

Riemannian approach to the Lindbladian dynamics of a locally purified tensor networkEmiliano Godinez-Ramirez ^{1,2,*} Richard M. Milbradt ^{1,†} and Christian B. Mendl ^{1,3,‡}¹*Technical University of Munich, School of Computation, Information and Technology, Boltzmannstraße 3, 85748 Garching, Germany*²*Institute for Quantum-Inspired and Quantum Optimization, Hamburg University of Technology, Hamburg, Germany*³*Technical University of Munich, Institute for Advanced Study, Lichtenbergstraße 2a, 85748 Garching, Germany*

(Received 3 February 2025; accepted 23 June 2025; published 28 July 2025)

Tensor networks offer a valuable framework for implementing Lindbladian dynamics in many-body open quantum systems with nearest-neighbor couplings. In particular, a tensor network *Ansatz* known as the locally purified density operator employs the local purification of the density matrix to guarantee the positivity of the state at all times. Within this framework, the dissipative evolution utilizes the Trotter-Suzuki splitting, yielding a second-order approximation error. However, due to the Lindbladian dynamics' nature, employing higher-order schemes results in nonphysical quantum channels. In this work, we leverage the gauge freedom inherent in the Kraus representation of quantum channels to improve the splitting error. To this end, we formulate an optimization problem on the Riemannian manifold of isometries and find a solution via the second-order trust-region algorithm. We validate our approach using two nearest-neighbor noise models and achieve an improvement of orders of magnitude compared to other positivity-preserving schemes. In addition, we demonstrate the usefulness of our method as a compression scheme, helping to control the exponential growth of computational resources, which thus far has limited the use of the locally purified *Ansatz*.

DOI: [10.1103/kztw-ywxr](https://doi.org/10.1103/kztw-ywxr)**I. INTRODUCTION**

The study of many-body quantum systems is of utmost importance in various branches of physics, including quantum information and quantum computing. Moreover, to understand realistic physical systems, we ought to consider how these interact with their environment, i.e., we need to understand how they evolve under the influence of noise. This shifts our focus to what is known as open quantum systems [1]. In particular, we are interested in the evolution under Markovian noise, giving rise to the Lindblad master equation [1,2]

Due to the limited analytical methods available to describe such systems, numerical methods play a pivotal role in understanding the so-called *dissipative dynamics*. In this context, tensor networks variational *Ansätze* emerge as an efficient approach capable of describing several systems of interest [3]. One such structure, known as the locally purified density operator (LPDO), exploits the main features of tensor networks *Ansätze* while preserving the physical properties of open quantum systems [4,5]. Werner *et al.* [6] introduced an algorithm yielding second-order approximation errors based on the ubiquitous Trotter-Suzuki splitting [7] to simulate the

Lindblad evolution with nearest-neighbor interactions using the LPDO *Ansatz*. However, due to the semigroup property of Markovian dynamics, it remains an open problem to simulate the dissipative evolution with higher-order accuracy.

In this work, we aim to improve the accuracy of the dissipative simulations under nearest-neighbor Lindbladian dynamics. To this end, we leverage the gauge freedom present in the Kraus representation of quantum channels to formulate an optimization problem on the Riemannian manifold of isometry matrices, i.e., the Stiefel manifold. We then proceed to solve it via the second-order trust-region algorithm, taking advantage of its global convergence properties [8] (cf. [9], which applies a similar approach to gate set tomography). We cover the core concepts needed to understand the main optimization problem addressed in this work in Sec. II. Then, we leverage this theory to formulate a real-valued cost function defined on the Stiefel manifold in Sec. III.

Once we formulate the optimization problem and define the Riemannian geometry to address it, we assemble these elements into the main numerical algorithm of this work. In Sec. IV, we outline the algorithm, discuss the creation of the optimization *Ansatz*, and introduce the framework for employing our algorithm as a compression scheme. We conclude this work by benchmarking our approach against other completely positive trace-preserving (CPTP) schemes, evaluating the performance of the compressed representation, and testing the applicability of the optimized operators when used within a larger system. These simulations are presented in Sec. V. For all of them, we have employed two nearest-neighbor models, namely, the *Kitaev wire* [10] and a modification of the sparse Pauli-Lindblad model [11] we call *pseudosparsely Pauli-Lindblad* (PSPL).

*Contact author: cristian.godinez@tum.de†Contact author: r.milbradt@tum.de‡Contact author: christian.mendl@tum.de

II. BACKGROUND

A. Tensor networks and open quantum systems

In this work, our goal is to simulate the dissipative Lindbladian dynamics of a quantum system governed by the master equation

$$\frac{d\rho}{dt} = \mathcal{L}(\rho), \quad (1)$$

where the density matrix ρ describes the systems' state and the Lindbladian generator \mathcal{L} takes the form

$$\mathcal{L}(\rho) = \sum_k \left(L_k \rho L_k^\dagger - \frac{1}{2} L_k^\dagger L_k \rho - \frac{1}{2} \rho L_k^\dagger L_k \right), \quad (2)$$

with the jump operators L_k modeling the interaction between the system and its environment. The solution of Eq. (1) for a finite-dimensional system with initial state $\rho_0 := \rho(t=0)$ is obtained by

$$|\rho\rangle\rangle = \Lambda |\rho_0\rangle\rangle, \quad \Lambda = e^{t\hat{\mathcal{L}}}, \quad t \geq 0 \quad (3)$$

where $|\rho\rangle\rangle$ and $\hat{\mathcal{L}}$ are the vectorizations of ρ and \mathcal{L} , respectively. The latter vectorized superoperator can be expressed as

$$\hat{\mathcal{L}} = \sum_k \left[L_k \otimes L_k^* - \frac{1}{2} (L_k^\dagger L_k \otimes \mathbb{I} + \mathbb{I} \otimes L_k^T L_k^*) \right]. \quad (4)$$

Due to the limited analytical methods available to describe open quantum systems, numerical methods play a pivotal role in understanding the dissipative dynamics arising from Eq. (1). In this light, tensor networks variational *Ansätze* [3] appear as an efficient approach capable of describing several systems of interest, with the matrix product states (MPS) [12,13] as the archetype class used to study one-dimensional pure states.

On the other hand, one-dimensional mixed states are still a relatively unexplored frontier by tensor network techniques. A first attempt at this employs the class of matrix product density operators (MPDO) [4,14]. However, locally verifying the positivity of the global MPDO is NP-hard [15], so performing local truncations usually destroys the positive definiteness of the state, yielding numerical instabilities in time-evolution algorithms. A second approach uses the so-called unraveling of the master equation [1,16] to simulate mixed states by means of MPS methods, together with stochastic quantum trajectories. This approach comes at the cost of a sampling overhead with the number of trajectories. In this work, we will leverage a third structure introduced in [4,5], known as the locally purified density operator (LPDO).

Definition 1 (Locally purified density operator). Given a density matrix ρ , its local purification operator F satisfies

$$\rho = FF^\dagger, \quad (5)$$

and its variational tensor network representation for N sites is

$$\begin{aligned} [F]_{r_1, \dots, r_N}^{s_1, \dots, s_N} &= \sum_{m_1, \dots, m_{N-1}} A_{m_0, m_1}^{[1]s_1, r_1} A_{m_1, m_2}^{[2]s_2, r_2} \dots A_{m_{N-1}, m_N}^{[N]s_N, r_N} \\ &= \begin{array}{c} \begin{array}{ccccccc} & s_1 & & s_2 & & & s_N \\ & \downarrow & & \downarrow & & & \downarrow \\ \begin{array}{|c|} \hline A_1 \\ \hline \end{array} & \begin{array}{|c|} \hline A_2 \\ \hline \end{array} & \dots & \begin{array}{|c|} \hline A_N \\ \hline \end{array} \\ \begin{array}{|c|} \hline m_0 \\ \hline \end{array} & \begin{array}{|c|} \hline m_1 \\ \hline \end{array} & \begin{array}{|c|} \hline m_2 \\ \hline \end{array} & \dots & \begin{array}{|c|} \hline m_{N-1} \\ \hline \end{array} & \begin{array}{|c|} \hline m_N \\ \hline \end{array} \\ \begin{array}{|c|} \hline r_1 \\ \hline \end{array} & \begin{array}{|c|} \hline r_2 \\ \hline \end{array} & & & & \begin{array}{|c|} \hline r_N \\ \hline \end{array} \end{array} \end{array} \quad (6)$$

We use the following notation for the tensor legs' dimensions:

$$[n] := \{1, 2, \dots, n\}, \quad (7)$$

thus, we can write each rank-4 tensor $\{A^{[\ell]}\}_{\ell=1}^N$ as

$$A^{[\ell]} = (A_{m_{\ell-1}, m_\ell}^{[\ell]s_\ell, r_\ell})_{s_\ell \in [d], r_\ell \in [K], m_{\ell-1} \in [M_{\ell-1}], m_\ell \in [M_\ell]}, \quad (8)$$

where s_ℓ is the *physical* index, r_ℓ is the *Kraus* (or *rank*) index, and m_ℓ is the *bond* index, with dimensions d , K , and M , respectively. For uniformity, the first and last tensors have bond dimensions $m_0 = m_N = 1$. Notice how the representation of ρ in terms of tensors $\{A^{[\ell]}\}$ is highly nonunique since, in addition to the usual gauge freedom present in MPS along the bond index, this structure has a gauge freedom along the Kraus index, i.e., Eq. (5) remains invariant under a unitary transformation of F .

This work focuses on the scenario where the jump operators L_k act nontrivially only on nearest-neighbor sites ℓ and $\ell + 1$. Therefore, the Lindbladian superoperator acting on all N sites can be written as

$$\hat{\mathcal{L}} = \sum_{\ell=1}^N \hat{\mathcal{L}}^{[\ell, \ell+1]}, \quad (9)$$

where each $\hat{\mathcal{L}}^{[\ell, \ell+1]}$ is of the form shown in Eq. (4). To simulate the Markovian dynamics of one-dimensional open quantum systems, as given by Eq. (3), we employ a practical algorithm described by Werner *et al.* [6] that leverages the positivity-preserving properties of the LPDO structure. Similar to the time-evolving block decimation (TEBD) [17] scheme, for a small time step τ , this algorithm splits $e^{\tau\hat{\mathcal{L}}}$ into several Trotter-Suzuki layers [18] made up of mutually commuting terms. From Lemma 4 in [6], the Trotter-Suzuki approximation for the Markovian dynamics for nearest-neighbor couplings is

$$e^{\tau\hat{\mathcal{L}}} = e^{\tau\hat{\mathcal{L}}_o/2} e^{\tau\hat{\mathcal{L}}_e} e^{\tau\hat{\mathcal{L}}_o/2} + O(\tau^3), \quad (10)$$

where the odd and even layers are made up of mutually commuting terms

$$\hat{\mathcal{L}}_o := \sum_{\ell} \hat{\mathcal{L}}^{[2\ell-1, 2\ell]} \quad \text{and} \quad \hat{\mathcal{L}}_e := \sum_{\ell} \hat{\mathcal{L}}^{[2\ell, 2\ell+1]}. \quad (11)$$

Thus, the exponentials become

$$e^{\tau\hat{\mathcal{L}}_o} = \prod_{\ell} e^{\tau\hat{\mathcal{L}}^{[2\ell-1, 2\ell]}} \quad \text{and} \quad e^{\tau\hat{\mathcal{L}}_e} = \prod_{\ell} e^{\tau\hat{\mathcal{L}}^{[2\ell, 2\ell+1]}}. \quad (12)$$

Each superoperator can be seen as a tensor acting on the vectorization of the LPDO in Eq. (5). In other words, they are rank- $4N$ tensors acting on the sites $[1, \dots, N]$ and their

To generalize the concept of the gradient of a function defined on the Stiefel manifold, we look at the *tangent space*.

Definition 3 (Tangent space). At a certain point $X \in \text{St}(n, p)$ on the Stiefel manifold, the tangent space $T_X \text{St}(n, p)$ is a vector space containing the directions we can move in tangent to the manifold. These directions are known as the tangent vectors $Z \in T_X \text{St}(n, p)$. The tangent space can be parametrized as

$$T_X \text{St}(n, p) = \{Z \in \mathbb{R}^{n \times p} : Z^T X + X^T Z = 0\}, \quad (19)$$

i.e., $X^T Z$ is skew symmetric [20].

The Riemannian gradient for a real-valued function defined on the Stiefel manifold, $f : \text{St}(n, p) \rightarrow \mathbb{R}$, can be evaluated using the projection onto the tangent space π_T as

$$\text{grad}f(X) = \pi_T[G_X^{-1} \text{grad}\bar{f}(X)], \quad (20)$$

where G_X^{-1} is the inverse metric operator, and $\text{grad}\bar{f}(x)$ is the *euclidean gradient* of f [21]. Similarly, to generalize the directional derivative, and with it the Hessian, to a Riemannian manifold, we look at the Riemannian connection ∇ .

