Estimate of the time required to perform a nonadiabatic holonomic quantum computation

Ole Sönnerborn [*]
Department of Mathematics and Computer Science, Karlstad University, 651 88 Karlstad, Sweden
and Department of Physics, Stockholm University, 106 91 Stockholm, Sweden

(Received 19 April 2024; accepted 31 May 2024; published 24 June 2024)

Nonadiabatic holonomic quantum computation has been proposed as a method to implement quantum logic gates with robustness comparable to that of adiabatic holonomic gates but with shorter execution times. In this paper, we establish an isoholonomic inequality for quantum gates, which provides a lower bound on the lengths of cyclic transformations of the computational space that generate a specific gate. Then, as a corollary, we derive a nonadiabatic execution time estimate for holonomic gates. In addition, we demonstrate that under certain dimensional conditions, the isoholonomic inequality is tight in the sense that every gate on the computational space can be implemented holonomically and unitarily in a time-optimal way. We illustrate the results by showing that the procedures for implementing a universal set of holonomic gates proposed in a pioneering paper on nonadiabatic holonomic quantum computation saturate the isoholonomic inequality and are thus time optimal.

DOI: 10.1103/PhysRevA.109.062433

I. INTRODUCTION

Adiabatic holonomic computation has been launched as a procedure to implement quantum gates resilient to certain types of errors [1–4]. However, the slow parametric control associated with adiabatic evolution makes adiabatic computations sensitive to external perturbations. To address this issue, an alternative method for realizing quantum gates using nonadiabatic holonomies has been proposed [5–9].[1] A nonadiabatic holonomic computation exploits the system's internal dynamics, which significantly shortens the execution time of the computation compared to the adiabatic case. However, fundamental properties of quantum mechanical systems preclude arbitrarily short execution times for holonomic quantum gates. In this paper, we derive an estimate of the time required to execute a holonomic quantum gate unitarily. This estimate builds upon and generalizes a corresponding estimate of the time it takes to generate an Aharonov-Anandan geometric phase, as reported in Ref. [12].

The main ingredient in the derivation of the execution time estimate is the isoholonomic inequality for quantum gates. The isoholonomic inequality establishes a minimum length for cyclic transformations of the computational space that holonomically generate a specific gate. This inequality, together with the results in Ref. [13], solves the isoholonomic problem for quantum gates formulated by Montgomery [14].

Holonomic gates are the building blocks of circuits in holonomic quantum computation. Since holonomic gates have a purely geometric origin, implementations of quantum gates through parallel transport operators are predicted to be highly robust against noise [4]. Nonadiabatic holonomic gates have been experimentally demonstrated in various physical systems [15–18]. We show that the scheme in the pioneering paper [6] for the implementation of a universal set of holonomic gates is time optimal.

The paper is organized as follows. Section II presents the main results. Section III introduces terminology and describes basic properties of Stiefel-Grassmann bundles. Section IV contains derivations of the main results. In Sec. V we apply the main results to a proposal on how to experimentally implement a universal set of holonomic quantum gates. Finally, in Sec. VI, we prove that the isoholonomic inequality is tight in a strong sense provided the dimension of the computational space is at most half of the dimension of the Hilbert space. The paper concludes with a summary.

II. RESULTS

Throughout, $\mathcal{R}$ denotes an n-dimensional subspace of a finite-dimensional Hilbert space $\mathcal{H}$. We write $P_{\mathcal{R}}$ for the orthogonal projection onto $\mathcal{R}$, and we use computational terminology and call $\mathcal{R}$ the computational space and unitary operators on $\mathcal{R}$ gates. Moreover, we assume all quantities have units such that $\hbar = 1$.

When considering a one-parameter family—a curve—of operators, vectors, or subspaces, we assume that the family depends smoothly on the parameter and that the parameter ranges from 0 to τ. Also, we refer to the parameter as time, even though it may not represent actual time. We say that the

[*]Contact author: ole.sonnerborn@kau.se

[1]The importance of developing nonadiabatic computational schemes has also been emphasized and demonstrated in the related field of geometric quantum computation; see Refs. [10,11].

curve is closed, cyclic, or a loop when the initial and final members of the curve are the same.

Inspired by a question from a colleague, Montgomery [14] formulated the isoholonomic problem for quantum gates as follows: Find the shortest cyclic transformation of a subspace whose holonomy is a given gate. In this paper, we provide a partial solution to this problem by deriving a lower bound on the length of a cyclic transformation of $\mathcal{R}$ in terms of its holonomy: Assume $\mathcal{R}_t$ is a curve of n-dimensional subspaces of $\mathcal{H}$ that starts and ends with $\mathcal{R}$. Let Γ be the holonomy of $\mathcal{R}_t$. Then the length of $\mathcal{R}_t$ is bounded from below by

$$L(\Gamma) = \sqrt{\sum_{j=1}^{n} |\theta_j|(2\pi - |\theta_j|)}, \tag{1}$$

where $\theta_1, \theta_2, \dots, \theta_n$ are the principal arguments of the eigenvalues of Γ.[2] We refer to the length estimate

$$\mathcal{L}[R_t] \geqslant L(\Gamma) \tag{2}$$

as the isoholonomic inequality and call $L(\Gamma)$ the isoholonomic bound of the gate Γ.

The isoholonomic inequality is tight when the dimension of $\mathcal{R}$ is at most half of the dimension of $\mathcal{H}$: Let Γ be any gate on $\mathcal{R}$, and k be the number of 1s in the spectrum of Γ. If the codimension of $\mathcal{R}$ is at least $n-k$, there is a parallel transporting Hamiltonian that drives $\mathcal{R}$ in a loop with holonomy Γ and length $L(\Gamma)$.

From the isoholonomic inequality one can derive an estimate of the time required to drive $\mathcal{R}$ unitarily in a loop with a given holonomy. Assume $\mathcal{R}_t = U_t(\mathcal{R})$, where U_t is the time propagator associated with a Hamiltonian H_t. The square of the speed of $\mathcal{R}_t$ equals

$$I(H_t;\mathcal{R}_t) = -\tfrac{1}{2}\operatorname{tr}([H_t, P_t]^2), \tag{3}$$

where P_t is the orthogonal projection onto $\mathcal{R}_t$. Write $\langle\!\langle I(H_t;\mathcal{R}_t)^{1/2}\rangle\!\rangle$ for the average speed of $\mathcal{R}_t$ over the evolution time interval. The isoholonomic inequality implies that the evolution time is not smaller than

$$\tau[H_t;\Gamma] = \frac{L(\Gamma)}{\langle\!\langle \sqrt{I(H_t;\mathcal{R}_t)}\rangle\!\rangle}. \tag{4}$$

The quantity $I(H_t;\mathcal{R}_t)$ measures the skewness of H_t relative to $\mathcal{R}_t$; see Refs. [19,20]. If the Hamiltonian is time independent, $H_t = H$, the skewness is a conserved quantity, and the evolution time is lower bounded by

$$\tau[H;\Gamma] = \frac{L(\Gamma)}{\sqrt{I(H;\mathcal{R})}}. \tag{5}$$

We derive the isoholonomic inequality (2) and the runtime bound (4) in Sec. IV, and we prove the tightness of the isoholonomic inequality in Sec. VI.

