

Equating quantum imaginary time evolution, Riemannian gradient flows, and stochastic implementations

Nathan A. McMahon,^{1,2,*} Mahum Pervez³ and Christian Arenz^{3,†}

¹*(aQa^L) Applied Quantum Algorithms, Leiden University, Leiden, The Netherlands*

²*LIACS, Leiden University, Leiden, The Netherlands*

³*School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, Arizona 85281, USA*



(Received 25 April 2025; revised 19 November 2025; accepted 25 February 2026; published 7 April 2026)

We identify quantum imaginary time evolution as a Riemannian gradient flow on the unitary group. We develop an upper bound for the error between the two evolutions that can be controlled through the step size of the Riemannian gradient descent that minimizes the energy of the system. We discuss implementations through adaptive quantum algorithms and present a stochastic Riemannian gradient descent algorithm in which each step is efficiently implementable on a quantum computer. We prove that for a sufficiently small step size, the stochastic evolution concentrates around the imaginary time evolution, thereby providing performance guarantees for cooling the system through stochastic Riemannian gradient descent. Finally, we show that Riemannian gradient descent can be understood as an approximation to quantum imaginary time evolution that does not employ classical optimization.

DOI: [10.1103/ht2m-lj91](https://doi.org/10.1103/ht2m-lj91)

I. INTRODUCTION

Imaginary time evolution (ITE) is a powerful classical method for determining properties of a many-body Hamiltonian H . For example, various quantum Monte Carlo methods [1] aim to find the spectral characteristics of H by solving the time imaginary Schrödinger equation [2] whose normalized solution is given by the state

$$|\psi(\beta)\rangle = \frac{e^{-\beta H} |\psi_0\rangle}{\|e^{-\beta H} |\psi_0\rangle\|}, \quad (1)$$

where β is the “imaginary time” and $|\psi_0\rangle$ is the initial state of the system. The appeal of such methods is the fact that if the initial state has nonzero overlap with the ground state $|E_0\rangle$ of H , i.e., $\langle\psi_0|E_0\rangle \neq 0$, then under sufficiently long evolution times β the state $|\psi(\beta)\rangle$ is given by $|E_0\rangle$.

Due to the exponential overhead in implementing imaginary time evolution on classical computers, over the last few years there has been a growing interest in developing quantum algorithms that aim to create the ITE state (1) on a quantum computer. Such methods face the challenge that the evolution $e^{-\beta H}$ that leads to the state (1) is not unitary, and therefore not directly implementable with unitary gates on a quantum device. Several approaches [3–14] have been proposed to overcome this challenge by designing unitary

evolutions that approximate $|\psi(\beta)\rangle$. For instance, the quantum imaginary time evolution (QITE) algorithm [3–11] and its probabilistic modifications [12,13] Trotterize ITE to approximate each Trotter step by a unitary that is identified through classical optimization. Similarly, in variational approaches to ITE [15,16] classical optimization routines are used in conjunction with a quantum computer to optimize a parametrized quantum circuit that prepares $|\psi(\beta)\rangle$. While these algorithms are often powerful heuristics, quantum algorithms that implement ITE with performance guarantees [3] remain scarce in the literature.

In this work, we address this challenge by identifying ITE as a Riemannian gradient flow [17–21] on the unitary group, which can be implemented through adaptive quantum algorithms [20–27] on quantum computers. The relationship between ITE and gradient flows has been observed before in the literature [28,29] and has recently been leveraged [30,31] to provide fidelity bounds for ground-state preparation and energy minimization. Here, we derive rigorous bounds for implementing the *full* ITE, thus providing performance guarantees for implementing the ITE state for generic imaginary evolution times β and for convergence to excited states (when $\langle\psi_0|E_0\rangle = 0$).

We first prove that the state created by the unitary evolution that describes the Riemannian gradient flow can be made arbitrarily close to the ITE state $|\psi(\beta)\rangle$. This is achieved by appropriately choosing the step size $\Delta\beta$ of the discretized Riemannian gradient flow evolution, which can be interpreted as a Riemannian gradient descent (RGD) algorithm minimizing the energy of the system, as depicted in Fig. 1. We go on to develop an upper bound (Theorem 1) for the error between the ITE state and the state that is created through Riemannian gradient descent, which can again be controlled through $\Delta\beta$. The developed equivalence between ITE and Riemannian gradient