Definition 4 (Riemannian connection). The Riemannian connection ∇ is a map connecting the different tangent spaces across the manifold, allowing us to differentiate a vector field \hat{Z} along another vector field \hat{W} . That is, at a point $X \in \text{St}(n, p)$, it is the map

$$\begin{aligned} \nabla : T_X \text{St}(n, p) \times \mathfrak{X}(X) &\rightarrow \mathfrak{X}(X) \\ (W, \hat{Z}) &\rightarrow \nabla_W \hat{Z}, \end{aligned} \quad (21)$$

where $\mathfrak{X}(X)$ is the set of vector fields on $\text{St}(n, p)$ whose domain includes X [8]. As shown in Theorem 3.1 of [21], the form of ∇ is metric dependent. See Appendix G for details on this dependence.

Then, the Riemannian Hessian (vector product) of the function f at X is the operator $\text{Hess}f(X)$ sending a tangent vector $Z \in T_X \text{St}(n, p)$ to the tangent vector

$$\text{Hess}f(X)[Z] = \nabla_Z \text{grad}f(X), \quad (22)$$

where $\text{grad}f$ is the Riemannian gradient from Eq. (20) [8]. Analogous to the Euclidean case, we can use the Riemannian gradient and Hessian to find the next iterate in an optimization scheme. However, following a tangent vector, generally, does not keep us in $\text{St}(n, p)$. Thus, we ought to use a retraction.

Definition 5 (Retraction). This update step is realized up to first order [22] by the smooth retraction map

$$R_X : T_X \text{St}(n, p) \rightarrow \text{St}(n, p) : Z \rightarrow R_X(X + Z). \quad (23)$$

For conciseness, we define $R_X(Z) := R_X(X + Z)$. Using the retraction, we can map an optimization problem defined on the Stiefel manifold onto the tangent space using the pullback \hat{f}_X :

$$\hat{f}_X : T_X \text{St}(n, p) \rightarrow \mathbb{R} : Z \rightarrow f(R_X(Z)). \quad (24)$$

Then, we can define a model \hat{m}_X as the Taylor expansion of \hat{f}_X around the origin element 0_X in the tangent space $T_X \text{St}(n, p)$. In particular, we look at the quadratic model [8]

$$\begin{aligned} \hat{m}_X(Z) &= f(X) + \langle \text{grad}f(X), Z \rangle_X \\ &+ \frac{1}{2} \langle \text{Hess}f(X)[Z], Z \rangle_X, \quad Z \in T_X \text{St}(n, p). \end{aligned} \quad (25)$$

To minimize \hat{m}_X , and consequently f , we use the trust-region algorithm as it combines the local superlinear rate of convergence of the Newton method with global convergence properties [8]. Starting at X_β , we obtain the next optimization iterate $X_{\beta+1}$ by finding an approximate solution Z_β to the minimization subproblem

$$\begin{aligned} \min_{Z \in T_{X_\beta} \text{St}(n, p)} \hat{m}_{X_\beta}(Z) \\ = f(X_\beta) + \langle \text{grad}f(X_\beta), Z \rangle_{X_\beta} \\ + \frac{1}{2} \langle \text{Hess}f(X_\beta)[Z], Z \rangle_{X_\beta}, \quad \langle Z, Z \rangle_{X_\beta} \leq \Delta^2, \end{aligned} \quad (26)$$

where Δ is known as the *trust-region radius*. The next iterate is then $X_{\beta+1} = R_{X_\beta}(Z_\beta)$. In this work, we choose to solve the trust-region subproblem approximately by means of a computationally light algorithm known as the conjugate-gradient (CG) method (see 7.3.2 of [8] for details). To assess the quality of the model, we define the quotient

$$\rho_\beta = \frac{f(X_\beta) - f(R_{X_\beta}(Z_\beta))}{\hat{m}_{X_\beta}(0_{X_\beta}) - \hat{m}_{X_\beta}(Z_\beta)} = \frac{\hat{f}_{X_\beta}(0_{X_\beta}) - \hat{f}_{X_\beta}(Z_\beta)}{\hat{m}_{X_\beta}(0_{X_\beta}) - \hat{m}_{X_\beta}(Z_\beta)}. \quad (27)$$

Based on the value of ρ_β we identify the following cases:

(i) If $\rho_\beta \ll 1$, the model is very inaccurate: we *reject* $R_{X_\beta}(Z_\beta)$ and *reduce* Δ .

(ii) If ρ_β is small but not as drastically: we *accept* $R_{X_\beta}(Z_\beta)$ but *reduce* Δ .

(iii) If $\rho_\beta \approx 1$, the model and function have a good agreement: we *accept* $R_{X_\beta}(Z_\beta)$ and *increase* Δ .

(iv) If $\rho_\beta \gg 1$, the model is inaccurate, but (luckily) the cost function is decreasing: in this case, we can try our luck and *increase* Δ to hope for further decrease in f .

III. FORMULATION

In Sec. II A, we discussed how to describe a mixed state as a locally purified density operator (LPDO) tensor network *Ansatz*. In particular, we saw how to simulate the dissipative dynamics of nearest-neighbor Lindblad operators, which yields the second-order Trotter approximation

$$e^{\tau \hat{L}} = e^{\tau \hat{L}_o/2} e^{\tau \hat{L}_e} e^{\tau \hat{L}_o/2} + O(\tau^3).$$

However, finding higher-order Trotter approximations for Markovian dynamics remains an open problem [6]. This challenge arises from the semigroup property of Lindblad superoperators since, contrary to the unitary evolution, the negative coefficients in higher-order Trotter splittings [7,23,24] would result in nonphysical maps. In this work, we aim to address this limitation by leveraging the gauge freedom of quantum channels along with the Riemannian optimization methods on the Stiefel manifold we described in Sec. II B. Our objective is to achieve a better approximation error for the dissipative dynamics of nearest-neighbor Lindbladians while maintaining the alternating structure of the layers shown by the second-order Trotter splitting in Eq. (17).

A. From superoperators to isometries

As we saw in Eq. (13), we can transform our expressions to focus on the local superoperators $e^{\tau \hat{D}^{[\ell, \ell+1]}}$. Then, we can employ these superoperators to reformulate our problem to

fit into a Riemannian optimization scheme. Using the Kraus operators' completeness relation

$$\sum_{q=1}^R E_q^\dagger E_q = \mathbb{I}, \quad (28)$$

we can follow the procedure outlined in [25] to cast these operators into the Stiefel manifold $\text{St}(n, p)$ we introduced in Definition 2. We will label the whole superoperator-isometry transformation as S_T , and define the map

$$S_T : \mathcal{T}(\mathcal{H}_1, \mathcal{H}_2) \rightarrow \text{St}(n, p) : \Lambda \rightarrow X = (\Lambda)^{S_T}, \quad (29)$$

where $\mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$ is the space of linear maps from operators in the Hilbert space \mathcal{H}_1 , to operators in the Hilbert space \mathcal{H}_2 (see Definition A3 of the Appendix). The inverse map from isometries to superoperators

$$S_P = (S_T)^{-1} \quad (30)$$

follows immediately from this definition.

B. Defining the cost function: Approximation error

Now that our operators belong to the Stiefel manifold, we can construct a real-valued cost function f on this manifold, i.e.,

$$f : \text{St}(n, p) \rightarrow \mathbb{R}. \quad (31)$$

To this end, we look at the difference between the approximation given by the right-hand side of Eq. (10) and the full superoperator on the left-hand side. In particular, we use the computationally relevant Frobenius norm induced by the Euclidean inner product

$$\left\| e^{\tau \hat{\mathcal{L}}} - e_{\text{approx}}^{\tau \hat{\mathcal{L}}} \right\|. \quad (32)$$

It is customary to divide the time evolution up to a final time τ into smaller time steps of size Δt , such that

$$\tau = n_\tau \Delta t, \quad (33)$$

with $n_\tau \in \mathbb{N}$ the number of time steps. If we use the second-order Trotter splitting from Eq. (10) for each Δt , we arrive at the following pattern for the approximate superoperators:

$$e_{\text{approx}}^{\tau \hat{\mathcal{L}}} = \underbrace{\begin{array}{ccccccc} \square & \square & \square & \square & \dots & \square & \square & \square \\ \square & \square & \square & \square & \dots & \square & \square & \square \\ \square & \square & \square & \square & \dots & \square & \square & \square \end{array}}_m \quad (34)$$

where the *blue* blocks correspond to the terms with half-time steps, while the *red* blocks correspond to the full time steps. The number of layers m in Eq. (34) is

$$m = 3n_\tau - (n_\tau - 1) = 2n_\tau + 1. \quad (35)$$

We see the final form of the splitting with n_τ time steps shows the same odd-even structure from Eq. (10). However, due to this new partition structure, the approximation error in terms of Δt becomes

$$e^{\tau \hat{\mathcal{L}}} = e_{\text{approx}}^{\tau \hat{\mathcal{L}}} + O(\Delta t^2). \quad (36)$$

Finally, in order to arrive at the final expression of the cost function, we lay out the assumptions used:

(i) To achieve translational symmetry, we assume the operators $\hat{D}^{[\ell, \ell+1]}$, and therefore the isometries X , are the same for all $\ell \in [1, N]$.

(ii) We use the splitting for n_τ time steps to simulate the evolution up to time τ with m *distinct* layers. Contrary to the repeating structure from the original Trotter splitting, we choose all the layers in our *Ansatz* to be independent of each other to increase the degrees of freedom of the optimization in hopes of achieving a higher expressivity.

(iii) We restrict our optimization to the *real* Stiefel manifold to avoid dealing with nonholomorphic functions defined on complex manifolds.

The composition of superoperators is achieved through usual matrix multiplication. Then, by labeling each of the isometries on the α th layer as X_α , and $\{X\}_\alpha$ as the list of m isometries, we can express the cost function as

$$f(\{X\}_\alpha) = \left\| e^{\tau \hat{\mathcal{L}}} - \prod_{\alpha} \left(\bigotimes_{\ell} (X_\alpha)^{S_P} \right)^{L_G} \right\|, \quad (37)$$

where S_P is the isometry-superoperator map from Eq. (30) and L_G is the local-to-global transformation, i.e., the inverse of the global-to-local map we introduced in Eq. (13) [see Eq. (B19) for details]. We can define the vector of isometries

$$\mathbf{X} := (X_1, \dots, X_m),$$

and with it, the superoperator

$$\mathcal{S}(\mathbf{X}) = \prod_{\alpha} \left(\bigotimes_{\ell} (X_\alpha)^{S_P} \right)^{L_G} \in \mathbb{R}^{d^{2N} \times d^{2N}}$$

acting on the full system. Thus, the main optimization problem of interest in this work can be formulated as

$$\min_{\mathbf{X}} f(\mathbf{X}) = \| e^{\tau \hat{\mathcal{L}}} - \mathcal{S}(\mathbf{X}) \|. \quad (38)$$

In other words, we want to find the list of isometries that minimize the approximation error of the dissipative dynamics.

IV. ALGORITHM

So far, we have covered the underlying theory required for this work and used it to formulate a real-valued cost function defined on the Stiefel manifold. It is now time to tackle the optimization problem. The main algorithm *A Riemannian Approach to Lindbladian Dynamics* used to solve Eq. (38) is outlined in Algorithm 1. The detailed construction of the Riemannian gradient, Hessian, and retraction, as well as the choice of the Riemannian metric, are out of the scope of this work but can be found in Appendixes G and I to K. Instead, in this section, we discuss the creation of the *Ansatz* employed in our optimization scheme and introduce the degree of freedom of the Kraus tensors that allows us to use our algorithm as a compression scheme.

ALGORITHM 1. A Riemannian approach to Lindbladian dynamics.

Input :

L_k : List of Lindblad operators acting on nearest neighbors
 τ : Final time of Lindblad dissipative evolution
 n_τ : Number of time steps for Trotter splitting
 N : Number of sites in the LPDO structure
 d : Physical dimension per site in the LPDO
 R : Desired rank of Kraus tensor acting on nearest neighbors in the LPDO tensors
 I : Number of iterations for trust-region optimization

Output :

f_β : Cost function history, evaluated after every iteration β of the optimization
 $\tilde{\mathbf{X}}$: List of isometries minimizing approximation error after I iterations

Require : All $L_k \in \mathbb{R}^{d^2 \times d^2}$

- 1: $e^{\tau \hat{\mathcal{L}}} \leftarrow$ Create reference superoperator from $\{L_k\}_k$ and Eq. (9). \triangleright Assume translational symmetry for all N sites.
 - 2: $\{\Lambda_\alpha\}_\alpha \leftarrow$ Create $m = 2n_\tau + 1$ superoperators using the Trotter splitting from Eq. (10) and the decomposition in Eq. (34).
 - 3: $\mathbf{X}^0 \leftarrow$ Generate *Ansatz* of isometries from superoperators $\{\Lambda_\alpha\}_\alpha$.
 - 4: $f(\mathbf{X}^\beta, e^{\tau \hat{\mathcal{L}}}) \leftarrow$ Define cost function taking as input the exact superoperator $e^{\tau \hat{\mathcal{L}}}$ and list of isometries \mathbf{X}^β .
 - 5: $\text{grad}f \leftarrow$ Define Riemannian gradient from cost function.
 - 6: $R_X \leftarrow$ Define retraction taking a tangent space vector Z back to Stiefel manifold.
 - 7: $\text{Hess}f \leftarrow$ Define Riemannian Hessian from cost function and Riemannian gradient.
 - 8: **for** $\beta \leftarrow 0$ **to** $I - 1$
 - 9: $\mathbf{X}^{\beta+1} \leftarrow$ Perform an iteration of the trust-region
 - 10: algorithm with $(f, \text{grad}f, \text{Hess}f, R_X, \mathbf{X}^\beta)$.
 - 11: **end for**
-
-

The entirety of this work has been implemented in Python [26] and can be found open-sourced on GitHub as a package called OPENTN [27]. As a back end for manipulating the tensors we have used NUMPY [28] and JAX [29]. The latter has also been utilized due to its automatic differentiation capabilities.

