Approximate dynamics leading to more optimal control: Efficient exact derivatives

Jesper Hasseriis Mohr Jensen,[1,*] Frederik Skovbo Møller,[1,2] Jens Jakob Sørensen,[1] and Jacob Friis Sherson[1,†]
[1]*Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, 8000 Aarhus C, Denmark*
[2]*Vienna Center for Quantum Science and Technology, Atominstitut, TU Wien, Stadionallee 2, 1020 Vienna, Austria*

(Received 6 November 2020; accepted 7 June 2021; published 21 June 2021)

Accurate derivatives are important for locally traversing and converging efficiently in quantum optimization landscapes. By deriving analytically exact control derivatives (gradient and Hessian) for unitary control tasks, we show here that the computational feasibility of meeting this accuracy requirement depends on the choice of propagation scheme and problem representation. Even when exact propagation is sufficiently cheap it is, perhaps surprisingly, much more efficient to optimize the (appropriately) approximate propagators: Approximations in the dynamics are traded off for significant complexity reductions in the exact derivative calculations. Importantly, past the initial analytical considerations, only standard numerical techniques are explicitly required with straightforward application to realistic systems. These results are numerically verified for two concrete problems of increasing Hilbert space dimensionality. The best schemes obtain unit fidelity to machine precision, whereas the results for other schemes are separated consistently by orders of magnitude in computation time and in worst-case ten orders of magnitude in achievable fidelity. Since these gaps continually increase with system size and complexity, this methodology allows numerically efficient optimization of very high-dimensional dynamics, e.g., in many-body contexts, operating in the high-fidelity regime.

DOI: 10.1103/PhysRevA.103.062612

I. INTRODUCTION

The demand for precise and fast quantum control extending into high-fidelity regimes places increasing emphasis on the role of optimization methodologies and their performance capacities. Identification and extraction of quantum optimal controls have enjoyed theoretical and experimental success in numerous research areas [1], such as superconducting qubits [2–5], nuclear magnetic resonance systems [6–10], nitrogen-vacancy centers [11–14], cold molecules [15–17], and cold atoms [18–24], to name a few. At the same time, an increasing array of algorithmic approaches is available in these arenas, counting among others derivative based (GRAPE [7,25,26], auxiliary matrix [27–29] or equivalently GOAT [30], GROUP [10,31], and Krotov [32–34]), derivative-free (Nelder-Mead CRAB [18,19,35], stochastic ascent [36], and genetic evolutionary [37]), and combinations thereof [38]. Along a separate axis lie additional choices of open-loop [39,40], closed-loop [41–43], and/or human-in-the-loop [36,44] control.

Irrespective of the physical platform and choice of optimization algorithm, a common denominator is inevitable: With growing problem complexity and numerical simulation efforts, the relative efficiency of each optimization cycle must be streamlined to allow convergence to, e.g., high-fidelity solutions within finite time. Accuracy and computational speed have been identified as important goals and challenges for modern control design [1,45]. In the context of derivative-based methods, i.e., update rules relying on local gradient and Hessian calculations of the optimization objective, this has been recognized at least since the seminal work presented in Ref. [7], where the analytical first-order approximation to the gradient was calculated. However, this first-order approximation is not suited for obtaining standard quasi-Newton search directions due to the rapid error accumulation in the Hessian approximation which is built iteratively from gradients [46]. The steepest descent direction is also only a minimally viable choice with the weakest convergence properties among the standard methods. The use of quasi-Newton methods with the more desirable convergence properties was enabled later by, e.g., Ref. [25], where the analytically exact gradient for an exact propagator was calculated at the expense of additional computational time per iteration. As system sizes increase, however, exact propagators and their exact derivatives become prohibitively resource intensive.

In this work we advance the theoretical toolbox for obtaining controls that satisfy high-performance criteria in arbitrary unitary quantum tasks. Following a discretize-then-optimize approach, we derive general analytically exact gradients and Hessians for different propagation schemes, specifically an exact exponentiation propagator and two Suzuki-Trotter propagators which we interpret in terms of optimization landscapes. This means that each choice of effective time evolution operator gives rise to its own, but not necessarily dynamically exact, optimization landscape as illustrated in Fig. 1 (Sec. II). Given, then, that the exact propagator approach is sometimes computationally infeasible, we thus examine the interplay between approximations in the landscape (i.e., dynamics) versus in the derivative calculations.

We show that the complexity of exact analytical derivatives strongly depends on the chosen propagation scheme,

*jhasseriis@phys.au.dk
†sherson@mgmt.au.dk

 ©2021 American Physical Society

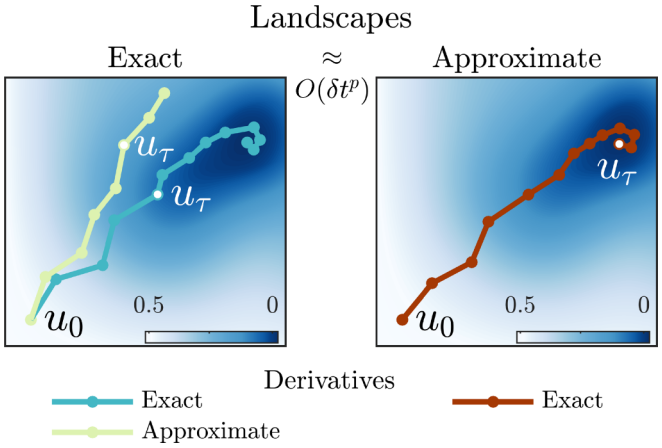

FIG. 1. Abstract illustration of local minimization in exact and approximate landscapes (equivalent to dynamical schemes), each with different analytical forms for the exact derivatives. The planes are spanned by control functions $u(t)$ with associated functional values given by the colormap and dots denote optimization iterates. In the exact landscape exact derivatives are expensive but ultimately lead to optimal results, while cheaper, approximate derivatives may not. In the approximate landscape exact derivatives and the dynamics itself are significantly cheaper and yield faithful optimal results under appropriate conditions. Thus, for the *same* optimization wall-clock time τ, a common initial u_0 leads to different u_τ (white dot centers) and hence manifests a performance gap.

corresponding to the specifics of the numerical implementation details, and representation of the problem: Solving the problem in a basis where the controllable part of the Hamiltonian is diagonal and simultaneously employing one of the Trotterized propagators greatly simplifies the derivative calculations (Sec. III). Analytically exact derivatives can thus be computed very efficiently, principally limited only by the time it takes to propagate states, which by virtue of the dynamical approximation is also particularly cheap. That is, our results and, e.g., Fig. 1 are not just trivial consequences of the reduced propagation time due to the dynamical approximation, but also the complexity reduction of the analytical exact derivatives. Through numerical experiments, remarks on implementation details, scaling comparisons, and generalizability analyses, our main goal is to show that the presented Trotter derivative methodologies are not only very efficient on realistic problems, but also straightforwardly applicable since they rely only on otherwise well-known ingredients. Another state-of-the-art approach to calculating exact gradients and Hessians is through the aforementioned so-called auxiliary matrix method [27–29] as implemented, e.g., in the SPINACH software library [47], and we include this methodology in our comparative studies.

We cement these findings and calculations[1] by first optimizing a minimal two-level Landau-Zener (LZ) problem and then a nine-level transmon system (Sec. IV). In both instances, we attain the performance hierarchy qualitatively captured in Fig. 1. We then show that this trend is exponentially monotonic in the face of more complex and larger systems. As the Hilbert space dimension scales exponentially in the number of constituents, this becomes especially relevant when, e.g., the system size enters the many-body regime where exact diagonalization, exact propagation, and associated exact derivatives are completely outside the realm of numerical feasibility. Finally, we show that the results generalize well to scenarios with more than one control (Sec. V) and touch on a few pertinent discussion points (Sec. VI).

[1]Only the exact gradient derivation for the exact propagator is similar to the calculations in Ref. [25].

II. EXACT DERIVATIVES FOR QUANTUM OPTIMAL CONTROL

A. Formulation of unitary control tasks

In quantum optimal control we seek to dynamically steer some quantum mechanical process in a controlled way such as to maximize a desired physical yield. For unitary evolution, any such task can be encoded as a minimization over an appropriate cost functional $J[\hat{\mathcal{U}}(T;0)]$, where

$$\hat{\mathcal{U}}(T;0) = \mathcal{T}\exp\left(-i\int_0^T \hat{H}(t)\,dt\right) \tag{1}$$

is the time evolution operator in units where $\hbar = 1$ from time $t = 0 \to T$, $\mathcal{T}$ denotes time ordering, and $\hat{H}$ is the system Hamiltonian carrying some generic time dependence. The cost functional is a purely mathematical and malleable construct that quantifies our desired set of success criteria or goals. Quite often these not only include the desired quantum dynamics but also experimental constraints. The individual criteria are typically represented by their own cost functional J_i and the total cost composed by $J = \sum_i J_i$. Typically, one will define the cost such that $J \geqslant 0$ and $J = 0$ is then guaranteed to be a global minimum. We briefly return to different potential choices of J_i at the end of this section. Minimizing J thus instructs us how to feasibly achieve the desired dynamics under the given experimental constraints.

The manipulatory access to the system dynamics in Eq. (1) is through a set of control parameters in the Hamiltonian. To preserve clarity of the presentation we initially consider the case of a single generic control $u(t)$ and we may without loss of generality separate the system Hamiltonian as

$$\hat{H} = \hat{H}(t, u(t)) = \hat{H}^d(t) + \hat{H}^c(t, u(t)). \tag{2}$$

In Sec. V we extend our analysis to more than one control and show that the generalization remains feasible. The drift Hamiltonian $\hat{H}^d(t)$ represents parts of the system dynamics that are uncontrollable. The control Hamiltonian $\hat{H}^c(t, u(t))$, on the other hand, depends on the control $u(t)$ and grants us a mandate to steer the dynamics. As an example, for a single-particle system the drift could be the kinetic energy $\hat{H}^d = \hat{T}$ and the control Hamiltonian the potential profile $\hat{H}^c = \hat{V}(x, u)$.

Similarly to the cost J, it is gainful to think of $u(t)$ as a malleable mathematical object that is tied to some physical quantity $p(t)$ in the system such as the intensity, trap center, or frequency of a laser that ultimately affects the system. In this general framing we can write $p(t) = g(u(t))$, where g is some suitable differentiable function and this allows a

certain degree of modeling convenience. For example, one might choose $g(u) = au$ for numerical reasons, where c is an appropriate scaling constant [46] such that appropriate values of u is of order 1. Other examples include letting g be a shifted and scaled sigmoid or arctan function [48] such that p is bounded within a range of values or letting g be a so-called transfer function such that p respects finite electronic response times [31]. This provides, e.g., alternative measures for incorporating experimental constraints without the use of additional cost functionals.

Notice that our definition of the control Hamiltonian in Eq. (2) subsumes the control itself, which is more general than the pervasive bilinear assumption, i.e., $\hat{H}^c(t, u(t)) \to u(t)\hat{H}^c(t)$. Linear physical dependences do indeed occur in many systems, e.g., in the form of spin couplings which can be subject to control. Nevertheless, assuming control linearity precludes both the use of nonlinear g as defined above and treatment of systems that are inherently not linear in the physical parameter. Examples of the latter include position-controlled Gaussian potential profiles $V(x, u) \propto \exp[2(x - u)^2/\sigma^2]$, where σ is a width [49], or phase-controlled optical lattices $V(x, u) \propto \cos^2(kx + u)$, where k is a wave number [50]. For these reasons and since the derivations presented here do not depend on it we do not impose linearity assumptions and simply note that this limiting case can always be taken at the end.

For numerical (and, as we will see, analytical) convenience it is natural to discretize time in regular δt intervals

$$t \in [t_1, t_2, \ldots, t_{N_t}] = [0, \delta t, \ldots, T], \quad t_j = (j-1)\delta t, \quad (3)$$

with time indices denoted by subscripts. Physical quantities evaluated at these grid points are similarly defined by $u_n = u(t_n)$ and $\hat{H}_n = \hat{H}(t_n, u_n)$ and similarly for $\hat{H}_n^c$ and $\hat{H}_n^d$. This leads to a product of time evolution operators

$$\hat{\mathcal{U}}(T;0) \approx \prod_{j=1}^{N_t-1} \hat{\mathcal{U}}_j = \hat{\mathcal{U}}_{N_t-1} \cdots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1, \quad (4a)$$

where

$$\hat{\mathcal{U}}_n = \exp\left(-i \int_{t_n}^{t_n+\delta t} \hat{H}(t, u(t)) dt\right) \quad (4b)$$

is the propagator across the time interval $[t_n, t_{n+1}] = [t_n, t_n + \delta t]$. As seen later, the error made with respect to Eq. (1) is given by $N_t - 1$ first-order Suzuki-Trotter expansions and vanishes when $\delta t \to 0 \leftrightarrow N_t \to \infty$ or if $[\hat{H}(t_j), \hat{H}(t_i)] = 0$ for all t_j and t_i. We return to the computation of $\hat{\mathcal{U}}_n$ in Sec. II D and for now simply note that $\hat{\mathcal{U}}(T;0)$ depends on the discretized control vector $\boldsymbol{u} = (u_1, \ldots, u_{N_t})^T$.

The optimal control task thus consists in finding appropriate control vector(s) $\boldsymbol{u}$ that correspond to local and hopefully global minima in the control landscape defined by the cost

$$J(\boldsymbol{u}) = J\big(\hat{\mathcal{U}}_{N_t-1} \cdots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1\big). \quad (5)$$

There is a plethora of techniques and prescriptions for maneuvering the landscape in search of such minima.

B. Derivative-based local optimization

In this paper we focus on derivative-based local optimization methodologies characterized by making informed decisions in traversing the control landscape $\boldsymbol{u}^{(k)} \to \boldsymbol{u}^{(k+1)}$ using local information at iteration k about the landscape topography. In prototypical line-search-based updates of the form

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \alpha^{(k)} \boldsymbol{p}^{(k)}, \quad \alpha^{(k)} \in \mathbb{R}^+, \quad (6a)$$

$$J(\boldsymbol{u}^{(k+1)}) \leqslant J(\boldsymbol{u}^{(k)}), \quad (6b)$$

the search direction $\boldsymbol{p}^{(k)}$ is calculated from the current local gradient (e.g., steepest descent, conjugate gradient, and quasi-Newton directions) and possibly also the Hessian (e.g., Newton direction). It is helpful to abstractly depict landscapes as in Fig. 1, where the colormap denotes the cost value $J(\boldsymbol{u})$, each dot is an iterate $\boldsymbol{u}^{(k)}$, and the line connecting it to the next is the step $\alpha^{(k)}\boldsymbol{p}^{(k)}$. The step size $\alpha^{(k)}$ is in practice determined by an inexact line search [46] and values largely depend on the chosen line-search algorithm, search direction, problem scaling, and how close to a minimum the iterate is. Far from the minimum, as is typical for the initial iterate, it can be on the order of hundreds or thousands, after which it usually becomes of order one or smaller as it approaches convergence.