*Contact author: n.a.mcmahon@liacs.leidenuniv.nl

†Contact author: carenz1@asu.edu

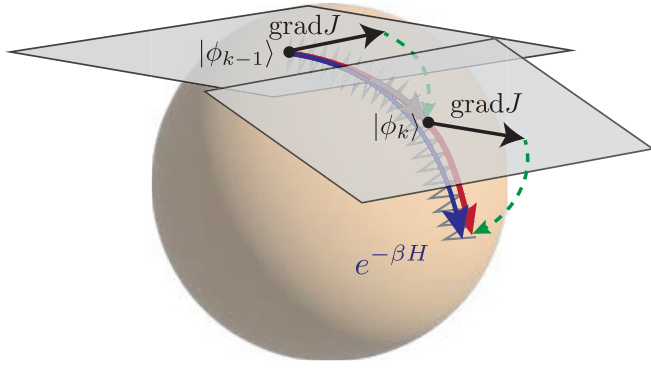


FIG. 1. Schematic representation of the correspondence between imaginary time evolution described by $e^{-\beta H}$ and Riemannian gradient descent minimizing the energy of the system. A unitary evolution $e^{-\Delta\beta \text{grad} J}$ (red) is created by the retraction (green) of the Riemannian gradient $\text{grad} J$ with respect to a cost function J , where $\Delta\beta$ is the step size. We show that the error between the ITE state and the state $|\phi_1\rangle$ created through one step of RGD is of the order $\mathcal{O}(\Delta\beta^{-2})$ (Lemma 1). Building upon this, we prove that after n steps the state $|\phi_n\rangle$ approximates the ITE state with an error $\mathcal{O}(\Delta\beta)$ (Theorem 1). We use this observation to develop a stochastic Riemannian gradient descent algorithm (gray line) whose evolution concentrates around ITE (Theorem 2).

flows has two key implications: (1) It motivates using the large literature on heuristic adaptive quantum algorithms [20–27] to design quantum algorithms for implementing ITE and (2) the developed upper bound gives performance guarantees for state preparation methods that are based on Riemannian gradient flows [19–21,32].

We demonstrate the utility of (1) and (2) by developing a stochastic implementation of the discretized Riemannian gradient flow. Each step in this stochastic Riemannian gradient descent (SRGD) algorithm is efficiently implementable on a quantum computer and the energy of the system is provably reduced on average in each step. We then derive performance bounds for using the stochastic Riemannian gradient descent algorithm to prepare the ITE state $|\psi(\beta)\rangle$ to arbitrary precision (Theorem 2). We show that the randomized evolution concentrates around ITE, converging at a rate determined by $\Delta\beta$. We proceed to use the developed bounds to obtain insights on how knowledge of the initial state and H may be exploited to develop quantum circuit implementations that prepare the ITE state efficiently.

II. QUANTUM IMAGINARY TIME EVOLUTION AS RIEMANNIAN GRADIENT DESCENT

The theory of Riemannian gradient flows is a rich framework for solving optimization problems defined on a Riemannian manifold, such as the unitary group [17]. Riemannian gradient flows have a long history in optimization. Applications range from diagonalizing a matrix through Brockett’s double bracket flow to solving least-squares-type problems [33,34] and finding the ground state of a many-body Hamiltonian [35,36]. Recently, Riemannian gradient flows have been utilized to develop quantum algorithms for ground-state problems [19–21]. Rather than optimizing over

parameters in a quantum circuit to minimize cost function J (e.g., as in variational quantum algorithms [37]), J is instead directly optimized over unitary transformations.

The Riemannian gradient flow on the special unitary group $\text{SU}(d)$ of dimension d is defined by the solution to the differential equation [17,18]

$$\frac{d}{dt}U = -\text{grad} J[U] \quad (2)$$

for the unitary operator $U \in \text{SU}(d)$. Here, $\text{grad} J[U] \in T_U \text{SU}(d)$ is the Riemannian gradient of a cost function $J[U]$ that lives in the tangent space $T_U \text{SU}(d)$ given at the identity $T_1 \text{SU}(d) = \mathfrak{su}(d)$ by the special unitary algebra $\mathfrak{su}(d)$. In this work, we consider a cost function that is given by the expectation value $J = \langle \phi | H | \phi \rangle$ of the Hamiltonian H with respect to the state $|\phi\rangle = U|\psi_0\rangle$. This state is created by applying the unitary transformation U to the initial state $|\phi_0\rangle$. The choice of the cost function is motivated by the fact that the solution to Eq. (2) prepares the ground state of H , by minimizing J , when t is sufficiently large [17,18].

The discretized solution to the differential equation (2) defining the Riemannian gradient flow is given by a sequence of unitary transformations of the form

$$U(\Delta t) = e^{-\Delta t \text{grad} J}, \quad (3)$$

where Δt is the step size and $\text{grad} J = [H, |\phi\rangle \langle \phi|] \in \mathfrak{su}(d)$ is the Riemannian gradient at the state $|\phi\rangle$ [38].