A. Creating the Ansatz: Compressed factorization

As for many other gradient-based optimization algorithms, the initial optimization point, i.e., the *Ansatz*, can play an important role in the number of iterations needed in the optimization scheme. Sometimes, it can even determine whether the algorithm reaches a local minimum or not. In this work, we choose to start at the even-odd splitting given by the second-order Trotter from Eq. (10). We have seen that for n_τ time steps, this *Ansatz* results in m layers made up of local superoperators

$$\Lambda_\alpha \in \mathbb{R}^{d^4 \times d^4}, \quad \forall \alpha \in [m]. \quad (39)$$

We can then convert each superoperator into its isometry representation using the S_T map from Eq. (29). The Cholesky decomposition of the Choi matrix is a crucial step of the S_T transformation (see Appendix C for details). The degree of

freedom in the choice of the channel rank R used in this factorization is of particular importance:

$$\mathcal{C}_R = \sum_{q=1}^R E_{s^{\text{out}}, s^{\text{in}}}^q (E_{s^{\text{out}}, s^{\text{in}}}^q)^\dagger, \quad (40)$$

with $s^{\text{out}}, s^{\text{in}} \in [d^2]$, and \mathcal{C}_R the rank- R approximation of the Choi matrix \mathcal{C} . The obvious choice is to set $R = \text{rank}(\mathcal{C})$ so that $\mathcal{C}_R = \mathcal{C}$. We call this the *natural Choi rank* of the channel. However, we can arbitrarily increase this number up to $R_{\text{max}} = d^4$ or compress it down to $R_{\text{min}} = 1$. The larger the Choi rank R , the higher the expressivity of each layer in the *Ansatz*, generally resulting in a lower approximation error. However, this increase in expressivity comes at the cost of an increase in the complexity of the optimization and, consequently, the computational cost. Another crucial factor to take into account is the effect of the channel rank on the growth of the Kraus dimension in the LPDO structure, where the indices r_ℓ get contracted with the channel index q of dimension R , as shown in Fig. 1. In Sec. V A, we explore how the choice of R affects our optimization algorithm and discuss the benefits of using a compressed representation of the quantum channels.

Once we choose a rank R , we apply the S_T transformation on the list of superoperators $\{\Lambda_\alpha\}_{\alpha \in [m]}$ to obtain a list of isometries

$$\mathbf{X}^0 := (X_1^0, \dots, X_m^0), \quad X_\alpha \in \mathbb{R}^{n \times p}, \quad \forall \alpha \in [m], \quad (41)$$

where we have defined $n = R d^2$ and $p = d^2$. The superscript 0 indicates that this corresponds to the zeroth iteration of the optimization scheme. The numerical implementation of the Cholesky decomposition is discussed in Appendix C.

B. Parametrizing the tangent space

To further understand the effect of the channel rank in the optimization scheme, it is useful to look at the number of parameters present in the *Ansatz*. In other words, we want to understand how R affects the expressivity. To this end, we look at the parametrized definition of the tangent space of the Stiefel manifold. The in-detail derivation following the outline from [30] is included in Appendix E.

Definition 6 (Tangent space: revisited). Given an element $X \in \text{St}(n, p)$ of the Stiefel manifold, the tangent space $T_X \text{St}(n, p)$ at X is

$$T_X \text{St}(n, p) = \{Z \in \mathbb{R}^{n \times p} : Z = XA + X_\perp B, A \in \mathbb{R}^{p \times p}, \\ A + A^T = 0, B \in \mathbb{R}^{(n-p) \times p}\}. \quad (42)$$

If we compare Eq. (42), we realize the product $X^T Z$ in the former is exactly the skew-symmetric matrix A of the latter, for $Z \in T_X \text{St}(n, p)$.

The arbitrary matrix B in Eq. (42) is made of $p(n-p)$ independent parameters, while the skew-symmetric matrix A is completely defined by its $p(p-1)/2$ upper diagonal elements. Thus, as a by-product of the parametrized formulation, we can verify the degrees of freedom of $T_X \text{St}(n, p)$ to be

$$\frac{p(p-1)}{2} + p(n-p) = np - \frac{p(p+1)}{2}. \quad (43)$$

Then, the total degrees of freedom present in our m -layer *Ansatz* are

$$\text{DOF} = \frac{mp}{2}(2n - p - 1) = \frac{md^2}{2}(2Rd^2 - d^2 - 1). \quad (44)$$

V. SIMULATIONS

In this chapter, we study the performance of the *Riemannian Approach to Lindbladian Dynamics* as introduced in Algorithm 1. As a test bed, we focus on two models of nearest-neighbor dissipative interactions, with real-valued jump operators, as our optimization is defined on the real Stiefel manifold. Namely, at all neighboring sites $[\ell, \ell + 1]$ of an N -sites quantum system we consider the Lindblad operators given by the following models:

(1) Kitaev wire:

$$L_1 = \frac{\sqrt{\gamma}}{4}(a^\dagger \otimes \mathbb{I} + \mathbb{I} \otimes a), \quad (45)$$

where a^\dagger and a are the two-level quantum harmonic oscillator creation and annihilation operators, respectively, and \mathbb{I} is the 2×2 identity matrix, as introduced in [10].

(2) Pseudospars Pauli-Lindblad (PSPL):

$$L_k = \sqrt{\gamma}(P_k - Q_k), \quad k \in [\kappa] \quad (46)$$

where P_k, Q_k are elements of the Pauli-2 group, such that the total number of Lindblad operators $\kappa \ll 4^2 - 1$, i.e., the Lindblad terms are sparse. Compared to the original sparse Pauli-Lindblad model introduced in [11], here we have the difference of two Pauli operators instead of a single Pauli. In particular, in this work, we choose to use two Lindblad operators with weight-one Pauli terms

$$\begin{aligned} L_1 &= \sqrt{\gamma}(X \otimes \mathbb{I} - \mathbb{I} \otimes X), \\ L_2 &= \sqrt{\gamma}(Z \otimes \mathbb{I} - \mathbb{I} \otimes Z). \end{aligned} \quad (47)$$

For both models, the prefactor γ corresponds to the noise strength, which we normalize to $\gamma = 1$ over all simulations. Instead, we vary the final evolution time τ over the experiments since this is equivalent to increasing the noise strength, so we can verify the validity of our results across different noise regimes. We have implicitly assumed the local Hilbert space dimension to be $d = 2$, i.e., we work with a system of qubits. We perform the simulations using the first nontrivial even number of sites $N = 4$ yielding an effective Hilbert space dimension $d^{2N} = 2^8$, as this keeps the computation tractable. In the following sections, we test our algorithm under the two aforementioned noise models, while exploring the influence of the channel rank R , introduced in Sec. IV. Further, we compare the performance of our algorithm against the second-order Trotter splitting [6] and a structure-preserving scheme of arbitrary order introduced in [31]. To end this chapter, we leverage the translational invariance and PBC assumed in our *Ansatz* to test the optimized isometries when embedding them into larger systems.

A. Rank up

A fundamental element in our optimization scheme was introduced in Sec. IV A: the rank R of the *Ansatz*. This quantity arises in the Cholesky decomposition from Eq. (40) and determines the column dimension of the Stiefel manifold $p = Rd^2$.

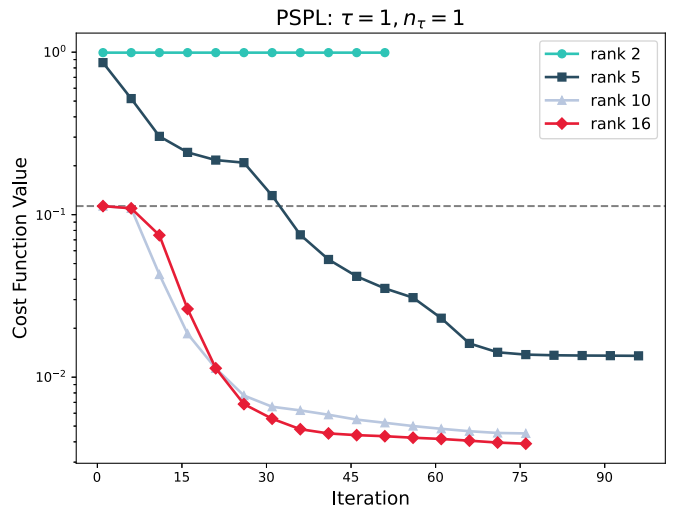


FIG. 2. Riemannian optimization with PSPL Lindblad model for a final time of $\tau = 1$ and $n_\tau = 1$ time steps. The four curves correspond to the use of different ranks for the *Ansatz*. The gray dashed line marks the value of the Trotter splitting error.

In this section, we will examine the conjectures presented in Sec. IV A, regarding the impact of the *Ansatz*'s rank on the performance of the Riemannian algorithm. To this end, we carry out the optimization scheme under the Kitaev and PSPL noise models, while choosing the *Ansatz*'s rank R as

$$R \in \{2, \dots, R_C, \dots, d^4\}, \quad (48)$$

where $R_C = \text{rank}(C)$ is the natural Choi rank, as introduced in Eq. (40). We do not use $R = 1$, as this would correspond to a unitary evolution.

We transform the local superoperator arising from the PSPL Lindblad operators in Eq. (47) into its Choi representation, to obtain the natural Choi rank of the channel

$$R_C^{\text{PSPL}} = 10.$$

In other words, the PSPL quantum channel can be described exactly with 10 Kraus operators. We then perform the Riemannian optimization for a final time of $\tau = 1$ with $n_\tau = 1$ time steps and visualize the optimization trajectory in Fig. 2 using four different ranks

$$R \in \{2, 5, 10, 16\}.$$

As expected, this comparison illustrates how the algorithm's performance improves with increasing rank. The highest ranks $R \in \{10, 16\}$ start the optimization at the Trotter splitting value, marked by the gray dashed line, and exhibit a continuous decrease with the iterations. Due to the increased parameter space, the optimization with $R = 16$ attains the lowest cost function value out of all. However, a more surprising result is yielded by the curve with $R = 5 < R_C^{\text{PSPL}}$. Since this *Ansatz* corresponds to a *compressed* representation of the channel, it initially performs worse than the higher-rank approximations, as we expected. Yet, after 30 iterations it already reaches the Trotter splitting value, and after 100 iterations, the cost function sees an improvement of nearly one order of magnitude with respect to the Trotter splitting and two orders of magnitude with respect to its *Ansatz* value.

TABLE I. Comparison of full Choi ranks for different approximations using the two toy models before and after optimization.

(a) Comparison for the PSPL model			(b) Comparison for the Kitaev model		
R	R_N	R_N^{opt}	R	R_N	R_N^{opt}
Exact	2^8		Exact	45	
2	53	53	2	45	117
5	2^8	2^8	4	45	205
10	2^8	2^8	8	45	94
16	2^8	2^8	16	45	202

This demonstrates that using this compressed representation not only enables a decrease in the approximation error of the dissipative dynamics but also helps control the exponential increase in the Kraus dimensions of the LPDO tensors after each dissipative step. Thus, we can directly employ our algorithm to alleviate the computational cost associated with the noisy evolution of the LPDO structure (cf. [32,33]).

The other end of the spectrum is shown by the turquoise curve with $R = 2$, where the improvement yielded by the optimization is negligible. To understand this behavior, we can look at the Choi rank of the superoperator acting on all N sites arising from the rank- R *Ansatz*, $\mathcal{S}^R(\mathbf{X})$. The Choi rank of $\mathcal{S}^R(\mathbf{X})$ is upper bounded as

$$R_N := \text{rank}\{(\mathcal{S}^R(\mathbf{X}))^{R_r}\} \leq R^{2m}. \quad (49)$$

To make this expression more tangible, we calculate numerically the Choi rank of the exact full superoperator $e^{\tau\hat{L}}$ and compare it to the rank of $\mathcal{S}^R(\mathbf{X})$ for each of the *Ansatz* ranks in our simulation.

Table I(a) shows both the initial rank R_N and the rank after the optimization R_N^{opt} . From this, we observe that, unlike the higher-rank approximations, the *Ansatz* with $R = 2$ reaches only $\approx \frac{1}{4}$ of the full rank of $e^{\tau\hat{L}}$, even after the attempt to optimize it. This result agrees with Eq. (49) and provides an explanation for the limited performance shown in Fig. 2.