Although the functional form of the cost J in Eq. (5) depends on the particular unitary task [34], e.g., gate synthesis, state transfer, or maximization of a given observable, and whether a pure state or density matrix description is considered, they all lead to the same principal form for the control derivative calculations, specifically

$$\text{gradient} \sim \frac{\partial}{\partial u_n}\big(\hat{\mathcal{U}}_{N_t-1} \cdots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1\big), \quad (7a)$$

$$\text{Hessian} \sim \frac{\partial^2}{\partial u_n \partial u_m}\big(\hat{\mathcal{U}}_{N_t-1} \cdots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1\big). \quad (7b)$$

Accurate control derivatives are paramount in successfully traversing the optimization landscape, since inaccuracies or willful approximations yield poor search directions and may significantly slow down, altogether prevent convergence, and/or limit the achievable fidelity (cf. the exact landscape in Fig. 1). Out of the standard search directions, the steepest descent direction is well known to have the comparatively worst general properties such as zigzag iterate trajectories and a linear convergence rate [46]. This is irrespective of the exactness of the gradient and inaccuracies will make this choice even more unattractive. The much more theoretically sound exact Newton direction exhibits quadratic convergence and a step size of $\alpha^k \to 1$ when approaching a minimum [46]. It is however typically several orders of magnitude more expensive to construct than a steepest descent direction because it relies on both the gradient and the Hessian. Inaccuracies in either derivative thus negate any of the potential gains from the extra computational effort. The class of quasi-Newton methods is generally accepted to be the most general-purpose class of search directions and exhibit superlinear convergence [46]. For example, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) direction is initialized as a steepest descent step and then builds up an approximation for the Hessian based on the gradients recorded in each iteration. If the individual gradients

are inaccurate, this error will accumulate in each iteration and make the Hessian approximation unreliable as the optimization progresses. This will in turn manifest as increasingly poor search directions that ultimately lead the optimization irreparably astray as depicted in Fig. 1. This is referred to as the slowdown problem in Ref. [25].

These considerations are generic and not limited to the field of quantum control. We calculate the exact derivatives that surpass these issues in Sec. III and verify the calculations and claims above by numerical demonstration in Sec. IV. For specificity, we consider the case of pure state transfer $|\psi_{\text{ini}}\rangle \to |\psi_{\text{tgt}}\rangle$, ideally obtaining unit fidelity given by

$$F = |\langle\psi_{\text{tgt}}|\psi(T)\rangle|^2 = |\langle\psi_{\text{tgt}}|\hat{\mathcal{U}}(T;0)|\psi_{\text{ini}}\rangle|^2, \tag{8}$$

where $|\psi_{\text{tgt}}\rangle$ is the target state and $|\psi_{\text{ini}}\rangle = |\psi_1\rangle$ is the initial state. The state $|\psi(T)\rangle = |\psi_{N_t}\rangle = \hat{\mathcal{U}}(T;0)\,|\psi_{\text{ini}}\rangle$ at final time T is produced by stepwise evolution according to $|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n\,|\psi_n\rangle$. The associated fidelity cost is $J_F = (1-F)/2$. It will be convenient to write the fidelity as $F = o^* o$ and thus the fidelity cost and its derivatives as

$$J_F = \frac{1}{2}(1 - o^* o), \tag{9a}$$

$$\frac{\partial J_F}{\partial u_n} = -\text{Re}\left(o^* \frac{\partial o}{\partial u_n}\right), \tag{9b}$$

$$\frac{\partial^2 J_F}{\partial u_n \partial u_m} = -\text{Re}\left[\left(\frac{\partial o}{\partial u_m}\right)^* \frac{\partial o}{\partial u_n} + o^* \frac{\partial^2 o}{\partial u_n \partial u_m}\right], \tag{9c}$$

with o being the overlap or transfer amplitude with derivatives

$$o = \langle\chi_{N_t}\,|\psi_{N_t}\rangle = \langle\chi_{N_t}\,|\hat{\mathcal{U}}_{N_t-1}\cdots\hat{\mathcal{U}}_n\cdots\hat{\mathcal{U}}_1|\psi_1\rangle, \tag{9d}$$

$$\frac{\partial o}{\partial u_n} = \langle\chi_{N_t}\,|\frac{\partial}{\partial u_n}(\hat{\mathcal{U}}_{N_t-1}\cdots\hat{\mathcal{U}}_n\cdots\hat{\mathcal{U}}_1)|\psi_1\rangle, \tag{9e}$$

$$\frac{\partial^2 o}{\partial u_n \partial u_m} = \langle\chi_{N_t}\,|\frac{\partial^2}{\partial u_n \partial u_m}(\hat{\mathcal{U}}_{N_t-1}\cdots\hat{\mathcal{U}}_n\cdots\hat{\mathcal{U}}_1)|\psi_1\rangle, \tag{9f}$$

where we introduced an auxiliary state $|\chi_{N_t}\rangle = |\psi_{\text{tgt}}\rangle$ with $|\chi_n\rangle = \hat{\mathcal{U}}_n^\dagger\,|\chi_{n+1}\rangle$. The overlap derivatives is in the form of Eqs. (7), as anticipated. The Hessian matrix is of course symmetric and allows reuse of gradient elements. To obtain numerically implementable expressions for the gradient and Hessian in Eqs. (9b) and (9c), our task is now to analytically calculate Eqs. (9e) and (9f). These calculations will clearly depend on the details of the propagator $\hat{\mathcal{U}}_n$.

We stress that, although the particular functional form for $J = J_F$ leads to "specialized" derivatives in Eqs. (9), dictated by the chain rule, our evaluations of Eqs. (7) are general and always constitute the by far largest numerical effort. Thus, obtaining exact derivatives for any other unitary control task mentioned above, e.g., gate synthesis and dynamics described by density matrices, is a trivial extension by appropriately applying the chain rule to the encoding functional. We therefore restrict our attention to the pure state transfer formulation in the remainder of this paper. For completeness we also include derivatives for common control regularization terms in the Appendix. These cost augmentations are, as mentioned earlier, often introduced for experimental reasons and they do not depend on the numerical propagation scheme.

C. Suzuki-Trotter expansions

To set the stage for the following section we briefly recall a few ubiquitous Suzuki-Trotter expansions for the operator exponential.

The exponential of the operator $\hat{X}$ or its matrix representation is defined in terms of its Taylor series

$$e^{\alpha\hat{X}} = \sum_{k=0}^{\infty} \frac{(\alpha\hat{X})^k}{k!} = 1 + \alpha\hat{X} + \frac{1}{2}\alpha^2\hat{X}^2 + O(\alpha^3), \tag{10}$$

where α is a scalar and $O(\alpha^3)$ denotes terms of order α^3 or higher. Operator exponentials appear in many scientific contexts and the literature surrounding its explicit and efficient evaluation is quite extensive [51]. Let us assume that additional structure is present, $\hat{X} = \hat{A} + \hat{B}$, in which case the expansion reads

$$e^{\alpha(\hat{A}+\hat{B})} = 1 + \alpha(\hat{A}+\hat{B}) + \tfrac{1}{2}\alpha^2(\hat{A}^2 + \hat{A}\hat{B} + \hat{B}\hat{A} + \hat{B}^2) + O(\alpha^3), \tag{11}$$

after performing the square. Properties of $\hat{A}$ and $\hat{B}$ typically ensure that individual evaluation of $e^{\alpha\hat{A}}$ and $e^{\alpha\hat{B}}$ is much simpler than the composite $e^{\alpha(\hat{A}+\hat{B})}$. For example, if $\hat{A}$ is represented in a diagonal basis by the $N \times N$ matrix $\boldsymbol{A}$, then the matrix exponential is also diagonal with elements $(e^{\alpha\boldsymbol{A}})_{n,n} = e^{\alpha\boldsymbol{A}_{n,n}}$ for all $n = 1, \ldots, N$.

Motivated by this fact and the rules for scalar exponentials, evaluating the product $e^{\alpha\hat{A}}e^{\alpha\hat{B}}$ using Eq. (10) twice leads to the simplest, first-order Suzuki-Trotter expansion [52] by comparing to Eq. (11),

$$e^{\alpha\hat{A}}e^{\alpha\hat{B}} = 1 + \alpha(\hat{A}+\hat{B}) + \frac{\alpha^2}{2}(\hat{A}^2 + 2\hat{A}\hat{B} + \hat{B}^2) + O(\alpha^3) = e^{\alpha(\hat{A}+\hat{B})} + O(\alpha^2). \tag{12a}$$

The first-order expansion evidently has an error scaling

$$e^{\alpha(\hat{A}+\hat{B})} - e^{\alpha\hat{A}}e^{\alpha\hat{B}} = \frac{\alpha^2}{2}[\hat{B},\hat{A}] + O(\alpha^3) = O(\alpha^2), \tag{12b}$$

which depends on the commutator $[\hat{B},\hat{A}]$. In fact it is well known that $e^{\alpha(\hat{A}+\hat{B})} = e^{\alpha\hat{A}}e^{\alpha\hat{B}}$ is exact if $[\hat{B},\hat{A}] = 0$. Suppose we instead consider an ansatz on the form $e^{\alpha\hat{A}_1}e^{\alpha\hat{B}}e^{\alpha\hat{A}_2}$ where

$$\hat{A} = \hat{A}_1 + \hat{A}_2 + O(\alpha^2). \tag{13a}$$

By applying Eq. (10) three times we obtain

$$e^{\alpha\hat{A}_1}e^{\alpha\hat{B}}e^{\alpha\hat{A}_2} = 1 + \alpha(\hat{A}_1 + \hat{A}_2 + \hat{B}) + \frac{\alpha^2}{2}\left(\hat{A}_1^2 + \hat{A}_2^2 + 2(\hat{A}_1\hat{A}_2 + \hat{A}_1\hat{B} + \hat{B}\hat{A}_2) + \hat{B}^2\right) + O(\alpha^3). \tag{13b}$$

This Suzuki-Trotter expansion has error

$$e^{\alpha(\hat{A}+\hat{B})} - e^{\alpha\hat{A}_1}e^{\alpha\hat{B}}e^{\alpha\hat{A}_2} = \tfrac{1}{2}\alpha^2([\hat{A}_2,\hat{A}_1] + [\hat{B},\hat{A}_1 - \hat{A}_2]) + O(\alpha^3), \tag{13c}$$

which is also to first order $O(\alpha^2)$ for arbitrary $\hat{A}_1$ and $\hat{A}_2$, but to second order $O(\alpha^3)$ when

$$[\hat{A}_1,\hat{A}_2] = 0, \tag{14a}$$

$$\hat{A}_1 - \hat{A}_2 = O(\alpha). \tag{14b}$$

More generally, it is possible to systematically construct Suzuki-Trotter expansion variants of arbitrarily high order by considering, e.g., the ansatz $e^{p_1\alpha\hat{A}}e^{p_2\alpha\hat{B}}e^{p_3\alpha\hat{A}}e^{p_4\alpha\hat{B}}\cdots e^{p_M\alpha\hat{B}} + O(\alpha^{m+1})$ and choosing suitable coefficients p_i [52]. Although arbitrarily low error is an enticing prospect, it also entails more computational time. Since in our context $\alpha = -i\delta t$ represents a small time step, it is sufficient to henceforth only consider expansions such as the ones above.

D. Dynamics and optimization landscapes

We now return to the numerical evaluation of Eqs. (4). To compute the propagator in Eq. (4b) we must first perform the integral in the exponential. We consider numerical quadratures based on the left-point rectangle rule and the trapezoidal rule

$$\int_{t_n}^{t_n+\delta t} \hat{H}(t,u(t))dt \approx \delta t \times \begin{cases} \hat{\overline{H}}_n & \text{(trapezoidal)} \\ \hat{H}_n & \text{(rectangle)}, \end{cases} \tag{15}$$

where $\hat{H}_n = \hat{H}(t_n, u_n)$ and $\hat{\overline{H}}_n = \frac{1}{2}(\hat{H}_{n+1} + \hat{H}_n)$. If the underlying time dependence is assumed continuous, the trapezoid and rectangle approximations have integration errors $O(\delta t^3)$ and $O(\delta t^2)$, respectively. If the time dependence is assumed piecewise constant, the rectangle approximation is exact. Both rules thus satisfy the condition in Eq. (13a).

Depending on the choice of quadrature, we will refer to the exact exponentiation propagators as

$$\hat{\mathcal{U}}_n^{\mathrm{ex}_1} = \exp(-i\hat{\overline{H}}_n\delta t), \tag{16a}$$

$$\hat{\mathcal{U}}_n^{\mathrm{ex}_2} = \exp(-i\hat{H}_n\delta t) \tag{16b}$$

and their corresponding Suzuki-Trotter (or Trotterized) propagators as, respectively,

$$\hat{\mathcal{U}}_n^{\mathrm{ST}_1} = \hat{\mathcal{U}}_{n+1}^{c/2}\hat{\mathcal{U}}_n^{d}\hat{\mathcal{U}}_n^{c/2} \approx \hat{\mathcal{U}}_n^{\mathrm{ex}_1}, \tag{16c}$$

$$\hat{\mathcal{U}}_n^{\mathrm{ST}_2} = \hat{\mathcal{U}}_n^{c/2}\,\hat{\mathcal{U}}_n^{d}\hat{\mathcal{U}}_n^{c/2} \approx \hat{\mathcal{U}}_n^{\mathrm{ex}_2}, \tag{16d}$$

with definitions for the control and drift exponentials

$$\hat{\mathcal{U}}_n^{c/2} \equiv \exp\left(-i\hat{H}_n^c\delta t/2\right), \tag{16e}$$

$$\hat{\mathcal{U}}_n^{d} \equiv \begin{cases} \exp\left(-i\hat{\overline{H}}_n^d\delta t\right) & \text{for } \mathrm{ST}_1 \\ \exp\left(-i\hat{H}_n^d\delta t\right) & \text{for } \mathrm{ST}_2. \end{cases} \tag{16f}$$

The operator splitting $\hat{\mathcal{U}}_n^{\mathrm{ST}_2}$ is achieved by utilizing Eqs. (13) with $B = \hat{H}_n^d$ and $\hat{A}_1 = \hat{A}_2 = \hat{H}_n^c/2$ and always has Trotterization error $O(\delta t^p)$ with $p = 3$ according to Eqs. (14). The operator splitting $\hat{\mathcal{U}}_n^{\mathrm{ST}_1}$ is achieved by letting $B = \hat{\overline{H}}_n^d$, $\hat{A}_1 = \hat{H}_{n+1}^c/2$, and $\hat{A}_2 = \hat{H}_n^c/2$ and has Trotterization error $O(\delta t^p)$ where $p = 2, 3$ depending on Eqs. (14). For unrelated reasons, we will later assume that the control Hamiltonians are diagonal, which leaves only the condition (14b). However, the precise Trotter error p is not of crucial importance since δt must under all circumstances be small enough that the errors made in going from Eq. (1) to Eqs. (4)[2] and in integrating Eq. (15) are small. What is much more important is that $\hat{\mathcal{U}}_n^{\mathrm{ST}_2}$ is fully local in n whereas $\hat{\mathcal{U}}_n^{\mathrm{ST}_1}$ depends on both n and $n+1$. The resulting derivative calculations with respect to u_n and final expressions are thus different. This underscores that the precise specification of the implementation is central for use in optimal control contexts. The derivatives should "match" the dynamics.

The local $O(\delta t^p)$ Trotterization errors accumulate throughout the $N_t - 1$ evolutions in Eq. (4a), yielding an overall error $(N_t - 1)\delta t^p \approx (T/\delta t)\delta t^p = T\delta t^{p-1} \sim O(\delta t^{p-1})$. It is convenient to interpret this as an approximation error with respect to the exact landscape (cf. Fig. 1)

$$J_F^{\mathrm{ST}}(\boldsymbol{u}) = J_F^{\mathrm{ex}}(\boldsymbol{u}) + O_{J_F^{\mathrm{ST}}}(\delta t^{p-1}) \tag{17}$$

for $\mathrm{ex} = \mathrm{ex}_1, \mathrm{ex}_2$ and associated $\mathrm{ST} = \mathrm{ST}_1, \mathrm{ST}_2$. The granularity of δt determines how faithful the representation is and it follows that geometric entities for the *same* point $\boldsymbol{u}$ are generally different in each landscape. This includes the height or cost value, derivatives, and thus also the search directions for optimization. Importantly, the optimal controls associated with optima in the Trotterized landscapes at large finite δt may not correspond to optima in the exact landscapes, which is equivalent to the target not being obtained when propagating said controls using Eqs. (16a) and (16b). As $\delta t \to 0$, however, the Trotterized landscapes continuously deform into the exact landscapes, and below some sufficiently small finite δt they represent it with only negligible perturbations. The landscapes transitively inherit the numerical implementation properties of their associated propagator.