The state that is created by successively applying the transformation (3) k times to the initial state $|\phi_0\rangle = |\psi_0\rangle$ is recursively given by

$$|\phi_k\rangle = U_k(\Delta t)|\phi_{k-1}\rangle. \quad (4)$$

The recursive update (4) can be understood as an RGD algorithm that minimizes J . We see that the unitary transformation $U_k(\Delta t) = e^{-\Delta t \text{grad} J}$, where $\text{grad} J = [H, |\phi_{k-1}\rangle \langle \phi_{k-1}|]$ creating the state $|\phi_k\rangle$ depends on the previous state $|\phi_{k-1}\rangle$. This observation indicates that designing a quantum algorithm that implements RGD requires some “feedback” mechanism that leverages the information about the previous state to inform the update step.

Before we discuss how a quantum algorithm can be designed that adaptively grows a quantum circuit to create $|\phi_n\rangle$ based on measurement data, we first discuss how RGD can be used to create the ITE state (1) arbitrarily well. We observe that for small step sizes $\Delta t = \Delta\beta$, the ITE state $|\psi(\Delta\beta)\rangle$ and the state $|\phi(\Delta t)\rangle = U_1(\Delta t)|\psi_0\rangle$ created through one step of RGD are close to each other, which is expressed through the relation $\frac{d}{dx}|\psi(x)\rangle|_{x=0} = \frac{d}{dx}|\phi(x)\rangle|_{x=0}$. As such, the Euclidean norm difference $\| |\psi(\Delta\beta)\rangle - |\phi(\Delta\beta)\rangle \|$ between the two states is of the order $\mathcal{O}(\Delta\beta^2)$. We remark that this observation is equivalent to the observation made in Ref. [30] that the ITE state solves Brockett’s double bracket flow equation [17,33], which describes the Riemannian gradient flow on the adjoint orbit of the unitary group.

An upper bound can be obtained for $\| |\psi(\Delta\beta)\rangle - |\phi(\Delta\beta)\rangle \|$ by bounding the remainder of the corresponding Taylor expansions. With further details found in Supplemental Material [39], we establish the following lemma.

Lemma 1. Let $|\psi(\Delta\beta)\rangle$ and $|\phi(\Delta\beta)\rangle = U(\Delta\beta)|\psi_0\rangle$ be the states created by ITE (1) and one step of RGD (3) with

step size $\Delta\beta$. Then, for any initial state $|\psi_0\rangle$,

$$\| |\psi(\Delta\beta)\rangle - |\phi(\Delta\beta)\rangle \| \leq 6\Delta\beta^2 \|H\|_\infty^2, \quad (5)$$

where $\|H\|_\infty$ denotes the spectral norm of H .

Consequently, if we divide β into n segments $\Delta\beta = \frac{\beta}{n}$, the error $\epsilon_n = \| |\psi(\beta)\rangle - |\phi_n\rangle \|$ between the state $|\phi_n\rangle$ created through n steps of RGD (4) and the ITE state $|\psi(\beta)\rangle$ is intuitively of the order $\mathcal{O}(1/n)$.

The main technical challenge in rigorously establishing this intuition is the fact that standard techniques used to show that the total error ϵ_n is upper bounded by n times the local error (i.e., bounded by Lemma 1) cannot be directly applied, due to the nonunitary and nonlinear nature of ITE and RGD. Instead, we show in Supplemental Material [39] that ϵ_k is given by a recursive relation of the form

$$\epsilon_k \leq \epsilon_{k-1}A + B, \quad (6)$$

where $A = 1 + 4\Delta\beta\|H\|_\infty$ and $B = 10\Delta\beta^2\|H\|_\infty^2$. As such, the error ϵ_n is upper bounded by a geometric series

$$\epsilon_n \leq B \sum_{k=0}^{n-1} A^k = B \frac{1 - A^n}{1 - A}, \quad (7)$$

which establishes the following bound.

Theorem 1. The error ϵ_n between the ITE state $|\psi(\beta)\rangle$ (1) and the state $|\phi_n\rangle$ (4) created by n steps of RGD with step size $\Delta\beta = \frac{\beta}{n}$ is upper bound by

$$\epsilon_n \leq \frac{5}{2}\Delta\beta\|H\|_\infty(e^{4\beta\|H\|_\infty} - 1). \quad (8)$$

Thus, for $\beta\|H\|_\infty \ll 1$, such that the second order of $e^{\beta\|H\|_\infty}$ can be neglected, the complexity for preparing the corresponding ITE state through RGD is of the order $\mathcal{O}(\frac{\beta^2\|H\|_\infty^2}{\epsilon})$. Such a short ITE has been used in primitive subroutines in other quantum algorithms, such as the quantum Lanczos or the quantum minimally entangled typical thermal states algorithm [3,40]. Thus, Theorem 1 provides error guarantees for the implementation of these subalgorithms via RGD.