Contrary to what we initially expected based on Fig. 2, increasing the rank of the *Ansätze* does not always lead to considerably better results. An example of this is the simulation under the Kitaev model shown in Fig. 3. For this computation, we use $\tau = 0.5$, $n_\tau = 4$, and choose R from $\{2, 4, 8, 16\}$. The first thing to notice is the natural Choi rank, which for this model is

$$R_C^{\text{Kitaev}} = 2.$$

For this reason, all the *Ansätze* in this simulation start at the Trotter splitting value. While we notice a consistent decrease in the cost function for all the curves in Fig. 3, it is also noteworthy that all of them reach a plateau at only 40 iterations, and none of them attain a cost function lower than 30% of the original value. Analogously to the computation for the PSPL model, we calculate the Choi rank of $\mathcal{S}^R(\mathbf{X})$ before and after the optimization, and show the results in Table I(b).

This comparison highlights a couple of things. First, the exact superoperator can be represented by a relatively low rank, $R_N = 45$, which is already achieved by all $\mathcal{S}^R(\mathbf{X})$ even before the optimization. This can serve as an explanation for the low approximation error at which we start the optimization scheme, hence making it harder to improve upon. Second, in

an attempt to improve on this approximation, the rank of all the *Ansätze* R_N^{opt} increases considerably. This would result in a rapid increase of the Kraus dimensions of the LPDO structure without necessarily achieving a substantial decrease in the approximation error of the dissipative dynamics.

B. Riemannian to the test

To benchmark the performance of our algorithm, we compare its scaling against the second-order Trotter splitting from Eq. (10) and the structure-preserving (SP) scheme introduced by Cao and Lu in [31]. In the latter, the authors introduce a family of unnormalized completely positive schemes of arbitrary order. To achieve a fair comparison against the SP methods, we use the normalized version of the schemes and compute the composite superoperator up to fourth order due to the computational cost of higher-order schemes. To this end, we compute the average error as

$$\mathbb{E}(\| * \| e^{\tau\hat{L}}(\rho_0) - \mathcal{S}_{n_\tau}(\rho_0)), \quad (50)$$

where \mathcal{S}_{n_τ} is the superoperator from n_τ time steps of either the normalized SP schemes, the Trotter splitting, or our Riemannian algorithm. The expectation value is taken over 500 randomly generated density matrices ρ_0 . We perform our simulations using the PSPL noise model for a final time of $\tau = 1$, *Ansatz* rank of $R = 10$, and plot the average error against the number of time steps n_τ in Fig. 4.

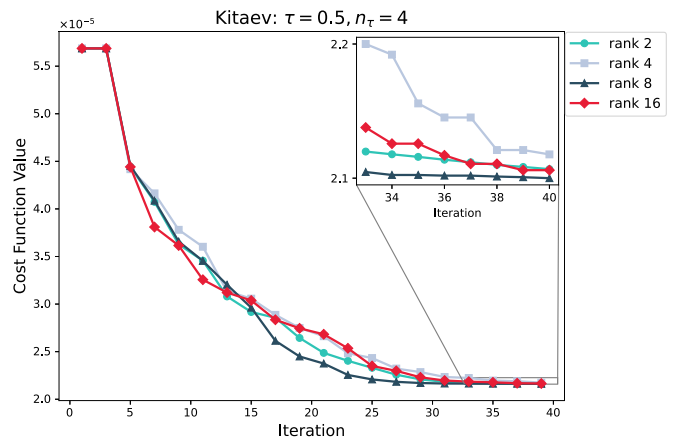


FIG. 3. Riemannian optimization with Kitaev Lindblad model for a final time of $\tau = 0.5$ and $n_\tau = 4$ time steps. The four curves correspond to the use of different ranks for the *Ansätze*. The inset shows a zoom-in version of the last 10 iterations to emphasize the difference between the *Ansätze*.

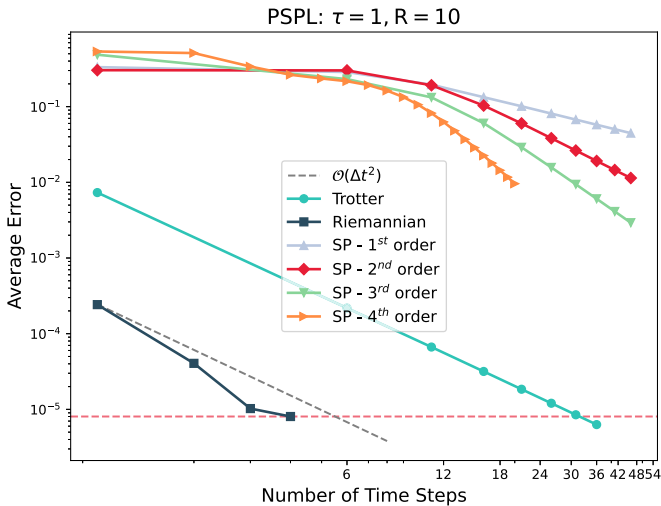


FIG. 4. Average error as a function of the number of time steps for different CPTP methods under the PSPL noise model, final time of $\tau = 1$, and *Ansatz* rank of $R = 10$. We use the normalized version of the structure-preserving schemes of orders 1, 2, 3, and 4. The red horizontal line is used to emphasize the number of time steps needed by the Trotter splitting to achieve the same accuracy as the Riemannian optimized method.

The results in Fig. 4 highlight the significant improvement attained by our optimization scheme, decreasing the error for every n_τ by more than one order of magnitude, compared to the Trotter splitting. As emphasized by the red horizontal line, the Trotter splitting would need $n_\tau > 30$ time steps to achieve the same accuracy obtained by only $n_\tau = 4$ using our method. In addition, comparing our method against the second-order curve showcases how, at some intervals, the Riemannian optimized error follows an order higher than two. This difference becomes more significant as we increase the number of time steps. We attribute this behavior to the increased expressivity of the *Ansatz* since, as shown in Eq. (44), the curve with $n_\tau = 4$ corresponds to $m = 9$ layers of isometries and 1350 degrees of freedom we can optimize over, compared to the 450 parameters available in the optimization using $n_\tau = 1$.

These results also showcase that while the structure-preserving schemes are capable of reaching approximations of arbitrarily high order, the error prefactor is larger than those for the Trotter and Riemannian methods. Moreover, the error order of these methods is only observed for a higher number of time steps. Both of these behaviors are expected and explained in detail in [31].

To end this section, we repeat the above comparison using the Kitaev model, final evolution time $\tau = 1$, and *Ansatz* rank $R = 2$. As shown in Fig. 5, in this setting, the improvement of the Riemannian optimization compared to the Trotter error is not as significant as for the PSPL model, even though at some intervals, the optimized curve again follows an order slightly higher than two. As previously discussed, we also observe that the three SP schemes employed suffer from a high error prefactor. However, the order-three SP scheme quickly catches up due to the higher approximation order. For values of n_τ close to 10, it attains an improvement of orders of magnitude com-

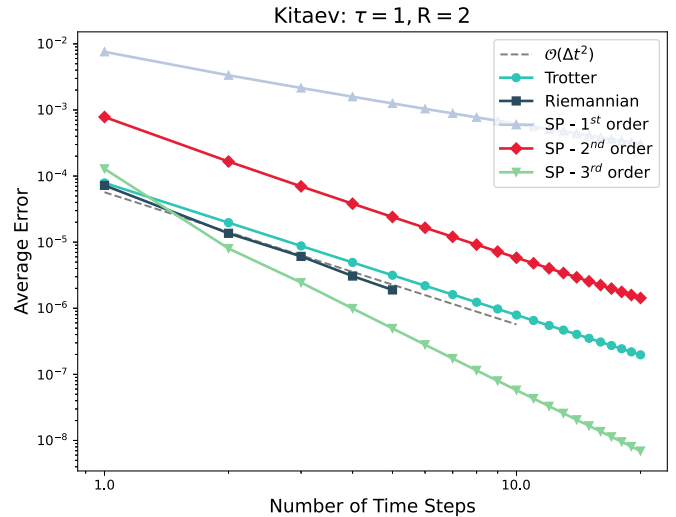


FIG. 5. Average error as a function of the number of time steps for different CPTP methods under the Kitaev model, with $\tau = 1$, and $R = 2$, plotted against increasing number of time steps. We use the normalized version of the structure-preserving schemes of orders 1, 2, and 3.

pared to the Trotter splitting. It is, however, worth highlighting that even in scenarios where the structure-preserving schemes outperform the Trotter and Riemannian approximations, they cannot be implemented directly on the LPDO *Ansatz*, as they do not preserve the local structure of the tensors.

For additional insight into the comparison between the superoperators achieved by each scheme, in Appendix D, we compare the Choi ranks R_N for each of the approximation methods using both noise models.

C. Increasing the system size

In Sec. III B, we discussed the assumptions we made when defining our cost function, one of which was the translational invariance of our *Ansatz*. Additionally, we assumed all odd layers to follow periodic boundary conditions, as shown in Eq. (34). Due to these assumptions, although we performed the optimizations for systems of size $N = 4$, it is possible to employ the optimized isometries within larger systems. In this section, we use the optimized layers from $N = 4$ to verify how well they approximate the dissipative dynamics of a system of $N = 6$ sites with an effective dimension of $d^{2N} = 2^{12}$.

For this comparison, we use the optimized isometries $\tilde{\mathbf{X}}$ obtained in Sec. V B under the Kitaev and PSPL models, using $\tau = 1$ and their natural ranks $R = 2$ and 10 , respectively. Figure 6 shows the approximation error against the number of time steps n_τ for the original $N = 4$ and the increased size $N = 6$. First, in Fig. 6(a) we examine the Kitaev model comparison, from which we observe that the error achieved by the optimized isometries within the larger system closely follows the original approximation curve. In other words, although the Trotter error increases with system size, so does the Riemannian error, resulting in very similar curves for both values of N .

On the other hand, Fig. 6(b) shows the comparison using the PSPL model, yielding somewhat different results. Here,

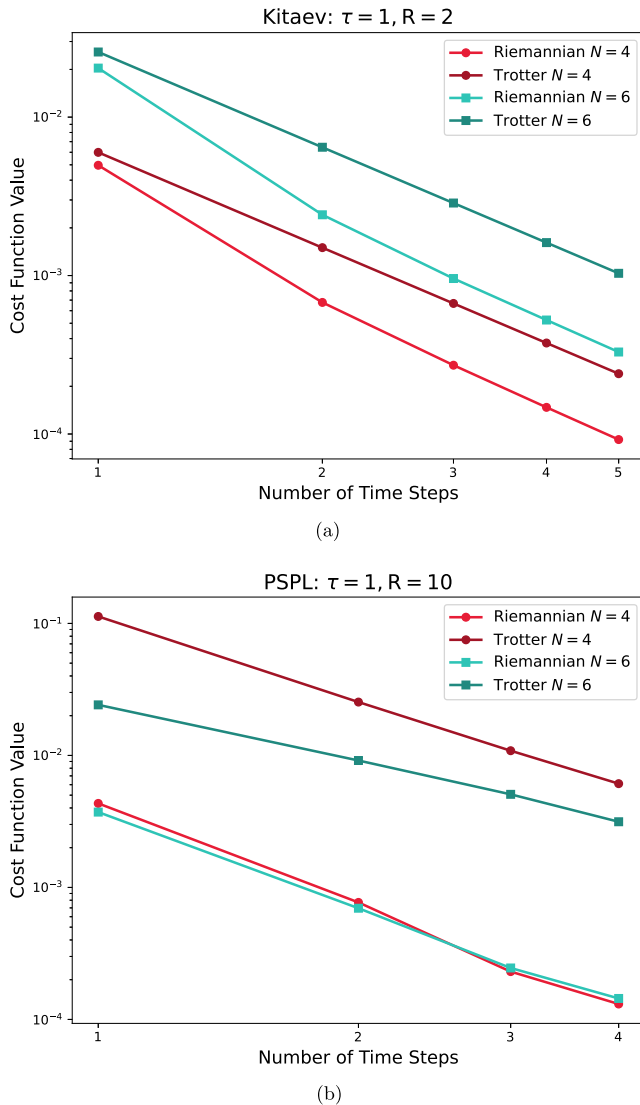


FIG. 6. Comparison of the cost function from Trotter and Riemannian schemes against the number of time steps n_τ , with evolution time $\tau = 1$, for system sizes $N = 4$ and 6 . The layers for the Riemannian scheme with $N = 6$ use the same isometries obtained with $N = 4$. (a) Shows the comparison for the Kitaev model with $R = 2$, while (b) shows the comparison using the PSPL model with $R = 10$.

we observe how the larger systems' Trotter and Riemannian errors are lower. This behavior is more pronounced for the Trotter scheme, as shown by the difference among the darker curves. However, for both methods with $N = 6$, we observe that as the number of time steps increases, the slope of the error decreases. From this observation, we can infer that the error for the system with $N = 4$ will be lower for higher n_τ . Despite this difference, the trajectory for both system sizes follows a similar trend, indicating an improvement of orders of magnitude regardless of the system size.

