}$$
$$= -(1-s)\langle [C,D]^\dagger, S\rangle - s\langle [B,D]^\dagger, S\rangle \tag{G3b}$$
$$= (1-s)\langle [C,D], S\rangle + s\langle [B,D], S\rangle \,. \tag{G3c}$$

Conditions (G1)-(G3) have to hold along a singular arc. In an algorithm to calculate the *dynamical Lie algebra* for the controllability of Eq. (15) [26], the matrices B and C are the matrices of "depth" zero in the calculation via iterated Lie brackets. The matrix $[C,B]$ is of depth one and the matrices $[[C,B],C]$ and $[[C,B],B]$ are of depth two. Now, one can have either (i) $\langle [C,D], S\rangle \neq \langle [B,D], S\rangle$ or (ii) $\langle [C,D], S\rangle = \langle [B,D], S\rangle$. In case (i) it follows from Eq. (G3) that

$$s = \frac{\langle [C,D], S\rangle}{\langle [C,D], S\rangle - \langle [B,D], S\rangle} \tag{G4}$$

(compare with Ref. [24, Eq. (12)]). This shows the continuity of s in the corresponding open set(s). Case (ii) implies $\langle [C,D], S\rangle \equiv \langle [B,D], S\rangle \equiv 0$ (in some closed set). One can further differentiate one of these equations and obtain an analog of Eq. (G3) at a higher order, at which point similar reasoning can be applied. In principle s can be defined in different intervals by equations such as Eq. (G4) or its higher order generalizations. In each interval s is continuous because of the continuity of S. We leave a more general proof of continuity of s on the *entire* singular arc as an open problem. In any case, equations (G2)-(G3) provide information on the dynamics along singular arc intervals. They are used in the example discussed in Appendix H.

Appendix H: Optimal control protocol for the spin-$\frac{1}{2}$ model

Here we analyze in detail the optimal control problem for the spin-$\frac{1}{2}$ model treated in Sec. IV C. Our goal is to give a simple but explicit example to show how the results developed in this paper can be used to find the optimal control. We shall use the same notation as in the example of Sec. IV C. To avoid the situation of an inactive terminal time constraint described in the example, we assume that the initial state is the ground state $x_0 = (-1,0,0)^T$, so that we can apply Theorem 3.

1. The global minimum is found using two non-singular arcs in time $t_f \geq \pi$

Recall that $C = \sigma^z/2$ and $B = \sigma^x/2$. The global minimum of the cost (7) is $J_{\min} = \mathrm{Tr}[C\rho_f] = -1/2$, achieved when $\rho_f = (\mathbb{1} - \sigma^z)/2$, the ground state of $C = \sigma^z/2$. Given that our initial condition is $\rho_0 = (\mathbb{1} - \sigma^x)/2$ (the ground state of B, corresponding to $\boldsymbol{v} = (-1,0,0)^T$), we can trivially reach $\rho_f = (\mathbb{1} - \sigma^z)/2$ by applying two consecutive bangs (i.e., unitary single-qubit gates): first $e^{-i(\pi/2)C}$ (rotation to $(\mathbb{1} - \sigma^y)/2$) with $s \equiv 0$, then $e^{-i(\pi/2)B}$ (rotation to $(\mathbb{1} - \sigma^z)/2$) with $s \equiv 1$. Each bang lasts for a time $\frac{\pi}{2}$, therefore the total bang-bang sequence last for a total time of π. *This sequence presents no singular arcs*. For any $t_f > \pi$ the constraint on t_f becomes inactive, i.e., increasing t_f cannot further lower the value of $J_{\min}$. Since we assume that the global ground state is not reached (recall the discussion in Sec. IV B), henceforth we assume that $t_f < \pi$. In principle, this setting could still allow for the appearance of singular arcs. However, we shall show that this is not the case.

2. Conditions on the singular arcs for the spin-$1/2$ model

Let us derive the conditions on the singular arcs in the present problem, which are a special case of the computations carried out in Appendix G. Using Eq. (20), we obtain $[C,B] = i\sigma^y/2$, $[[C,B],C] = -\sigma^x/2$, $[[C,B],B] = \sigma^z/2$. Using these, Eqs. (40b), (41), and (G1) for the switching operator S become:

$$2x_C = \mathrm{Tr}\,(S\sigma^z) \equiv \mathrm{Tr}\,(S\sigma^x) = 2x_B \,. \tag{H1}$$

Condition (G2) becomes:

$$\mathrm{Tr}\,(S\sigma^y) \equiv 0 \,, \tag{H2}$$

and condition (G3) becomes $(1-s)\,\mathrm{Tr}(\sigma^x S) - s\,\mathrm{Tr}(\sigma^z S) \equiv 0$, which using Eq. (H1) gives:

$$(1-2s)\,\mathrm{Tr}\,(\sigma^x S) \equiv 0 \,. \tag{H3}$$

Thus, either $s = 1/2$ or $\mathrm{Tr}(\sigma^x S) \equiv 0$. Let assume the latter. From Eqs. (H1) and (H2) we obtain $\mathrm{Tr}(\sigma^z S) = \mathrm{Tr}(\sigma^y S) \equiv 0$. Since we can expand $S = \frac{1}{2}\sum_{i=1}^3 \mathrm{Tr}(S\sigma_i)\sigma_i$ (S is traceless since it is defined as a commutator), this would then imply that $S \equiv 0$ on a singular interval. However, since S satisfies the linear equation (39), this would imply $S \equiv 0$ on the whole interval $[0, t_f]$ and, in particular, $[p_0, \rho_0] = [p_f, \rho_f] = -[C, \rho_f] = 0$. This would imply that ρ_f is a linear combination of eigenprojectors of C, but since ρ_f is a pure state it must in fact be equal to a single eigenprojector. Moreover, this must be the ground state of C since $J(t)$ is minimized at $t = t_f$. But, since $J(t_f) \leq J(0)$, the only possibility is that $J = \mathrm{Tr}(C\rho)$ reaches its global minimum at t_f, which contradicts our assumption that t_f is smaller than a value that would allow the global minimum to be reached. Hence we conclude that $\mathrm{Tr}(\sigma^x S) \neq 0$ in Eq. (H3), *which yields* $s \equiv 1/2$ *on the singular arcs*. However, we shall see in Proposition 3 that singular arcs are in fact not possible in this case.