We further remark that since by the triangle inequality we have $\| |E_0\rangle - |\phi_n\rangle \| \leq \| |E_0\rangle - |\psi(\beta)\rangle \| + \epsilon_n$, Theorem 1 allows us to bound the error for preparing the ground state $|E_0\rangle$ of H via RGD [19–21]. However, Theorem 1 goes beyond ground-state preparation as it bounds the error for preparing a generic ITE state. This includes preparing excited states, which can be achieved by initializing the system in a state that has zero overlap with the ground state.

III. GRADIENT FLOW-BASED QUANTUM ALGORITHMS FOR QUANTUM IMAGINARY TIME EVOLUTION

The RGD update step $U_k = e^{-\Delta\beta \text{grad} J}$ is, in general, not efficiently implementable on a quantum computer. As the Riemannian gradient $\text{grad} J \in \mathfrak{su}(2^N)$ generally lives in an exponentially large Lie algebra $\mathfrak{su}(2^N)$ of dimension $2^{2N} - 1$, where N is the number of qubits, we can in general not hope for implementing U_k with polynomially many gates. However, several approximation schemes have been introduced in the literature that project the Riemannian gradient into smaller dimensional subspaces to make the update step efficiently implementable, e.g., through Trotterization [19–22]. The key

idea is to pick a polynomially sized subspace $\mathcal{A}_k \subset \mathfrak{su}(2^N)$ in each update step k . The Riemannian gradient is then projected into this subspace to obtain an approximate Riemannian gradient $\widetilde{\text{grad} J} = i \sum_{P_j \in \mathcal{A}_k} C_{k,j} P_j$ in which the coefficients $C_{k,j}$ can be estimated via measurements of the gradients

$$\begin{aligned} C_{k,j} &= \left. \frac{\partial}{\partial \theta} \langle \phi_k | e^{i\theta P_j} H e^{-i\theta P_j} | \phi_k \rangle \right|_{\theta=0} \\ &= \langle \text{grad} J, i P_j \rangle, \end{aligned} \quad (9)$$

e.g., through the parameter shift rule [41,42] or finite differences $C_{k,j} \approx (\langle \phi_k | e^{i\Delta\theta P_j} H e^{-i\Delta\theta P_j} | \phi_k \rangle - \langle \phi_k | H | \phi_k \rangle) / \Delta\theta$. Here, $\langle X, Y \rangle = \text{Tr}\{X^\dagger Y\}$ denotes the Hilbert-Schmidt inner product.

A particularly appealing strategy is to project into a randomly chosen directions P_j (in the tangent space) in each update step [21]. For ground-state problems, this randomization strategy converges to the ground state almost surely despite the existence of saddle points [19]. Here, we employ a similar randomization strategy to efficiently implement U_k on average.

Consider sampling in each update step k a gate $V_k^{(j)} = e^{-iC_{k,j}D\Delta\beta P_j}$ by picking uniformly random a (normalized) Pauli operator $iP_j \in \mathfrak{su}(d)$ where $D = \dim(\mathfrak{su}(d)) = d^2 - 1$. We recall that the coefficients $C_{k,j}$ are determined by Eq. (9) and depend on the state of the previous step. To gain some intuition on how this randomization procedure gives rise to implementing ITE, we start by considering the energy change $\Delta E_j = \langle \phi | H | \phi \rangle - \langle \phi | V_k^{(j)\dagger} H V_k^{(j)} | \phi \rangle$ at the step k for some state $|\phi\rangle$. For a step size $\Delta\beta = 1/(4D\|H\|_\infty)$, this change is lower bounded by $\Delta E_j \geq \frac{1}{8\|H\|_\infty} \langle \text{grad} J, i P_j \rangle^2$ [21]. Sampling Pauli operators P_j from $\mathfrak{su}(d)$ uniformly gives, for any A , that $\frac{1}{D} \sum_j P_j A P_j = \frac{\text{Tr}[A]}{d}$. As such, we find that the average energy change $\overline{\Delta E} = \mathbb{E}_j[\Delta E_j]$ at a single random step k is lower bound by

$$\overline{\Delta E} \geq \frac{\|\text{grad} J\|_{\text{HS}}^2}{8d\|H\|_\infty}. \quad (10)$$

The Hilbert-Schmidt norm $\|\cdot\|_{\text{HS}}$ of the Riemannian gradient $\text{grad} J = [H, |\phi\rangle\langle\phi|]$ is given by the variance $\|\text{grad} J\|_{\text{HS}} = \sqrt{\langle \phi | H^2 | \phi \rangle - \langle \phi | H | \phi \rangle^2}$, which is zero only at eigenstates of the Hamiltonian. As such, on average a randomized step corresponds to the energy change that is obtained for one step of RGD. In fact, to first order in the step size, the averaged evolution is described by the unitary quantum channel that corresponds to one step of RGD.