III. PARALLEL TRANSPORT AND HOLONOMIC GATES

Cyclic transformations of $\mathcal{R}$ correspond to curves in the Grassmann manifold of n-dimensional subspaces of $\mathcal{H}$ that start and end at $\mathcal{R}$. The Grassmann manifold can be identified with the manifold of orthogonal projections on $\mathcal{H}$ of rank n by identifying each n-dimensional subspace of $\mathcal{H}$ with the orthogonal projection onto that subspace.[3] A cyclic transformation of $\mathcal{R}$ is then represented by a curve of orthogonal projections that starts and ends at $P_{\mathcal{R}}$. We will use the same notation, $\mathcal{G}(n;\mathcal{H})$, for the space of n-dimensional subspaces of $\mathcal{H}$ and the space of orthogonal projection operators of rank n on $\mathcal{H}$.

Remark 1. The elements of $\mathcal{G}(1;\mathcal{H})$ represent the pure states of a quantum system modeled on $\mathcal{H}$. In Ref. [12], we derived time estimates for cyclic transformations of pure states in terms of their Aharonov-Anandan geometric phase. Here, we generalize one of these to an estimate of the time required to execute a holonomic gate.

An n frame in $\mathcal{H}$ is an ordered sequence of n orthonormal vectors in $\mathcal{H}$. It will prove convenient to represent an n-frame F as a row matrix,

$$F = (|u_1\rangle\,|u_2\rangle \cdots |u_n\rangle). \tag{6}$$

We will only consider frames of the unspecified but fixed length n, and will, therefore, only write frame when referring to an n frame.

We can act on F with an operator A defined on its span. The result is the row matrix AF whose elements are the images of the vectors of F under A,

$$AF = (A|u_1\rangle\,A|u_2\rangle \cdots A|u_n\rangle). \tag{7}$$

The matrix AF is a frame if and only if A is an isometry on the span of F. We can also act on F from the right by an $n\times k$ numerical matrix $M = (m_{ij})$. The result is a row matrix FM whose elements are linear combinations of the vectors of F:

$$FM = \left(\sum_{i=1}^{n} m_{i1}|u_i\rangle\ \sum_{i=1}^{n} m_{i2}|u_i\rangle \cdots \sum_{i=1}^{n} m_{ik}|u_i\rangle\right). \tag{8}$$

If $k=1$, FM is a linear combination of the vectors in F; if $k=n$ and M is unitary, FM is a frame that spans the same subspace as F.

We can also multiply frames by conjugates of frames. Depending on how we multiply them, we get either an operator or a matrix of numbers: If $F_1 = (|u_1\rangle\,|u_2\rangle \cdots |u_n\rangle)$ and $F_2 = (|v_1\rangle\,|v_2\rangle \cdots |v_n\rangle)$, then

$$F_1F_2^\dagger = |u_1\rangle\langle v_1| + |u_2\rangle\langle v_2| + \cdots + |u_n\rangle\langle v_n| \tag{9}$$

and

$$F_1^\dagger F_2 = \begin{pmatrix} \langle u_1|v_1\rangle & \langle u_1|v_2\rangle & \dots & \langle u_1|v_n\rangle \\ \langle u_2|v_1\rangle & \langle u_2|v_2\rangle & \dots & \langle u_2|v_n\rangle \\ \vdots & \vdots & & \vdots \\ \langle u_n|v_1\rangle & \langle u_n|v_2\rangle & \dots & \langle u_n|v_n\rangle \end{pmatrix}. \tag{10}$$

The frames in $\mathcal{H}$ form the Stiefel manifold $\mathcal{V}(n;\mathcal{H})$. If F is a frame, $FF^\dagger$ is the orthogonal projection operator onto the span of F. Thus, $FF^\dagger$ belongs to $\mathcal{G}(n;\mathcal{H})$. The assignment

$$\mathcal{V}(n;\mathcal{H}) \ni F \to FF^\dagger \in \mathcal{G}(n;\mathcal{H}) \tag{11}$$

[2]The principal argument of a nonzero complex number z is the phase θ in the interval $(-\pi,\pi]$ for which $z = |z|e^{i\theta}$.

[3]This identification is used to induce a topology and a geometry on the Grassmann manifold from the space of Hermitian operators on $\mathcal{H}$.

is a principal fiber bundle called the Stiefel-Grassmann bundle. This bundle has gauge group the group of unitary matrices $\mathrm{U}(n)$, which means that frames F_1 and F_2 project onto the same projection operator if and only if $F_2 = F_1U$ for an $n \times n$ unitary matrix U. We recommend the two-volume work [21] as a reference for the theory of principal fiber bundles.

A. Parallel transport operators

Parallel transport operators parallel transport frames. To specify what this means we need to introduce a connection on the Stiefel manifold.

Suppose $\dot{F}$ is a tangent vector at the frame F. If we represent $\dot{F}$ as a row matrix of vectors, $F^\dagger\dot{F}$ is an element of the Lie algebra $\mathfrak{u}(n)$ of $n \times n$ skew-Hermitian matrices. We define $\mathcal{A}$ as the $\mathfrak{u}(n)$-valued connection on $\mathcal{V}(n;\mathcal{H})$ sending $\dot{F}$ to $F^\dagger\dot{F}$,

$$\mathcal{A}(\dot{F}) = F^\dagger\dot{F}. \tag{12}$$

Using standard terminology, we say that $\dot{F}$ is horizontal if $\mathcal{A}(\dot{F}) = 0$. We also say that a curve of frames F_t is horizontal if all its velocity vectors are horizontal.

Consider a curve of n-dimensional subspaces $\mathcal{R}_t$ in $\mathcal{H}$ starting with $\mathcal{R}$. Let P_t be the corresponding curve of projection operators. According to a fundamental result from the theory of fiber bundles, there exists a unique one-parameter family of isometries $\Pi_t : \mathcal{R} \to \mathcal{R}_t$ such that for each frame F for $\mathcal{R}$, the curve $F_t = \Pi_tF$ is a horizontal lift of P_t, that is, a horizontal curve of frames projecting onto P_t. The isometries Π_t are the parallel transport operators associated with $\mathcal{R}_t$; see Ref. [21] for details.

We can express the parallel transport operators in terms of the horizontal lift F_t as $\Pi_t = F_tF^\dagger$. More generally, if F_t is any curve of frames projecting onto P_t,

$$\Pi_t = F_t\overleftarrow{\mathcal{T}}\exp\left(-\int_0^t \mathcal{A}(\dot{F}_s)\,ds\right)F_0^\dagger. \tag{13}$$

The symbol $\overleftarrow{\mathcal{T}}$ indicates that the exponential is forward time ordered.

B. Holonomic gates

If $\mathcal{R}_t$ describes a cyclic transformation of $\mathcal{R}$, the final parallel transport operator Π_τ maps $\mathcal{R}$ isometrically onto itself. This operator is called the holonomy of $\mathcal{R}_t$. We will henceforth write $\Gamma[\mathcal{R}_t]$ for the holonomy of $\mathcal{R}_t$. A holonomic gate on $\mathcal{R}$ is a gate implemented as the holonomy of a cyclic transformation of $\mathcal{R}$.