3. Candidate optimal controls with a singular arc

Using conditions (H1) and (H2) along with Eq. (42) equated to λ, we have that in the time interval of a singular arc

$$S \equiv \lambda\,(\sigma^x + \sigma^z)\ . \tag{H4}$$

$\lambda = 0$ is impossible because according to the argument at the end of the previous subsection $S \equiv 0$ is to be excluded. Since $\lambda \neq 0$, we can apply all the conclusions of Theorem 2 and affirm that the optimal control starts with an $s \equiv 0$ bang arc and ends with an $s \equiv 1$ bang arc. Therefore, preceding or following a singular arc we must have $s \equiv 0$ or $s \equiv 1$, respectively. Let us show that after a singular arc we cannot go to a switching point, i.e., where Eq. (H1) holds [(λ, λ) in Fig. 2]. (Analogously, changing the sign of time, we can show that a singular arc cannot be preceded by a switching point.) Assume that after the singular arc we have $s \equiv 0$. The switching operator $S = S(t)$, with $t = 0$ at the end of this singular arc, is then the solution of Eq. (39) with the initial condition (H4), i.e., $S(t) = \lambda e^{-it\sigma^z/2}\sigma^x e^{it\sigma^z/2} + \lambda\sigma^z$. The minimum time needed for it to return to a switching point is $t = 2\pi$. This contradicts the fact that $t_f < \pi$ and therefore is impossible. Similar reasoning shows that we cannot go back to a switching point with $s \equiv 1$. Therefore, we have learned the following fact about optimal control in the single-qubit case:

Proposition 2. *The optimal control has at most one singular arc and, if it does, the optimal control is the sequence $s \equiv 0$, $s \equiv \frac{1}{2}$, $s \equiv 1$.*

Consider now the initial switching operator S_0, which together with the differential equation Eq. (39) determines the control sequence. Since S is traceless, we can write $S_0 = \boldsymbol{r}_0 \cdot \boldsymbol{\sigma}$, but using Eq. (37) and $\rho_0 = \frac{1}{2}(\mathbb{1} - \sigma^x)$ we see by expanding p in the Pauli matrix basis that S_0 cannot contain σ^x, i.e., we find that S_0 has the form

$$S_0 = r_{0y}\sigma^y + r_{0z}\sigma^z\ . \tag{H5}$$

In the first interval $s \equiv 0$ and therefore, from Eq. (39):

$$S(t) = r_{0z}\sigma^z + r_{0y}e^{-it\sigma^z/2}\sigma^y e^{it\sigma^z/2} \tag{H6a}$$

$$= r_{0z}\sigma^z + r_{0y}\cos(t)\sigma^y - r_{0y}\sin(t)\sigma^x\ . \tag{H6b}$$

If there is a singular arc and therefore S takes the form (H4), then we must have $t = \frac{\pi}{2}$ and $r_{0y} = -r_{0z}$ or $t = \frac{3\pi}{2}$ and $r_{0y} = r_{0z}$. The second case is to be excluded since $t_f < \pi$. After the singular arc we would have $s \equiv 1$, which, using Eq. (39) again would give:

$$S(t) = r_{0z}\sigma^x + r_{0z}e^{-it\sigma^x/2}\sigma^z e^{it\sigma^x/2} \tag{H7a}$$

$$= r_{0z}\sigma^x + r_{0z}(\cos(t)\sigma^z - \sin(t)\sigma^y)\ . \tag{H7b}$$

Since we have to reach the point $(x_C, x_B) = (0, \lambda)$ in Fig. 2, we must have $\cos(t) = 0$, i.e., $t = \frac{\pi}{2}$ or $t = \frac{3\pi}{2}$, which has to be added to the time used before the last interval. Therefore the total time is greater than or equal to π, which we have excluded.

In conclusion we have:

Proposition 3. *No singular arc exists in the optimal control for the spin-1/2 example with $t_f < \pi$.*

4. Candidate optimal controls without singular arcs

Now we consider the optimal control candidates knowing that they must be free of singular arcs, i.e., they can consist only of bangs. Since $S(t)$ is traceless we again use the parametrization $S(t) = \boldsymbol{r}(t) \cdot \boldsymbol{\sigma}$. We already know [Eq. (H5)] that:

$$\boldsymbol{r}(0) = \boldsymbol{r}_0 = (0, r_{0y}, r_{0z})^T\ . \tag{H8}$$

We know from Eq. (39) that the vector $\boldsymbol{r}$ evolves according to Eqs. (22a)-(22b). More explicitly, let $\boldsymbol{X} := \boldsymbol{\mathcal{K}}_B$ and $\boldsymbol{Z} := \boldsymbol{\mathcal{K}}_C$. When $s \equiv 0$, $H = C$ and $\boldsymbol{r}$ evolves according to $e^{t\boldsymbol{Z}}$, and likewise when $s \equiv 1$ it evolves according to $e^{t\boldsymbol{X}}$, where, using Eq. (21):

$$e^{t\boldsymbol{X}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(t) & -\sin(t) \\ 0 & \sin(t) & \cos(t) \end{pmatrix} \tag{H9a}$$

$$e^{t\boldsymbol{Z}} = \begin{pmatrix} \cos(t) & -\sin(t) & 0 \\ \sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{pmatrix}\ . \tag{H9b}$$