We denote by $|\chi_k^{(\gamma)}\rangle$ a particular random state generated by a sequence of random unitaries, $V_k^{(j)}$, corresponding to the choice of random Pauli operators $\gamma = \{iP_1, iP_2, \dots, iP_k\}$. We refer to the sequence of random unitaries $V_k^{(j)}$ that creates the states $|\chi_k^{(\gamma)}\rangle$ as *stochastic Riemannian gradient descent* [19,43]. In Theorem 2, we establish an upper bound for the average of the fidelity error

$$\epsilon_k^{(\gamma)} = 1 - |\langle \psi(\beta) | \chi_k^{(\gamma)} \rangle|^2 \quad (11)$$

and show that the variance can be controlled through the step size of SRGD.

Theorem 2. The average fidelity error $\bar{\varepsilon}_n = \mathbb{E}_\gamma[\varepsilon_n^{(\gamma)}]$ after n steps of SRGD with step size $\Delta\beta = \frac{\beta}{n}$ is upper bounded by

$$b_n = \frac{9}{2}\sqrt{2\Delta\beta\|H\|_\infty}D(e^{8\beta\|H\|_\infty} - 1)^{\frac{1}{2}}, \quad (12)$$

for sufficiently large n so that $\Delta\beta\|H\|_\infty \leq 1$. For any $\delta > 0$, the probability that a state created through SRGD will give rise to a fidelity error greater than $b_n + \delta$ is upper bounded by

$$\Pr(\varepsilon_n^{(\gamma)} > b_n + \delta) \leq \frac{8\Delta\beta\|H\|_\infty D^2}{\delta^2}(e^{8\beta\|H\|_\infty} - 1). \quad (13)$$

We prove Theorem 2 in Supplemental Material [39] by establishing a recursion relation for the variance similar to Eq. (6) and then applying the Chebyshev inequality.

Theorem 2 shows that by choosing the step size $\Delta\beta$, we can control the deviation δ from the average fidelity error upper bound b_n . Since b_n can also be controlled in the same way, we see that a random state $|\chi_n^{(\gamma)}\rangle$ created through SRGD may be made arbitrarily close to the ITE state. This observation provides an interesting perspective of SRGD as “cooling” the initial state $|\psi_0\rangle$ to the ground state of H .

IV. DISCUSSION

Each step of SRGD is efficiently implementable on a quantum computer. The gates $V_k^{(j)}$, generated by Pauli operators, can be implemented by a quantum circuit of depth at most linear in the number of qubits. However, the overall run-time $\mathcal{O}(\beta\|H\|_\infty D^2 \exp(8\beta\|H\|_\infty))$ of the randomized algorithm cannot be efficient in general for complexity reasons [44], which is reflected in Theorem 2 by the dependence on $D = 2^{2N} - 1$.

In contrast, the lower bound for the average energy change (10) suggests that the complexity of SRGD scales inversely with the smallest coefficient $C_j = \langle \text{grad } J, iP_j \rangle^2$ of the Riemannian gradient (taken over all steps). Consequently, we expect that if there exists a constant $\alpha > 0$ such that C_j does not vanish faster than $\mathcal{O}(N^{-\alpha})$, then the ITE state can be prepared in polynomially many steps. As the Riemannian gradient depends on the Hamiltonian H and the state $|\chi_k^{(\gamma)}\rangle$, a more efficient sampling strategy should exploit additional properties of H and $|\chi_k^{(\gamma)}\rangle$.

For example, consider the Riemannian gradient $[H, |\chi_k^{(\gamma)}\rangle\langle\chi_k^{(\gamma)}|]$ at a random state $|\chi_k^{(\gamma)}\rangle$ that has support on only a polynomially (in N) sized subset of Pauli operators at each step k . We then may instead sample only over this subset, giving rise to an effective sample space of dimension $D = \text{poly}(N)$. The commutator $[H, |\chi_k^{(\gamma)}\rangle\langle\chi_k^{(\gamma)}|]$ is then approximated by linear combinations of only polynomially many Pauli operators. We observe that such a scenario is possible if the state $|\chi_k^{(\gamma)}\rangle$ has finite correlation length, and H is a geometrically local Hamiltonian. The finite correlation length suggests that expectation values of Pauli operators P , $\langle P \rangle = \langle \chi_k^{(\gamma)} | P | \chi_k^{(\gamma)} \rangle$, will decay exponentially with the support of P . This implies that $|\chi_k^{(\gamma)}\rangle\langle\chi_k^{(\gamma)}| \approx \sum_{P \in \mathcal{S}} \langle P \rangle P$, where $\mathcal{S}$ is a $\text{poly}(N)$ sized set of all Pauli operators with support under some fixed constant. In turn, this would imply that $[H, |\chi_k^{(\gamma)}\rangle\langle\chi_k^{(\gamma)}|]$ would be well approximated using only $\text{poly}(N)$ Pauli operators, providing a regime where a gradient