The parallel transport operators Π_t associated with an evolution $\mathcal{R}_t$ of $\mathcal{R}$ move every vector $|\psi\rangle$ in $\mathcal{R}$ in such a way that, at every t, the velocity vector of the curve $|\psi_t\rangle = \Pi_t|\psi\rangle$ is orthogonal to $\mathcal{R}_t$. Geometrically, this means that the parallel transport operators cause no time-local rotation within $\mathcal{R}$. A holonomic gate on $\mathcal{R}$ is thus a consequence solely of the translational motion of $\mathcal{R}$ in the Grassmann manifold. This observation underlies the hypothetical claim that holonomic gates should be particularly robust against noise and certain types of implementation errors [4,9].

C. Parallel transporting Hamiltonians

We say that a Hamiltonian is parallel transporting if the associated time propagator parallel translates frames for $\mathcal{R}$. For any Hamiltonian H_t, we can define a parallel transporting Hamiltonian $\bar{H}_t$ that drives $\mathcal{R}$ along the same path and at the same speed as H_t: Let P_t be the curve of orthogonal projectors generated from $P_\mathcal{R}$ by H_t and define $\bar{H}_t$ as

$$\bar{H}_t = H_tP_t + P_tH_t - 2P_tH_tP_t. \tag{14}$$

Then $[\bar{H}_t, P_t] = [H_t, P_t]$, which shows that $\bar{H}_t$ propagates $\mathcal{R}$ in the same way as H_t, and if F is any frame for $\mathcal{R}$, and $F_t = \bar{U}_tF$, where $\bar{U}_t$ is the time propagator of $\bar{H}_t$, then $F_t^\dagger\bar{H}_tF_t = 0$, which shows that $\bar{H}_t$ is parallel transporting. For a time-independent Hamiltonian H, the corresponding parallel transporting Hamiltonian is

$$\bar{H}_t = e^{-itH}(HP_\mathcal{R} + P_\mathcal{R}H - 2P_\mathcal{R}HP_\mathcal{R})e^{itH}. \tag{15}$$

Although H is time independent, the parallel transporting Hamiltonian need not be time independent.

D. Dynamical operators

The total phase acquired during a cyclic unitary evolution of a pure state can be divided into a geometric part (the holonomy) and a dynamic part [22]. It was previously believed that a corresponding division was generally not possible for cyclic unitary evolutions of subspaces. However, Yu and Tong [23] recently showed that such a division is always possible. Here, we derive the result of Yu and Tong using the framework presented above.

Suppose H_t drives $\mathcal{R}$ in a loop $\mathcal{R}_t$. Let U_t be the time propagator associated with H_t. Choose a frame F for $\mathcal{R}$ and define a curve of frames as $F_t = U_tF$. According to Eq. (13), the holonomy of the loop is

$$\Gamma[\mathcal{R}_t] = U_\tau F\,\overleftarrow{\mathcal{T}}\exp\left(i\int_0^\tau F_t^\dagger H_tF_t\,dt\right)F^\dagger. \tag{16}$$

We define the dynamical operator of H_t on $\mathcal{R}$ as

$$D[H_t] = F\,\overrightarrow{\mathcal{T}}\exp\left(-i\int_0^\tau F_t^\dagger H_tF_t\,dt\right)F^\dagger, \tag{17}$$

where $\overrightarrow{\mathcal{T}}$ indicates that the exponential is backward time ordered. By Eq. (16), the restriction of U_τ to $\mathcal{R}$ decomposes as

$$U_\tau\big|_\mathcal{R} = \Gamma[\mathcal{R}_t]D[H_t]. \tag{18}$$

IV. ISOHOLONOMIC INEQUALITY

We equip the Grassmann and Stiefel manifolds with the Riemannian metrics

$$g_\mathcal{G}(\dot{P}_1, \dot{P}_2) = \tfrac{1}{2}\operatorname{tr}\big(\dot{P}_1\dot{P}_2\big), \tag{19}$$

$$g_\mathcal{V}(\dot{F}_1, \dot{F}_2) = \tfrac{1}{2}\operatorname{tr}(\dot{F}_1^\dagger\dot{F}_2 + \dot{F}_2^\dagger\dot{F}_1). \tag{20}$$

Furthermore, we define the length of a curve of orthogonal projectors P_t and the length of a curve of frames F_t as

$$\mathcal{L}[P_t] = \int_0^\tau \sqrt{g_{\mathcal{G}}(\dot{P}_t, \dot{P}_t)}\, dt, \tag{21}$$

$$\mathcal{L}[F_t] = \int_0^\tau \sqrt{g_{\mathcal{V}}(\dot{F}_t, \dot{F}_t)}\, dt. \tag{22}$$

We also define the kinetic energies of P_t and F_t as

$$\mathcal{E}[P_t] = \frac{1}{2}\int_0^\tau g_{\mathcal{G}}(\dot{P}_t, \dot{P}_t)\, dt, \tag{23}$$

$$\mathcal{E}[F_t] = \frac{1}{2}\int_0^\tau g_{\mathcal{V}}(\dot{F}_t, \dot{F}_t)\, dt. \tag{24}$$

From the Cauchy-Schwarz inequality we get

$$2\tau\mathcal{E}[P_t] \geqslant \mathcal{L}[P_t]^2, \tag{25}$$

$$2\tau\mathcal{E}[F_t] \geqslant \mathcal{L}[F_t]^2, \tag{26}$$

with the inequalities being equalities if P_t and F_t have constant speeds.

The Stiefel-Grassmann bundle projection is a Riemannian submersion, which means that the tangent map of the projection preserves the inner product between horizontal vectors. Consequently, the length of a curve in the Grassmannian and the lengths of all of its horizontal lifts are the same. The same is true for the kinetic energy.

A. Isoholonomic inequality for states

If $n = 1$, the Grassmannian is the projective space of density operators representing pure states of quantum systems modeled on $\mathcal{H}$, and $g_{\mathcal{G}}$ is the Fubini-Study metric.

The holonomy of a closed curve of pure states ρ_t multiplies unit vectors over the common initial and final state by a phase factor. The argument of the holonomy is the Aharonov-Anandan geometric phase of ρ_t [22]. The isoholonomic inequality for states says that the Fubini-Study length of ρ_t is bounded from below as follows:

$$\mathcal{L}[\rho_t] \geqslant L(\theta), \quad L(\theta) = \sqrt{|\theta|(2\pi - |\theta|)}, \tag{27}$$

where θ is the principal argument of the holonomy of ρ_t [12,14]. Below, we extend this inequality to an estimate of the length of a closed curve of subspaces of $\mathcal{H}$ of arbitrary dimension in terms of the holonomy of the curve. The derivation uses the estimate (27). For convenience, we have included a slightly rewritten version of the derivation of the estimate (27) found in Ref. [12] in Appendix.