Since we have shown that $\lambda > 0$, from Theorem 2, the control law will start with an $s \equiv 0$ bang arc and end with an $s \equiv 1$ bang arc. The control law is determined by a sequence of intervals of lengths $\{\tau_1, \tau_2, \dots\}$ where for k odd (even) τ_k marks the switch from $s \equiv 0$ to $s \equiv 1$ ($s \equiv 1$ to $s \equiv 0$), that is $\boldsymbol{Z} \to \boldsymbol{X}$ ($\boldsymbol{X} \to \boldsymbol{Z}$). That is,

$$\boldsymbol{r}(t_k) = \begin{cases} e^{\tau_k \boldsymbol{Z}}\boldsymbol{r}(t_{k-1}) & k \text{ odd} \\ e^{\tau_k \boldsymbol{X}}\boldsymbol{r}(t_{k-1}) & k \text{ even} \end{cases}\ , \tag{H10}$$

where $t_k = \sum_{i=1}^{k} \tau_i$ is the total time after k intervals. Note that in principle there is no guarantee that such a switching sequence is finite, even if the total control interval is finite; this is known in the control theory literature as the *Fuller phenomenon* (see, e.g., Ref. [84]). We shall see in Remark 1 that this does not happen in our case and we have a *finite* sequence of intervals of lengths $\{\tau_1, \tau_2, \dots, \tau_N\}$ with N even (according to Theorem 2). Given our definitions, $t_f = \sum_{i=1}^{N} \tau_i$, where τ_1 and τ_N are the lengths of the initial ($\boldsymbol{Z}$) and final ($\boldsymbol{X}$) arcs, respectively.

5. Characterization of the switching times

Note that the vector $\boldsymbol{r} = (r_x, r_y, r_z)^T$ consists of the components of S along the Pauli basis, and that $r_x = x_B$ and $r_z = x_C$ [Eq. (H1)]. Recall also that, as argued in Sec. VI A (see Fig. 2), $(x_C, x_B) = (\lambda, \lambda)$ at every switching point. Hence $r_x = r_z = \lambda$ at every switching point between non-singular arcs in our discussion below.

The optimal candidate control law is characterized by a sequence of intervals of length τ_1, τ_2, etc. Define the sequence $\{\Delta_k\}$ recursively from the sequence $\{\tau_k\}$ via $\Delta_0 = 0$, $\Delta_k = \tau_k - \Delta_{k-1}$, for $k = 1, 2, \dots$. Then:

Lemma 1. *For $n = 1, 2, \ldots$, except for the n corresponding to the last control interval*

$$(-1)^n \sin(\Delta_n)\, r_{0y} = r_{0z}, \tag{H11}$$

[cf. Eq. (H5)] and

$$\boldsymbol{r}(t_n) = (r_{0z}, \cos(\Delta_n) r_{0y}, r_{0z})^T \ . \tag{H12}$$

Proof. The proof is by induction on n. For $n = 1$ we have $\Delta_1 = t_1 = \tau_1$. At the end of the first arc we must reach the point $(x_C, x_B) = (\lambda, \lambda)$, which means that $\boldsymbol{r}(\tau_1) = (r_{0z}, *, r_{0z})^T$. Thus, using Eq. (H10) and $e^{\tau_1 \boldsymbol{Z}}$ in Eq. (H9b), we obtain Eqs. (H11) and (H12).

Now assume Eq. (H11) and Eq. (H12) hold for $n-1$. If n is even we have

$$\boldsymbol{r}(t_n) = e^{\tau_n \boldsymbol{X}} \boldsymbol{r}(t_{n-1}) \ . \tag{H13}$$

Next, use Eq. (H9a), and impose $\boldsymbol{r}(t_n) = (r_{0z}, *, r_{0z})^T$, since $(x_C, x_B) = (\lambda, \lambda)$ at every switching point. Equality of the z component then gives $\sin(\tau_n)\cos(\Delta_{n-1}) r_{0y} + \cos(\tau_n) r_{0z} = r_{0z}$, and using Eq. (H11) with n replaced by $n-1$, we obtain $r_{0z} = \sin(\tau_n)\cos(\Delta_{n-1}) r_{0y} - \cos(\tau_n)\sin(\Delta_{n-1}) r_{0y} = \sin(\tau_n - \Delta_{n-1}) r_{0y} = \sin(\Delta_n) r_{0y}$. Calculating the y component of $\boldsymbol{r}(t_n)$, we obtain $\cos(\tau_n)\cos(\Delta_{n-1}) r_{0y} - \sin(\tau_n) r_{0z} = \cos(\tau_n)\cos(\Delta_{n-1}) r_{0y} + \sin(\tau_n)\sin(\Delta_{n-1}) r_{0y} = \cos(\Delta_n) r_{0y}$, using again the inductive assumption Eq. (H11). A similar calculation with $\boldsymbol{Z}$ replacing $\boldsymbol{X}$ in Eq. (H13) gives the result when n is odd. □

6. Determination of the Optimal Control

We now use the formulas in the above Lemma to determine the optimal control. Define $\mu = \arcsin\left(\frac{r_{0z}}{r_{0y}}\right)$ and notice that from Eq. (H11) for $n = 1$ and $\Delta_1 = \tau_1$ we have $\mu = \arcsin(-\sin(\tau_1))$. Since $0 < \tau_1 < \pi$, we have $\mu = -\tau_1$ for $\tau_1 \in (0, \pi/2]$ and $\mu = -\pi + \tau_1$ for $\tau_1 \in [\pi/2, \pi)$, and in particular $\mu < 0$.

Let us consider first the possibility that $0 < \tau_1 \leq \frac{\pi}{2}$. We also have $\tau_1 = \Delta_1 = -\mu$. If there is more than one switch (i.e., $\tau_2 > 0$), then we can derive Δ_2 from Eq. (H11). We have either $\Delta_2 = \pi - \mu + 2l\pi$ or $\Delta_2 = \mu + 2l\pi$ for integer l. Recalling that $\tau_2 = \tau_1 + \Delta_2$, in the first case we have $\tau_2 = \pi - 2\mu + 2l\pi$, and in the second case $\tau_2 = 2l\pi$. The second case is not possible because τ_2 must be in $(0, \pi)$. The first case is not possible either because $l \geq 0$ would contradict that the total time must be less than π while $l < 0$ would give a negative or zero interval τ_2. Therefore, in the case $0 < \tau_1 \leq \frac{\pi}{2}$ there exists only one switch and the control is simply the sequence of two bangs, one corresponding to $s \equiv 0$ followed by one corresponding to $s \equiv 1$. Before determining where the switch must occur, let us consider the case $\frac{\pi}{2} < \tau_1 < \pi$.