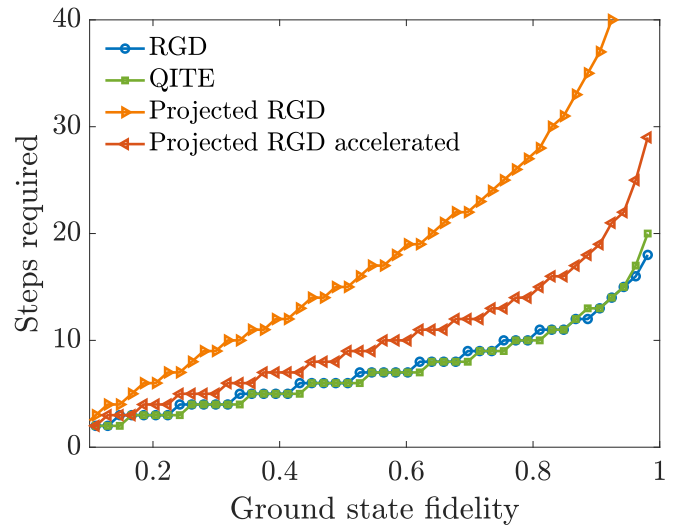


FIG. 2. Number of steps required to prepare the ground state of a transverse field Ising Hamiltonian with six spins as a function of the ground-state fidelity. For the QITE implementation (green), we used the same set $\mathcal{S}$ of geometrically local Pauli operators up to weight 4 that we used to approximate the Riemannian gradient $\text{grad } J$ by projecting $\text{grad } J$ into the subspace $\mathcal{S}$ (orange, red). The orange curve corresponds to a projected RGD with the same step size as in QITE and RGD, while the red curve corresponds to an accelerated RGD with a scaled steps size. The blue curve shows the behavior of the exact RGD algorithm (4).

descent algorithm should efficiently converge to the optimal point (or close to it) as an implementation of ITE would. The use of correlation length is similar to the strategy employed in QITE. In fact, we show below that QITE implements RGD when “enough” Pauli operators are available to approximate the unitary evolution that creates the ITE state.

V. RELATIONS TO QITE

The QITE algorithms aim to approximate the ITE state on a quantum computer for short imaginary times $\Delta\beta$ by finding the unitary evolution $W = e^{-i\Delta\beta A}$ that prepares a state that is close to the (Trotterized) ITE state. This is achieved by expanding the operator $A = \sum_{P \in \mathcal{S}} \lambda_P P$ using geometrically local Pauli operators P and optimizing the coefficients λ_P to best approximate the ITE state by solving a quadratic optimization problem offline. Now, when all Paulis spanning $\mathfrak{su}(2^N)$ are used to express A , Lemma 1 shows that then $iA = \text{grad } J$ can be identified as the Riemannian gradient. As such, solving a quadratic optimization problem offline is then not needed as we can directly identify the coefficients $\lambda_P = \langle \text{grad } J, iP \rangle$ as the elements of the Riemannian gradient.

However, to implement U efficiently, e.g., via Hamiltonian simulation techniques, only polynomially many Pauli operators $|\mathcal{S}| = \text{poly}(N)$ can be used to expand A . With further details found in Supplemental Material [39], in Fig. 2 we compare the behavior of QITE and RGD for preparing the ground state of a transverse field Ising chain Hamiltonian with six spins when only polynomially many Pauli operators are used to expand A . We observe that the behavior of RGD matches

remarkably well the behavior of QITE. This suggests that even when not all Pauli operators are available to expand A , the quadratic optimization in QITE finds a “good” approximation to RGD. In turn, this suggest that implementing QITE can be achieved by implementing RGD. Indeed, when the Riemannian gradient $\text{grad} J = i \sum_{P \in \mathcal{S}} \langle \text{grad} J, iP \rangle P$ is projected into the polynomially sized subspace $\mathcal{S}$ used in QITE, a similar scaling behavior (orange, red) is observed as a function of the precision/ground-state fidelity.

VI. CONCLUSION

In this work, we showed that quantum imaginary time evolution can be understood as Riemannian gradient descent that minimizes the energy of the system. We derived a rigorous bound for the error between the states created through imaginary time evolution and Riemannian gradient descent that can be controlled by the step size. We went on to prove bounds for stochastic implementations, in which each step of the Riemannian gradient descent algorithm is efficiently implementable on quantum computers. We characterized how much the energy decreases on average in a randomized step. Interestingly, due to a concentration result, the stochastic evolutions follow with high probability the imaginary time evolution. This implies that, for sufficiently small step sizes,

the system is cooled by moving into random tangent space directions of the Riemannian gradient descent.