Example 1. Consider a qubit with Hamiltonian

$$H = \epsilon_0|0\rangle\langle 0| + \epsilon_1|1\rangle\langle 1|, \quad \epsilon_0 < \epsilon_1. \tag{28}$$

Assume the qubit is initially in the pure state ρ and evolves unitarily as $\rho_t = e^{-itH}\rho e^{itH}$. Let $|\psi\rangle = a|0\rangle + b|1\rangle$ be a unit vector such that $\rho = |\psi\rangle\langle\psi|$, and let $|\psi_t\rangle = e^{-itH}|\psi\rangle$. The curve ρ_t is periodic with period $\tau = 2\pi/(\epsilon_1 - \epsilon_0)$, and according to Eq. (13), the holonomy of ρ_t is

$$\begin{aligned}\Gamma[\rho_t] &= \langle\psi|\psi_\tau\rangle \exp\left(-\int_0^\tau \langle\psi_t|\dot{\psi}_t\rangle\, dt\right)\\ &= (|a|^2 e^{-i\epsilon_0\tau} + |b|^2 e^{-i\epsilon_1\tau})e^{i\tau(\epsilon_0|a|^2+\epsilon_1|b|^2)}\\ &= e^{2\pi i|b|^2}.\end{aligned} \tag{29}$$

Furthermore, the evolution has the speed

$$\sqrt{\tfrac{1}{2}\operatorname{tr}\left(\dot{\rho}_t^2\right)} = (\epsilon_1 - \epsilon_0)|a||b| \tag{30}$$

and, thus, the length

$$\mathcal{L}[\rho_t] = \tau(\epsilon_1 - \epsilon_0)|a||b| = 2\pi|a||b|. \tag{31}$$

Let θ be the principal argument of the holonomy. Then,

$$\mathcal{L}[\rho_t]^2 = 2\pi|b|^2(2\pi - 2\pi|b|^2) = |\theta|(2\pi - |\theta|). \tag{32}$$

We conclude that a qubit with time-independent Hamiltonian saturates the isoholonomic inequality.

B. Isoholonomic inequality for gates

In this section, we show that the length of a closed curve in $\mathcal{G}(n;\mathcal{H})$ with holonomy Γ is bounded from below by $L(\Gamma)$ as defined in Eq. (1). Then, in Sec. VI, we show, inspired by Ref. [13], that $L(\Gamma)$ is a tight bound when the dimension of $\mathcal{H}$ is greater than or equal to $2n - k$, where k is the number of 1s in the spectrum of Γ. We do this by constructing a Hamiltonian that drives $\mathcal{R}$ in a loop with holonomy Γ and length $L(\Gamma)$. The question of whether $L(\Gamma)$ is a tight bound when the dimension of $\mathcal{H}$ is less than $2n - k$ is still open.

Assume P_t is a closed curve of rank n orthogonal projection operators at $P_{\mathcal{R}}$ having holonomy Γ. Since length and holonomy are parametrization invariant quantities, we can assume that P_t has a constant speed and returns to $P_{\mathcal{R}}$ at time $\tau = 1$.

Let $e^{i\theta_1}, e^{i\theta_2}, \dots, e^{i\theta_n}$ be the eigenvalues of Γ, with θ_j being the principal argument of the jth eigenvalue. Let F be a frame consisting of eigenvectors of Γ,

$$F = (|u_1\rangle\, |u_2\rangle \dots |u_n\rangle), \quad \Gamma|u_j\rangle = e^{i\theta_j}|u_j\rangle, \tag{33}$$

and let F_t be the horizontal lift of P_t starting at F,

$$F_t = (|u_{1;t}\rangle\, |u_{2;t}\rangle \dots |u_{n;t}\rangle), \quad |u_{j;0}\rangle = |u_j\rangle. \tag{34}$$

Since the Stiefel-Grassmann bundle is a Riemannian submersion and P_t has a constant speed, so does F_t, and the square of the length of P_t is

$$\mathcal{L}[P_t]^2 = 2\mathcal{E}[P_t] = 2\mathcal{E}[F_t] = 2\sum_{j=1}^n \mathcal{E}(|u_{j;t}\rangle). \tag{35}$$

Furthermore, since F_t is horizontal, each curve $|u_{j;t}\rangle$ is Aharonov-Anandan horizontal,

$$\langle u_{j;t}|\dot{u}_{j;t}\rangle = \langle u_j|FF_t^\dagger\dot{F}_tF^\dagger|u_j\rangle = 0, \tag{36}$$

and $|u_{j;t}\rangle$ projects onto a closed curve of pure states $\rho_{j;t}$ with Aharonov-Anandan geometric phase θ_j,

$$|u_{j;1}\rangle = F_1F^\dagger|u_j\rangle = \Gamma|u_j\rangle = e^{i\theta_j}|u_j\rangle. \tag{37}$$

The curves $|u_{j;t}\rangle$ and $\rho_{j;t}$ have the same kinetic energies, and by the isoholonomic inequality for states (27), the length of

TABLE I. Isoholonomic bounds for a complete set of qubit gates: the one-qubit Hadamard gate H, phase gate S, $\pi/8$ gate T, and the two-qubit CNOT gate.

Gate	Matrix representation	Isoholonomic bound
Hadamard	$H=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$	$L(H)=\pi$
phase gate	$S=\begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix}$	$L(S)=\frac{\pi\sqrt{3}}{2}$
$\pi/8$ gate	$T=\begin{pmatrix}1 & 0\\ 0 & e^{\frac{i\pi}{4}}\end{pmatrix}$	$L(T)=\frac{\pi\sqrt{7}}{4}$
CNOT	$\text{CNOT}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}$	$L(\text{CNOT})=\pi$

$\rho_{j;t}$ is lower bounded by $L(\theta_j)$. Thus,

$$2\mathcal{E}[|u_{j;t}\rangle]=2\mathcal{E}[\rho_{j;t}]\geqslant L(\theta_j)^2. \tag{38}$$

Equations (35) and (38) imply that

$$\mathcal{L}[P_t]^2\geqslant\sum_{j=1}^{n}L(\theta_j)^2=L(\Gamma)^2. \tag{39}$$

This proves the isoholonomic inequality (2).

Example 2. The quantum Fourier transform

$$F_n|u_j\rangle=\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}e^{2\pi ijk}|u_k\rangle \tag{40}$$

is used in many quantum algorithms [24]. The quantum Fourier transform has characteristic polynomial

$$(x-1)^{\lfloor\frac{n+4}{4}\rfloor}(x+1)^{\lfloor\frac{n+2}{4}\rfloor}(x+i)^{\lfloor\frac{n+1}{4}\rfloor}(x-i)^{\lfloor\frac{n-1}{4}\rfloor}, \tag{41}$$

from which we can read off the eigenvalues of the transform and their multiplicities. We conclude that the Fourier transform has isoholonomic bound

$$L(F_n)=\pi\sqrt{\left\lfloor\frac{n+2}{4}\right\rfloor+\frac{3}{4}\left(\left\lfloor\frac{n+1}{4}\right\rfloor+\left\lfloor\frac{n-1}{4}\right\rfloor\right)}. \tag{42}$$

For $n=2$, the Fourier transform equals the Hadamard gate H. The Hadamard gate thus has isoholonomic bound $L(H)=\pi$. We have listed the isoholonomic bounds for a universal set of qubit gates in Table I.