Lemma 2. *Assume that $\tau_1 \in (\pi/2, \pi)$ and define $\mu = \arcsin\left(\frac{r_{0z}}{r_{0y}}\right) < 0$. Then $\Delta_n = \pi + \mu$ for n odd and $\Delta_n = \mu$ for n even.*

Proof. The claim follows by induction from Eq. (H11). Applying it for $n = 1$, since $\Delta_1 = \tau_1$, we have $\Delta_1 = \pi + \mu$. Now, assume that the claim is true for n even. From Eq. (H11) applied for n even, we obtain $\Delta_n = \mu + 2l\pi$ or $\Delta_n = \pi - \mu + 2l\pi$, for integer l. Since $\tau_n = \Delta_{n-1} + \Delta_n$, using the inductive assumption, we obtain in the two cases, $\tau_n = \pi + 2\mu + 2l\pi$ and $\tau_n = 2\pi + 2l\pi$, respectively. The latter case is impossible because it would mean that $\tau_n > \pi$. The first case is only possible with $l = 0$, because $l > 0$ would imply $\tau_n > 2\pi$ while $l < 0$ would give a negative time interval. Since we saw above that $\Delta_n = \mu + 2l\pi$ for n even, this gives $\Delta_n = \mu$ for such n.

Let us now prove that $\Delta_n = \pi + \mu$ for $n > 1$ and odd. Again using Eq. (H11) we obtain either $\Delta_n = -\mu + 2l\pi$ or $\Delta_n = \pi + \mu + 2l\pi$. The first case is impossible because it would mean $\tau_n = \Delta_{n-1} + \Delta_n = 2l\pi$ (using the inductive assumption). The second case would give $\tau_n = \Delta_{n-1} + \Delta_n = \pi + 2\mu + 2l\pi$ which is only possible for $l = 0$. This gives $\Delta_n = \pi + \mu$. □

Remark 1. *One of the consequences of the above lemma is that the switching sequence is finite, i.e., we do not have intervals between two switches which become arbitrarily small and hence the Fuller phenomenon [84] is ruled out in our case. In particular if $0 < \tau_1 < \frac{\pi}{2}$ there is only one switch, as we have seen, while if $\frac{\pi}{2} < \tau_1 < \pi$ then we have multiple switches with $\tau_k = \Delta_k + \Delta_{k-1} = \pi + 2\mu$, which is a constant independent of k.*

In order to learn more about the optimal control, and rule out the second case of $\frac{\pi}{2} < \tau_1 < \pi$, we examine the final arc, which is of the form $e^{\tau_N \boldsymbol{X}}$ ($s \equiv 1$). Recall that with the final arc we have to reach the point $(x_C, x_B) = (0, \lambda)$, which imposes that the final switching operator is of the form $\boldsymbol{r}(t_N) = (r_{0z}, *, 0)^T$ [see Eq. (H7b) and the discussion immediately below it]. Thus, using Eq. (H13) with $n = N$ and Eq. (H9a), we obtain $\sin(\tau_N)\cos(\Delta_{N-1}) r_{0y} + \cos(t_N) r_{0z} = 0$. Using Eq. (H11) and $r_{0y} \neq 0$, we obtain $\sin(\tau_N)\cos(\Delta_{N-1}) - \sin(\Delta_{N-1})\cos(\tau_N) = \sin(\Delta_N) = 0$, where $\Delta_N = \tau_N - \Delta_{N-1}$. Therefore $\tau_N = \Delta_{N-1} + l\pi$ for l integer. Now there are two cases: Multiple switches or only one switch. In the case of multiple switches, we are in the situation described in Lemma 2. We have $\Delta_{N-1} = \Delta_1 = \pi + \mu$. Therefore $\tau_N = \pi + \mu + l\pi$. The integer l must be zero because if it is positive we have $\tau_N > \pi$ and if it is negative, we have a negative interval τ_N. Therefore $\tau_N = \pi + \mu$. However $t_f \geq \tau_1 + \tau_N = 2\Delta_1 = 2(\pi + \mu) > \pi$ which is impossible. Therefore the situation $\frac{\pi}{2} < \tau_1 < \pi$ cannot occur. The only possibility is the situation with $0 < \tau_1 < \frac{\pi}{2}$ with one switch only. In this case, as above we have $\tau_2 = \Delta_1 + l\pi = \tau_1 + l\pi$ with $l = 0$ since again $l < 0$ will give a negative time interval and l positive will give total time greater than π. So $\tau_2 = \tau_1$ and the optimal control is the simplest one. This completes the proof of Proposition 4.

Appendix I: Additional considerations regarding the shortening of the initial and final arcs

In principle one can be more quantitative about the shortening of the arcs discussed in Sec. VI B by repeating the analysis of Ref. [85] for the quantity $x_B(\Delta t)$ [Eq. (45)]. The latter work provided a detailed analysis of the return probability $\mathcal{F}(t) := \left|\langle\psi|e^{-itH}|\psi\rangle\right|^2$, where H is a time-independent Hamiltonian and $|\psi\rangle$ an initial state. One of the results of Ref. [85] was an explicit form for the average number of zeroes of the equation $\mathcal{F}(t) = v$. Indeed, $\mathcal{F}$ is a particular case of the left-hand side of Eq. (45) and can be written in that form with $B = S_0 = |\psi\rangle\langle\psi|$, $C = H$. Let $N_{x_B}(v)$ be the average number of solutions of the equation $x_B(t) = v$. The number of zeroes N in a large interval of length T turns out to be proportional to T: $N_{x_B}(v) = TD_{x_B}(v)$ where D_{x_B} can be computed using the methods of Ref. [85]. Then $\Delta t \simeq T/N_{x_B}(\lambda) = 1/D_{x_B}(\lambda)$. Using [85, Eq. (6)], which applies for the special case mentioned above, we then obtain:

$$\Delta t(\lambda) \simeq \frac{\sqrt{\pi}}{2}\frac{1}{\Delta E}\sqrt{\frac{\langle x_B\rangle}{\lambda}}e^{\lambda/\langle x_B\rangle}\ , \tag{I1}$$

where $\langle x_B\rangle := \lim_{T\to\infty}\frac{1}{T}\int_0^T x_B(t)dt$ and ΔE is the standard deviation of the energies $\{E_k\}$ with respect to the distribution $\{p_k = \|\Pi_k|\psi\rangle\|^4 / \sum_n \|\Pi_n|\psi\rangle\|^4\}$ (recall that $C = \sum_k E_k\Pi_k$ is the spectral resolution of C). The derivation of Eq. (I1) is subtle and is carried out in Ref. [85]. Since it only represents a special case in our context, we do not pursue it further here, and present Eq. (I1) mainly to stimulate the interest of the reader and establish a possible entry point towards a rigorous treatment of the shortening of the initial and final arcs.