We discussed potential ways forward to leverage information about the Hamiltonian and the initial state to efficiently prepare the imaginary time state on a quantum computer. Problem-dependent knowledge could inform when learning an efficient representation of the Riemannian gradient (i.e., in a lower dimensional subspace) may be possible. Ultimately, the derived bounds and discussions provide a jumping off point for utilizing the vast literature around adaptive quantum algorithms for performing imaginary time evolution on quantum computers. Conversely, the bounds also provide a starting point to study convergence of gradient descent algorithms through comparison to ITE.

ACKNOWLEDGMENTS

N.A.M. was cofunded by the European Union (ERC CoG, BeMAIQuantum, 101124342). M.P. and C.A. acknowledge support from the National Science Foundation (Grant No. 2231328).

DATA AVAILABILITY

The data that support the findings of this article are openly available [45].

-
- [1] M. Suzuki, Quantum Monte Carlo methods in equilibrium and nonequilibrium systems, in *Proceedings of the Ninth Taniguchi International Symposium, Susono, Japan, 1986*, Springer Series in Solid-State Sciences Vol. 74 (Springer, Berlin, Heidelberg, 2012).
 - [2] A. Goldberg and J. L. Schwartz, Integration of the Schrödinger equation in imaginary time, *J. Comput. Phys.* **1**, 433 (1967).
 - [3] M. Motta, C. Sun, A. T. Tan, M. J. O’Rourke, E. Ye, A. J. Minnich, F. G. Brandao, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, *Nat. Phys.* **16**, 205 (2020).
 - [4] K. Yeter-Aydeniz, R. C. Pooser, and G. Siopsis, Practical quantum computation of chemical and nuclear energy levels using quantum imaginary time evolution and Lanczos algorithms, *npj Quantum Inf.* **6**, 63 (2020).
 - [5] N. Gomes, F. Zhang, N. F. Berthussen, C.-Z. Wang, K.-M. Ho, P. P. Orth, and Y. Yao, Efficient step-merged quantum imaginary time evolution algorithm for quantum chemistry, *J. Chem. Theory Comput.* **16**, 6256 (2020).
 - [6] S.-N. Sun, M. Motta, R. N. Tazhigulov, A. T. Tan, G. K.-L. Chan, and A. J. Minnich, Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution, *PRX Quantum* **2**, 010317 (2021).
 - [7] S.-H. Lin, R. Dilip, A. G. Green, A. Smith, and F. Pollmann, Real- and imaginary-time evolution with compressed quantum circuits, *PRX Quantum* **2**, 010342 (2021).
 - [8] H. Kazuki, T. Kosugi, and Y.-I. Matsushita, Implementation of quantum imaginary-time evolution method on NISQ devices by introducing nonlocal approximation, *npj Quantum Inf.* **7**, 85 (2021).
 - [9] K. Yeter-Aydeniz, E. Moschandreou, and G. Siopsis, Quantum imaginary-time evolution algorithm for quantum field theories with continuous variables, *Phys. Rev. A* **105**, 012412 (2022).
 - [10] T. Tsuchimochi, Y. Ryo, S. L. Ten-No, and K. Sasasako, Improved algorithms of quantum imaginary time evolution for ground and excited states of molecular systems, *J. Chem. Theory Comput.* **19**, 503 (2023).
 - [11] K. Hejazi, M. Motta, and G. K.-L. Chan, Adiabatic quantum imaginary time evolution, *Phys. Rev. Res.* **6**, 033084 (2024).
 - [12] T. Liu, J.-G. Liu, and H. Fan, Probabilistic nonunitary gate in imaginary time evolution, *Quantum Inf. Proc.* **20**, 204 (2021).
 - [13] T. Kosugi, Y. Nishiya, H. Nishi, and Y.-I. Matsushita, Imaginary-time evolution using forward and backward real-time evolution with a single ancilla: First-quantized eigensolver algorithm for quantum chemistry, *Phys. Rev. Res.* **4**, 033121 (2022).
 - [14] J. M. Leamer, D. I. Bondar, and G. McCaul, Quantum dynamical emulation, *arXiv:2403.03350*.
 - [15] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, *npj Quantum Inf.* **5**, 75 (2019).
 - [16] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, Variational quantum algorithms for discovering Hamiltonian spectra, *Phys. Rev. A* **99**, 062304 (2019).
 - [17] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Communications and Control Engineering (Springer, London, 2012).
 - [18] T. Schulte-Herbrüggen, S. J. Glaser, G. Dirr, and U. Helmke, Gradient flows for optimization in quantum information and quantum dynamics: Foundations and applications, *Rev. Math. Phys.* **22**, 597 (2010).