C. Runtime bound

The evolution time estimate $\tau\geqslant\tau[H_t;\Gamma]$ follows immediately from the isoholonomic inequality for gates and the observation that if $\mathcal{R}$ is transported in a loop $\mathcal{R}_t$ by the Hamiltonian H_t, and P_t is the corresponding curve of projection operators, the square of the speed of $\mathcal{R}_t$ equals the skewness of H_t relative to $\mathcal{R}_t$,

$$g_{\mathcal{G}}(\dot{P}_t,\dot{P}_t)=\tfrac{1}{2}\operatorname{tr}((-i[H_t,P_t])^2)=I(H_t;\mathcal{R}_t). \tag{43}$$

If the Hamiltonian is time-independent, $H_t=H$, the skewness is conserved and the speed is constant,

$$\begin{aligned}2I(H;\mathcal{R}_t)&=\operatorname{tr}((-i[H,e^{-itH}P_{\mathcal{R}}e^{itH}])^2)\\&=\operatorname{tr}(e^{-itH}(-i[H,P_{\mathcal{R}}])^2e^{itH})\\&=2I(H;\mathcal{R}).\end{aligned} \tag{44}$$

For a parallel transporting Hamiltonian, we have that

$$I(H_t;\mathcal{R}_t)=\operatorname{tr}\left(H_t^2P_t\right). \tag{45}$$

V. TIME-OPTIMAL UNIVERSAL GATES

In the standard description of nonadiabatic holonomic quantum computation [9,25], input states are prepared in a space $\mathcal{R}$ associated with a register of qubits. The states are then manipulated with holonomic gates implemented by parallel transporting Hamiltonians.

Typically, a frame of product vectors

$$F=(|100\cdots0\rangle\,|010\cdots0\rangle\ldots|111\cdots1\rangle) \tag{46}$$

is used as a reference frame in $\mathcal{R}$ where, at each position, $|0\rangle$ and $|1\rangle$ are orthonormal vectors that span the marginal Hilbert space of the corresponding qubit. The space $\mathcal{R}$ is referred to as the computational space, F is the computational basis, and states and gates are represented as matrices relative to F.

A universal set of quantum gates can approximate any other quantum gate to any desired precision. For a computational system manipulating qubits, the one-qubit Hadamard gate, phase gate, $\pi/8$ gate, and two-qubit CNOT gate form a universal set [24]. The isoholonomic bounds for these are listed in Table I.

Reference [6] by Sjöqvist *et al.* contains proposals for how to holonomically implement the one-qubit gates

$$\Gamma_1(\alpha,\beta)=\begin{pmatrix}\cos\alpha & e^{-i\beta}\sin\alpha\\ e^{i\beta}\sin\alpha & -\cos\alpha\end{pmatrix} \tag{47}$$

and the two-qubit gates

$$\Gamma_2(\alpha,\beta)=\begin{pmatrix}\cos\alpha & 0 & 0 & e^{-i\beta}\sin\alpha\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ e^{i\beta}\sin\alpha & 0 & 0 & -\cos\alpha\end{pmatrix}, \tag{48}$$

which together form a universal set [6,26]. We demonstrate below that the proposals in Ref. [6] are time optimal in the sense that the length of the trajectory of the computational space equals the isoholonomic bound of the implemented quantum gate.

A. One-qubit gates

Following Sjöqvist *et al.* [6] we consider a system with three bare energy levels in a Λ configuration. We assume that the lower, closely spaced levels are represented by $|0\rangle$ and $|1\rangle$ and that the excited energy level, whose energy we set to 0, is represented by $|e\rangle$; see Fig. 1. Furthermore, we assume that resonant laser pulses drive the transitions $|0\rangle\leftrightarrow|e\rangle$ and $|1\rangle\leftrightarrow|e\rangle$ and that the dipole and rotating wave approximations are applicable. In the rotating frame of the laser fields,

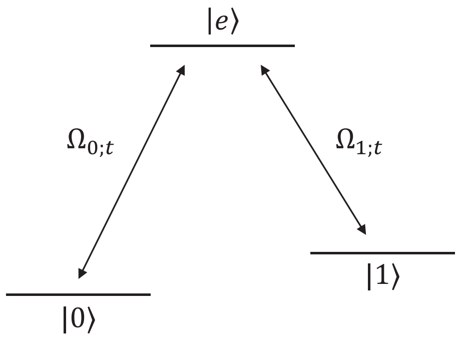

FIG. 1. A three-level system in a Λ configuration with two closely spaced levels $|0\rangle$ and $|1\rangle$ coupled to an excited level $|e\rangle$ by resonant pulsed laser beams with Rabi frequencies $\Omega_{0;t}$ and $\Omega_{1;t}$.

the Hamiltonian can then be written

$$H_t = \Omega_{0;t}|e\rangle\langle 0| + \Omega_{0;t}^*|0\rangle\langle e| + \Omega_{1;t}|e\rangle\langle 1| + \Omega_{1;t}^*|1\rangle\langle e|. \quad (49)$$

We take the sum of the lower energy levels as the computational space and $|0\rangle$ and $|1\rangle$ as the computational basis. If the laser pulses have a common envelope,

$$\Omega_{j;t} = \Omega(t)\omega_j, \quad |\omega_0|^2 + |\omega_1|^2 = 1, \quad (50)$$

the Hamiltonian is parallel transporting, and if the support of the envelope is $[0, \tau]$ and

$$\int_0^\tau \Omega(t)\,dt = \pi, \quad (51)$$

the Hamiltonian drives the computational space in a loop $\mathcal{R}_t$ in time τ. The holonomy of the loop is

$$\Gamma[\mathcal{R}_t] = \begin{pmatrix} |\omega_1|^2 - |\omega_0|^2 & -2\omega_0^*\omega_1 \\ -2\omega_0\omega_1^* & |\omega_0|^2 - |\omega_1|^2 \end{pmatrix}, \quad (52)$$

and if we adjust the laser pulses so that $\omega_0 = \sin(\alpha/2)e^{i\beta/2}$ and $\omega_1 = -\cos(\alpha/2)e^{-i\beta/2}$, then $\Gamma[\mathcal{R}_t] = \Gamma_1(\alpha, \beta)$.

The gate $\Gamma_1(\alpha, \beta)$ has eigenvalues 1 and -1, and hence the isoholonomic bound $L(\Gamma_1(\alpha, \beta)) = \pi$. The loop of the computational space has the length

$$\tau\langle\!\langle\sqrt{I(H_t;\mathcal{R}_t)}\,\rangle\!\rangle = \tau\langle\!\langle\Omega(t)\rangle\!\rangle = \pi. \quad (53)$$

Since the length and the isoholonomic bound agree, the implementation is time optimal.

Remark 2. For $\alpha = \pi/4$ and $\beta = 0$, the above scheme implements the Hadamard gate time optimally. However, it cannot generate the phase and $\pi/8$ gates. References [27,28] contain proposals on how to generate these gates in a Λ system with off-resonant driving. Strictly speaking, the Hamiltonians in these proposals are not parallel transporting as they give rise to (irrelevant) dynamical phases. This issue will be addressed in a forthcoming paper.