As a final comment, note that when C is a classical Ising Hamiltonian of the form $C = J\sum_{i,j=1}^n \sigma_i^z\sigma_j^z$, the energies are integer multiples of J and highly degenerate (therefore the spectrum is commensurate: $\{E_k\} \subset \{0, \pm J, \pm 2J, \ldots\}$). This implies that the function $x_B(t)$ in Eq. (46) is periodic with period not larger than $2\pi/J$ (as opposed to being almost periodic) and this has implications for $D_{x_B}(\lambda)$, but the general considerations we have outlined still hold.

[1] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, Aug 2018. URL: http://dx.doi.org/10.22331/q-2018-08-06-79, doi:10.22331/q-2018-08-06-79.

[2] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse Ising model. 58(5):5355, November 1998. URL: http://journals.aps.org/pre/abstract/10.1103/PhysRevE.58.5355.

[3] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm, 2014. URL: https://arxiv.org/abs/1411.4028, arXiv:1411.4028.

[4] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum Computation by Adiabatic Evolution. *arXiv:quant-ph/0001106*, January 2000. URL: http://arxiv.org/abs/quant-ph/0001106.

[5] T. Kato. On the adiabatic theorem of quantum mechanics. *J. Phys. Soc. Jap.*, 5:435, 1950. doi:10.1143/JPSJ.5.435.

[6] Sabine Jansen, Mary-Beth Ruskai, and Ruedi Seiler. Bounds for the adiabatic approximation with applications to quantum computation. *J. Math. Phys.*, 48(10):102111, 2007. URL: http://scitation.aip.org/content/aip/journal/jmp/48/10/10.1063/1.2798382.

[7] Daniel A. Lidar, Ali T. Rezakhani, and Alioscia Hamma. Adiabatic approximation with exponential accuracy for many-body systems and quantum computation. *J. Math. Phys.*, 50(10):102106, 2009. URL: http://scitation.aip.org/content/aip/journal/jmp/50/10/10.1063/1.3236685.

[8] Yimin Ge, András Molnár, and J. Ignacio Cirac. Rapid adiabatic preparation of injective projected entangled pair states and gibbs states. *Physical Review Letters*, 116(8):080503–, 02 2016. URL: http://link.aps.org/doi/10.1103/PhysRevLett.116.080503.

[9] Aniruddha Bapat and Stephen Jordan. Bang-bang control as a design principle for classical and quantum optimization algorithms. *Quantum Information & Computation*, 19(5 & 6):424–446, 2019. URL: https://arxiv.org/abs/1812.02746.

[10] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D. Lukin. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Physical Review X*, 10(2):021067–, 06 2020. URL: https://link.aps.org/doi/10.1103/PhysRevX.10.021067.

[11] Michael Streif and Martin Leib. Comparison of qaoa with quantum and simulated annealing. *arXiv preprint arXiv:1901.01903*, 2019. URL: https://arxiv.org/abs/1901.01903.

[12] Guido Pagano, Aniruddha Bapat, Patrick Becker, Katherine S. Collins, Arinjoy De, Paul W. Hess, Harvey B. Kaplan, Antonis Kyprianidis, Wen Lin Tan, Christopher Baldwin, Lucas T. Brady, Abhinav Deshpande, Fangli Liu, Stephen Jordan, Alexey V. Gorshkov, and Christopher Monroe. Quantum approximate optimization of the long-range ising model with a trapped-ion quantum simulator. *Proceedings of the National Academy of Sciences*, 117(41):25396, 10 2020. URL: http://www.pnas.org/content/117/41/25396.abstract.

[13] Jérémie Roland and Nicolas J. Cerf. Quantum search by local adiabatic evolution. *Phys. Rev. A*, 65(4):042308–, 03 2002. URL: http://link.aps.org/doi/10.1103/PhysRevA.65.042308.

[14] A. T. Rezakhani, A. K. Pimachev, and D. A. Lidar. Accuracy versus run time in an adiabatic quantum search. *Phys. Rev. A*, 82(5):052305–, 11 2010. URL: http://link.aps.org/doi/10.1103/PhysRevA.82.052305.

[15] A. T. Rezakhani, W. J. Kuo, A. Hamma, D. A. Lidar, and P. Zanardi. Quantum adiabatic brachistochrone. *Phys. Rev. Lett.*, 103(8):080502–, 08 2009. URL: http://link.aps.org/doi/10.1103/PhysRevLett.103.080502.

[16] Satoshi Morita and Hidetoshi Nishimori. Mathematical foundation of quantum annealing. *J. Math. Phys.*, 49(12):125210–47, 12 2008. doi:10.1063/1.2995837.

[17] O. Galindo and V. Kreinovich. What is the optimal annealing schedule in quantum annealing. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 963–967, 2020. doi:10.1109/SSCI47803.2020.9308407.

[18] Mario Szegedy. What do QAOA energies reveal about graphs?, 2019. URL: https://arxiv.org/abs/1912.12277, arXiv:1912.12277.

[19] Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamon. Optimizing variational quantum algorithms using pontryagin's minimum principle. *Physical Review X*, 7(2):021027–, 05 2017. URL: https://link.aps.org/doi/10.1103/PhysRevX.7.021027, doi:10.1103/PhysRevX.7.021027.