VI. CONCLUSION AND OUTLOOK

In this work, we employed a framework based on the locally purified density operator tensor network *Ansatz* to

formulate the approximation error of the nearest-neighbor Lindbladian dynamics as an optimization problem on the Stiefel manifold. We showed that our method is capable of achieving an improvement of orders of magnitude with respect to other known schemes. In addition, this formulation allowed us to devise a compressed representation of the quantum channels. This result alone is of particular interest as it would allow us to alleviate the exponential increase in the Kraus index of the LPDO when performing noisy evolution while still improving the approximation error. Finally, we also showed that our method can improve the approximation error of larger system sizes by optimizing computationally tractable systems. This is especially relevant as the dimension of the Lindblad superoperators grows as d^{2N} with the system size N .

A potential area for improvement is the wall time needed by our algorithm, with the most computationally expensive step being the construction of the Riemannian Hessian. To address this, we can implement a more efficient Hessian-vector product computation that does not need to construct the full Hessian. We leave this task for future work. Conversely, one might argue that the wall time of our algorithm should not be considered a strict constraint. This is especially true for a time-independent noise model, where our scheme can be seen as a single preprocessing step, entailing a *nonscaling* overhead.

It remains an open task to study the effect of the specific *Ansatz* in the optimization algorithm, as discussed in [34], since this might influence the complexity of the optimization landscape and, therefore, affect the convergence point and speed. A similar open issue is to develop a rigorous mathematical understanding of the dependence of the algorithm's performance on the properties of the noise model. In this light, studying other physically meaningful noise models with high rank would provide further insight.

Our numerical implementation of the Riemannian gradient and Hessian is based on an automatic differentiation software framework, providing us great flexibility when defining the cost function. Therefore, generalizing the *Ansatz* to a nontranslational invariant *Ansatz* would be relatively straightforward. Finally, note that we considered only the real Stiefel manifold. This restricts the optimization to real Kraus channels. Thus, a further generalization of this work could encompass the extension to the complex Stiefel manifold. This would require recounting the tangent space dimensions and the correct formulation of the nonholomorphic complex derivatives.

ACKNOWLEDGMENTS

We would like to thank F. Fröhler for the discussions regarding the numerical implementation of the structure-preserving schemes. The research is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus.

DATA AVAILABILITY

The data that support the findings of this article are openly available [27].

and

$$\begin{aligned} \mathcal{D}_{2,1}^{k[\ell,\ell+1]*} &= L_k^*, \\ \mathcal{D}_{2,2}^{k[\ell,\ell+1]*} &= \frac{1}{\sqrt{2}} L_k^T L_k^*, \\ \mathcal{D}_{2,3}^{k[\ell,\ell+1]*} &= \frac{1}{\sqrt{2}} \mathbb{I}, \end{aligned} \quad (\text{B8})$$

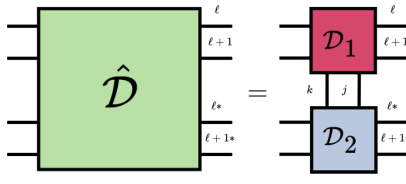
respectively. Using these tensors, we arrive at the expression

$$\hat{\mathcal{D}}_k^{[\ell,\ell+1]} = \sum_j \mathcal{D}_{1,j}^{k[\ell,\ell+1]} \otimes \mathcal{D}_{2,j}^{k[\ell,\ell+1]*}. \quad (\text{B9})$$

By summing over all k we obtain

$$\hat{\mathcal{D}}^{[\ell,\ell+1]} = \sum_k \mathcal{D}_k^{[\ell,\ell+1]} = \sum_{j,k} \mathcal{D}_{1,j}^{k[\ell,\ell+1]} \otimes \mathcal{D}_{2,j}^{k[\ell,\ell+1]*}, \quad (\text{B10})$$

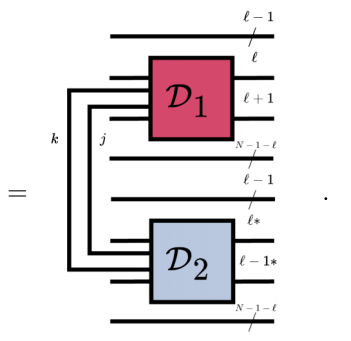
with the tensor diagram representation



$$(\text{B11})$$

Using these relations, we can rewrite the global superoperator $\hat{\mathcal{L}}^{[\ell,\ell+1]}$ as a tensor product of the form

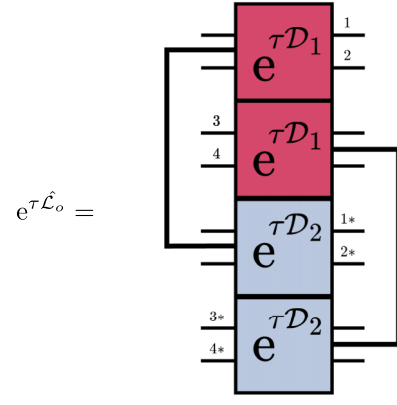
$$\begin{aligned} \hat{\mathcal{L}}^{[\ell,\ell+1]} &= \sum_{j,k} \mathbb{I}^{\otimes \ell-1} \otimes \mathcal{D}_{1,j}^{k[\ell,\ell+1]} \otimes \mathbb{I}^{\otimes N-1-\ell} \\ &\quad \otimes \mathbb{I}^{\otimes \ell-1} \otimes \mathcal{D}_{2,j}^{k[\ell,\ell+1]*} \otimes \mathbb{I}^{\otimes N-1-\ell} \end{aligned}$$



$$(\text{B12})$$

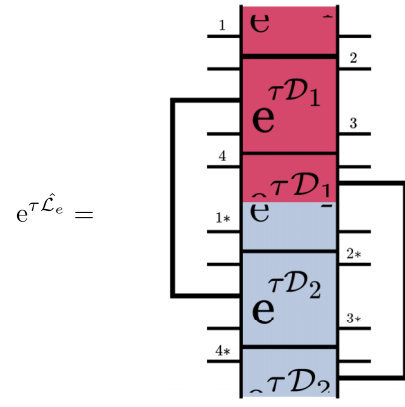
Each of the terms in the odd and even layers are of this form and, therefore, commute. This allows us to derive an expression for the odd $e^{\tau \hat{\mathcal{L}}_o}$ and even $e^{\tau \hat{\mathcal{L}}_e}$ exponentials. Namely, using Eqs. (B10) and (B12) we can depict the tensor diagrams

for $N = 4$ as



$$e^{\tau \hat{\mathcal{L}}_o} = \quad (\text{B13})$$

and

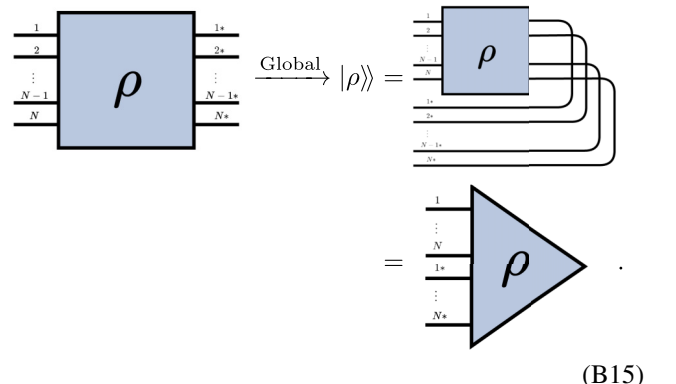


$$e^{\tau \hat{\mathcal{L}}_e} = \quad (\text{B14})$$

where the single thicker leg going from $e^{\tau \mathcal{D}_1}$ to $e^{\tau \mathcal{D}_2}$ is the channel index q shown in Eq. (16). We have used Eq. (B4) for the PBC in $e^{\tau \hat{\mathcal{L}}_e}$, i.e., the term acting on sites 4 and 1 is shown as having half the operator on each site. A reshuffle of the legs, which we call *global-to-local* transformation, allows us to bring $\mathcal{D}_1^{[\ell,\ell+1]}$ and $\mathcal{D}_2^{[\ell,\ell+1]*}$ next to each other (see Definition B1). This yields the expressions in Eq. (14):

$$e^{\tau \hat{\mathcal{L}}_{o,L}} = \bigotimes_{\ell} e^{\tau \hat{\mathcal{D}}^{[2\ell-1,2\ell]}} \quad \text{and} \quad e^{\tau \hat{\mathcal{L}}_{e,L}} = \bigotimes_{\ell} e^{\tau \hat{\mathcal{D}}^{[2\ell,2\ell+1]}}.$$

The standard ordering of the legs like the one in $\hat{\mathcal{L}}^{[\ell,\ell+1]}$ is based on the *global* vectorization of the N -sites density matrix



$$(\text{B15})$$

In other words, we have sites $[1, \dots, N]$ followed by their conjugated counterpart $[1^*, \dots, N^*]$, as shown in Eq. (B6). On the other hand, Eq. (14) is based on a *local* vectorization

where every two sites $[\ell, \ell + 1]$ are followed by their conjugate versions $[\ell^*, \ell + 1^*]$. The local vectorization of the density matrix looks like

$$\text{Local} \rightarrow |\rho\rangle\rangle = \rho \quad (\text{B16})$$

We can define a transformation that converts between these two vectorizations.

Definition B1 (Global to local). Given a multipartite superoperator $\Lambda_G \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2 \dots \otimes \mathcal{H}_N \otimes \mathcal{H}_{1^*} \otimes \mathcal{H}_{2^*} \otimes \dots \otimes \mathcal{H}_{N^*})$, the global-to-local transformation G_L is defined as the map

$$\begin{aligned} G_L : \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2 \dots \otimes \mathcal{H}_N \otimes \mathcal{H}_{1^*} \otimes \mathcal{H}_{2^*} \otimes \dots \otimes \mathcal{H}_{N^*}) \\ \rightarrow \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2 \dots \otimes \mathcal{H}_{N-1} \otimes \mathcal{H}_N \otimes \mathcal{H}_{N-1^*} \otimes \mathcal{H}_{N^*}) \\ : \Lambda_G \rightarrow \Lambda_L = (\Lambda_G)^{G_L}. \end{aligned} \quad (\text{B17})$$

Applying this transformation to a superoperator Λ is depicted as

$$\left(\Lambda_G \right)^{G_L} = \Lambda_L \quad (\text{B18})$$

The inverse map follows trivially, and we call it the *local-to-global* transformation

$$L_G : \Lambda_L \rightarrow \Lambda_G = (\Lambda_L)^{L_G}. \quad (\text{B19})$$

Remark B1. Notice that the localized version of the superoperator arising from the global-to-local operation is defined in this way because the operators L_k act on two sites. A similar construction would be obtained if instead each operator acted on m sites. In this case, we would have sites $[\ell, \dots, \ell + m - 1]$ followed by their conjugated version $[\ell^*, \dots, \ell + m - 1^*]$.

The local vectorization is the natural ordering that arises from the tensor product in Eq. (14) where each local superoperator looks like

$$e^{\tau \hat{D}}^{[\ell, \ell+1]} = e^{\tau \hat{D}} \quad (\text{B20})$$

Therefore, if we tensor these operators across all sites, we obtain the alternating pattern of the legs we see in Eq. (B16).

For $N = 4$, the localized odd-layer diagram is

$$e^{\tau \hat{L}_{o,L}} = \begin{matrix} \begin{matrix} 1 \\ 2 \\ 1^* \\ 2^* \end{matrix} & \begin{matrix} 1 \\ 2 \\ 1^* \\ 2^* \end{matrix} \\ \begin{matrix} 3 \\ 4 \\ 3^* \\ 4^* \end{matrix} & \begin{matrix} 3 \\ 4 \\ 3^* \\ 4^* \end{matrix} \end{matrix} \quad (\text{B21})$$

The even layer follows a similar pattern by taking care of the PBC.

APPENDIX C: NUMERICAL CHOLESKY FACTORIZATION

As explained in Sec. IV A, we factorize the Choi matrix corresponding to the Lindbladian quantum channels by employing the Cholesky decomposition [35] of a positive-semidefinite matrix \mathcal{C} into its factors \mathcal{F} and \mathcal{F}^\dagger :

$$\mathcal{C} = \mathcal{F} \mathcal{F}^\dagger. \quad (\text{C1})$$

In practice, we achieve the Cholesky factorization through the truncated SVD decomposition

$$\mathcal{C}_R = V D V^\dagger = V \sqrt{D} \sqrt{D} V^\dagger = V' V'^\dagger, \quad (\text{C2})$$

where V and V' are $d^4 \times R$ isometries, D is a $R \times R$ diagonal matrix, and \sqrt{D} is its elementwise square root. The SVD scheme is preferred in this work over the direct Cholesky decomposition implemented in NUMPY, as the latter suffers from numerical instabilities when the Choi matrix has negative eigenvalues that are zero up to numerical precision.