Remark 3. For high-speed computations, the rotating wave approximation may not be valid. Reference [8] considers the scheme in Ref. [6] without the rotating wave approximation. The proposals in Ref. [8] are also time optimal.

B. Two-qubit gates

To implement the two-qubit gate $\Gamma_2(\alpha, \beta)$, Sjöqvist *et al.* [6] consider an array of ions, each of which exhibits an internal Λ structure as in Fig. 1. The laser pulses that drive the transitions $|j\rangle \leftrightarrow |e\rangle$ are controlled so that the dynamics of a pair of ions is governed by an effective Hamiltonian of the form $H_t = \Omega(t)(H_0 + H_1)$, where

$$H_0 = \omega_{00}|ee\rangle\langle 00| + \omega_{11}|ee\rangle\langle 11| + \text{H.c.}, \quad (54)$$

$$H_1 = \omega_{0e}|e0\rangle\langle 0e| + \omega_{1e}|e1\rangle\langle 1e| + \text{H.c.} \quad (55)$$

We take $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ as the computational basis and the span of these vectors as the computational space, and we adjust the laser pulses so that

$$\omega_{00} = \sin(\alpha/2)e^{i\beta/2}, \quad (56)$$

$$\omega_{11} = -\cos(\alpha/2)e^{-i\beta/2}, \quad (57)$$

$$\omega_{0e} = \sin(\alpha/2), \quad (58)$$

$$\omega_{1e} = -\cos(\alpha/2), \quad (59)$$

and so that the envelope function satisfies

$$\int_0^\tau \Omega(t)\,dt = \pi. \quad (60)$$

The Hamiltonian then parallel transports the computational space in a loop, $\mathcal{R}_t$, in time τ and thereby implements the gate $\Gamma_2(\alpha, \beta)$. This gate has eigenvalues 1 of multiplicity 3 and -1 of multiplicity 1, and thus the isoholonomic bound $L(\Gamma_2(\alpha, \beta)) = \pi$. Since the length of the loop of the computational space agrees with this bound,

$$\tau\langle\!\langle\sqrt{I(H_t;\mathcal{R}_t)}\,\rangle\!\rangle = \tau\langle\!\langle\Omega(t)\rangle\!\rangle = \pi, \quad (61)$$

the implementation is time optimal.

VI. TIGHTNESS OF THE ISOHOLONOMIC BOUND

We can generalize the qubit example in Sec. IV to a proof that $L(\Gamma)$ is tight when the dimension of $\mathcal{H}$ is at least $2n - k$, where k is the number of 1s in the spectrum of Γ. To do this, we arrange the eigenvalues of Γ so that

$$\theta_{n-k+1} = \theta_{n-k+2} = \cdots = \theta_n = 0. \quad (62)$$

Let F be a frame for $\mathcal{R}$ of eigenvectors of Γ as shown in Eq. (33), and let $|v_1\rangle, |v_2\rangle, \ldots, |v_{n-k}\rangle$ be $n-k$ pairwise orthogonal unit vectors in the orthogonal complement of $\mathcal{R}$. Within the span of $|u_j\rangle$ and $|v_j\rangle$ choose two orthonormal vectors $|0_j\rangle$ and $|1_j\rangle$, where the latter vector satisfies the condition

$$2\pi|\langle 1_j|u_j\rangle|^2 = \theta_j \mod 2\pi. \quad (63)$$

Let $\epsilon_0 < \epsilon_1$, and define a Hamiltonian H as

$$H = \sum_{j=1}^{n-k} \epsilon_0|0_j\rangle\langle 0_j| + \epsilon_1|1_j\rangle\langle 1_j|. \quad (64)$$

Furthermore, define

$$F_t = (|u_{1;t}\rangle\, |u_{2;t}\rangle \dots |u_{n;t}\rangle) \tag{65}$$

as $F_t = e^{-itH}F$ and let $P_t = F_tF_t^\dagger$.

For $j = 1, 2, \dots, n-k$, the vector $|u_j\rangle$ rotates within the span of $|u_j\rangle$ and $|v_j\rangle$ and returns to $\mathcal{R}$ for the first time at $\tau = 2\pi/(\epsilon_1 - \epsilon_0)$; for $j = n-k+1, n-k+2, \dots, n$, the vector $|u_j\rangle$ is held fixed. Thus, $\mathcal{R}$ is driven in a loop $\mathcal{R}_t$ with period $\tau = 2\pi/(\epsilon_1 - \epsilon_0)$.

According to Eq. (13), the holonomy of $\mathcal{R}_t$ is represented by the matrix

$$F^\dagger\Gamma[\mathcal{R}_t]F = F^\dagger F_\tau \overleftarrow{\mathcal{T}}\exp\left(-\int_0^\tau F_t^\dagger \dot{F}_t\, dt\right) \tag{66}$$

relative to F. Since the $|u_j\rangle$s rotate in pairwise perpendicular subspaces, $F^\dagger F_\tau$ and $F_t^\dagger \dot{F}_t$ are diagonal matrices,

$$F^\dagger F_\tau = \mathrm{diag}(\langle u_1|u_{1;\tau}\rangle, \dots, \langle u_n|u_{n;\tau}\rangle), \tag{67}$$

$$F_t^\dagger \dot{F}_t = \mathrm{diag}(\langle u_{1;t}|\dot{u}_{1;t}\rangle, \dots, \langle u_{n;t}|\dot{u}_{n;t}\rangle). \tag{68}$$

For $j = 1, 2, \dots, n-k$ write $|u_j\rangle = a_j|0_j\rangle + b_j|1_j\rangle$. Then, as in Eq. (29),

$$\langle u_j|u_{j;\tau}\rangle = |a_j|^2 e^{-i\epsilon_0\tau} + |b_j|^2 e^{-i\epsilon_1\tau}, \tag{69}$$

$$\langle u_{j;t}|\dot{u}_{j;t}\rangle = -i(\epsilon_0|a_j|^2 + \epsilon_1|b_j|^2). \tag{70}$$

Also, $\langle u_j|u_{j;\tau}\rangle = 1$ and $\langle u_{j;t}|\dot{u}_{j;t}\rangle = 0$ for $j = n-k+1, n-k+2, \dots, n$. We conclude that

$$F^\dagger\Gamma[\mathcal{R}_t]F = \mathrm{diag}(e^{2\pi i|b_1|^2}, \dots, e^{2\pi i|b_{n-k}|^2}, 1, \dots, 1) \tag{71}$$

which, by assumptions (62) and (63), shows that the holonomy of $\mathcal{R}_t$ is Γ.

To calculate the length of $\mathcal{R}_t$ we write $\rho_{j;t} = |u_{j;t}\rangle\langle u_{j;t}|$ and observe that the square of the speed of $\mathcal{R}_t$ is

$$\frac{1}{2}\,\mathrm{tr}\left(\dot{P}_t^2\right) = \sum_{j=1}^{n}\frac{1}{2}\,\mathrm{tr}\left(\dot{\rho}_{j;t}^2\right) = \sum_{j=1}^{n-k}(\epsilon_1-\epsilon_0)^2|a_j|^2|b_j|^2. \tag{72}$$

Since the speed is constant, the length of $\mathcal{R}_t$ squared is

$$\tau^2\sum_{j=1}^{n-k}\frac{1}{2}\,\mathrm{tr}\left(\dot{\rho}_{j;t}^2\right) = \sum_{j=1}^{n-k}4\pi^2|a_j|^2|b_j|^2 = \sum_{j=1}^{n-k}L(\theta_j)^2. \tag{73}$$

The second identity follows from Eq. (32) and the assumption (63). We conclude that $\mathcal{R}_t$ has length $L(\Gamma)$.