[20] Robert J. Kipka and Yuri S. Ledyaev. Pontryagin Maximum Principle for Control Systems on Infinite Dimensional Manifolds. *Set-Valued Var. Anal*, 23(1):133–147, March 2015. doi:10.1007/s11228-014-0301-8.

[21] Chungwei Lin, Yebin Wang, Grigory Kolesov, and Uroš Kalabić. Application of pontryagin's minimum principle to grover's quantum search problem. *Physical Review A*, 100(2):022327–, 08 2019. URL: https://link.aps.org/doi/10.1103/PhysRevA.100.022327, doi:10.1103/PhysRevA.100.022327.

[22] Glen Bigan Mbeng, Rosario Fazio, and Giuseppe E. Santoro. Optimal quantum control with digitized quantum annealing, 2019. URL: https://arxiv.org/abs/1911.12259, arXiv:1911.12259.

[23] Glen Bigan Mbeng, Rosario Fazio, and Giuseppe Santoro. Quantum annealing: a journey through digitalization, control, and hybrid quantum variational schemes, 2019. URL: https://arxiv.org/abs/1906.08948, arXiv:1906.08948.

[24] Lucas T. Brady, Christopher L. Baldwin, Aniruddha Bapat, Yaroslav Kharkov, and Alexey V. Gorshkov. Optimal Protocols in Quantum Annealing and Quantum Approximate Optimization Algorithm Problems. *Phys. Rev. Lett.*, 126(7):070505, February 2021. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.070505.

[25] In the control literature the notation u or $u(t)$ is used to denote the control function, rather than s or $s(t)$. Here we choose to use the notation that is more familiar in the quantum computing community.

[26] Domenico D'Alessandro. *Introduction to Quantum Control and Dynamics*. Chapman and Hall/CRC, Boca Raton, 1 edition edition, August 2007.

[27] Velimir Jurdjevic and Héctor J Sussmann. Control systems on Lie groups. *Journal of Differential Equations*, 12(2):313–329, September 1972. URL: https://www.sciencedirect.com/science/article/pii/0022039672900356, doi:10.1016/0022-0396(72)90035-6.

[28] We use the notation $f \equiv c$ to mean that the function f is "identically" equal to c in some interval Θ, i.e., $f \equiv c$ is equivalent to $f(t) = c$ for $t \in \Theta$.

[29] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Reviews of Modern Physics*, 90(1):015002–, 01 2018. URL: https://link.aps.org/doi/10.1103/RevModPhys.90.015002.

[30] Philipp Hauke, Helmut G Katzgraber, Wolfgang Lechner, Hidetoshi Nishimori, and William D Oliver. Perspectives of quantum annealing: Methods and implementations. *Reports on Progress in Physics*, 2020. URL: https://iopscience.iop.org/article/10.1088/1361-6633/ab85b8.

[31] Adolfo del Campo. Shortcuts to adiabaticity by counterdiabatic driving. *Physical Review Letters*, 111(10):100502–, 09 2013. URL: http://link.aps.org/doi/10.1103/PhysRevLett.111.100502.

[32] Kazutaka Takahashi. Shortcuts to adiabaticity for quantum annealing. *Physical Review A*, 95(1):012309–, 01 2017. URL: http://link.aps.org/doi/10.1103/PhysRevA.95.012309.

[33] E. J. Crosson and D. A. Lidar. Prospects for quantum enhancement with diabatic quantum annealing. *Nature Reviews Physics*, 2021. URL: https://doi.org/10.1038/s42254-021-00313-6.

[34] The form we have assumed is called a time-convolutionless master equation. The most general master equation is in Nakajima-Zwanzig form and includes a memory kernel superoperator $\mathcal{K}(t,t')$ acting jointly on the system and the environment E, such that (for a factorized initial condition) $\dot{\rho} = \mathrm{Tr}_E \int_0^t \mathcal{K}(t,t')\rho(t')\otimes\rho_E dt'$, with ρ_E a fixed environment state and Tr_E denoting the partial trace over the environment [43].

[35] We also use an asterisk to denote complex conjugation; the meaning will always be clear by context.

[36] $\mathcal{L}^\dagger$ in Eq. (8) indicates the Hilbert-Schmidt adjoint of $\mathcal{L}$, defined via $\langle \mathcal{L}^\dagger(A), B\rangle := \langle A, \mathcal{L}(B)\rangle = \mathrm{Tr}[A^\dagger \mathcal{L}(B)]$, $\forall A, B$; see Appendix C.

[37] Víctor Ayala, Heriberto Román-Flores, and Adriano Da Silva. About the continuity of reachable sets of restricted affine control systems. *Chaos, Solitons & Fractals*, 94:37–43, January 2017. URL: https://www.sciencedirect.com/science/article/pii/S0960077916303332, doi:10.1016/j.chaos.2016.11.006.

[38] Armin Rahmani, Takuya Kitagawa, Eugene Demler, and Claudio Chamon. Cooling through optimal control of quantum evolution. *Physical Review A*, 87(4):043607–, 04 2013. URL: https://link.aps.org/doi/10.1103/PhysRevA.87.043607, doi:10.1103/PhysRevA.87.043607.

[39] Seraph Bao, Silken Kleer, Ruoyu Wang, and Armin Rahmani. Optimal control of superconducting gmon qubits using pontryagin's minimum principle: Preparing a maximally entangled state with singular bang-bang protocols. *Physical Review A*, 97(6):062343–, 06 2018. URL: https://link.aps.org/doi/10.1103/PhysRevA.97.062343, doi:10.1103/PhysRevA.97.062343.

[40] Torsten Karzig, Armin Rahmani, Felix von Oppen, and Gil Refael. Optimal control of majorana zero modes. *Physical Review B*, 91(20):201404–, 05 2015. URL: https://link.aps.org/doi/10.1103/PhysRevB.91.201404, doi:10.1103/PhysRevB.91.201404.

[41] E.B. Davies. *Quantum Theory of Open Systems*. Academic Press, London, 1976.

[42] Robert Alicki and K. Lendi. *Quantum Dynamical Semigroups and Applications*. Springer Science & Business Media, April 2007.