Numerically, the exact propagator corresponds to direct exponentiation of the Hamiltonian matrix, an operation that scales extremely poorly with increasing Hilbert space dimension $D_{\mathcal{H}}$. On the other hand, the Trotterized propagators lend themselves more readily to a variety of very efficient, problem-dependent implementations through, e.g., the use of sparsity structures, and these have a much more benign Hilbert space scaling, extending its applicability far beyond the exact propagator approach. For example, we will assume that the control Hamiltonian is diagonal and this significantly boosts runtime performance. This is because $\hat{\mathcal{U}}_n^{c/2}$ is relatively cheap to represent and calculate since diagonal matrix exponentiation is the elementwise exponentiation of the diagonal. Additionally, the relatively very expensive calculations of $\hat{\mathcal{U}}_n^{d}$ for all n can be performed once and stored in memory or on the disk as they are by definition independent of the choice of $\boldsymbol{u}$. We can also write vectorized forms useful for state transfer and unitary synthesis, respectively, as

$$\mathcal{U}_n^{\mathrm{ST}_1} \times \boldsymbol{\psi}_n = \mathcal{U}_{n+1}^{c/2} \odot \left[\mathcal{U}_n^{d} \times \left(\mathcal{U}_n^{c/2} \odot \boldsymbol{\psi}_n\right)\right], \tag{18a}$$

$$\mathcal{U}_n^{\mathrm{ST}_1} = \mathcal{U}_n^{d} \odot \left[\mathcal{U}_{n+1}^{c/2} \times \left(\mathcal{U}_n^{c/2}\right)^{\mathrm{T}}\right], \tag{18b}$$

where $\mathcal{U}_n^{\mathrm{ST}_1}$ and $\mathcal{U}_n^{d}$ are dense matrices, $\mathcal{U}_n^{c/2}$ and $\boldsymbol{\psi}_n$ are a vectors, $\times$ ($\odot$) denotes regular (elementwise) matrix multiplication, and T is transposition. These numerical techniques are the original reasons for approximating the dynamics. In the next section we show that also more subtle and important simplifications occur when calculating the exact derivatives.

[2]This corresponds to $N_t - 1$ applications of the first-order expansion in Eqs. (12).

III. ANALYTICAL RESULTS

Detailed calculations of the analytically exact derivatives stated in Eqs. (9) using the propagators in Eqs. (16) are given in the Appendix. Here we focus on the central equations and contexts. Of the exact propagators, we consider for simplicity only $\hat{\mathcal{U}}_n^{\rm ex} = \hat{\mathcal{U}}_n^{\rm ex_2}$ given by Eq. (16b) from here on.

By defining the recursive commutator

$$[X, Y]_k = [X, [X, Y]_{k-1}], \quad [X, Y]_0 = Y, \tag{19}$$

a central calculation shows that

$$\frac{\partial \hat{\mathcal{U}}_n^{\rm ex}}{\partial u_n} = \hat{\mathcal{U}}_n^{\rm ex}(-i\delta t) \sum_{k=0}^{\infty} \frac{i^k \delta t^k}{(k+1)!} \left[\hat{H}_n, \frac{\partial \hat{H}_n^c}{\partial u_n}\right]_k, \tag{20}$$

where $\partial \hat{H}_n^c / \partial u_n = \partial \hat{H}_n / \partial u_n$ is the control derivative Hamiltonian. The presence of the infinite sum means that in the context of optimization it may not be desirable to use exact time evolution even if it is readily available. In numerical application, the summation continues until machine precision or to a desired accuracy, corresponding to some $k_{\max}$. This is necessary because the Hamiltonian and its control derivative generally do not commute, $[\hat{H}_n, \frac{\partial \hat{H}_n^c}{\partial u_n}] \neq 0$, and the recursive commutator is not guaranteed to terminate. A few examples include $\hat{H} = \hat{\sigma}_x + u(t) \cdot \hat{\sigma}_z$, where $\hat{\sigma}_i$ are Pauli spin-$\frac{1}{2}$ operators, $\hat{H} = \hat{T} + \hat{V}(u(t))$, where $\hat{V}$ ($\hat{T}$) is the potential (kinetic) energy operator for a single particle, or the Bose-Hubbard Hamiltonian $\hat{H} = -\hat{J} + \hat{U}(u(t))$, where $\hat{J}$ ($\hat{U}$) is the tunneling (on-site interaction) operator.

The propagator derivatives for the Trotterizations in Eq. (16c) and (16d) contain an infinite series of the same structure as in Eq. (20), but with $\hat{H}_n \to \hat{H}_n^c$ in the first argument of the recursive commutator. Thus, by additionally assuming that the control Hamiltonian is expressed in its diagonal representation, we find

$$\frac{\partial}{\partial u_n}(\hat{\mathcal{U}}_n^{\rm ST_1} \hat{\mathcal{U}}_{n-1}^{\rm ST_1}) = (-i\delta t)\hat{\mathcal{U}}_n^{\rm ST_1} \frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_{n-1}^{\rm ST_1}, \tag{21a}$$

$$\frac{\partial \hat{\mathcal{U}}_n^{\rm ST_2}}{\partial u_n} = -\frac{i\delta t}{2}\left(\frac{\partial \hat{H}_n^c}{\partial u_n}\hat{\mathcal{U}}_n^{\rm ST_2} + \hat{\mathcal{U}}_n^{\rm ST_2}\frac{\partial \hat{H}_n^c}{\partial u_n}\right), \tag{21b}$$

since the series terminate exactly after $k_{\max} = 0$. This huge simplification occurs because two diagonal matrices always commute, $[\hat{H}_n^c, \frac{\partial \hat{H}_n^c}{\partial u_n}]_k = \frac{\partial \hat{H}_n^c}{\partial u_n}\delta_{0,k}$. Incidentally, in many cases the "natural" basis states for computations are already the ones that diagonalize $\hat{H}_n^c$, e.g., spin eigenstates ($|\uparrow\rangle$ and $|\downarrow\rangle$), position eigenstates ($|\boldsymbol{x}\rangle$), or site-occupation eigenstates ($|n_i\rangle$), respectively, for the Hamiltonians mentioned above.

Inserting the results in Eqs. (20) and (21) into Eqs. (9), we obtain the gradient elements

$$\frac{\partial J_F^{\rm ex}}{\partial u_n} = {\rm Re}\left(io^*\langle\chi_n|\frac{\partial \hat{H}_n^c}{\partial u_n}|\psi_n\rangle\right)\delta t + O_{\nabla J_F^{\rm ex}}(\delta t^2), \tag{22a}$$

$$\frac{\partial J_F^{\rm ST_1}}{\partial u_n} = {\rm Re}\left(io^*\langle\chi_n|\frac{\partial \hat{H}_n^c}{\partial u_n}|\psi_n\rangle\right)\delta t, \tag{22b}$$

$$\frac{\partial J_F^{\rm ST_2}}{\partial u_n} = {\rm Re}\left(\frac{io^*}{2}\left\{\langle\chi_{n+1}|\frac{\partial \hat{H}_n^c}{\partial u_n}|\psi_{n+1}\rangle + \langle\chi_n|\frac{\partial \hat{H}_n^c}{\partial u_n}|\psi_n\rangle\right\}\right)\delta t. \tag{22c}$$

The states in each equation are understood to be evolved according to the propagation scheme denoted on the left-hand side, but this notational completeness is omitted here for brevity. It is important to remember that these Trotter gradients assume that the control Hamiltonian $\hat{H}^c$ is diagonal. The remainder term for the exact propagator inherits the infinite series of Eq. (20),

$$O_{\nabla J_F^{\rm ex}}(\delta t^2) \propto \langle\chi_n|\left(\sum_{k=1}^{\infty}\frac{i^k\delta t^k}{(k+1)!}\left[\hat{H}_n, \frac{\partial \hat{H}_n^c}{\partial u_n}\right]_k\right)|\psi_n\rangle\delta t, \tag{23}$$

where the first-order approximation in Eq. (22a) is the $k = 0$ term.

Thus, the gradients for $J^{\rm ex}$ and $J^{\rm ST_1}$ are in the same form[3] only to first order in δt: Whereas $\nabla J_F^{\rm ST_1}$ is analytically exact with just the δt term, $\nabla J_F^{\rm ex}$ entails an expensive remainder term beyond the first-order approximation. This is noteworthy and nontrivial since, while $\rm ST_2$ is the most commonly encountered type of Trotterization, $\nabla J_F^{\rm ex}$ and $\nabla J_F^{\rm ST_2}$ do *not* coincide even to first order. Generally, neither do the derivatives for other dynamical approximations such as the Krylov-Lanczos method, finite Taylor expansions, the Crank-Nicolson method, and Chebyshev schemes [53–55]. Nevertheless, the gradient $\nabla J_F^{\rm ST_2}$ is only slightly more involved than $\nabla J_F^{\rm ST_1}$ due to two sum terms and this is negligible compared to the computational effort of the time evolution and overlap calculations can be reused. Preference towards either may therefore rely on which of the integration quadratures in Eq. (15) is most appropriate in a given situation. Similar calculations and arguments apply to the analytically exact Hessians, but in this case, the expressions for $\nabla^2 J_F^{\rm ex}$ and $\nabla^2 J_F^{\rm ST_2}$ are much more complicated than $\nabla^2 J_F^{\rm ST_1}$, as shown in the Appendix.

An alternative way of calculating the exact propagator derivatives is through the auxiliary matrix method [28,29]. Focusing on the first derivative, the relation

$$\begin{pmatrix} \hat{\mathcal{U}}_n & \frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} \\ 0 & \hat{\mathcal{U}}_n \end{pmatrix} = \exp\left[-i\begin{pmatrix} \hat{H}_n & \frac{\partial \hat{H}_n}{\partial u_n} \\ 0 & \hat{H}_n \end{pmatrix}\delta t\right] \tag{24}$$

allows extraction of both $\hat{\mathcal{U}}_n$ and $\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n}$ by explicitly calculating the right-hand-side 2×2 block matrix exponential. The expression can be augmented to a 3×3 block matrix to also include the Hessian. This approach provides exact derivatives while elegantly circumventing the cumbersome commutator series in Eq. (23) associated with the exact propagator, but the required $2D_{\mathcal{H}}$ (or $3D_{\mathcal{H}}$ with Hessian) square matrix exponentials becomes similarly expensive.

What happens if we combine the auxiliary matrix with our Trotterization with diagonal controls? The derivative of u_n-dependent propagators is, e.g., $\frac{\partial}{\partial u_n}(\hat{\mathcal{U}}_n^{\rm ST_1}\hat{\mathcal{U}}_{n-1}^{\rm ST_1}) = (\hat{\mathcal{U}}_{n+1}^{c/2}\hat{\mathcal{U}}_n^d)\frac{\partial \hat{\mathcal{U}}_n^c}{\partial u_n}(\hat{\mathcal{U}}_{n-1}^d\hat{\mathcal{U}}_{n-1}^{c/2})$ and we can then employ Eq. (24) with $\hat{\mathcal{U}}_n \to \hat{\mathcal{U}}_n^c$ and $\hat{H}_n \to \hat{H}_n^c$. Since $\hat{H}_n^c$ is assumed diagonal, the right-hand-side exponent is a very sparse, *almost* diagonal matrix with a single dense off-diagonal. Although very efficient numerical sparse solvers exist, the computation remains nontrivial, and it is always cheaper to instead utilize either

[3] Except at the end points (see the Appendix).

Eqs. (22b) or (22c) in this case since the exponential itself is analytically trivialized and $\frac{\partial \hat{H}_n^c}{\partial u_n}$ is also an analytically closed form.

In Sec. VI we also discuss the results in comparison to the Krylov-Lanczos propagation scheme, another common type of approximate time evolution for quantum states. Other types of techniques for extending simulation capabilities such as the reduced density matrix in, e.g., nuclear magnetic resonance systems [56] will not be further discussed here.

IV. NUMERICAL RESULTS

To numerically verify the analytical conclusions and assertions made in Sec. III, we initially examine the methods' performance capacities on two concrete state-transfer problems with dimensionalities of $D_{\mathcal{H}} = 2$ and 9, respectively. We then investigate the gradient evaluation time as $D_{\mathcal{H}}$ continually increases. In general, the particular parameters chosen for these studies are not central to the overall methodological conclusions and will only be discussed to the extent that they are relevant. The process duration T for the presented results is chosen such that $T_{\text{min}}^{F=1} \lesssim T \ll T_{\text{adiabatic}}$, where $T_{\text{min}}^{F=1}$ is the minimal duration that $F = 1$ solutions exist and $T_{\text{adiabatic}}$ is the adiabatic limit. We also briefly discuss the behavior at 0.5, 0.75, 1.25, and 1.5 times this T. All results were generated on a 2017 Macbook Pro laptop with 16 GB RAM using a single 2.8-GHz Intel Core i7 processor.

A. Two-level system

We first consider the canonical LZ model with $\hat{H}_n = \hat{H}^d + \hat{H}_n^c = \frac{1}{2}(\hat{\sigma}_x + u_n \cdot \hat{\sigma}_z)$ and the state transfer $|\psi_{\text{ini}}\rangle = |\uparrow\rangle \rightarrow |\psi_{\text{tgt}}\rangle = |\downarrow\rangle$. Employing the natural basis $\{|\uparrow\rangle, |\downarrow\rangle\}$, this problem is already represented in the necessary control-diagonal form. The reason for choosing this problem is twofold. First, it represents the smallest possible nontrivial type of problem ($D_{\mathcal{H}} = 2$). Second, it has well-understood solutions [57] with a single, analytical π-pulse solution $u_n = 0$ at the minimal $F = 1$ duration $T_{\text{min}}^{F=1} = \pi$ and remains solvable beyond this duration. Yet despite its simplicity, the model remains prototypical even in the context of the many-body arena. For example, many-body dynamics can in certain scenarios be thought of as a cascade of independent LZ transitions and similar characteristics between LZ and some many-body control problems have been identified [58–60].

We optimize the same 100 uniformly randomly generated seeds $u_n = \text{uniform}(-10, 10)$ at $T = 1.01\pi \gtrsim T_{\text{min}}^{F=1}$ with $\delta t = 0.075$ using the BFGS search direction implemented in MATLAB's `fminunc` in five different scenarios: exact propagator with first-order ($k_{\text{max}} = 0$) and exact ($k_{\text{max}} = 9$ and auxiliary method) gradients, and both Trotterized propagators with exact gradients. Figure 2 shows the optimization results. Only 14 seeds did not converge to machine precision within 400 iterations when using exact gradients.