- [19] E. Malvetti, C. Arenz, G. Dirr, and T. Schulte-Herbrüggen, Randomized gradient descents on Riemannian manifolds: Almost sure convergence to global minima in and beyond quantum optimization, [arXiv:2405.12039](#).
- [20] R. Wiersema and N. Killoran, Optimizing quantum circuits with Riemannian gradient flow, *Phys. Rev. A* **107**, 062421 (2023).
- [21] A. B. Magann, S. E. Economou, and C. Arenz, Randomized adaptive quantum state preparation, *Phys. Rev. Res.* **5**, 033227 (2023).
- [22] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, *Nat. Commun.* **10**, 3007 (2019).
- [23] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-adaptive: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor, *PRX Quantum* **2**, 020310 (2021).
- [24] A. B. Magann, K. M. Rudinger, M. D. Grace, and M. Sarovar, Feedback-based quantum optimization, *Phys. Rev. Lett.* **129**, 250502 (2022).
- [25] J. B. Larsen, M. D. Grace, A. D. Baczewski, and A. B. Magann, Feedback-based quantum algorithms for ground state preparation, *Phys. Rev. Res.* **6**, 033336 (2024).
- [26] A. B. Magann, K. M. Rudinger, M. D. Grace, and M. Sarovar, Lyapunov-control-inspired strategies for quantum combinatorial optimization, *Phys. Rev. A* **106**, 062414 (2022).
- [27] H. L. Tang, Y. Chen, P. Biswas, A. B. Magann, C. Arenz, and S. E. Economou, Nonvariational ADAPT algorithm for quantum simulations, *Phys. Rev. Res.* **7**, 023275 (2025).
- [28] J. Stokes, B. Chen, and S. Veerapaneni, Numerical and geometrical aspects of flow-based variational quantum Monte Carlo, *Mach. Learn.: Sci. Technol.* **4**, 021001 (2023).
- [29] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, *Quantum* **4**, 269 (2020).
- [30] M. Gluza, J. Son, B. H. Tiang, Y. Suzuki, Z. Holmes, and N. H. Ng, Double-bracket quantum algorithms for quantum imaginary-time evolution, *Phys. Rev. Lett.* **136**, 020601 (2026).
- [31] R. Zander, R. Seidel, L. Xiaoyue, and M. Gluza, Role of Riemannian geometry in double-bracket quantum imaginary-time evolution, in *Geometric Science of Information: GSI 2025*, edited by F. Nielsen and F. Barbaresco, Lecture Notes in Computer Science Vol. 16035 (Springer, Cham, 2026).
- [32] M. Gluza, Double-bracket quantum algorithms for diagonalization, *Quantum* **8**, 1316 (2024).
- [33] R. W. Brockett, Least squares matching problems, *Lin. Alg. Appl.* **122-124**, 761 (1989).
- [34] R. W. Brockett, Differential geometry and the design of gradient algorithms, in *Proceedings of Symposia in Pure Mathematics*, edited by R. Greene and S. T. Yau (American Mathematical Society, Providence, Rhode Island, 1993), Vol. 54.1, pp. 69–92.
- [35] C. M. Dawson, J. Eisert, and T. J. Osborne, Unifying variational methods for simulating quantum many-body systems, *Phys. Rev. Lett.* **100**, 130501 (2008).
- [36] M. B. Hastings, On Lieb-Robinson bounds for the double bracket flow, in *The Physics and Mathematics of Elliott Lieb*, edited by R. L. Frank, A. Laptev, M. Lewin, and R. Seiringer (EMS Press, Berlin, 2022), p. 515.
- [37] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [38] With a slight abuse of notation, from now on we denote the Riemannian gradient by $\text{grad} J = [H, |\phi\rangle\langle\phi|]$, noting that due to the invariance of the Hilbert-Schmidt inner product $\langle X, Y \rangle = \text{Tr}\{X^\dagger Y\}$ with respect to U , the inner product between tangent space elements $X, Y \in T_U \text{SU}(d)$ and the inner product between elements $X, Y \in \mathfrak{su}(d)$ is the same.
- [39] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/ht2m-1j91> for detailed proofs of Lemma 1, Theorem 1, and Theorem 2 as well as details of simulations for projected RGD.
- [40] S. R. White, Minimally entangled typical quantum states at finite temperature, *Phys. Rev. Lett.* **102**, 190601 (2009).
- [41] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [42] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [43] D. H. Gutman and N. Ho-Nguyen, Coordinate descent without coordinates: Tangent subspace descent on Riemannian manifolds, *Math. Oper. Res.* **48**, 127 (2023).
- [44] J. Kempe, A. Kitaev, and O. Regev, The complexity of the local Hamiltonian problem, *SIAM J. Comput.* **35**, 1070 (2006).
- [45] N. McMahon, ProjRGD_ITE, Code for paper Equating quantum imaginary time evolution, Riemannian gradient flows, and stochastic implementations (2025), https://github.com/qnla/ProjRGD_ITE.