[43] H.-P. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, Oxford, 2002. URL: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199213900.001.0001/acprof-9780199213900.

[44] Angel Rivas and Susana F. Huelga. *Open Quantum Systems: An Introduction*. SpringerBriefs in Physics. Springer-Verlag, Berlin Heidelberg, 2012. URL: https://doi.org/10.1007/978-3-642-23354-8.

[45] Daniel A Lidar. Lecture notes on the theory of open quantum systems. *arXiv preprint arXiv:1902.00967*, 2019. URL:

https://arxiv.org/abs/1902.00967.

[46] Antonella Fiacca, Nikolaos S. Papageorgiou, and Francesca Papalini. On the existence of optimal controls for nonlinear infinite dimensional systems. *Czechoslovak Mathematical Journal*, 48(2):291–312, June 1998. doi:10.1023/A:1022889404413.

[47] Lorenzo Campos Venuti and Daniel A. Lidar. Error reduction in quantum annealing using boundary cancellation: Only the end matters. *Phys. Rev. A*, 98(2):022315, August 2018. URL: https://link.aps.org/doi/10.1103/PhysRevA.98.022315, doi:10.1103/PhysRevA.98.022315.

[48] Andrew M. Childs, Edward Farhi, and John Preskill. Robustness of adiabatic quantum computation. 65(1):012322, December 2001. URL: http://journals.aps.org/pra/abstract/10.1103/PhysRevA.65.012322.

[49] Robert Alicki, Daniel A. Lidar, and Paolo Zanardi. Internal consistency of fault-tolerant quantum error correction in light of rigorous derivations of the quantum markovian limit. *Phys. Rev. A*, 73:052311, May 2006. doi:10.1103/PhysRevA.73.052311.

[50] Tameem Albash, Sergio Boixo, Daniel A. Lidar, and Paolo Zanardi. Quantum adiabatic Markovian master equations. *New J. Phys.*, 14(12):123016, December 2012. URL: http://iopscience.iop.org/1367-2630/14/12/123016, doi:10.1088/1367-2630/14/12/123016.

[51] Evgeny Mozgunov and Daniel Lidar. Completely positive master equation for arbitrary driving and small level spacing. *Quantum*, 4:227, February 2020. doi:10.22331/q-2020-02-06-227.

[52] Makoto Yamaguchi, Tatsuro Yuge, and Tetsuo Ogawa. Markovian quantum master equation beyond adiabatic regime. *Physical Review E*, 95(1):012136–, 01 2017. URL: https://link.aps.org/doi/10.1103/PhysRevE.95.012136, doi:10.1103/PhysRevE.95.012136.

[53] Roie Dann, Amikam Levy, and Ronnie Kosloff. Time-dependent markovian quantum master equation. *Physical Review A*, 98(5):052129–, 11 2018. URL: https://link.aps.org/doi/10.1103/PhysRevA.98.052129, doi:10.1103/PhysRevA.98.052129.

[54] Frederik Nathan and Mark S. Rudner. Universal lindblad equation for open quantum systems. *Physical Review B*, 102(11):115109–, 09 2020. URL: https://link.aps.org/doi/10.1103/PhysRevB.102.115109, doi:10.1103/PhysRevB.102.115109.

[55] Dragomir Davidović. Completely Positive, Simple, and Possibly Highly Accurate Approximation of the Redfield Equation. *Quantum*, 4:326, September 2020. doi:10.22331/q-2020-09-21-326.

[56] Marek Winczewski, Antonio Mandarino, Michał Horodecki, and Robert Alicki. Bypassing the intermediate times dilemma for open quantum system, 2021. URL: https://arxiv.org/abs/2106.05776, arXiv:2106.05776.

[57] Vittorio Gorini, Andrzej Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of n-level systems. *J. Math. Phys.*, 17(5):821–825, 05 1976. URL: dx.doi.org/10.1063/1.522979.

[58] G. Lindblad. On the generators of quantum dynamical semigroups. *Comm. Math. Phys.*, 48(2):119–130, 1976. URL: dx.doi.org/10.1007/BF01608499, doi:10.1007/BF01608499.

[59] Dariusz Chruściński and Saverio Pascazio. A brief history of the gkls equation. *Open Systems & Information Dynamics*, 24(03):1740001, 2018/04/06 2017. doi:10.1142/S1230161217400017.

[60] It can be shown that the conditions of Filippov's theorem (see Ref. [73, Th. 2.1]) for the existence of the optimal control solution are satisfied also in this case.

[61] In writing these expressions we have adapted the results of Ref. [55] to the adiabatic limit by using the instantaneous eigenbasis of H_S, and also performed the "$\int^t \to \int^\infty$ approximation" (otherwise γ would have a t dependence, which would prevent us from being able to apply Theorem 1); see Ref. [55] for complete details.

[62] Of course $[S_\alpha, C] = 0 \;\Rightarrow [S_\alpha \circ \sqrt{\gamma}, C] = 0$ when $\gamma_{nm} = c$ $\forall n, m$, but this is a highly nongeneric scenario.

[63] The commutant of an algebra $\mathcal{A} = \mathrm{Alg}\{A\}$ is defined as the set $\mathcal{A}' := \{X | [X, A] = 0 \; \forall A \in \mathcal{A}\}$.

[64] Paolo Zanardi. Dissipation and decoherence in a quantum register. *Physical Review A*, 57(5):3276–3284, 05 1998. URL: https://link.aps.org/doi/10.1103/PhysRevA.57.3276, doi:10.1103/PhysRevA.57.3276.

[65] D. A. Lidar, I. L. Chuang, and K. B. Whaley. Decoherence-free subspaces for quantum computation. *Phys. Rev. Lett.*, 81(12):2594–2597, 09 1998. URL: http://link.aps.org/doi/10.1103/PhysRevLett.81.2594.

[66] D. Bacon, J. Kempe, D. A. Lidar, and K. B. Whaley. Universal fault-tolerant quantum computation on decoherence-free subspaces. *Phys. Rev. Lett.*, 85(8):1758–1761, 08 2000. URL: http://link.aps.org/doi/10.1103/PhysRevLett.85.1758.