From the $1 - F$ iteration median trajectories we find that unit fidelities to machine precision are easily obtainable only when utilizing any of the exact gradients, whereas the inexact approximate variant leads to poor results. Solutions of the former exhibit very rapid convergence when they approach the optimum with a variance in the low tens for the number of iterations needed. Looking however at the optimization wall-clock time trajectories provides a definite per-

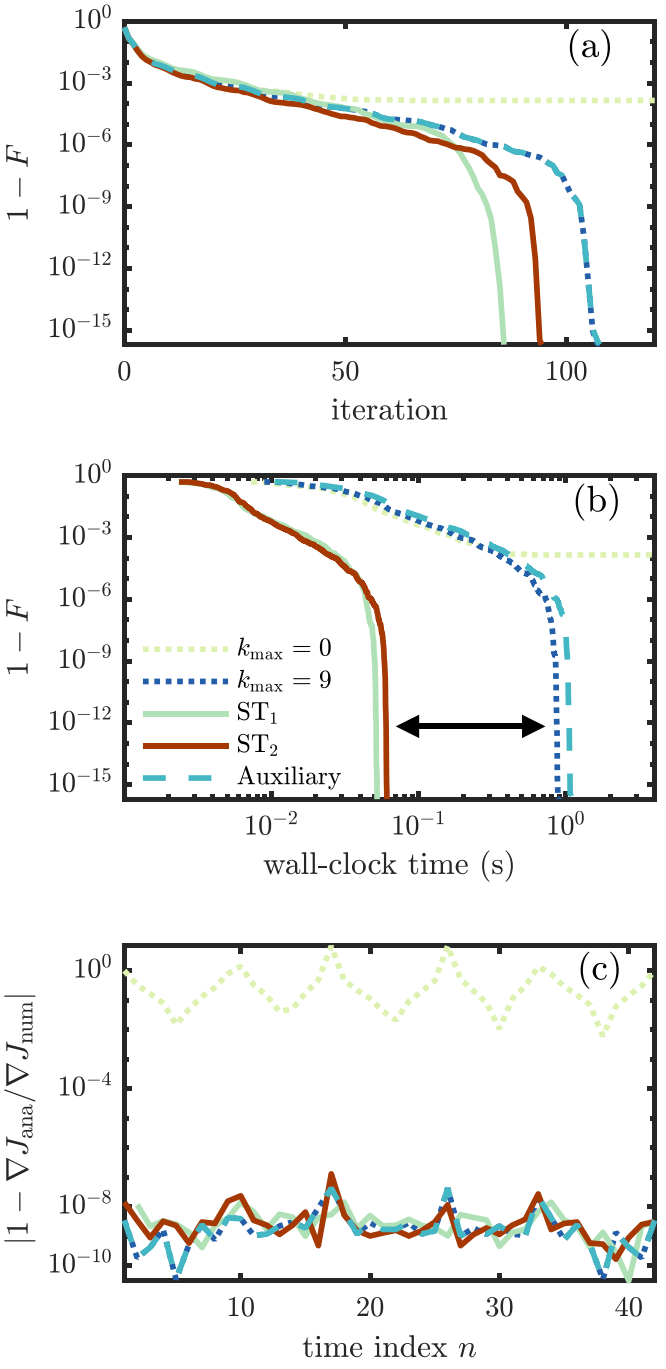

FIG. 2. Demonstration of optimization in exact and Trotterized landscapes for the simple two-level LZ problem. The $1 - F$ median is shown as a function of (a) iteration and (b) wall-clock time. The performance gap between the exact methods is roughly denoted by the horizontal arrow. (c) Analytical gradients ∇J_{ana} relative to their numerical central finite-difference approximation ∇J_{num} with perturbation $\epsilon_{\text{pert}} = \epsilon_{\text{mach}}^{1/3}$ for a constant example control with $u_n = 5$. Note that the auxiliary and $k_{\text{max}} = 9$ gradients are identical to machine precision and thus yield (nearly) the same iteration trajectories and, incidentally, nearly the same wall-clock time trajectories.

formance hierarchy with negligible variance. The mint and red trajectories are separated by more than an order of magnitude in computation time from the dark and light blue trajectories. These medians are associated with the Trotterized and exact landscapes qualitatively shown in Fig. 1, respectively, and the separation is due to both the difference in propagation computation time and either the $k_{\max} > 0$ tail in Eq. (23) or the $2D_{\mathcal{H}} \times 2D_{\mathcal{H}}$ matrix exponential in Eq. (24). As mentioned, the first-order gradient approximation $k_{\max} = 0$ in the exact landscape performs significantly worse while also being slower compared to iterations in the Trotterized landscapes. Note however that its performance is decent up until around 50 iterations, at which point it prevents convergence to unit fidelities by more than ten orders of magnitude. This is because the Hessian approximation eventually becomes completely unreliable as discussed in Sec. II.

As a further verification of Eqs. (22) we compare the analytical gradients to their respective central finite difference gradients and find relatively close agreement. The latter are themselves associated with errors of $O_{\nabla J_F}(\delta t^2)$, and the relative differences are on the order $\sqrt{\epsilon_{\text{mach}}}$, where $\epsilon_{\text{mach}} = 2.22 \times 10^{-16}$ is the machine precision for the double-precision floating-point format. We also quantitatively find that the first-order approximation is unsurprisingly poor.

The truncation parameter necessary for exact gradients is roughly bounded $\delta t^{k_{\max}+1}/(k_{\max}+1)! \lesssim \epsilon_{\text{mach}}$. With decreasing δt, the necessary $k_{\max}$ for exact gradients also decreases and the $k = 0$ term becomes increasingly dominant. Indeed, running the same optimizations as shown in Fig. 2 for $\delta t = 0.025$, the $k_{\max} = 0$ optimization yields final results that are two to three orders of magnitude better relative to $k_{\max} = 0$ in Fig. 2 with $\delta t = 0.075$. At $\delta t = 0.01$ this first-order approximation is sufficient for finding machine precision unit fidelities. That is, reducing the number of k terms required for accurate gradients is traded off for increased computation time per iteration due to additional time evolutions. Consequently, even though such $k_{\max} = 0$ trajectories may now be sufficient in terms of final results, they are stretched to much higher wall-clock times than any of the results shown in Fig. 2.

We also performed optimizations at durations 0.5, 0.75, 1.25, and 1.5 times $T = 1.01\pi$. For both the smaller and larger durations we find empirically that the control landscape becomes very easy in the sense that only a few iterations (fewer than ten) are required for convergence and the optimized fidelities are nearly identical. This is not surprising since the problem is relatively simple even when the control is heavily constrained [61]. Due to this simplicity and the associated low number of required iterations, the approximate gradient actually becomes competitive with the exact ones in the exact landscape in terms of wall-clock time at these durations. That is, convergence is achieved before the Hessian approximation becomes very unreliable. This simplicity should not be expected for the majority of control problems, and a detailed study of this behavior and at intermediate T is outside the scope of this paper. The performance gap to the exact gradient methods in the approximate landscape seen in Fig. 2 persists across all the different tested T values regardless. Together with the discussion above, this shows that the optimal choice of $k_{\max}$ is nontrivial and dependent on the other problem parameters. This suggests a methodological simplicity of the Trotter approach where gradient exactness is always ensured by just the first-order term.

B. Transmon system

We now turn to a second example of higher dimensionality ($D_{\mathcal{H}} = 9$), a superconducting transmon system with a two-qutrit computational basis $\{|00\rangle, |10\rangle, \ldots |12\rangle, |22\rangle\}$ described by the Hamiltonian

$$\hat{H}_n = \left[\Delta \hat{b}_1^\dagger \hat{b}_1 + \frac{1}{2}\sum_{j=1,2} \delta_j \hat{b}_j^\dagger \hat{b}_j (\hat{b}_j^\dagger \hat{b}_j - 1) + \kappa(\hat{b}_1^\dagger \hat{b}_2 + \hat{b}_1 \hat{b}_2^\dagger)\right] + u_n(\hat{b}_1^\dagger + \hat{b}_1) = \hat{H}^d + \hat{H}_n^c, \tag{25}$$

with the same parameter values as in Ref. [54], here with the relabeling $J \to \kappa$ to avoid ambiguity with the cost functional. We consider the state transfer $|\psi_{\text{ini}}\rangle = |10\rangle \to |\psi_{\text{tgt}}\rangle = |11\rangle$, i.e., a single state mapping of a controlled-NOT (CNOT) gate in the qubit subspace $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, without control constraints.[4] Note that the control Hamiltonian is not diagonal in the natural computational basis. To obtain a proper representation for the Trotter exact derivatives, we therefore numerically diagonalize $\hat{H}_n^c = u_n(\hat{b}_1^\dagger + \hat{b}_1)$. The eigenvectors are identical for all nonzero values of the control and we take $u_n = 1$ for simplicity. Storing these eigenvectors as columns in the basis transformation operator $\hat{\mathcal{R}}$, we perform the basis change

$$\hat{H}_n^c \leftarrow \hat{\mathcal{R}}^\dagger \hat{H}_n^c \hat{\mathcal{R}}, \qquad \hat{H}^d \leftarrow \hat{\mathcal{R}}^\dagger \hat{H}^d \hat{\mathcal{R}}, \tag{26}$$

$$|\psi_{\text{ini}}\rangle \leftarrow \hat{\mathcal{R}}^\dagger |\psi_{\text{ini}}\rangle, \quad |\psi_{\text{tgt}}\rangle \leftarrow \hat{\mathcal{R}}^\dagger |\psi_{\text{tgt}}\rangle. \tag{27}$$

The results of optimizing this control-diagonalized problem are shown in Fig. 3. We find a nearly identical situation to Fig. 2, except the gap between the control-diagonal Trotter and exact propagator gradient methods has significantly increased (note the logarithmic scale) due to the increased $D_{\mathcal{H}}$. Further, more iterations are generally needed and the first-order approximation "falsely" converges to even worse fidelities. This indicates that this problem is somewhat more challenging than that of Fig. 2.

Performing optimizations at durations 0.5, 0.75, 1.25, and 1.5 times $T = 2.83$ did not change the results outside of scaling the performance gap and, for the lower values, the overall best attainable fidelity. In contrast to the two-level problem, this problem was never so simple that the approximate gradient could yield results competitive results with the exact gradient in the exact landscape and we expect this to be representative of most control problems.

[4] This process is expected to have a lower quantum speed limit than what we found for the full gate [54]. The full CNOT gate could have been considered in the state transfer formulation by optimizing a composite cost, e.g., $J_F^{|00\rangle\to|00\rangle} + J_F^{|01\rangle\to|01\rangle} + J_F^{|11\rangle\to|10\rangle} + J_F^{|10\rangle\to|11\rangle}$.

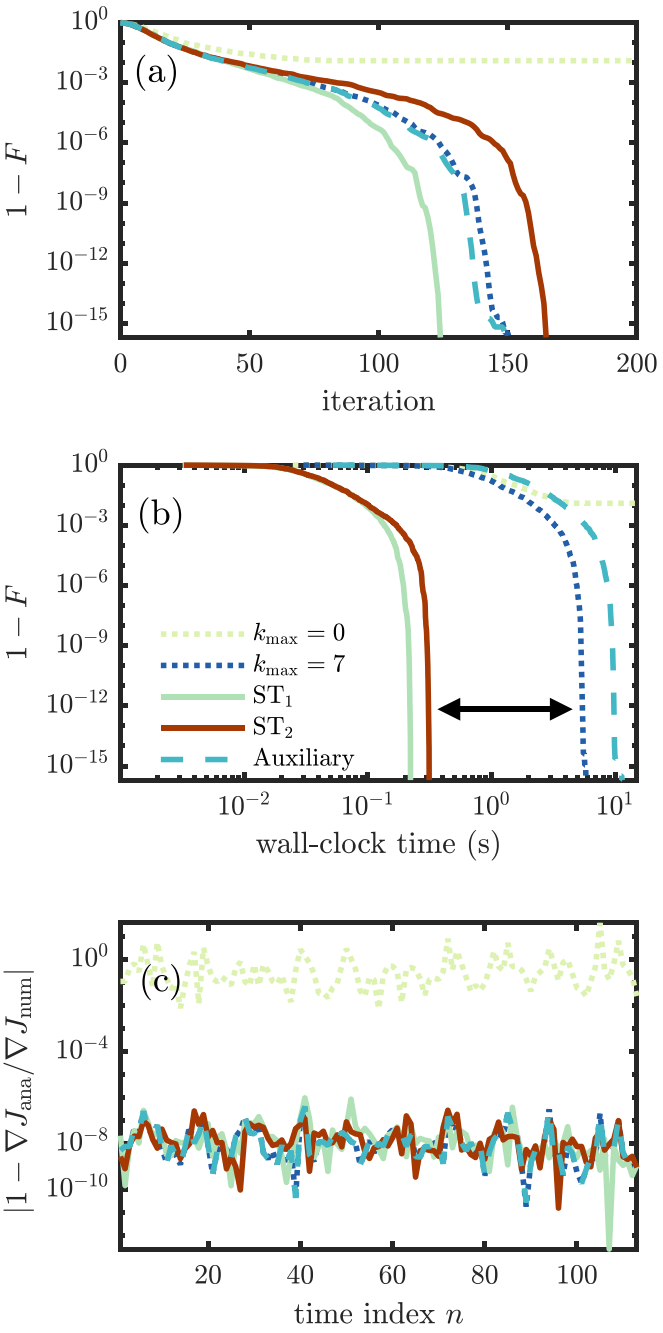

FIG. 3. Same as Fig. 2 for 50 seeds, but for the transmon system (25)–(27) with $T = 2.83$ (50 ns) and $\delta t = 0.025$ (0.442 ns) given in nondimensionalized numerical units and SI units, respectively (energy is measured in units of $|\kappa| \approx 5.97 \times 10^{-27}$ J and time in $\hbar/|\kappa| \approx 17.7$ ns).

C. Gradient evaluation time

Next we record the wall-clock time for calculating the various gradients as a function of $D_{\mathcal{H}}$, including $D_{\mathcal{H}} = 2$ and 9 associated with Figs. 2 and 3. For each $D_{\mathcal{H}}$ we generate ten random Hamiltonian matrices and controls of length $N_t = 400$. Figure 4 shows the median wall-clock time consumption and these results reveal that the performance gap between the exact derivative methods is monotonically and exponentially increasing (note the logarithmic scale). This is

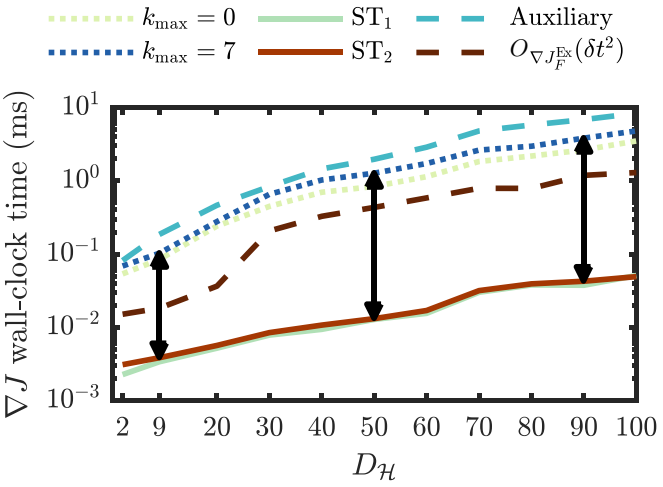

FIG. 4. Gradient and recursive commutator tail (23) median calculation time with negligible variance as a function of Hilbert space dimensionality. The exponentially increasing performance gap between exact derivative methods is roughly indicated by the arrows.

not inherently surprising due to the exact propagation itself being trivially much slower. However, even when subtracting this contribution, calculating the recursive commutator tail in Eq. (23) to ensure gradient exactness exhibits a similar scaling with orders of magnitude in separation to the full Trotter calculations. That is, the smaller gaps between the three upper methods also grow with $D_{\mathcal{H}}$. This trend was already visible in Figs. 2 and 3 by comparing the relative wall-clock time distance between the dark and light blue trajectories.

A natural question is, then, where the exact derivatives for other approximate dynamical schemes such as the aforementioned Krylov-Lanczos, finite Taylor series, Crank-Nicolson, and Chebyshev methods would manifest in Fig. 4. Leaving a complete numerical study of this for future work, we nevertheless argue based on scaling properties in Sec. VI that Krylov-Lanczos scheme, perhaps the most prominent general alternative, would lie somewhere in the shown performance gap for $D_{\mathcal{H}} \gg 1$ and above otherwise. We have performed similar mathematical analysis for the remaining mentioned schemes but consider it beyond the scope of this paper to include it explicitly since they yield comparable or worse results. In addition to δt, most of these also depend on series truncation parameters similar to $k_{\max}$ in Eq. (23).