APPENDIX D: COMPARISON OF RANKS FOR DIFFERENT APPROXIMATION SCHEMES

In this Appendix, we present additional data demonstrating the increase in the Choi rank R_N of the superoperator acting on all $N = 4$ sites as the number of time steps n_τ increases. We provide a comparison of the ranks obtained using the Trotter, Riemannian, and SP schemes, which were introduced in Sec. V B. The comparison for the Kitaev model with $\tau = 1$ and $R = 2$ is shown in Table II.

Analogously, in Table III, we show the comparison of the ranks when using the PSPL model with $\tau = 1$ and $R = 10$.

APPENDIX E: PARAMETRIZING THE TANGENT SPACE

The choice of the Riemannian metric, inducing an inner product on the tangent space, determines the form of the Riemannian elements. Thus, it might impact the performance of the optimization algorithm. Therefore, in the next section, we discuss the use of different tangent space metrics. To this end, let us start this Appendix by following the procedure outlined in [30] to derive parametrized definitions for the tangent and normal spaces of the Stiefel manifold that will come in handy both for the metric discussions and the numerical implementations. We start by looking at the projection operator $\pi : \mathbb{R}^{n \times p} \rightarrow \text{St}(n, p)$. This operation acts on an arbitrary

TABLE II. Comparison of full Choi ranks for different approximation schemes using the Kitaev model for increasing number of time steps.

n_τ	Trotter	Riemannian	SP first order	SP second order	SP third order
1	36	46	5	19	29
2	45	68	19	45	45
3	45	87	45	45	45
4	45	130	45	45	45
5	45	160	45	45	45

rank- p matrix $Y \in \mathbb{R}^{n \times p}$ and projects it onto the closest point on $\text{St}(n, p)$,

$$\pi(Y) = \arg \min_{X \in \text{St}(n, p)} \|* \|Y - X\|^2, \tag{E1}$$

where $\|* \|$ is the Frobenius norm induced by the Euclidean inner product. Let us now look at the perturbed point $\pi(X + \epsilon Y)$, for $X \in \text{St}(n, p)$, $Y \in \mathbb{R}^{n \times p}$, and $\epsilon \in \mathbb{R}$ a scalar. Since the Stiefel manifold does not fill the entire $\mathbb{R}^{n \times p}$ space, there are some directions Y that satisfy

$$\pi(X + \epsilon Y) = X + O(\epsilon^2). \tag{E2}$$

In other words, moving in the direction of Y does not take $\pi(X + \epsilon Y)$ away from X . We recognize the collection of such directions Y as the *normal space*. Using this practical definition, the tangent space can be defined as the orthogonal complement of the normal space. Now, to find the concrete parametrization of these spaces, let us first define the column orthogonal complement of the isometry $X \in \text{St}(n, p)$ as the matrix $X_\perp \in \mathbb{R}^{n \times (n-p)}$, satisfying

$$[X X_\perp]^T [X X_\perp] = \mathbb{I}_n, \tag{E3}$$

$$X X^T + X_\perp X_\perp^T = \mathbb{I}_n, \tag{E4}$$

where \mathbb{I}_n is the $n \times n$ identity matrix and the side-by-side stacking of matrices $[AB]$ is a block matrix yielding a row vector with two matrix elements (A, B) . Analogously, the column vector $[AB]^T$ arises from the vertical stacking of the matrices. We can use these relations to define the decomposition of an arbitrary matrix $Y \in \mathbb{R}^{n \times p}$,

$$Y = XA + X_\perp B + XC, \tag{E5}$$

in terms of a skew-symmetric matrix $A \in \mathbb{R}^{p \times p}$, a symmetric matrix $C \in \mathbb{R}^{p \times p}$, and an arbitrary matrix $B \in \mathbb{R}^{(n-p) \times p}$. This important decomposition is used in Lemma 8 of [30] to prove the relation

$$\pi(X + \epsilon Y) = X + \epsilon(XA + X_\perp B) + O(\epsilon^2). \tag{E6}$$

It then follows that using $Y = XC$ we get $\pi(X + \epsilon XC) = X + O(\epsilon^2)$, and for a small enough $|\epsilon| > 0$, this yields

$$\pi(X + \epsilon XC) = X. \tag{E7}$$

We recognize this expression as the definition of the normal space.

Definition E1 (Normal space). Given an element $X \in \text{St}(n, p)$ of the Stiefel manifold, the normal space $N_X \text{St}(n, p)$ at X is

$$N_X \text{St}(n, p) = \{N \in \mathbb{R}^{n \times p} : N = XC, C \in \mathbb{R}^{p \times p}, C = C^T\}. \tag{E8}$$

Inspired by Eq. (E6), we can infer the form of the tangent space vectors to be $XA + X_\perp B$, which yields the parametrized definition of the tangent space.

Definition E2 (Tangent space: revisited). Given an element $X \in \text{St}(n, p)$ of the Stiefel manifold, the tangent space $T_X \text{St}(n, p)$ at X is

$$T_X \text{St}(n, p) = \{Z \in \mathbb{R}^{n \times p} : Z = XA + X_\perp B, A \in \mathbb{R}^{p \times p}, A + A^T = 0, B \in \mathbb{R}^{(n-p) \times p}\}. \tag{E9}$$

If we compare Eqs. (19) and (E9), we realize the product $X^T Z$ in the former is exactly the skew-symmetric matrix A of the latter, for $Z \in T_X \text{St}(n, p)$.

It is straightforward to verify that $XA + X_\perp B$ is orthogonal to XC under the Euclidean inner product

$$(XA + X_\perp B, XC)_e = \text{tr}\{(XA + X_\perp B)^T XC\} = 0, \tag{E10}$$

where we used $\text{tr}(A^T C) = 0$. From the revisited definitions

TABLE III. Comparison of full Choi ranks for different approximation schemes using the PSPL model for increasing number of time steps.

n_τ	Trotter	Riemannian	SP first order	SP second order	SP third order	SP fourth order
1	256	256	7	38	96	190
2	256	256	38	241	256	256
3	256	256	145	256	256	256
4	256	256	241	256	256	256
5	256	256	256	256	256	256

of the tangent and normal space, we can verify that at $X \in \text{St}(n, p)$

$$\pi_T(Y) = (I - XX^T)Y + X \text{skew}(X^T Y) \quad (\text{E11})$$

is the projection of an arbitrary matrix $Y \in \mathbb{R}^{n \times p}$ onto the tangent space with the skew operator

$$\text{skew}(M) = \frac{1}{2}(M - M^T). \quad (\text{E12})$$

The first term comes from using Eqs. (E3)–(E5) as

$$(I - XX^T)Y = X_{\perp} X_{\perp}^T Y = X_{\perp} B. \quad (\text{E13})$$

The second term uses Eqs. (E3) and (E5) to obtain $X^T Y = A + C$, followed by the skew operator defined in Eq. (E12) to “pick” the skew-symmetric factor

$$X \text{skew}(X^T Y) = XA. \quad (\text{E14})$$

Analogously, the projection onto the normal space is

$$\pi_N(Y) = X \text{sym}(X^T Y) = \frac{X}{2}(Y^T X + X^T Y) \quad (\text{E15})$$

using the symmetric operator

$$\text{sym}A = \frac{1}{2}(A + A^T) \quad (\text{E16})$$

that picks the symmetric factor $\text{sym}(X^T Y) = C$. Finally, we multiply by X , yielding the parametrization in Eq. (E8). We can simplify Eq. (E11) to obtain a relation between π_N and π_T :

$$\pi_T(Y) = Y - \frac{X}{2}(Y^T X + X^T Y) = Y - \pi_N(Y). \quad (\text{E17})$$

APPENDIX F: RIEMANNIAN CONNECTION AND METRIC

As we saw in Eq. (22), the Riemannian connection is a fundamental element in our second-order optimization scheme. Here, we describe the relation between the Riemannian connection ∇ and the Riemannian metric G_X . Let $\bar{\nabla}$ and ∇ be the Riemannian connections of the Euclidean space \mathcal{E} and the Stiefel manifold $\text{St}(n, p)$, respectively. By proposition 5.3.2 of [8] we obtain the relation

$$\nabla_W \hat{Z} = \pi_T(\bar{\nabla}_W \hat{Z}), \quad (\text{F1})$$

for all $W \in T_X \text{St}(n, p)$ and $\hat{Z} \in \mathfrak{X}(X)$. The ambient Riemannian connection $\bar{\nabla}$ is given in Theorem 3.1 of [21] as

$$\bar{\nabla}_W \hat{Z} := D\hat{Z}(X)[W] + G_X^{-1} \mathcal{K}(W, \hat{Z}) \quad (\text{F2})$$

in terms of the differential map (a classical directional derivative)

$$D\hat{Z}(X)[W] = \lim_{t \rightarrow 0} \frac{\hat{Z}(X + tW) - \hat{Z}(X)}{t}, \quad (\text{F3})$$

and the Christoffel metric term $\mathcal{K} \in \mathcal{E}$

$$\mathcal{K}(W, \hat{Z}) := \frac{1}{2}[(DG_X[W])\hat{Z} + (DG_X[\hat{Z}]W - \mathcal{X}(W, \hat{Z})]. \quad (\text{F4})$$

The cross term $\mathcal{X}(W, \hat{Z}) \in \mathcal{E}$ in this expression is a bilinear form such that for any vector field \hat{Z}_0

$$\langle \mathcal{X}(W, \hat{Z}), \hat{Z}_0 \rangle_e = \langle W, (DG_X[\hat{Z}_0])\hat{Z} \rangle_e. \quad (\text{F5})$$

APPENDIX G: CHOOSING A RIEMANNIAN METRIC

Let us employ the parametrization and dimension counting tools we have just described in Appendix E for the discussion of the Riemannian metric choice. The Euclidean metric $\langle \cdot, \cdot \rangle_e$ we have used thus far seems like the natural choice for a submanifold of the Euclidean space. However, it turns out this inner product does not weigh the degrees of freedom of the Stiefel tangent space equally. To see this, let us look at the Euclidean inner product of $Z \in T_X \text{St}(n, p)$:

$$\langle Z, Z \rangle_e = \text{tr}A^T A + \text{tr}B^T B = 2 \sum_{i < j} a_{ij}^2 + \sum_{i, j} b_{ij}^2, \quad (\text{G1})$$

where we have summed only over the upper diagonal for A . We can see from this equation that the Euclidean metric counts the degrees of freedom of A twice as much as the ones for B . As shown in [20], an alternative metric is the *canonical metric* defined for $Z, W \in T_X \text{St}(n, p)$ as

$$\langle Z, W \rangle_c = \text{tr}Z^T (\mathbb{I} - \frac{1}{2}XX^T)W. \quad (\text{G2})$$

Using this metric, the independent coordinates of A and B are weighted equally at every $X \in \text{St}(n, p)$:

$$\langle Z, Z \rangle_c = \frac{1}{2} \text{tr}A^T A + \text{tr}B^T B = \sum_{i < j} a_{ij}^2 + \sum_{i, j} b_{ij}^2. \quad (\text{G3})$$

So far, we have studied the Euclidean and canonical metrics. However, it turns out these are not the only inner products that induce a Riemannian metric on the Stiefel manifold. In [21], a general formulation for the action of the ambient metric G_X at $X \in \text{St}(n, p)$ is given by

$$G_X Z = \alpha_0(\mathbb{I} - XX^T)Z + \alpha_1 XX^T Z. \quad (\text{G4})$$

For $\alpha_0, \alpha_1 > 0$, this metric induces an inner product in $T_X \text{St}(n, p)$, and hence it induces a Riemannian metric in $\text{St}(n, p)$. The inverse matrix is

$$G_X^{-1} Z = \frac{1}{\alpha_0}(\mathbb{I} - XX^T)Z + \frac{1}{\alpha_1} XX^T Z. \quad (\text{G5})$$

Then, we conclude the metrics described thus far are special cases of the general formulation in Eq. (G4), with $\alpha_0 = \alpha_1 = 1$ and $\alpha_0 = 1, \alpha_1 = \frac{1}{2}$ for the Euclidean and the canonical metric, respectively. As anticipated, the specific form of the gradient depends on the metric of choice. Using Eqs. (E11) and (G5) we can rewrite the Riemannian gradient expression in Eq. (20) as

$$\begin{aligned} \text{grad}f(x) &= \pi_T(G^{-1} \text{grad}\bar{f}(x)) \\ &= \frac{\text{grad}\bar{f}(x)}{\alpha_0} + \frac{\alpha_1^{-1} - 2\alpha_0^{-1}}{2} XX^T \text{grad}\bar{f}(x) \\ &\quad - \frac{X \text{grad}\bar{f}(x)^T X}{2\alpha_1}. \end{aligned} \quad (\text{G6})$$

As shown in Eq. (F1), the metric choice also affects the Riemannian connection and, with it, the Riemannian Hessian calculation. We can use Eq. (G4) to obtain the cross term in Eq. (F4) for a point $X \in \text{St}(n, p)$ and tangent vectors $Z, W \in T_X \text{St}(n, p)$ [21]:

$$\mathcal{X}(Z, W) = (\alpha_1 - \alpha_0)(ZW^T + WZ^T)X. \quad (\text{G7})$$

Then, using Eqs. (F1), (F2), and (F4), the general connection for the Stiefel manifold is given by [21]

$$\begin{aligned} \nabla_Z W = & DW[Z] + \frac{1}{2}X(W^T Z + Z^T W) \\ & + \frac{\alpha_0 - \alpha_1}{\alpha_0}(\mathbb{I} - XX^T)(WZ^T + ZW^T)X. \end{aligned} \quad (\text{G8})$$

APPENDIX H: PRODUCT OF ISOMETRY MANIFOLDS

In Eq. (38), we established the main cost function we seek to minimize in this work, namely,

$$f(\mathbf{X}) = \|\ast\|e^{\tau\hat{L}} - \mathcal{S}(\mathbf{X}).$$

Since this function depends on m isometries, its domain is not a single Stiefel manifold, but rather a product manifold: $\text{St}(n_1, p_1) \times \text{St}(n_2, p_2) \dots \times \text{St}(n_m, p_m)$. We will assume all the individual manifolds are of the same dimensions and label the Cartesian product as $\text{St}(n, p)^{\times m}$. We can then define the cost function f as the map acting on this product manifold:

$$f : \text{St}(n, p)^{\times m} \rightarrow \mathbb{R} : f(\mathbf{X}) \rightarrow \|\ast\|e^{\tau\hat{L}} - \mathcal{S}(\mathbf{X}), \quad (\text{H1})$$

with $\mathbf{X} \in \text{St}(n, p)^{\times m}$. The tangent space of a product manifold is, in general, the Cartesian product of the individual tangent spaces, but for our finite-dimensional tensors, it reduces to the direct sum of them [36]. We will see what this means in practice for our Riemannian gradient, Hessian, and retraction.