[67] Morten Kjaergaard, Mollie E. Schwartz, Jochen Braumüller, Philip Krantz, Joel I.-J. Wang, Simon Gustavsson, and William D. Oliver. Superconducting qubits: Current state of play. *Annual Review of Condensed Matter Physics*, 11(1):369–395, Mar 2020. URL: http://dx.doi.org/10.1146/annurev-conmatphys-031119-050605, doi:10.1146/annurev-conmatphys-031119-050605.

[68] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson, M. H. S. Amin, N. Dickson, K. Karimi, B. Macready, C. J. S. Truncik, and G. Rose. Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor. *Phys. Rev. B*, 82:024511, Jul 2010. doi:10.1103/PhysRevB.82.024511.

[69] Fei Yan, Simon Gustavsson, Archana Kamal, Jeffrey Birenbaum, Adam P Sears, David Hover, Ted J. Gudmundsen, Danna Rosenberg, Gabriel Samach, S Weber, Jonilyn L. Yoder, Terry P. Orlando, John Clarke, Andrew J. Kerman, and William D. Oliver. The flux qubit revisited to enhance coherence and reproducibility. *Nature Communications*, 7:12964 EP –, 11 2016. URL: http://dx.doi.org/10.1038/ncomms12964.

[70] Anatoly Yu Smirnov and Mohammad H. Amin. Theory of open quantum dynamics with hybrid noise. *New Journal of Physics*, 20(10):103037, October 2018. doi:10.1088/1367-2630/aae79c.

[71] Mostafa Khezri, Jeffrey A. Grover, James I. Basham, Steven M. Disseler, Huo Chen, Sergey Novikov, Kenneth M. Zick, and Daniel A. Lidar. Anneal-path correction in flux qubits. *npj Quantum Information*, 7(1):36, 2021. URL: https://www.nature.com/articles/s41534-021-00371-9.

[72] Matthew P. Harrigan, Kevin J. Sung, Matthew Neeley, Kevin J. Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas

Bushnell, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Daniel Eppens, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Alan Ho, Sabrina Hong, Trent Huang, L. B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Martin Leib, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Charles Neill, Florian Neukart, Murphy Yuezhen Niu, Thomas E. O'Brien, Bryan O'Gorman, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Andrea Skolik, Vadim Smelyanskiy, Doug Strain, Michael Streif, Marco Szalay, Amit Vainsencher, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Leo Zhou, Hartmut Neven, Dave Bacon, Erik Lucero, Edward Farhi, and Ryan Babbush. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, 2021. doi:10.1038/s41567-020-01105-y.

[73] Wendell H. Fleming and Raymond W. Rishel. *Deterministic and Stochastic Optimal Control*. Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 1975. URL: https://www.springer.com/gp/book/9780387901558, doi:10.1007/978-1-4612-6380-7.

[74] W. Marshall. Antiferromagnetism. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 232(1188):48–68, October 1955. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1955.0200, doi:10.1098/rspa.1955.0200.

[75] Elliott Lieb and Daniel Mattis. Ordering Energy Levels of Interacting Spin Systems. *Journal of Mathematical Physics*, 3(4):749–751, July 1962. URL: https://aip.scitation.org/doi/10.1063/1.1724276, doi:10.1063/1.1724276.

[76] In our case x is either a wavefunction ψ ($N = n$) or a density matrix ρ ($N = n^2$), after the coordinatization described in Appendix C.

[77] In control theory, problems with a cost that depends only on the final values t_f and $x(t_f)$, are called of Meyer type. In our case the cost does not depend explicitly on time so we simply consider cost of the form $\phi(x(t_f))$ to simplify the notation.

[78] Andrei A. Agrachev and Yuri Sachkov. *Control Theory from the Geometric Viewpoint*. Encyclopaedia of Mathematical Sciences, Control Theory and Optimization. Springer-Verlag, Berlin Heidelberg, 2004. URL: https://www.springer.com/gp/book/9783540210191, doi:10.1007/978-3-662-06404-7.

[79] Filippov's existence theorem can also be applied in the version of Theorem 2.1 in [73, Ch. III] if we assume [as for Eq. (15)] that f in Eq. (B1) is linear in both the control s and state x, the cost J in Eq. (7) is smooth, and the set $\tilde{S}$ is compact. Application of this theorem also takes into account that ρ in Eq. (7) is bounded. The only condition that is not directly satisfied for Theorem 2.1 in [73, Ch. III] is condition (c) for which we can, however, use the alternative (c') of Corollary 2.2 in [73, Ch. III], i.e., the existence of a constant μ_1 such that $\{x_1 \,|\, J = \phi(x_1) \leq \mu_1\} \subseteq \mathbb{R}^n$ is compact.

[80] For a linear function $\phi(x) = k^T x$, if the minimum is achieved at an interior point setting the derivatives equal to zero would imply $k = 0$.

[81] This representation is closely related to the familiar "coherence vector" $\boldsymbol{v}$, wherein one represents ρ as $\rho = \frac{1}{n}(\mathbb{1} + \sum_{j=1}^{n^2-1} \boldsymbol{v}_j F_j)$, where $F_{n^2} := \mathbb{1}$ and the other orthonormal basis elements $\{F_j\}_{j=1}^{n^2-1}$ are in addition traceless; see, e.g., Ref. [86].

[82] Recall that here the asterisk denotes optimality, not complex conjugation.

[83] In the notation of Appendix B the co-state vector satisfies the adjoint equation (B3), and in the present case $\frac{\partial f}{\partial x} = \boldsymbol{\mathcal{L}}$.

[84] V. F. Borisov. Fuller's phenomenon: Review. *J Math Sci*, 100(4):2311–2354, July 2000. doi:10.1007/s10958-000-0001-9.

[85] Lorenzo Campos Venuti. The recurrence time in quantum mechanics, 2015. URL: http://arxiv.org/abs/1509.04352, arXiv:1509.04352.

[86] M.S. Byrd and N. Khaneja. Characterization of the positivity of the density matrix in terms of the coherence vector representation. 68:062322, 2003. URL: http://link.aps.org/abstract/PRA/v68/e062322.