V. GENERALIZABILITY OF MULTIPLE CONTROLS

Up to now our analyses assumed a single, generically parametrized control Hamiltonian $\hat{H}^c(t, u(t))$. When including more than one control, our approach depends on the computational feasibility of maintaining the diagonality criteria for the control Hamiltonians as follows.

We define a set of K controls and their corresponding control Hamiltonians by

$$C(t) = \{\hat{H}^c_{(k)}(t, u_{(k)}(t))\}_{k=1}^{K} = \cup_{q=1}^{Q} C_q(t), \tag{28}$$

which we have sorted into Q sets C_q of mutually commuting elements with $K = \sum_q |C_q(t)|$. Let $\hat{\mathcal{R}}_q(t)$ be the unitary

basis change operator that simultaneously diagonalizes the elements of $C_q(t)$ from a chosen reference basis, e.g., a natural or the computational. That is, $[\hat{X}, \hat{Y}] = 0$ and $\hat{\mathcal{R}}_q(t)^\dagger \hat{X} \hat{\mathcal{R}}_q(t)$ is diagonal for all $\hat{X}, \hat{Y} \in C_q(t)$.

Upon temporal discretization, consider the time evolution operator at time index n with, e.g., the ST_2 expansion of Eq. (16d),

$$\begin{aligned}\hat{\mathcal{U}}_n^{\mathrm{ex}} &= \exp\left[-i\delta t\left(\hat{H}_n^d + \sum_{k=1}^{K} \hat{H}_{n,(k)}^c\right)\right] \\ &\approx \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \\ &= \exp\left[-\frac{i\delta t}{2}\left(\sum_k \hat{H}_{n,(k)}^c\right)\right] e^{-i\delta t \hat{H}_n^d} \\ &\quad \times \exp\left[-\frac{i\delta t}{2}\left(\sum_k \hat{H}_{n,(k)}^c\right)\right]. \end{aligned} \tag{29}$$

The sum can be grouped as $\sum_k = \sum_q \sum_{C_{n,q}}$ and the exponential can then be split into the sets of mutually commuting elements with further first-order Trotterizations as in Sec. II C,

$$\begin{aligned}&\exp\left[-\frac{i\delta t}{2}\sum_q\left(\sum_{C_{n,q}} \hat{H}_{n,(k)}^c\right)\right] \\ &\approx \prod_{q=1}^{Q} \exp\left[-\frac{i\delta t}{2}\left(\sum_{C_{n,q}} \hat{H}_{n,(k)}^c\right)\right] \\ &= \prod_{q=1}^{Q} \hat{\mathcal{R}}_{n,q}\left(\exp\left[-\frac{i\delta t}{2}\left(\sum_{C_{n,q}} \hat{\mathcal{R}}_{n,q}^\dagger \hat{H}_{n,(k)}^c \hat{\mathcal{R}}_{n,q}\right)\right]\right)\hat{\mathcal{R}}_{n,q}^\dagger, \end{aligned} \tag{30}$$

which is exact if $Q = 1$ and of $O(\delta t^2)$ otherwise. Also, we utilized that $e^{\hat{A}} = \hat{B} e^{\hat{B}^\dagger \hat{A} \hat{B}} \hat{B}^\dagger$ for any unitary $\hat{B}$. Here each $\hat{\mathcal{R}}_{n,q}^\dagger \hat{H}_{n,(k)}^c \hat{\mathcal{R}}_{n,q}$ is by definition diagonal and Eq. (30) thus sequentially transforms into the respective control-diagonal bases where it is trivial to compute the exponentials and their exact derivatives. This is the same core idea as in the split-step Fourier transform [62] for propagating real-space wave functions.

For this to be an effective approach for $Q > 1$, performing the basis change $\hat{\mathcal{R}}_{n,q}$ must be significantly cheaper than the original dense exponentiation. The one exceptional instance where this condition is not met occurs when $\hat{\mathcal{R}}_{n,q}$ depends on the control value and simultaneously no closed analytical solution to the transformations is known. This implies that $\hat{\mathcal{R}}_{n,q}$ must be obtained anew in each iteration by numerical diagonalization, which is as expensive as dense exponentiation. Otherwise, $\hat{\mathcal{R}}_{n,q}$ and products involving these need only be calculated maximally once and can be stored on the disk and loaded into memory at runtime. Note in particular that the exceptional case can be categorically ruled out upon additionally including the pervasive assumption of bilinear controls, i.e., $\hat{H}_{(k)}^c(t, u(t)) \to u_{(k)}(t)\hat{H}_{(k)}^c(t)$. As in the transmon example, this is because the control value is simply a scaling factor to the matrix diagonalization.

In certain $Q > 1$ cases it may not possible to numerically diagonalize for $\hat{\mathcal{R}}_{n,q}$ even once due to very large $D_{\mathcal{H}}$, for example, in a many-body setting. One is then seemingly restricted to $Q = 1$, which incidentally covers a broad range of rich and realistic problems: for example, combined control over (i) individual and similar site-site couplings in spin chains [63] [$\sum_k u_k(t)\hat{\sigma}_k^i$ and $\sum_k u_k(t)\hat{\sigma}_k^i\hat{\sigma}_{k+1}^i$, respectively, for any $i = x, y, z$] or (ii) on-site interaction and single-site potentials in the Bose-Hubbard model [18,64] [$u(t)\sum_k \hat{n}_k(\hat{n}_k - 1)$ and $\sum_k u_k(t)\hat{n}_k$, respectively]. However, we may yet entertain the capability of treating $Q > 1$. For example, a $Q = 2$ spin chain in the form $\hat{H} = u_1(t)\sum \hat{\sigma}_k^x\hat{\sigma}_{k+1}^x + u_2(t)\sum \hat{\sigma}_k^z\hat{\sigma}_{k+1}^z$ can be represented in terms of either $\hat{\sigma}^x$ or $\hat{\sigma}^z$ eigenstates which yield control diagonal representations for the respective terms. We can then proceed to optimize one of the mutually commuting sets while momentarily considering the rest a drift contribution, so effectively $Q = 1$ for a number of iterations. That is, by numerically constructing the bases, Hamiltonians, and states for each set individually and choosing either of the representations at the start of each iteration, we circumvent the need for explicitly calculating $\hat{\mathcal{R}}_{n,q}$ that transforms between them. The controls $u_1(t)$ and $u_2(t)$ in the example may then be sequentially optimized and in the optimization literature this is known as coordinate descent [46].

The principal computational cost associated with considering multiple controls in Eq. (28) is therefore not the number of controls K itself but the number of mutually commuting sets $Q \leqslant K$ they distribute into. Importantly, the number of basis transformations scales only linearly with $Q - 1$ and typical values are $Q = 1, 2$. For the simplest case $Q = 1$ and time-independent $\hat{\mathcal{R}} = \hat{\mathcal{R}}_{n,1}$ the full time evolution $\hat{\mathcal{U}}(T;0)$ is further simplified since $\hat{\mathcal{R}}^\dagger\hat{\mathcal{R}} = 1$. The two concrete problems studied in Sec. II fall into this category. For the two-level (transmon) system, the natural basis did (did not) diagonalize the control and thus $\hat{\mathcal{R}} = 1$ ($\hat{\mathcal{R}}$ was numerically obtained).

The presented control-diagonal Trotter methodology's results and relative performance capacity therefore generalize well to more than one control.

VI. DISCUSSION AND OUTLOOK

We discussed how and why accurate derivatives are central to achieve high fidelities and convergence rates in derivative-based methods for generic optimization tasks. Examining common choices of time evolution scheme, the exact and two Suzuki-Trotter expanded propagators, showed how these can be interpreted and related in terms of optimization landscapes. We found their resulting analytically exact derivatives to differ vastly in complexity: Assuming a diagonal control Hamiltonian for the Trotterized landscapes, we circumvented a detrimental infinite series and highlighted many additional attractive properties compared to the exact landscape leading to the performance gap qualitatively illustrated in Fig. 1.

Additionally, only one of the Trotter gradients' analytical forms coincides with that of the first-order approximation in the exact landscape. This is quite exceptional, since the second Trotter gradient of equivalent complexity, as well as those due to other, standard dynamical approximations with increased derivative complexities, does not. When balancing respective errors in the dynamics and in the derivative calculations, the

latter is in a certain sense more important as the derivatives should match the landscape. We demonstrated the main ideas by considering two problems of varying Hilbert space size and in both instances found the expected performance hierarchies. The control-diagonal Trotter exact derivatives lead to a several orders of magnitude increase in computational speed and high-fidelity results with zero error to machine precision. This trend was verified to be monotonic and exponentially growing due to separate numerical complexity differences in both dynamics and derivatives with the Hilbert space size. Finally, it was shown that the control-diagonal Trotter methodology generalizes well to more than one control.

The immediate advantages of optimizing in the Trotterized landscapes over the exact landscape are twofold: (i) They are applicable to much larger systems and (ii) their analytically exact control derivatives and thus search directions essential for optimization convergence are greatly simplified. We assumed that the control Hamiltonian is diagonal and this can practically always be, and often automatically is, fulfilled. In particular, the main computational effort in the optimization lies in propagating the state, i.e., the dynamics, and is not subject to severely scaling bottlenecks like matrix exponentiation, recursive commutators, or diagonalization. We did not include the exact gradient obtained in the diagonalization paradigm [54] for our comparative studies. The principal cost is N_t diagonalizations of size $D_{\mathcal{H}} \times D_{\mathcal{H}}$ in each iteration and it would thus exhibit a similar performance gap to the Trotter methods.

The only implicit requirement is that δt is small enough for the Trotterization to faithfully approximate the exact dynamics or, equivalently, the exact landscape in Fig. 1 and Eq. (17). We did not explicitly include this consideration in the discussion of the presented examples since it is not central or important to the overall methodological performance hierarchy. Of course, to obtain meaningful results in practical application, it is essential to establish an upper bound for the time resolution also with respect to the initial approximation of Eq. (1) by Eqs. (4). Note, however, that the exact Trotter derivatives are irrespective of δt in terms of complexity. This allows for the effective use of homotopy methods [64,65] (among others) in δt, that is, optimization with an increasingly finer time resolution. This can be used to significantly speed up initial iterations without loss of accuracy in the final dynamics, which typically has an error of $O(\delta t^2)$, because the (quasi)continuous deformation of δt can be made arbitrarily small at the end. The exact propagator can nonetheless still be used if the Hilbert space is sufficiently low dimensional such that direct exponentiation and series summation or alternatively auxiliary matrix exponentiation or diagonalization is feasible (cf. Fig. 4). However, even for the simplest possible nontrivial two-level LZ problem illustrated in Fig. 2, this approach is seen to be much slower than the alternative, controllably approximate methods. This difference in computational feasibility increases monotonically and exponentially with the Hilbert space dimensionality as evidenced by Fig. 4. The aggregate computational performance of our control-diagonal Trotter methodology provides a scaffolding for efficient derivative-based optimization of very-high-dimensional many-body dynamics in the high-fidelity limit. We pursue this in parallel work [64] for a system far beyond exact diagonalization approaches, necessitating a matrix product state description. Below we expand on a few pertinent discussion points.

Krylov-Lanczos methods. Another common way of approximating the time evolution for extended applicability is through the use of Krylov-Lanczos subspace methods [53,66,67] where exponential operator applications are performed without explicit construction. Notice that the auxiliary matrix method (24) can be adapted to this setting by multiplying from the right by $(\mathbf{0}, |\psi_n\rangle)^T$, yielding two separate Krylov-Lanczos calculations $(\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} |\psi_n\rangle , \hat{\mathcal{U}}_n |\psi_n\rangle)^T$ for each of the K controls in Eq. (30). Here we briefly compare this approach with our Trotterized control-diagonal scheme for $K = 1$. For example, numerically stepping forward in time $|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n |\psi_n\rangle$ with Krylov-Lanczos entails (i) iterative construction of $k \leqslant D_{\mathcal{H}}$ Lanczos vectors q_i of dimension $D_{\mathcal{H}}$, each requiring a matrix-vector multiplication of the form $\hat{H}_n q_i$ as the most expensive operation, and (ii) matrix exponentiation of a $k \times k$ matrix and at least another matrix multiplication. This turns out to be computationally efficient compared to exact propagation when $1 < k \ll D_{\mathcal{H}}$, where k controls the approximation accuracy. The control-diagonal Trotter steps, e.g., $|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} |\psi_n\rangle$, naively require numerically a total of three matrix multiplications, but the vectorized form (18a) reduces this significantly. Since $\hat{\mathcal{U}}_n^{c/2}$ is diagonal, the exponentiation of each diagonal element can be efficiently stored in a vector, and a total of two elementwise vector-vector multiplications need to be performed for the control part. For the drift part, recall that $\hat{\mathcal{U}}_n^d$ only needs to be calculated once and can be cached indefinitely, leaving only a matrix-vector multiplication with the same cost as constructing a single Lanczos vector. Thus, simply constructing the Lanczos vectors [step (i)] is more costly than performing the full Trotter step. Further, the Krylov-Lanczos procedure obfuscates the direct analytical dependence on the control u_n, disallowing a straightforward analytical derivative calculation. As opposed to the control-diagonal Trotterization, this leads to a mismatch between the optimization landscape and the derivative calculations unless k is large enough since the exactness of both is linked to an approximation parameter. Nevertheless, Krylov propagation is strongly preferable to exact propagation even for moderate values of $D_{\mathcal{H}}$ and we expect that further studies would place it somewhere inside the performance gaps in Figs. 2–4.

Robustness. Optimal controls extracted from open-loop methodologies may be sensitive to variations and uncertainties in the underlying physical model. Let x be a physical and possibly time-dependent quantity related to the control or any of the uncontrolled system parameters. If x was modeled to have value x_{theory} but in actual experimental implementation has value x_{expt}, then the fidelity is likely to degrade as a result, $F_{\text{expt}} < F_{\text{theory}}$. These modeling errors could originate from many sources, for example, imperfect equipment fabrication, drifting or fluctuating noisy signals, or nondeterministic run-to-run system preparation.

It is however possible to account for such errors by including these uncertainties through, e.g., ensemble optimization [10,68]. The cost function is then taken as an ensemble average $\bar{J} = \sum_{l=1}^{L} J[x_l]$, possibly weighted, over L

realizations of the physical system. In each realization, the uncertain parameter is taken to be $x = x_l$, which could be randomly or regularly sampled from a suitable model distribution, for example, Gaussian or bounded uniform. If the noise on x is adequately characterized, the optimal controls achieved by minimizing $\bar{J}$ will have built-in robustness to the parameter fluctuations.

Although this idea is simple and straightforward to implement, each iteration now requires the calculation of L gradients since $\frac{\partial \bar{J}}{\partial u_n} = \sum_{l=1}^{L} \frac{\partial J[x_l]}{\partial u_n}$. This places an increased emphasis on both gradient computation speed and exactness. Reaching a certain number of iterations increases the wall-clock time by roughly a factor of L. Additionally, derivative errors are compounded much more severely between iterations. For example, Hessian approximation errors for quasi-Newton methods would lead to false convergence at roughly L times the normal rate or equivalently at $1/L$ times the normal iterations.

As an example, suppose one of the physical parameters associated with Figs. 2–3 was uncertain and optimized with $L = 10$ ensemble members. The $k_{\max} = 0$ approximation in Figs. 2 and 3 stagnated at fewer than 50 iterations would now stagnate at fewer than 10. The exact gradient wall-clock time performance gap between the Suzuki-Trotter and exact propagator methods would increase by another order of magnitude in absolute time since all iterations take $L = 10$ longer.

The herein presented control-diagonal Trotter methods are thus prospectively useful not only for handling larger Hilbert spaces but also for more efficiently incorporating robustness.