Remark 4. The calculations above show that if the dimension of $\mathcal{R}$ is at most half the dimension of $\mathcal{H}$, and every direct sum of qubit Hamiltonians can be generated, then every gate on $\mathcal{R}$ can be implemented time optimally.

Remark 5. The Hamiltonian in Eq. (64) need not be parallel transporting. To define a parallel transporting Hamiltonian $\bar{H}_t$ that drives $\mathcal{R}$ along the same trajectory as H we use Eq. (15) and define $\bar{H}_t$ as

$$\bar{H}_t = e^{-itH}(HP_{\mathcal{R}} + P_{\mathcal{R}}H - 2P_{\mathcal{R}}HP_{\mathcal{R}})e^{itH}. \tag{74}$$

To simplify the expression for $\bar{H}_t$, we assume the phases of $|0_j\rangle$ and $|1_j\rangle$ are such that a_j and b_j are real. Then,

$$\bar{H}_t = (\epsilon_1-\epsilon_0)\sum_{j=1}^{n-k}a_jb_j\big(2a_jb_j(|1_j\rangle\langle 1_j| - |0_j\rangle\langle 0_j|) + \big(a_j^2 - b_j^2\big)(e^{it(\epsilon_1-\epsilon_0)}|0_j\rangle\langle 1_j| + e^{it(\epsilon_0-\epsilon_1)}|1_j\rangle\langle 0_j|)\big). \tag{75}$$

VII. SUMMARY

We have derived an estimate, called the isoholonomic inequality, for the length of a cyclic transformation of a subspace of a Hilbert space with a given holonomy. The isoholonomic inequality constitutes half of the solution to the isoholonomic problem for holonomic quantum gates, as formulated in Ref. [14] (see Ref. [13] for the other half). We have also converted the isoholonomic inequality into an estimate of the time required to execute a holonomic quantum gate unitarily. As an illustration, we have shown that the implementation scheme in Ref. [6] for a universal set of holonomic gates is time optimal. The paper ended with a proof that the isoholonomic inequality is tight if the dimension of the subspace being transformed is at most half of the dimension of the Hilbert space.

ACKNOWLEDGMENTS

The author thanks Niklas Hörnedal for fruitful discussions and for proofreading early drafts.

APPENDIX: DERIVATION OF THE ISOHOLONOMIC INEQUALITY FOR STATES

Let $L(\theta)$ be the shortest length a closed curve of pure states can have, given that its Aharonov-Anandan holonomy is $e^{i\theta}$, where $-\pi < \theta \leqslant \pi$. We show that

$$L(\theta) = \sqrt{|\theta|(2\pi - |\theta|)}. \tag{A1}$$

Since there is nothing to prove if $\theta = 0$, we assume $\theta \neq 0$.

Let ρ be an arbitrary pure state, and let ρ_t be a closed curve of pure states at ρ with holonomy $e^{i\theta}$ and length $L(\theta)$.[4] Since a reparameterization of ρ_t does not change its length and holonomy, we can assume that ρ_t has a constant speed and returns to its initial state ρ at $\tau = 1$.

Let $|\psi\rangle$ be a unit vector projecting onto ρ, and let $|\psi_t\rangle$ be the horizontal lift of ρ_t starting from $|\psi\rangle$. The curve $|\psi_t\rangle$ extends from $|\psi\rangle$ to $e^{i\theta}|\psi\rangle$, both of which project onto ρ. Since the Hopf bundle is a Riemannian submersion, $|\psi_t\rangle$ and ρ_t have the same length. Furthermore, $|\psi_t\rangle$ has the same constant speed as ρ_t. Thus,

$$L(\theta) = \int_0^1\sqrt{\langle\dot{\psi}_t|\dot{\psi}_t\rangle}\,dt = \sqrt{\langle\dot{\psi}_0|\dot{\psi}_0\rangle}. \tag{A2}$$

[4] $L(\theta)$ does not depend on the choice of initial state ρ because length and Aharonov-Anandan holonomy are unitarily invariant quantities. Furthermore, each θ in $(-\pi, \pi]$ is the Aharonov-Anandan geometric phase of a closed curve of pure states at ρ, and at least one of these has length $L(\theta)$ since $\mathcal{G}(1;\mathcal{H})$ is compact.

The curve $|\psi_t\rangle$ is an extremal for the augmented kinetic energy functional

$$\mathcal{E}[|\phi_t\rangle, \lambda_t] = \frac{1}{2}\int_0^1 (\langle\dot{\phi}_t|\dot{\phi}_t\rangle + 2i\lambda_t\langle\phi_t|\dot{\phi}_t\rangle)\,dt, \tag{A3}$$

with λ_t being a Lagrange multiplier that forces horizontality. This is because ρ_t is a closed curve of minimal length among those having holonomy $e^{i\theta}$ [14]. The kinetic energy functional is defined on the space of curves in $\mathcal{V}(1;\mathcal{H})$ extending from $|\psi\rangle$ to $e^{i\theta}|\psi\rangle$ over the time interval [0,1].

Each variational vector field of $|\psi_t\rangle$ that fixes the endpoints of $|\psi_t\rangle$ has the form $-iX_t|\psi_t\rangle$, where X_t is an arbitrary curve of Hermitian operators that vanishes for $t = 0$ and $t = 1$. A variation of $|\psi_t\rangle$ with this variation vector field is $|\psi_{\epsilon,t}\rangle = U_{\epsilon,t}|\psi_t\rangle$ where, for each ϵ in $(-1, 1)$, $U_{\epsilon,t}$ is the backward time-ordered exponential of $-i\epsilon\dot{X}_t$,

$$U_{\epsilon,t} = \vec{\mathcal{T}}\exp\left(-i\epsilon\int_0^t \dot{X}_s\,ds\right). \tag{A4}$$

A partial integration shows that the variational derivative of the augmented kinetic energy of the variation of $|\psi_t\rangle$ is

$$\frac{d}{d\epsilon}\mathcal{E}[|\psi_{\epsilon,t}\rangle, \lambda_t]\bigg|_{\epsilon=0} = \frac{1}{2}\int_0^1 \mathrm{tr}\left(X_t\frac{d}{dt}(i|\psi_t\rangle\langle\dot{\psi}_t| - i|\dot{\psi}_t\rangle\langle\psi_t| - 2\lambda_t|\psi_t\rangle\langle\psi_t|)\right)dt. \tag{A5}$$

The Lagrange equation for $|\psi_t\rangle$ thus reads

$$\frac{d}{dt}(i|\psi_t\rangle\langle\dot{\psi}_t| - i|\dot{\psi}_t\rangle\langle\psi_t| - 2\lambda_t|\psi_t\rangle\langle\psi_t|) = 0. \tag{A6}$$