APPENDIX I: AN APPROPRIATE RETRACTION

Back in Definition 5, we introduced another of the cornerstone concepts we employ in our optimization algorithm: the retraction R_X . This is used within the trust-region algorithm to define the first-order model in Eq. (25) and to generate the next candidate at the β th step in the optimization algorithm $X^{\beta+1} = R_{X^\beta}(Z)$, with Z the solution of the inner iteration in Eq. (26). As discussed in Appendix H, the tangent space $T_X \text{St}(n, p)^{\times m}$ of the product manifold $\text{St}(n, p)^{\times m}$ results from the direct sum of the individual tangent spaces. Therefore, to implement a retraction from $Z \in T_X \text{St}(n, p)^{\times m}$ onto $\mathbf{X} \in \text{St}(n, p)^{\times m}$, all we need is to apply the retraction to each pair $(X_\alpha \in \mathbf{X}, Z_\alpha \in \mathbf{Z})$ individually. The first-order retraction for the Stiefel manifold used in this work is the *canonical generalized polar decomposition* for any rectangular matrix $M \in \mathbb{R}^{n \times p}$,

$$M = XC, \quad (\text{I1})$$

in terms of an isometry $X \in \mathbb{R}^{n \times p}$ and a positive-semidefinite matrix $C \in \mathbb{R}^{p \times p}$ [37]. Given the SVD decomposition of M ,

$$M = UDV^T,$$

the isometry factor X is

$$X = U \mathbb{I}_{n,p} V^T, \quad (\text{I2})$$

with $\mathbb{I}_{n,p}$ the first p columns of the $n \times n$ identity matrix.

APPENDIX J: BUILDING THE RIEMANNIAN GRADIENT

A central element for the trust-region method used in Algorithm 1 is the Riemannian gradient of the function f from

Eq. (H1). As mentioned in Appendix H, the tangent space of the product manifold is the direct sum of the individual tangent spaces. In practice, this means we can treat the tangent vectors Z , living on the tangent space $T_X \text{St}(n, p)$ at $X \in \text{St}(n, p)$, as independent of the rest and concatenate them to obtain $Z \in T_X \text{St}(n, p)^{\times m}$. Based on the arguments developed in Appendix G, we will employ the equitative canonical metric in this work. Consequently, the Riemannian gradient for every $X \in \mathbf{X}$ takes the form

$$\text{grad}f(X) = \text{grad}\bar{f}(x) + X\text{grad}\bar{f}(x)^T X. \quad (\text{J1})$$

The ambient gradient $\text{grad}\bar{f}(X)$ coincides with the Euclidean gradient in the above equation. As aforementioned, to obtain $\text{grad}f(\mathbf{X}) \in T_X \text{St}(n, p)^{\times m}$, we will concatenate the m individual Riemannian gradients; in other words, we will *stack* them.

After using the ambient gradients together with Eq. (J1) to obtain the Riemannian gradients, a further intermediate step is the parametrization of the tangent vectors. As we saw in Eq. (44), the tangent spaces $T_X \text{St}(n, p)$ are parametrized by

$$\text{DOF} = \frac{mp}{2}(2n - p - 1) = mnp - \frac{mp}{2}(p + 1)$$

degrees of freedom. Therefore, we can store and manipulate $mnp - \frac{mp}{2}(p + 1)$ parameters, instead of mnp . To achieve this, we use Eqs. (E13) and (E14) to obtain A and B for each $\text{grad}f(X)$, respectively, retrieving the $p(p - 1)/2$ upper triangular elements of A and stacking them on top of all $p(n - p)$ elements of B . For convenience, we reshape all the stacked parametrized gradients into a vector of length DOF.

APPENDIX K: BUILDING THE RIEMANNIAN HESSIAN

The last missing piece in our optimization algorithm is the calculation of the Riemannian Hessian. As we saw in Eq. (25), this linear mapping from $T_X \text{St}(n, p)$ to $T_X \text{St}(n, p)$ allows us to define an approximation of the cost function f , which makes it a fundamental element of the trust-region algorithm.

We have discussed the general form of the Riemannian connection of the Stiefel manifold in Eq. (G8). Using the canonical metric $G_X^{[1,1/2]}$, this becomes [21]

$$\begin{aligned} \nabla_Z W = & DW[Z] + \frac{1}{2}X(W^T Z + Z^T W) \\ & + \frac{1}{2}(\mathbb{I} - XX^T)(WZ^T + ZW^T)X. \end{aligned} \quad (\text{K1})$$

Together with the Riemannian Hessian-vector product for $Z \in T_X \text{St}(n, p)$ defined as

$$\text{Hess}f(X)[Z] = \nabla_Z \text{grad}f(X), \quad (\text{K2})$$

we obtain the relation

$$\begin{aligned} \text{Hess}f(X)[Z] = & D\text{grad}f(X)[Z] + \frac{1}{2}X[\text{grad}f(X)^T Z \\ & + Z^T \text{grad}f(X)] + \frac{1}{2}(\mathbb{I} - XX^T) \\ & \times [\text{grad}f(X)Z^T + Z\text{grad}f(X)^T]X. \end{aligned} \quad (\text{K3})$$

In order to compute the full Riemannian Hessian $\text{Hess}f(X)$, we take inspiration from the Euclidean counterpart, where each column of the Euclidean Hessian $H_f(x)_k$ is revealed by

the Hessian-vector product with the basis vectors $e_k \in \mathbb{R}^n$:

$$H_f(x)_k = H_f(x)e_k.$$

Here, we will follow a similar approach, and generalize the concept of a basis vector on \mathbb{R}^n by means of the coordinate vector fields or elementary tangent directions on the tangent space of a matrix manifold [8]. Using the canonical metric, the elementary tangent directions are [30]

$$E_{ij}^A = X(E_{ij} - E_{ji}), \quad E_{ij}, E_{ji} \in \mathbb{R}^{p \times p}, \quad (\text{K4})$$

$$E_{ij}^B = X_{\perp} E_{ij}, \quad E_{ij} \in \mathbb{R}^{(n-p) \times p}, \quad (\text{K5})$$

where E_{ij} is a unit matrix defined as one at the element (i, j) and zero everywhere else, for appropriate (i, j) index values. We notice Eq. (K4) corresponds to the antisymmetric component XA of the tangent vectors, so (i, j) should traverse the $p(p-1)/2$ upper diagonal elements. On the other hand, Eq. (K5) corresponds to the arbitrary component $X_{\perp}B$, so the indices (i, j) traverse all $p(n-p)$ values. We define the flattened index $\text{idx} \in [p(2n-p-1)/2]$, and the corresponding elementary direction $E^{[\text{idx}]}$, such that

$$E^{[\text{idx}]} = \begin{cases} E_{ij}^A & \text{if } \text{idx} \leq \frac{p(p-1)}{2}, \\ E_{ij}^B & \text{else.} \end{cases} \quad (\text{K6})$$

Thus, for a single isometry X , we obtain the relation

$$\text{Hess}f(X)_{\text{idx}} = \text{Hess}f(X)[E^{[\text{idx}]}]. \quad (\text{K7})$$

The classical directional derivative $D\text{grad}f(X)[Z]$ of the vector-valued function $\text{grad}f(X)$ is implemented through the Jacobian-vector product. Once again, we take advantage of the automatic differentiation capabilities of JAX, and in particular we employ the forward-mode differentiation function `jvp` [29].

In order to accommodate this procedure to the product manifold structure, we compute the (idx, α) element of the Riemannian Hessian $\text{Hess}f(\mathbf{X})$ using the tangent directions constructed as the m -length list:

$$\text{tangents} = [O_{n \times p}, \dots, E^{[\text{idx}]}, \dots, O_{n \times p}], \quad (\text{K8})$$

where the only nonzero matrix is the α th element $E^{[\text{idx}]}$. Then, each of the resulting elements from `JAX.jvp` is projected using the Riemannian connection in Eq. (G8) to obtain the corresponding tangent vector in $T_X \text{St}(n, p)$. We store the corresponding parametrization on $\text{Hess}f(\mathbf{X})_{\text{idx}, \alpha}$. To obtain the full Riemannian Hessian, we repeat the process for all $\alpha \in [m]$ and $\text{idx} \in [p(2n-p-1)/2]$, and reshape the Riemannian Hessian into a $\text{DOF} \times \text{DOF}$ dimensional matrix.

APPENDIX L: THE IMPORTANCE OF A GOOD PARAMETRIZATION

In Appendix E, we discussed how to parametrize the tangent spaces of the Stiefel manifold. Later, in Appendix G, we used these ideas to derive expressions that showcase the influence of the Riemannian metric in calculating the gradient and Hessian. We now explore how the choice of the metric affects the performance of the Riemannian optimization in practice.

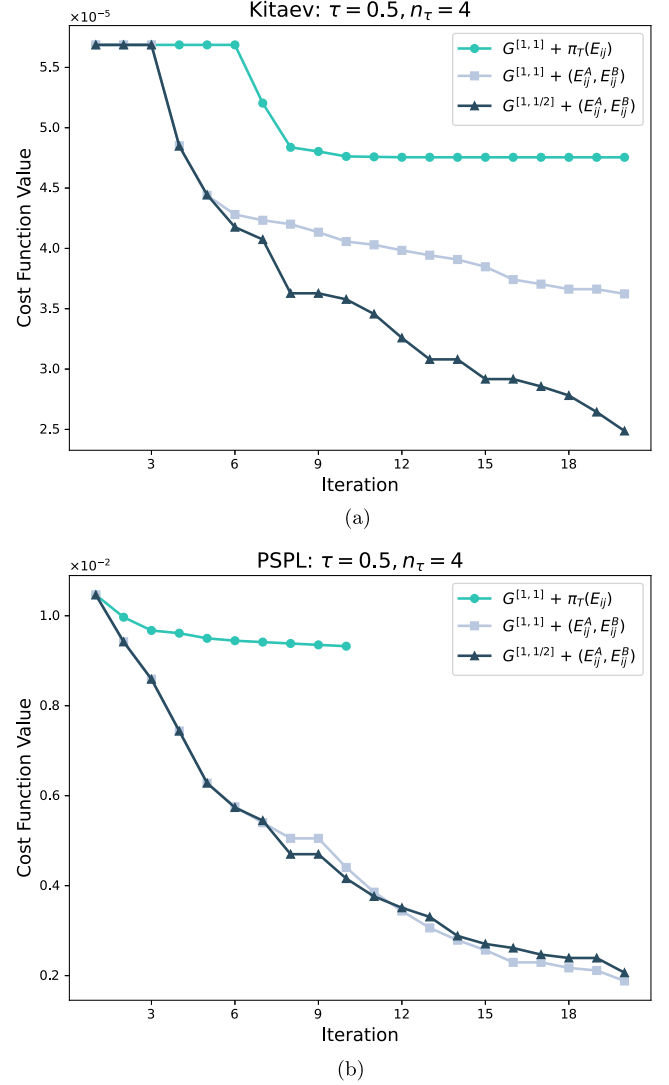


FIG. 7. Comparison of the cost function for a final evolution time of $\tau = 0.5$ and $n_{\tau} = 4$ time steps, after 20 optimization iterations using different combinations of metric G and elementary tangent directions $E_{i,j}$ (see main text). The dark blue curve shows the choice used in this work. (a) Optimization for the Kitaev model. (b) Optimization for the PSPL model. The curve with $\pi_{\tau}(E_{ij})$ is only shown for 10 time steps as it reaches a saddle point early on.