Hessian. The exact Hessian has strong theoretical properties as discussed in Sec. II, and although our calculations of the exact Hessian have been verified numerically, we did not perform comparative studies. Nevertheless, we have found in a parallel similar work that a calculation of the exact Hessian within the diagonalization paradigm [54] outperforms a gradient-only quasi-Newton approach in terms of statistics and best results in certain domains. This suggests similar possibilities in the present case.

Discretize-then-optimize. Finally, we point out that the results in this paper followed a discretize-then-optimize (time discretization before ordinary vector derivatives of a cost function) rather than optimize-then-discretize [48] (time discretization after continuous Gâteaux derivatives of a cost functional) approach. Since these approaches in general do not necessarily yield the same derivative expressions, the former approach is preferable because it specifically takes into account the chosen propagation scheme implementation and the derivatives match the landscape equivalently dynamics which has been a main point throughout this paper.[5] It is therefore quite fortuitous that, e.g., (i) the exact propagator gradient ∇J_F^{ex} calculated by the optimize-then-discretize approach [48] yields exactly the same expression as $\nabla J_F^{\text{ST}_1}$ and (ii) $\hat{\mathcal{U}}^{\text{ST}_1}$ is a standard propagator for some systems, e.g., for wave functions in real space [62,65,69]. The combined effect is that the sought-after gradient exactness is obtained by virtue of standard methods alone in these situations. The same would not be true if either $\hat{\mathcal{U}}^{\text{ex}}$ or $\hat{\mathcal{U}}^{\text{ST}_2}$ was used [cf. Eqs. (22)].

[5]Similarly, application of Krylov subspace methods for the time evolution while using exact gradients for the exact propagator constitutes another potential mismatch between the derivatives and landscape equivalently dynamics.

ACKNOWLEDGMENTS

We thank I. Kuprov for useful discussions and C. A. Weidner and M. Dalgaard for feedback. This work was funded by the ERC, H2020 Grant No. 639560 (MECTRL), and the John Templeton and Carlsberg Foundations.

APPENDIX: DERIVATION OF EXACT GRADIENTS AND HESSIANS

Here we present the calculations leading to the exact gradient and Hessian expressions for one of the exact exponentiation propagators and the two Trotterized propagators with a diagonal control Hamiltonian defined in Eqs. (16). We also define the regularization cost functionals and likewise calculate their derivatives after discretization. Emphasis is put on thoroughness of the steps, and relevant equations for the derivations are restated for convenience where applicable so as to be self-contained.

We assume that the control $u(t)$ is discretized on a regularly spaced time grid $t \in [t_1, t_2, \ldots, t_{N_t}] = [0, \delta t, \ldots, T]$. Recall that $|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n |\psi_n\rangle$ with $|\psi_1\rangle = |\psi_{\text{ini}}\rangle$ and $|\psi_{N_t}\rangle = |\psi(T)\rangle$, and $\hat{H}(u_n) = \hat{H}_n = \hat{H}_n^d + \hat{H}_n^c$, where $\hat{H}_n^c$ ($\hat{H}_n^d$) is the control (drift) Hamiltonian. Define the auxiliary state $|\chi_{N_t}\rangle = |\psi_{\text{tgt}}\rangle$ with $|\chi_n\rangle = \hat{\mathcal{U}}_n^\dagger |\chi_{n+1}\rangle$. The derivatives given in Eqs. (9) are to be evaluated for the propagators $\hat{\mathcal{U}}^{\text{ex}} = \hat{\mathcal{U}}^{\text{ex}_2}$, $\hat{\mathcal{U}}^{\text{ST}_1}$, and $\hat{\mathcal{U}}^{\text{ST}_2}$ given in Eqs. (16).

As mentioned in the main text, the following results are trivially extended to situations other than pure state transfer: The central calculations of Eqs. (7) are the same. The primary differences lie in how these enter a given cost functional and if the cached objects are states or either unitary or density matrices.

1. Derivatives for exact propagators

The exact exponentiation propagator has the form $\hat{\mathcal{U}}_n^{\text{ex}} = \exp[-i\hat{H}(u_n)\delta t]$ and the derivative overlap reads

$$\frac{\partial o}{\partial u_n} = \langle \chi_{N_t} \mid \frac{\partial}{\partial u_n}\big(\hat{\mathcal{U}}_{N_t-1}^{\text{ex}} \cdots \hat{\mathcal{U}}_n^{\text{ex}} \cdots \hat{\mathcal{U}}_1^{\text{ex}}\big)|\psi_1\rangle$$
$$= \langle \chi_{n+1}| \frac{\partial \hat{\mathcal{U}}_n^{\text{ex}}}{\partial u_n} |\psi_n\rangle. \tag{A1}$$

The task is then to calculate $\frac{\partial \hat{\mathcal{U}}_n^{\text{ex}}}{\partial u_n}$ and be careful with ordering. The ex superscript is omitted for brevity in most of the steps below. We expand the exponential as

$$\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} = \frac{\partial}{\partial u_n}(e^{-i\hat{H}_n\delta t})$$
$$= \sum_{p=0}^{\infty} \frac{(-i\delta t)^p}{p!} \frac{\partial}{\partial u_n}\big(\hat{H}_n^p\big)$$
$$= \sum_{p=1}^{\infty} \frac{(-i\delta t)^p}{p!} \sum_{q=0}^{p-1} \hat{H}_n^q \left(\frac{\partial \hat{H}_n}{\partial u_n}\right) \hat{H}^{p-q-1}. \tag{A2}$$

Define for momentary simplicity $A \equiv -i\hat{H}_n\delta t$ and $B \equiv -i\frac{\partial \hat{H}_n}{\partial u_n}\delta t$. Then it can be shown that

$$\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{A^p B A^q}{(p+q+1)!}. \tag{A3}$$

Using now the relations for the beta and Gamma functions [70]

$$\Gamma(a) = (n-1)! \quad \text{for } a \in \mathbb{Z}^+, \tag{A4}$$

$$\begin{aligned}\beta(a,b) &= \int_0^1 (1-\alpha)^{a-1}\alpha^{b-1}d\alpha \\ &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ &= \frac{(a-1)!(b-1)!}{(a+b-1)!} \\ &= \frac{p!q!}{(p+q+1)!}\end{aligned} \tag{A5}$$

and taking $a = p+1$ and $b = q+1$, we obtain

$$\frac{1}{(p+q+1)!} = \frac{1}{p!q!}\int_0^1 (1-\alpha)^p\alpha^q d\alpha. \tag{A6}$$

Inserting this and initially pulling out the integral, we obtain

$$\begin{aligned}\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} &= \sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{A^p B A^q}{p!q!}\int_0^1 \alpha^p(1-\alpha)^q d\alpha \\ &= \int_0^1\left(\sum_{p=0}^{\infty}\frac{((1-\alpha)A)^p}{p!}\right)B\sum_{q=0}^{\infty}\frac{(\alpha A)^q}{q!}d\alpha \\ &= \int_0^1 e^{(1-\alpha)A}Be^{\alpha A}d\alpha \\ &= e^A\int_0^1 e^{-\alpha A}Be^{\alpha A}d\alpha \\ &= \hat{\mathcal{U}}_n\int_0^1 e^{(i\alpha\delta t)\hat{H}_n}\left(-i\delta t\frac{\partial \hat{H}_n}{\partial u_n}\right)e^{-(i\alpha\delta t)\hat{H}_n}d\alpha.\end{aligned} \tag{A7}$$

The integrand can be evaluated by defining the recursive commutator in Eq. (19) with the base case $[c_xX, c_yY]_0 = c_yY$ and using Baker-Campbell-Hausdorff relations [71]

$$[c_xX, c_yY]_k = [c_xX, [c_xX, c_yY]_{k-1}] = c_x^k c_y[X,Y]_k, \tag{A8}$$

$$e^{c_xX}Ye^{-c_xX} = \sum_{k=0}^{\infty}\frac{[c_xX, c_yY]_k}{k!} = \sum_{k=0}^{\infty}\frac{c_x^k c_y}{k!}[X,Y]_k; \tag{A9}$$

by evaluating these with scalars $c_x = i\alpha\delta t$ and $c_y = -i\delta t$, we obtain

$$\begin{aligned}\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} &= \hat{\mathcal{U}}_n\int_0^1\left(\sum_{k=0}^{\infty}\frac{(i\alpha\delta t)^k(-i\delta t)}{k!}\left[\hat{H}_n, \frac{\partial \hat{H}_n}{\partial u_n}\right]_k d\alpha\right) \\ &= \hat{\mathcal{U}}_n\sum_{k=0}^{\infty}(-i\delta t)\frac{i^k\delta t^k}{k!}\left[\hat{H}_n, \frac{\partial \hat{H}_n}{\partial u_n}\right]_k\left(\int_0^1\alpha^k d\alpha\right) \\ &= \hat{\mathcal{U}}_n(-i\delta t)\sum_{k=0}^{\infty}\frac{i^k\delta t^k}{(k+1)!}\left[\hat{H}_n, \frac{\partial \hat{H}_n}{\partial u_n}\right]_k.\end{aligned} \tag{A10}$$

Substituting this into Eq. (A1), then inserting the resulting expression into Eqs. (9), and using $\frac{\partial \hat{H}_n}{\partial u_n} = \frac{\partial \hat{H}_n^c}{\partial u_n}$ gives

$$\begin{aligned}\frac{\partial J_F^{\text{ex}}}{\partial u_n} &= -\text{Re}\left(o^*\langle\chi_{n+1}|\frac{\partial \hat{\mathcal{U}}_n^{\text{ex}}}{\partial u_n}|\psi_n\rangle\right) \\ &= \text{Re}\left\{io^*\langle\chi_n|\left(\sum_{k=0}^{\infty}\frac{i^k\delta t^k}{(k+1)!}\left[\hat{H}_n, \frac{\partial \hat{H}_n^c}{\partial u_n}\right]_k\right)|\psi_n\rangle\right\}\delta t \\ &= \text{Re}\left(io^*\langle\chi_n|\frac{\partial \hat{H}_n^c}{\partial u_n}|\psi_n\rangle\right)\delta t + O_{\nabla J_F^{\text{ex}}}(\delta t^2),\end{aligned} \tag{A11}$$

which is the expression stated in Eqs. (22a) and (23).

With the gradient in hand, the Hessian calculation only needs additional evaluation of the second derivatives of o,

$$\begin{aligned}\frac{\partial^2 o}{\partial u_n\partial u_m} &= \langle\chi_{N_t}|\hat{\mathcal{U}}_{N_t-1}\cdots\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n}\cdots\frac{\partial \hat{\mathcal{U}}_m}{\partial u_m}\cdots\hat{\mathcal{U}}_1|\psi_1\rangle \\ &= \langle\chi_{n+1}|\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n}\left(\prod_{j=m+1}^{n-1}\hat{\mathcal{U}}_j\right)\frac{\partial \hat{\mathcal{U}}_m}{\partial u_m}|\psi_m\rangle \quad \text{for } n > m,\end{aligned} \tag{A12a}$$

$$\begin{aligned}\frac{\partial^2 o}{\partial u_n\partial u_m} &= \langle\chi_{N_t}|\hat{\mathcal{U}}_{N_t-1}\cdots\frac{\partial^2 \hat{\mathcal{U}}_n}{\partial u_n^2}\cdots\hat{\mathcal{U}}_1|\psi_1\rangle \\ &= \langle\chi_{n+1}|\frac{\partial^2 \hat{\mathcal{U}}_n}{\partial u_n^2}|\psi_n\rangle \quad \text{for} \quad n = m.\end{aligned} \tag{A12b}$$

The case $m > n$ is the same as $n > m$ with indices $n \rightleftarrows m$ and we thus need only calculate one of the cases due to this symmetry. Inserting these expressions in Eqs. (9), we obtain the exact Hessian elements $n \geqslant m$ without loss of generality for the exact propagator $\hat{\mathcal{U}}_n = \hat{\mathcal{U}}_n^{\text{ex}}$,

$$\begin{aligned}&\frac{\partial^2 J_F^{\text{ex}}}{\partial u_n\partial u_m} \\ &= -\text{Re}\left[\left(\langle\chi_{m+1}|\frac{\partial \hat{\mathcal{U}}_m}{\partial u_m}|\psi_m\rangle\right)^*\langle\chi_{n+1}|\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n}|\psi_n\rangle\right] \\ &\quad -\text{Re}\left[o^*\langle\chi_{n+1}|\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n}\left(\prod_{j=m+1}^{n-1}\hat{\mathcal{U}}_j\right)\frac{\partial \hat{\mathcal{U}}_m}{\partial u_m}|\psi_m\rangle\right](1-\delta_{n,m}) \\ &\quad -\text{Re}\left(o^*\langle\chi_{n+1}|\frac{\partial^2 \hat{\mathcal{U}}_n}{\partial u_n^2}|\psi_n\rangle\right)\delta_{n,m},\end{aligned} \tag{A13}$$

where $\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n}$ is given by Eq. (A10). Note that the third and second terms appear only on the diagonal and off-diagonal, respectively. In optimization contexts the propagator gradient is also always computed and those elements can thus be reused here in practical applications. Evaluating $\frac{\partial^2 \hat{\mathcal{U}}_n}{\partial u_n^2}$ is straightforward and gives

$$\begin{aligned}\frac{\partial^2 \hat{\mathcal{U}}_n}{\partial u_n^2} = \hat{\mathcal{U}}_n\Bigg\{&\left(-i\delta t\sum_{k=0}^{\infty}\frac{i^k\delta t^k}{(k+1)!}\left[\hat{H}_n, \frac{\partial \hat{H}_n^c}{\partial u_n}\right]_k\right)^2 \\ &-i\delta t\sum_{k=0}^{\infty}\frac{i^k\delta t^k}{(k+1)!}\frac{\partial}{\partial u_n}\left(\left[\hat{H}_n, \frac{\partial \hat{H}_n^c}{\partial u_n}\right]_k\right)\Bigg\},\end{aligned} \tag{A14}$$

but the recursive commutator derivative is cumbersome

$$\begin{aligned}
[\hat{H}_n, \hat{H}_n']_k' &= [\hat{H}_n, [\hat{H}_n, \hat{H}_n']_{k-1}'] - [\hat{H}_n', [\hat{H}_n, \hat{H}_n']_{k-1}] \\
[\hat{H}_n, \hat{H}_n']_0' &= \hat{H}_n'', \\
[\hat{H}_n, \hat{H}_n']_1' &= [\hat{H}_n, \hat{H}_n''], \\
[\hat{H}_n, \hat{H}_n']_2' &= [\hat{H}_n, [\hat{H}_n, \hat{H}_n'']] + [\hat{H}_n', [\hat{H}_n, \hat{H}_n']], \\
&\vdots
\end{aligned} \tag{A15}$$

where we explicitly evaluated the first few terms and $\hat{H}_n' \equiv \partial \hat{H}_n / \partial u_n$. Note that the exact derivatives entail an infinite summation or to machine precision for finite arithmetic.