From the Lagrange equation follows that

$$A = i|\psi_t\rangle\langle\dot{\psi}_t| - i|\dot{\psi}_t\rangle\langle\psi_t| - 2\lambda_t|\psi_t\rangle\langle\psi_t| \tag{A7}$$

is a time-independent Hermitian operator, and from Eq. (A7) follows that the Lagrange multiplier is time independent, $2\dot{\lambda}_t = \langle\dot{\psi}_t|A|\psi_t\rangle + \langle\psi_t|A|\dot{\psi}_t\rangle = 0$. We write λ for the value of the Lagrange multiplier. By Eq. (A7),

$$|\dot{\psi}_t\rangle = -i(A - 2\lambda)|\psi_t\rangle. \tag{A8}$$

Equation (A7) also tells us that the support of A is two dimensional and is spanned by $|\psi\rangle$ and $|\dot{\psi}_0\rangle$. From this observation and Eq. (A8) we can conclude that $|\psi\rangle$, and thus the entire curve $|\psi_t\rangle$, is contained in the sum of two eigenspaces of $A - 2\lambda$. The eigenvalues are

$$a_\pm = -\lambda \pm \sqrt{\lambda^2 + \langle\dot{\psi}_0|\dot{\psi}_0\rangle}. \tag{A9}$$

The holonomy condition $|\psi_1\rangle = e^{i\theta}|\psi\rangle$ is satisfied if and only if $a_+ = 2\pi k - \theta$ for an integer $k \geqslant 0$ and $a_- = 2\pi l - \theta$ for an integer $l \leqslant 0$. Hence,

$$\begin{aligned} L(\theta)^2 &= \langle\dot{\psi}_0|\dot{\psi}_0\rangle \\ &= -a_+a_- \\ &= -(2\pi k - \theta)(2\pi l - \theta) \\ &\geqslant |\theta|(2\pi - |\theta|). \end{aligned} \tag{A10}$$

That the last inequality is an equality follows, for example, from the observation that every cyclic evolution of a qubit generated by a time-independent Hamiltonian saturates the inequality, as was shown in Sec. IV.

[1] F. Wilczek and A. Zee, Appearance of gauge structure in simple dynamical systems, Phys. Rev. Lett. **52**, 2111 (1984).

[2] P. Zanardi and M. Rasetti, Holonomic quantum computation, Phys. Lett. A **264**, 94 (1999).

[3] J. Pachos, P. Zanardi, and M. Rasetti, Non-Abelian Berry connections for quantum computation, Phys. Rev. A **61**, 010305(R) (1999).

[4] J. Pachos and P. Zanardi, Quantum holonomies for quantum computing, Int. J. Mod. Phys. B **15**, 1257 (2001).

[5] J. Anandan, Non-adiabatic non-abelian geometric phase, Phys. Lett. A **133**, 171 (1988).

[6] E. Sjöqvist, D. M. Tong, L. M. Andersson, B. Hessmo, M. Johansson, and K. Singh, Non-adiabatic holonomic quantum computation, New J. Phys. **14**, 103035 (2012).

[7] G. F. Xu, J. Zhang, D. M. Tong, E. Sjöqvist, and L. C. Kwek, Nonadiabatic holonomic quantum computation in decoherence-free subspaces, Phys. Rev. Lett. **109**, 170501 (2012).

[8] G. O. Alves and E. Sjöqvist, Time-optimal holonomic quantum computation, Phys. Rev. A **106**, 032406 (2022).

[9] J. Zhang, T. H. Kyaw, S. Filipp, L.-C. Kwek, E. Sjöqvist, and D. Tong, Geometric and holonomic quantum computation, Phys. Rep. **1027**, 1 (2023).

[10] W. Xiang-Bin and M. Keiji, Nonadiabatic conditional geometric phase shift with NMR, Phys. Rev. Lett. **87**, 097901 (2001).

[11] S.-L. Zhu and Z. D. Wang, Implementation of universal quantum gates based on nonadiabatic geometric phases, Phys. Rev. Lett. **89**, 097902 (2002).

[12] N. Hörnedal and O. Sönnerborn, Tight lower bounds on the time it takes to generate a geometric phase, Phys. Scr. **98**, 105108 (2023).

[13] S. Tanimura, M. Nakahara, and D. Hayashi, Exact solutions of the isoholonomic problem and the optimal control problem in holonomic quantum computation, J. Math. Phys. **46**, 022101 (2005).

[14] R. Montgomery, Isoholonomic problems and some applications, Commun. Math. Phys. **128**, 565 (1990).

[15] A. A. Abdumalikov, Jr., J. M. Fink, K. Juliusson, M. Pechal, S. Berger, A. Wallraff, and S. Filipp, Experimental realization of non-Abelian non-adiabatic geometric gates, Nature (London) **496**, 482 (2013).

[16] G. Feng, G. Xu, and G. Long, Experimental realization of nonadiabatic holonomic quantum computation, Phys. Rev. Lett. **110**, 190501 (2013).

[17] S. Arroyo-Camejo, A. Lazariev, S. W. Hell, and G. Balasubramanian, Room temperature high-fidelity holonomic

single-qubit gate on a solid-state spin, Nat Commun. **5**, 4870 (2014).

[18] C. Zu, W.-B. Wang, L. He, W.-G. Zhang, C.-Y. Dai, F. Wang, and L.-M. Duan, Experimental realization of universal geometric quantum gates with solid-state spins, Nature (London) **514**, 72 (2014).

[19] D. Girolami, Observable measure of quantum coherence in finite dimensional systems, Phys. Rev. Lett. **113**, 170401 (2014).

[20] S. Luo and Y. Sun, Skew information revisited: Its variants and a comparison of them, Theor. Math. Phys. **202**, 104 (2020).

[21] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vols. I, II, Wiley Classics Library (Wiley, New York, 1996).

[22] Y. Aharonov and J. Anandan, Phase change during a cyclic quantum evolution, Phys. Rev. Lett. **58**, 1593 (1987).

[23] X.-D. Yu and D. M. Tong, Evolution operator can always be separated into the product of holonomy and dynamic operators, Phys. Rev. Lett. **131**, 200202 (2023).

[24] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Edition (Cambridge University Press, New York, 2010).

[25] E. Sjöqvist, V. Azimi Mousolou, and C. M. Canali, Conceptual aspects of geometric quantum computation, Quantum Inf. Process. **15**, 3995 (2016).

[26] M. J. Bremner, C. M. Dawson, J. L. Dodd, A. Gilchrist, A. W. Harrow, D. Mortimer, M. A. Nielsen, and T. J. Osborne, Practical scheme for quantum computation with any two-qubit entangling gate, Phys. Rev. Lett. **89**, 247902 (2002).

[27] E. Sjöqvist, Nonadiabatic holonomic single-qubit gates in off-resonant Λ systems, Phys. Lett. A **380**, 65 (2016).

[28] G. F. Xu, C. L. Liu, P. Z. Zhao, and D. M. Tong, Nonadiabatic holonomic gates realized by a single-shot implementation, Phys. Rev. A **92**, 052302 (2015).