To this end, we will carry our optimization scheme for three different combinations of metric and elementary tangent directions:

- (1) Euclidean metric $G^{[1,1]}$ and a suboptimal elementary tangent direction $\pi_{\tau}(E_{ij})$, i.e., the unit matrix E_{ij} projected onto the tangent space.
- (2) Euclidean metric $G^{[1,1]}$ and the elementary tangent directions (E_{ij}^A, E_{ij}^B) introduced in Eqs. (K4) and (K5).
- (3) Canonical metric $G^{[1,1/2]}$ and elementary tangent directions (E_{ij}^A, E_{ij}^B) .

Figure 7(a) shows the value of the cost function $f(\mathbf{X}^{\beta})$ for 20 iterations under the Kitaev model with $n_{\tau} = 4$ time steps and evolution time of $\tau = 0.5$. This comparison validates the choice of the canonical metric as the Riemannian metric of the Stiefel manifold since it clearly outperforms the optimization

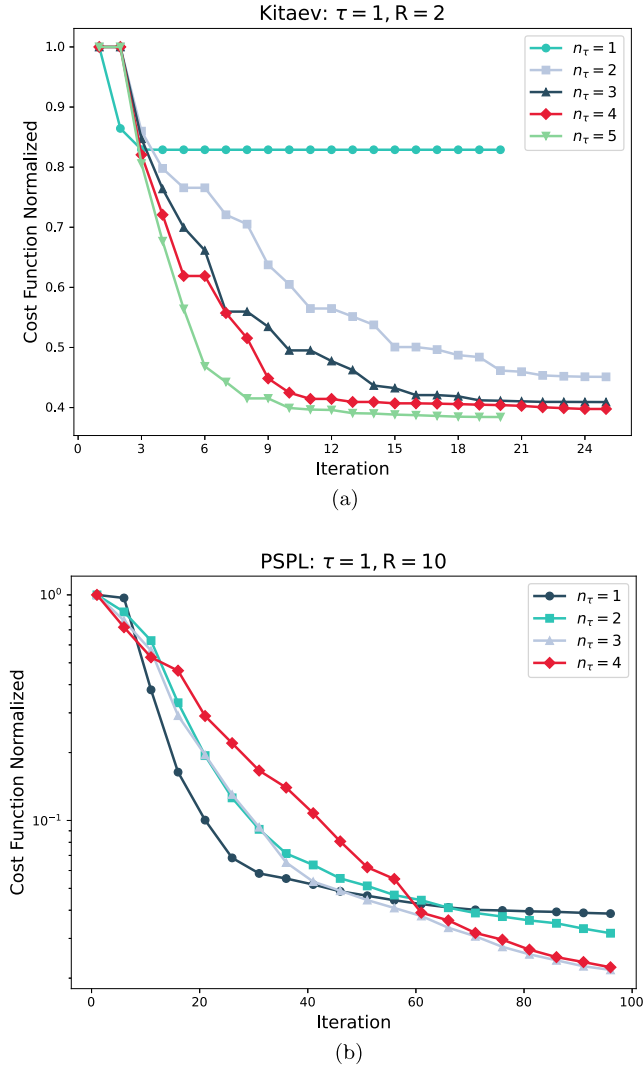


FIG. 8. Normalized approximation error as a function of the number of iterations for both models, while varying the number of time steps n_τ . (a) Normalized approximation error for the Kitaev model, with $\tau = 1$ and $R = 2$. (b) Normalized approximation error for the PSPL model, with $\tau = 1$ and $R = 10$.

using the Euclidean counterpart. This difference, as argued in Appendix G, can be attributed to the correct weight assigned by the canonical metric to each degree of freedom parametrizing the tangent space vectors. On the contrary, the Euclidean metric assigns twice as much weight to the anti-symmetric components, as shown in Eq. (G1). In addition, the poor performance of the elementary directions $\pi_\tau(E_{ij})$, shown by the turquoise curve, confirms the importance of choosing properly how to traverse the tangent space when creating the Riemannian Hessian using Eq. (K7). In Fig. 7(a), we also see that using this elementary direction results in a fast convergence to a suboptimal local minimum.

However, it turns out the canonical metric is not always the obvious choice. Figure 7(b) shows the cost function optimization under the PSPL model and final evolution time $\tau = 0.5$ with a time step $n_\tau = 4$. Contrary to the results from Fig. 7(a), the comparison in Fig. 7(b) shows that the Euclidean and canonical metrics yield very similar results when using the same elementary tangent directions. Moreover, the Euclidean metric even outperforms the canonical metric at some intervals of the iterations.

APPENDIX M: EXPRESSIVITY VS ITERATIONS

In Sec. V A, we carried out the Riemannian optimization using $n_\tau = 1$ and 4 as the number of time steps for the PSPL and Kitaev models, respectively. However, an important question to ask when working with iterative optimization methods is as follows: “How many iterations do we need for the algorithm to converge to a minimum?” In this Appendix, we will try to address this question numerically by looking at the behavior of the cost function when increasing the number of time steps. This dependence is particularly interesting to us, as the degrees of freedom in the optimization grow linearly with the number of time steps n_τ , i.e., the number of time steps increases the expressivity.

First, we look at the optimization with the Kitaev model, using $\tau = 1$ and $R = 2$ (its natural Choi rank), and vary the number of time steps n_τ . In order to compare the decrease in the cost function for each value of n_τ , we look at the *normalized* cost function trajectory

$$f_{\text{Norm}}(\mathbf{X}^\beta) = \frac{f(\mathbf{X}^\beta)}{f(\mathbf{X}^0)}. \quad (\text{M1})$$

The results shown in Fig. 8(a) provide insight to answer the question we posed at the beginning of this Appendix. First, we notice that by increasing the number of time steps, thus increasing the expressivity, we can achieve a lower relative approximation error. We also observe that all the curves converge after less than 30 iterations. The curve with $n_\tau = 1$ shows the least improvement, as it reaches a plateau after only 5 iterations. While the curves for $n_\tau = 3, 4, 5$ all converge around the same point, surprisingly, the optimization using $n_\tau = 2$ takes the longest to settle.

We repeat an analogous simulation using the PSPL model, $\tau = 1$, and $R = 10$. Compared to the Kitaev scenario, the trajectories in Fig. 8(b) show that only the optimization using the lowest number of time steps $n_\tau = 1$ converges under 100 iterations. Even after decreasing the cost function over an order of magnitude, the curves for $n_\tau = 3$ and 4 still depict a steep descent towards the final iterations. We can attribute this behavior to the big parameter space induced by the high-rank value $R = 10$, compared to the Kitaev rank $R = 2$, causing an increase in the complexity of the optimization. By a similar argument, we observe how the red curve corresponding to $n_\tau = 4$ exhibits the worse performance up to the 60th iteration, as it requires more iterations to find a minimum.

[1] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2010).

[2] D. Manzano, A short introduction to the Lindblad master equation, *AIP Adv.* **10**, 025106 (2020).

- [3] M. B. Hastings, Solving gapped Hamiltonians locally, *Phys. Rev. B* **73**, 085115 (2006).
- [4] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Matrix product density operators: Simulation of finite-temperature and dissipative systems, *Phys. Rev. Lett.* **93**, 207204 (2004).
- [5] G. D. I. Cuevas, N. Schuch, D. Pérez-García, and J. Ignacio Cirac, Purifications of multipartite states: Limitations and constructive methods, *New J. Phys.* **15**, 123021 (2013).
- [6] A. H. Werner, D. Jaschke, P. Silvi, M. Kliesch, T. Calarco, J. Eisert, and S. Montangero, Positive tensor network approach for simulating open quantum many-body systems, *Phys. Rev. Lett.* **116**, 237201 (2016).
- [7] M. Suzuki, General theory of fractal path integrals with applications to many-body theories and statistical physics, *J. Math. Phys.* **32**, 400 (1991).
- [8] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, Princeton, NJ, 2008).
- [9] R. Brieger, I. Roth, and M. Kliesch, Compressive gate set tomography, *PRX Quantum* **4**, 010325 (2023).
- [10] S. Diehl, E. Rico, M. A. Baranov, and P. Zoller, Topology by dissipation in atomic quantum wires, *Nat. Phys.* **7**, 971 (2011).
- [11] E. van den Berg, Z. K. Mineev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors, *Nat. Phys.* **19**, 1116 (2023).
- [12] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, *Commun. Math. Phys.* **144**, 443 (1992).
- [13] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, *Quantum Inf. Comput.* **7**, 401 (2007).
- [14] M. Zwolek and G. Vidal, Mixed-state dynamics in one-dimensional quantum lattice systems: A time-dependent super-operator renormalization algorithm, *Phys. Rev. Lett.* **93**, 207205 (2004).
- [15] M. Kliesch, D. Gross, and J. Eisert, Matrix-product operators and states: NP-hardness and undecidability, *Phys. Rev. Lett.* **113**, 160503 (2014).
- [16] H. Pichler, J. Schachenmayer, A. J. Daley, and P. Zoller, Heating dynamics of bosonic atoms in a noisy optical lattice, *Phys. Rev. A* **87**, 033606 (2013).
- [17] G. Vidal, Efficient classical simulation of slightly entangled quantum computations, *Phys. Rev. Lett.* **91**, 147902 (2003).
- [18] M. Kliesch, T. Barthel, C. Gogolin, M. Kastoryano, and J. Eisert, Dissipative quantum Church-Turing theorem, *Phys. Rev. Lett.* **107**, 120501 (2011).
- [19] C. J. Wood, J. D. Biamonte, and D. G. Cory, Tensor networks and graphical calculus for open quantum systems, *Quantum Inf. Comput.* **15**, 0759 (2015).
- [20] A. Edelman, T. A. Arias, and S. T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.* **20**, 303 (1998).
- [21] D. Nguyen, Operator valued formulas for riemannian gradient and Hessian and families of tractable metrics in Riemannian optimization, *J. Optim. Theory Appl.* **198**, 135 (2023).
- [22] R. Zhang, Newton retraction as approximate geodesics on submanifolds, [arXiv:2006.14751](https://arxiv.org/abs/2006.14751).
- [23] H. Yoshida, Construction of higher order symplectic integrators, *Phys. Lett. A* **150**, 262 (1990).
- [24] S. Blanes and P. Moan, Practical symplectic partitioned Runge–Kutta and Runge–Kutta–Nyström methods, *J. Comput. Appl. Math.* **142**, 313 (2002).
- [25] I. Luchnikov, A. Ryzhov, S. Filippov, and H. Ouerdane, QGOpt: Riemannian optimization for quantum technologies, *SciPost Phys.* **10**, 079 (2021).
- [26] Python software foundation, Python Language Reference, version 3.11 (2022), <https://www.python.org>.
- [27] E. Godinez, OpenTN: A Riemannian approach to the Lindbladian dynamics of a locally purified tensor network, <https://github.com/EmilianoG-byte/openTN>.
- [28] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant *et al.*, Array programming with NumPy, *Nature (London)* **585**, 357 (2020).
- [29] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: composable transformations of Python+NumPy programs, version 0.3.13 (2018), <http://github.com/google/jax>.
- [30] J. Manton, Optimization algorithms exploiting unitary constraints, *IEEE Trans. Signal Process.* **50**, 635 (2002).
- [31] Y. Cao and J. Lu, Structure-preserving numerical schemes for Lindblad equations, *J. Sci. Comput.* **102**, 27 (2025).
- [32] S. Cheng, C. Cao, C. Zhang, Y. Liu, S.-Y. Hou, P. Xu, and B. Zeng, Simulating noisy quantum circuits with matrix product density operators, *Phys. Rev. Res.* **3**, 023005 (2021).
- [33] A. Müller, T. Ayril, and C. Bertrand, Enabling large-depth simulation of noisy quantum circuits with positive tensor networks, [arXiv:2403.00152](https://arxiv.org/abs/2403.00152).
- [34] A. Kotil, R. Banerjee, Q. Huang, and C. B. Mendl, Riemannian quantum circuit optimization for Hamiltonian simulation, *J. Phys. A: Math. Theor.* **57**, 135303 (2024).
- [35] D. Dereniowski and M. Kubale, Cholesky factorization of matrices in parallel and ranking of graphs, in *Parallel Processing and Applied Mathematics*, edited by R. Wyrzykowski, J. Dongarra, M. Paprzycki, and J. Waśniewski (Springer, Berlin, 2004), pp. 985–992.
- [36] M. Hauru, M. Van Damme, and J. Haegeman, Riemannian optimization of isometric tensor networks, *SciPost Phys.* **10**, 040 (2021).
- [37] N. J. Higham, C. Mehl, and F. Tisseur, The canonical generalized polar decomposition, *SIAM J. Matrix Anal. Appl.* **31**, 2163 (2010).