2. Derivatives for Trotterized propagators

We consider now in turn the Suzuki-Trotter expansions $\hat{\mathcal{U}}_n^{\mathrm{ST}_1} = \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2}$ and $\hat{\mathcal{U}}_n^{\mathrm{ST}_2} = \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2}$, where $\hat{\mathcal{U}}_n^{c/2} \equiv \exp(-i\hat{H}_n^c \delta t/2)$ and $\hat{\mathcal{U}}_n^d$ is given by Eq. (16f).

a. Derivatives of $\hat{\mathcal{U}}_n^{\mathrm{ST}_1}$

For the Suzuki-Trotter expansion $\hat{\mathcal{U}}_n^{\mathrm{ST}_1} = \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2}$, the u_n control dependence is distributed among n and $n-1$ (except at the end points $n = 1, N$), yielding

$$\begin{aligned}
\frac{\partial o}{\partial u_n} &= \langle \chi_{N_t} | \frac{\partial}{\partial u_n} \big(\hat{\mathcal{U}}_{N_t-1}^{\mathrm{ST}_1} \cdots \hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1} \cdots \hat{\mathcal{U}}_1^{\mathrm{ST}_1} \big) | \psi_1 \rangle \\
&= \langle \chi_{n+1} | \frac{\partial}{\partial u_n} \big(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1} \big) | \psi_{n-1} \rangle.
\end{aligned} \tag{A16}$$

Additionally, assume that the control Hamiltonian is diagonal. We use Eq. (A10) to take the derivative of $(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1})$,

$$\begin{aligned}
\frac{\partial}{\partial u_n} \big(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1} \big) &= \frac{\partial}{\partial u_n} \big[\big(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \big) \big(\hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \big) \big] \\
&= \big(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \big) \frac{\partial \hat{\mathcal{U}}_n^c}{\partial u_n} \big(\hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \big) \\
&= \big(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \big) \left\{ \hat{\mathcal{U}}_n^c \left(-i\delta t \frac{\partial \hat{H}_n^c}{\partial u_n} \right) \right\} \big(\hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \big) \\
&= (-i\delta t) \big(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \big) \frac{\partial \hat{H}_n^c}{\partial u_n} \big(\hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \big) \\
&= (-i\delta t) \hat{\mathcal{U}}_n^{\mathrm{ST}_1} \frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1}.
\end{aligned} \tag{A17}$$

Here we also used the fact that two diagonal matrices always commute, first to evaluate the recursive commutator $[\hat{H}_n^c, \frac{\partial \hat{H}_n^c}{\partial u_n}]_k = \frac{\partial \hat{H}_n^c}{\partial u_n} \delta_{0,k}$ from Eq. (A10) and second to recombine the initial propagators since $[\frac{\partial \hat{H}_n^c}{\partial u_n}, \hat{\mathcal{U}}_n^{c/2}] = 0$. Inserting this into Eqs. (A16) and (9) yields the exact gradient stated in Eq. (22b),

$$\frac{\partial J_F^{\mathrm{ST}_1}}{\partial u_n} = \mathrm{Re}\left(i o^* \langle \chi_n | \frac{\partial H_n^c}{\partial u_n} | \psi_n \rangle \right) \delta t, \tag{A18}$$

with an additional factor $\frac{1}{2}$ at the end points ($n = 1, N_t$). Apart from these, this is identical in structure to the exact propagator gradient (22a) when in the first-order approximation $k_{\max} = 0$, i.e., discarding the otherwise expensive $O(\delta t^2)$ tail.

The second derivatives to be calculated for the Hessian are

$$\frac{\partial^2 o}{\partial u_n \partial u_m} = \langle \chi_{n+1} | \frac{\partial}{\partial u_n} \big(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1} \big) \left(\prod_{j=m+1}^{n-2} \hat{\mathcal{U}}_j \right) \frac{\partial}{\partial u_m} \big(\hat{\mathcal{U}}_m^{\mathrm{ST}_1} \hat{\mathcal{U}}_{m-1}^{\mathrm{ST}_1} \big) | \psi_{m-1} \rangle \quad \text{for} \quad n > m, \tag{A19a}$$

$$\frac{\partial^2 o}{\partial u_n \partial u_m} = \langle \chi_{n+1} | \frac{\partial^2}{\partial u_n^2} \big(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1} \big) | \psi_{n-1} \rangle \quad \text{for} \quad n = m. \tag{A19b}$$

For $n > m$ we only need the first derivative given in Eq. (A17). For $n = m$ we take the second derivative of $(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1})$ using Eq. (A17) and $[\frac{\partial \hat{H}_n^c}{\partial u_n}, \hat{\mathcal{U}}_n^{c/2}] = 0$,

$$\begin{aligned}
\frac{\partial^2}{\partial u_n^2} \big(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1} \big) &= (-i\delta t) \frac{\partial}{\partial u_n} \left(\hat{\mathcal{U}}_n^{\mathrm{ST}_1} \frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1} \right) \\
&= (-i\delta t) \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \frac{\partial}{\partial u_n} \left(\hat{\mathcal{U}}_n^c \frac{\partial \hat{H}_n^c}{\partial u_n} \right) \hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \\
&= (-i\delta t) \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \left(\hat{\mathcal{U}}_n^c \left(-i\delta t \frac{\partial \hat{H}_n^c}{\partial u_n} \right) \frac{\partial \hat{H}_n^c}{\partial u_n} + \hat{\mathcal{U}}_n^c \frac{\partial^2 \hat{H}_n^c}{\partial u_n^2} \right) \hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \\
&= (-i\delta t) \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \left(\frac{\partial^2 \hat{H}_n^c}{\partial u_n^2} - i\delta t \left(\frac{\partial \hat{H}_n^c}{\partial u_n} \right)^2 \right) \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \\
&= (-i\delta t) \hat{\mathcal{U}}_n^{\mathrm{ST}_1} \left(\frac{\partial^2 \hat{H}_n^c}{\partial u_n^2} - i\delta t \left(\frac{\partial \hat{H}_n^c}{\partial u_n} \right)^2 \right) \hat{\mathcal{U}}_{n-1}^{\mathrm{ST}_1}.
\end{aligned} \tag{A20}$$

Inserting into Eqs. (9), the exact Hessian elements $n \geqslant m$ for this Trotterization scheme are therefore

$$\frac{\partial^2 J_F^{\mathrm{ST}_1}}{\partial u_n \partial u_m} = -\mathrm{Re}\left(\langle\psi_m| \frac{\partial \hat{H}_m^c}{\partial u_m} |\chi_m\rangle \langle\chi_n| \frac{\partial \hat{H}_n^c}{\partial u_n} |\psi_n\rangle \right) \delta t^2 + \mathrm{Re}\left[o^* \langle\chi_n| \frac{\partial \hat{H}_n^c}{\partial u_n} \left(\prod_{j=m}^{n-1} \hat{\mathcal{U}}_j^{\mathrm{ST}_1} \right) \frac{\partial \hat{H}_m^c}{\partial u_m} |\psi_m\rangle \right] (1-\delta_{n,m}) \delta t^2$$
$$+ \mathrm{Re}\left\{ i o^* \langle\chi_n| \left[\frac{\partial^2 \hat{H}_n^c}{\partial u_n^2} - i\delta t \left(\frac{\partial \hat{H}_n^c}{\partial u_n} \right)^2 \right] |\psi_n\rangle \right\} \delta_{n,m} \delta t. \tag{A21}$$

Derivatives of the end points corresponding to the outer "rim" of the Hessian matrix carry an additional factor $\frac{1}{2}$ each, for a total of $\frac{1}{4}$ in the corners and $\frac{1}{2}$ on the edges. Note that the third and second terms appear only on the diagonal and off-diagonal, respectively. As with the gradient, the Hessian is similarly identical to Eq. (A13) when retaining only the $k = k_{\max} = 0$ term. As an implementation detail, note that the propagated states and operator-state products from (22b) may be reused here. The second term is the most costly to evaluate because of additional state propagations. The order of evaluation should be done row by row to further increase reusability of computations.

b. Derivatives of $\hat{\mathcal{U}}_n^{\mathrm{ST}_2}$

The Suzuki-Trotter expansion reads $\hat{\mathcal{U}}_n^{\mathrm{ST}_2} = \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2}$ and the overlap derivative reads

$$\frac{\partial o}{\partial u_n} = \frac{\partial}{\partial u_n} (\hat{\mathcal{U}}_{N_t-1}^{\mathrm{ST}_2} \cdots \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \cdots \hat{\mathcal{U}}_1^{\mathrm{ST}_2}) = \langle\chi_{n+1}| \frac{\partial \hat{\mathcal{U}}_n^{\mathrm{ST}_2}}{\partial u_n} |\psi_n\rangle. \tag{A22}$$

Invoking Eq. (A10) for $\hat{\mathcal{U}}_n^{c/2}$, we find

$$\frac{\partial \hat{\mathcal{U}}_n^{\mathrm{ST}_2}}{\partial u_n} = \frac{\partial \hat{\mathcal{U}}_n^{c/2}}{\partial u_n} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} + \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \frac{\partial \hat{\mathcal{U}}_n^{c/2}}{\partial u_n}$$
$$= -\frac{i\delta t}{2} \left(\hat{\mathcal{U}}_n^{c/2} \frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} + \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \frac{\partial \hat{H}_n^c}{\partial u_n} \right)$$
$$= -\frac{i\delta t}{2} \left(\frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_n^{\mathrm{ST}_2} + \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \frac{\partial \hat{H}_n^c}{\partial u_n} \right). \tag{A23}$$

Here we also used that two diagonal matrices always commute, first to evaluate the recursive commutator $[\hat{H}_n^c, \frac{\partial \hat{H}_n^c}{\partial u_n}]_k = \frac{\partial \hat{H}_n^c}{\partial u_n} \delta_{0,k}$ from Eq. (A10) and second to recombine the initial propagators since $[\frac{\partial \hat{H}_n^c}{\partial u_n}, \hat{\mathcal{U}}_n^{c/2}] = 0$. Substituting back into Eqs. (A22) and (9), we find

$$\frac{\partial o}{\partial u_n} = -\frac{i\delta t}{2} \langle\chi_{n+1}| \left(\frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_n^{\mathrm{ST}_2} + \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \frac{\partial \hat{H}_n^c}{\partial u_n} \right) |\psi_n\rangle$$
$$= -\frac{i\delta t}{2} \left(\langle\chi_{n+1}| \frac{\partial \hat{H}_n^c}{\partial u_n} |\psi_{n+1}\rangle + \langle\chi_n| \frac{\partial \hat{H}_n^c}{\partial u_n} |\psi_n\rangle \right)$$
$$= -\frac{i\delta t}{2} \sum_{p=n}^{n+1} \langle\chi_p| \frac{\partial \hat{H}_n^c}{\partial u_n} |\psi_p\rangle \tag{A24}$$
$$\Rightarrow \frac{\partial J_F^{\mathrm{ST}_2}}{\partial u_n} = \mathrm{Re}\left(\frac{i o^*}{2} \sum_{p=n}^{n+1} \langle\chi_p| \frac{\partial H_n^c}{\partial u_n} |\psi_p\rangle \right) \delta t \tag{A25}$$

for all $n = 1, \ldots, N_t - 1$, which is the expression in Eq. (22c).

After some lines of calculation, the second derivatives of o for the Hessian evaluate to

$$\frac{\partial^2 o}{\partial u_n \partial u_m} = -\frac{\delta t^2}{4} \sum_{p=n}^{n+1} \sum_{q=m}^{m+1} \langle\chi_p| \frac{\partial \hat{H}_n^c}{\partial u_n} \left(\prod_{j=q}^{p-1} \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \right) \times \frac{\partial \hat{H}_m^c}{\partial u_m} |\psi_q\rangle \quad \text{for} \quad n > m, \tag{A26a}$$

$$\frac{\partial^2 o}{\partial u_n \partial u_m} = -\frac{i\delta t}{2} \left\{ \sum_{j=n}^{n+1} \langle\chi_j| \left[\frac{\partial^2 \hat{H}_n^c}{\partial u_n^2} - \frac{i\delta t}{2} \left(\frac{\partial \hat{H}_n^c}{\partial u_n} \right)^2 \right] |\psi_j\rangle - i\delta t \langle\chi_{n+1}| \frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \frac{\partial \hat{H}_n^c}{\partial u_n} |\psi_n\rangle \right\} \quad \text{for } n = m. \tag{A26b}$$

Substituting Eqs. (A24), (A26a), and (A26b) into Eq. (9) yields the final result

$$\frac{\partial^2 J_F^{\mathrm{ST}_2}}{\partial u_n \partial u_m} = \mathrm{Re}\left(\frac{1}{4} \sum_{p=n}^{n+1} \sum_{q=m}^{m+1} \langle\psi_q| \frac{\partial \hat{H}_m^c}{\partial u_m} |\chi_q\rangle \langle\chi_p| \frac{\partial \hat{H}_n^c}{\partial u_n} |\psi_p\rangle \delta t^2 \right.$$
$$+ \frac{o^*}{4} \sum_{p=n}^{n+1} \sum_{q=m}^{m+1} \langle\chi_p| \frac{\partial \hat{H}_n^c}{\partial u_n} \prod_{j=q}^{p-1} \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \frac{\partial \hat{H}_m^c}{\partial u_m} |\psi_q\rangle (1-\delta_{n,m}) \delta t^2$$
$$+ \frac{i o^*}{2} \left\{ \sum_{j=n}^{n+1} \langle\chi_j| \left[\frac{\partial^2 \hat{H}_n^c}{\partial u_n^2} - \frac{i\delta t}{2} \left(\frac{\partial \hat{H}_n^c}{\partial u_n} \right)^2 \right] |\psi_j\rangle \right.$$
$$\left. \left. - i\delta t \langle\chi_{n+1}| \frac{\partial \hat{H}_n^c}{\partial u_n} \hat{\mathcal{U}}_n^{\mathrm{ST}_2} \frac{\partial \hat{H}_n^c}{\partial u_n} |\psi_n\rangle \right\} \delta_{n,m} \delta t \right). \tag{A27}$$

The Hessian expression for $\hat{\mathcal{U}}_n^{\mathrm{ST}_2}$ is thus much more cumbersome than that for $\hat{\mathcal{U}}_n^{\mathrm{ST}_1}$.

3. Derivatives for regularizations

In many applications, it is advantageous to regularize the control amplitude and/or its temporal derivative. This requires additional terms in the cost functional objective, imposition of discretization, and calculation of the respective derivatives. As with the propagator, the chosen form of the implementation scheme changes the derivative calculations.

The amplitude regularization is straightforward,

$$J_\alpha = \frac{\alpha}{2}\int_0^T u(t)^2 dt \to \frac{\alpha}{2}\delta t \sum_{i=1}^{N_t} u_i^2, \tag{A28}$$

$$\frac{\partial J_\alpha}{\partial u_n} = \alpha \delta t u_n, \quad \frac{\partial^2 J_\alpha}{\partial u_m \partial u_n} = \alpha \delta t \delta_{n,m}, \tag{A29}$$

where α is a weighting factor. The derivative regularization is a bit more involved because of the end points

$$J_\gamma = \frac{\gamma}{2}\int_0^T \dot{u}(t)^2 dt \to \frac{\gamma}{8\delta t}\left((-3u_1 + 4u_2 - u_3)^2 + \sum_{i=2}^{N_t-1}(u_{i+1} - u_{i-1})^2 + (3u_{N_t} - 4u_{N_t-1} + u_{N_t-2})^2\right), \tag{A30}$$

where we used forward (backward) difference approximations for the first (last) point and center approximations for the bulk, all to $O(\delta t^2)$. The derivative with respect to the first and last three indices is different from the bulk. The resulting gradient written in vector form is

$$\nabla J_\gamma = \frac{\gamma}{4\delta t}\begin{bmatrix} 10u_1 - 12u_2 + 2u_3 \\ -12u_1 + 17u_2 - 4u_3 - u_4 \\ 2u_1 - 4u_2 + 3u_3 - u_5 \\ \vdots \\ 2u_n - u_{n-2} + u_{n+2} \\ \vdots \\ 2u_{N_t} - 4u_{N_t-1} + 3u_{N_t-2} - u_{N_t-4} \\ -12u_{N_t} + 17u_{N_t-1} - 4u_{N_t-2} - u_{N_t-3} \\ 10u_{N_t} - 12u_{N_t-1} + 2u_{N_t-2} \end{bmatrix}, \tag{A31}$$

where the vertical dots extend over the bulk points. Similarly, the three first and last Hessian rows are different from the bulk. In a stacking notation where the indices denote the rows, we obtain

$$\nabla^2 J_\gamma = \left[\begin{array}{c} [\nabla^2 J_\gamma]_{1:3s} \\ \hline [\nabla^2 J_\gamma]_{4:N_t-3} \\ \hline [\nabla^2 J_\gamma]_{N_t-2:N_t} \end{array}\right], \tag{A32}$$

where the matrices evaluate to

$$[\nabla^2 J_\gamma]_{1:3} = \frac{\gamma}{4\delta t}\begin{bmatrix} 10 & -12 & 2 & 0 & 0 & 0 & \cdots \\ -12 & 17 & -4 & -1 & 0 & 0 & \cdots \\ 2 & -4 & 3 & 0 & -1 & 0 & \cdots \end{bmatrix}, \tag{A33}$$

$$[\nabla^2 J_\gamma]_{4:N_t-3} = \frac{\gamma}{4\delta t}\begin{bmatrix} 0 & -1 & 0 & 2 & 0 & -1 & 0 & 0 & \cdots \\ & & \ddots & & \ddots & & \ddots & & \\ \cdots & 0 & 0 & -1 & 0 & 2 & 0 & -1 & 0 \end{bmatrix}, \tag{A34}$$

$$[\nabla^2 J_\gamma]_{N_t-2:N_t} = \frac{\gamma}{4\delta t}\begin{bmatrix} \cdots & 0 & -1 & 0 & 3 & -4 & 2 \\ \cdots & 0 & 0 & -1 & -4 & 17 & -12 \\ \cdots & 0 & 0 & 0 & 2 & -12 & 10 \end{bmatrix}, \tag{A35}$$

with the dots denoting continuation of the number they point to.

[1] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen *et al.*, Training Schrödinger's cat: Quantum optimal control, Eur. Phys. J. D **69**, 279 (2015).

[2] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits, Phys. Rev. Lett. **103**, 110501 (2009).

[3] D. J. Egger and F. K. Wilhelm, Optimized controlled-Z gates for two superconducting qubits coupled through a resonator, Supercond. Sci. Technol. **27**, 014001 (2013).

[4] M. H. Goerz, F. Motzoi, K. B. Whaley, and C. P. Koch, Charting the circuit QED design landscape using optimal control theory, npj Quantum Inf. **3**, 37 (2017).

[5] M. Dalgaard, F. Motzoi, J. J. Sørensen, and J. Sherson, Global optimization of quantum dynamics with alphazero deep exploration, npj Quantum Inf. **6**, 6 (2020).

[6] C. T. Kehlet, A. C. Sivertsen, M. Bjerring, T. O. Reiss, N. Khaneja, S. J. Glaser, and N. C. Nielsen, Improving solid-state NMR dipolar recoupling by optimal control, J. Am. Chem. Soc. **126**, 10202 (2004).

[7] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms, J. Magn. Reson. **172**, 296 (2005).

[8] N. C. Nielsen, C. Kehlet, S. J. Glaser, and N. Khaneja, Optimal control methods in NMR spectroscopy, in *eMagRes*, edited by R. K. Harris and R. L. Wasylishen (Wiley, New York, 2010).

[9] W. Kallies and S. J. Glaser, Cooperative broadband spin echoes through optimal control, J. Magn. Reson. **286**, 115 (2018).

[10] J. J. Sørensen, J. S. Nyemann, F. Motzoi, J. Sherson, and T. Vosegaard, Optimization of pulses with low bandwidth for improved excitation of multiple-quantum coherences in NMR of quadrupolar nuclei, J. Chem. Phys. **152**, 054104 (2020).

[11] J. Scheuer, X. Kong, R. S. Said, J. Chen, A. Kurz, L. Marseglia, J. Du, P. R. Hemmer, S. Montangero, T. Calarco *et al.*, Precise qubit control beyond the rotating wave approximation, New J. Phys. **16**, 093022 (2014).

[12] F. Dolde, V. Bergholm, Y. Wang, I. Jakobi, B. Naydenov, S. Pezzagna, J. Meijer, F. Jelezko, P. Neumann, T. Schulte-Herbrüggen *et al.*, High-fidelity spin entanglement using optimal control, Nat. Commun. **5**, 3371 (2014).

[13] G. Waldherr, Y. Wang, S. Zaiser, M. Jamali, T. Schulte-Herbrüggen, H. Abe, T. Ohshima, J. Isoya, J. Du, P. Neumann *et al.*, Quantum error correction in a solid-state hybrid spin register, Nature (London) **506**, 204 (2014).

[14] Y. Chou, S.-Y. Huang, and H.-S. Goan, Optimal control of fast and high-fidelity quantum gates with electron and nuclear spins of a nitrogen-vacancy center in diamond, Phys. Rev. A **91**, 052315 (2015).

[15] C. P. Koch, J. P. Palao, R. Kosloff, and F. Masnou-Seeuws, Stabilization of ultracold molecules using optimal control theory, Phys. Rev. A **70**, 013402 (2004).

[16] E. F. De Lima, T.-S. Ho, and H. Rabitz, Optimal laser control of molecular photoassociation along with vibrational stabilization, Chem. Phys. Lett. **501**, 267 (2011).

[17] K. M. Tibbetts, X. Xing, and H. Rabitz, Optimal control of molecular fragmentation with homologous families of photonic reagents and chemical substrates, Phys. Chem. Chem. Phys. **15**, 18012 (2013).

[18] P. Doria, T. Calarco, and S. Montangero, Optimal Control Technique for Many-Body Quantum Dynamics, Phys. Rev. Lett. **106**, 190501 (2011).

[19] S. van Frank, M. Bonneau, J. Schmiedmayer, S. Hild, C. Gross, M. Cheneau, I. Bloch, T. Pichler, A. Negretti, T. Calarco *et al.*, Optimal control of complex atomic quantum systems, Sci. Rep. **6**, 34187 (2016).

[20] M. Mundt and D. J. Tannor, Optimal control of interacting particles: a multi-configuration time-dependent Hartree-Fock approach, New J. Phys. **11**, 105038 (2009).

[21] G. Jäger and U. Hohenester, Optimal quantum control of Bose-Einstein condensates in magnetic microtraps: Consideration of filter effects, Phys. Rev. A **88**, 035601 (2013).

[22] J. Cui, R. van Bijnen, T. Pohl, S. Montangero, and T. Calarco, Optimal control of Rydberg lattice gases, Quantum Sci. Technol. **2**, 035006 (2017).

[23] S. Patsch, D. M. Reich, J.-M. Raimond, M. Brune, S. Gleyzes, and C. P. Koch, Fast and accurate circularization of a Rydberg atom, Phys. Rev. A **97**, 053418 (2018).

[24] A. Larrouy, S. Patsch, R. Richaud, J.-M. Raimond, M. Brune, C. P. Koch, and S. Gleyzes, Fast Navigation in a Large Hilbert Space Using Quantum Optimal Control, Phys. Rev. X **10**, 021058 (2020).

[25] P. De Fouquieres, S. Schirmer, S. Glaser, and I. Kuprov, Second order gradient ascent pulse engineering, J. Magn. Reson. **212**, 412 (2011).

[26] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Comparing, optimizing, and benchmarking quantum-control algorithms in a unifying programming framework, Phys. Rev. A **84**, 022305 (2011).

[27] F. F. Floether, P. De Fouquieres, and S. G. Schirmer, Robust quantum gates for open systems via optimal control: Markovian versus non-Markovian dynamics, New J. Phys. **14**, 073023 (2012).

[28] D. Goodwin and I. Kuprov, Auxiliary matrix formalism for interaction representation transformations, optimal control, and spin relaxation theories, J. Chem. Phys. **143**, 084113 (2015).

[29] D. Goodwin and I. Kuprov, Modified Newton-Raphson GRAPE methods for optimal control of spin systems, J. Chem. Phys. **144**, 204107 (2016).

[30] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, Tunable, Flexible, and Efficient Optimization of Control Pulses for Practical Qubits, Phys. Rev. Lett. **120**, 150401 (2018).

[31] J. J. W. H. Sørensen, M. O. Aranburu, T. Heinzel, and J. F. Sherson, Quantum optimal control in a chopped basis: Applications in control of Bose-Einstein condensates, Phys. Rev. A **98**, 022119 (2018).

[32] D. J. Tannor, V. Kazakov, and V. Orlov, in *Time-Dependent Quantum Molecular Dynamics*, edited by J. Broeckhove and L. Lathouwers, NATO Advanced Studies Institute, Series B: Physics (Springer, Boston, 1992), Vol. 299, pp. 347–360.

[33] J. P. Palao and R. Kosloff, Quantum Computing by an Optimal Control Algorithm for Unitary Transformations, Phys. Rev. Lett. **89**, 188301 (2002).

[34] S. G. Schirmer and P. de Fouquieres, Efficient algorithms for optimal control of quantum dynamics: The Krotov method unencumbered, New J. Phys. **13**, 073029 (2011).

[35] T. Caneva, T. Calarco, and S. Montangero, Chopped random-basis quantum optimization, Phys. Rev. A **84**, 022326 (2011).

[36] J. H. M. Jensen, M. Gajdacz, S. Z. Ahmed, J. H. Czarkowski, C. Weidner, J. Rafner, J. J. Sørensen, K. Mølmer, and J. F. Sherson, Crowdsourcing human common sense for quantum control, Phys. Rev. Research **3**, 013057 (2021).

[37] X. Li, D. Pęcak, T. Sowiński, J. Sherson, and A. E. B. Nielsen, Global optimization for quantum dynamics of few-fermion systems, Phys. Rev. A **97**, 033602 (2018).

[38] J. J. Sørensen, M. Aranburu, T. Heinzel, and J. Sherson, Approaching the quantum speed limit with global-local optimization, arXiv:1802.07521.

[39] D. Dong and I. R. Petersen, Quantum control theory and applications: A survey, IET Control Theory Appl. **4**, 2651 (2010).

[40] F. Caruso, S. Montangero, T. Calarco, S. F. Huelga, and M. B. Plenio, Coherent optimal control of photosynthetic molecules, Phys. Rev. A **85**, 042331 (2012).

[41] I. Walmsley and H. Rabitz, Quantum physics under control, Phys. Today **56**(8), 43 (2003).

[42] S. Rosi, A. Bernard, N. Fabbri, L. Fallani, C. Fort, M. Inguscio, T. Calarco, and S. Montangero, Fast closed-loop optimal control of ultracold atoms in an optical lattice, Phys. Rev. A **88**, 021601(R) (2013).

[43] G. Feng, F. H. Cho, H. Katiyar, J. Li, D. Lu, J. Baugh, and R. Laflamme, Gradient-based closed-loop quantum optimal control in a solid-state two-qubit system, Phys. Rev. A **98**, 052341 (2018).

[44] R. Heck, O. Vuculescu, J. J. Sørensen, J. Zoller, M. G. Andreasen, M. G. Bason, P. Ejlertsen, O. Elíasson, P. Haikka, J. S. Laustsen, L. L. Nielsen, A. Mao, R. Müller, M. Napolitano, M. K. Pedersen, A. R. Thorsen, C. Bergenholtz, T. Calarco, S. Montangero, and J. F. Sherson, Remote optimization of an ultracold atoms experiment by experts and citizen scientists, Proc. Natl. Acad. Sci. U.S.A. **115**, E11231 (2018).

[45] A. Acín, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko *et al.*, The quantum technologies roadmap: A european community view, New J. Phys. **20**, 080201 (2018).

[46] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer Science+Business Media, New York, 2006).

[47] H. Hogben, M. Krzystyniak, G. Charnock, P. Hore, and I. Kuprov, *Spinach*–A software library for simulation of spin dynamics in large spin systems, J. Magn. Reson. **208**, 179 (2011).

[48] J. J. Sørensen, J. H. M. Jensen, T. Heinzel, and J. F. Sherson, QEngine: A C++ library for quantum optimal control of ultracold atoms, Comput. Phys. Commun. **243**, 135 (2019).

[49] C. Weitenberg, S. Kuhr, K. Mølmer, and J. F. Sherson, Quantum computation architecture using optical tweezers, Phys. Rev. A **84**, 032322 (2011).

[50] O. Mandel, M. Greiner, A. Widera, T. Rom, T. W. Hänsch, and I. Bloch, Coherent Transport of Neutral Atoms in Spin-Dependent Optical Lattice Potentials, Phys. Rev. Lett. **91**, 010407 (2003).

[51] C. Moler and C. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, SIAM Rev. **45**, 3 (2003).

[52] N. Hatano and M. Suzuki, in *Quantum Annealing and Other Optimization Methods*, edited by A. Das and B. K. Chakrabarti, Lecture Notes in Physics Vol. 679 (Springer, Berlin, 2005), pp. 37–68.

[53] R. Beerwerth and H. Bauke, Krylov subspace methods for the Dirac equation, Comput. Phys. Commun. **188**, 189 (2015).

[54] M. Dalgaard, F. Motzoi, J. H. M. Jensen, and J. Sherson, Hessian-based optimization of constrained quantum control, Phys. Rev. A **102**, 042612 (2020).

[55] H. Fehske, J. Schleede, G. Schubert, G. Wellein, V. S. Filinov, and A. R. Bishop, Numerical approaches to time evolution of complex quantum systems, Phys. Lett. A **373**, 2182 (2009).

[56] I. Kuprov, N. Wagner-Rundell, and P. Hore, Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction, J. Magn. Reson. **189**, 241 (2007).

[57] G. C. Hegerfeldt, Driving at the Quantum Speed Limit: Optimal Control of a Two-Level System, Phys. Rev. Lett. **111**, 260501 (2013).

[58] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, Theory of quantum annealing of an Ising spin glass, Science **295**, 2427 (2002).

[59] T. Caneva, T. Calarco, R. Fazio, G. E. Santoro, and S. Montangero, Speeding up critical system dynamics through optimized evolution, Phys. Rev. A **84**, 012312 (2011).

[60] T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Optimal Control at the Quantum Speed Limit, Phys. Rev. Lett. **103**, 240501 (2009).

[61] M. Larocca, P. M. Poggi, and D. A. Wisniacki, Quantum control landscape for a two-level system near the quantum speed limit, J. Phys. A: Math. Theor. **51**, 385305 (2018).

[62] G. Von Winckel and A. Borzì, Computational techniques for a quantum control problem with H^1-cost, Inverse Probl. **24**, 034007 (2008).

[63] M. Murphy, S. Montangero, V. Giovannetti, and T. Calarco, Communication at the quantum speed limit along a spin chain, Phys. Rev. A **82**, 022318 (2010).

[64] J. H. M. Jensen, F. S. Møller, J. J. Sørensen, and J. F. Sherson, Achieving fast high-fidelity control of many-body dynamics, arXiv:2008.06076.

[65] A. Borzì, G. Ciaramella, and M. Sprengel, *Computational Science & Engineering* (Society for Industrial and Applied Mathematics, Philadelphia, 2017), p. 397.

[66] T. J. Park and J. Light, Unitary quantum time evolution by iterative Lanczos reduction, J. Chem. Phys. **85**, 5870 (1986).

[67] M. Hochbruck and C. Lubich, On Krylov subspace approximations to the matrix exponential operator, SIAM J. Numer. Anal. **34**, 1911 (1997).

[68] M. H. Goerz, E. J. Halperin, J. M. Aytac, C. P. Koch, and K. B. Whaley, Robustness of high-fidelity Rydberg gates with single-site addressability, Phys. Rev. A **90**, 032329 (2014).

[69] U. Hohenester, P. K. Rekdal, A. Borzì, and J. Schmiedmayer, Optimal quantum control of Bose-Einstein condensates in magnetic microtraps, Phys. Rev. A **75**, 023602 (2007).

[70] M. L. Boas, *Mathematical Methods in the Physical Sciences* (Wiley, New York, 2006).

[71] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics* (Cambridge University Press, Cambridge, 2017).