Letter

Accelerated Newton-Raphson GRAPE methods for optimal control

David L. Goodwin [1,2,*] and Mads Sloth Vinding [3,†]

[1] *Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, United Kingdom*

[2] *Institute for Biological Interfaces 4 – Magnetic Resonance, Karlsruhe Institute for Technology (KIT), Karlsruhe, Fritz-Haber-Weg 6, 76131, Germany*

[3] *Center of Functionally Integrative Neuroscience (CFIN), Department of Clinical Medicine, Aarhus University, DK-8000, Aarhus, Denmark*

(Received 21 July 2022; accepted 22 December 2022; published 22 March 2023)

A Hessian-based state-to-state optimal control method in Liouville space is presented to mitigate previously undesirable polynomial scaling of Hessian computation time. This method, an improvement to the state-of-the-art Newton-Raphson gradient ascent pulse engineering (GRAPE) method, is derived with respect to two exact time-propagator derivative calculation techniques, auxiliary matrix and efficient spin control using analytical Lie algebraic derivatives (ESCALADE) methods. We observed that compared to the best current implementation of the Newton-Raphson GRAPE method, for an ensemble of two-level systems, with realistic conditions, our auxiliary matrix and ESCALADE Hessians can be 4–200 and 70–600 times faster, respectively. Additionally, the Newton-Raphson GRAPE method using ESCALADE is presented in a Liouville space for higher-level systems and with the derivation of x-, y-, and z-control propagator derivatives, also extending the application of ESCALADE and the recent quantum optimal control by adaptive low-cost algorithm (QOALA) method for coupled systems.

DOI: 10.1103/PhysRevResearch.5.L012042

I. INTRODUCTION

The problem of transferring the state of a dynamical system to a desired target state while minimizing the remaining distance and costs is often solved with optimal control theory [1–3]. Applications include quantum sensing [4–7], quantum computing [8–10], and nuclear magnetic resonance (NMR). spectroscopy [11–14] and imaging (MRI) [15,16].

A number of different approaches to optimal control (OC) has led to the development of different methods: Lagrangian methods [17–20]; minimal-time OC [11]; gradient ascent pulse engineering (GRAPE) [13]; sophisticated gradient-free searches [21,22]; Krylov-Newton methods [23]; OC with a basis of analytic controls [24,25]; and a tensor product approach for large quantum systems [26]. The method outlined in this Letter is based on a piecewise-constant control pulse approximation [27–30] of GRAPE [13,31,32] using a gradient-following numerical optimization.

Although finding an optimal solution to the problem of controlling a single two-level system from a defined initial system state to a desired target state, a *state-to-state* problem, is considered straightforward and computationally inexpensive with modern methods and computing power, an OC problem can become numerically and computationally arduous [33,34], particularly for applications that account for practical hardware configurations and limitations [12,35,36]. Additionally, the computational expense can increase dramatically when optimizing over an ensemble of systems, such as the case in solid-state NMR [14,37,38], where the ensemble also includes crystalline orientations of a powder average [39]. The particular application of OC of interest to the authors is that of a neural network-based method for MRI [40–42] and a method of morphic OC [43,44], requiring hundreds of thousands of optimized pulse shapes to form their optimal libraries.

Keeping the example of MRI, the utility of OC is highlighted when considering the legal constraints of power deposition safeguards [45] and the obvious financial rewards of reducing the time a patient stays inside the MRI machine.

Modern techniques can mitigate protracted numerical convergence with a quadratically convergent optimization method, requiring the calculation of a Hessian matrix, giving large savings in the number of required serial optimization iterations [32,37,46]. However, it is known within the community that calculation of the Hessian matrix does not scale well to a control problem with a large number of controllable amplitudes [47] (also shown in Fig. 1).

This Letter presents a jump in computational efficiency of multiple orders of magnitude with a reworking of the original method, devised to calculate the exact Hessian matrix [32,37,48], and the recently published flavor of GRAPE with exact, matrix-free calculations, named efficient spin control using analytical Lie algebraic derivatives (ESCALADE) [46]. This Letter will present the mathematical formulation of these

*Present address: Interdisciplinary Nanoscience Center (iNANO), Department of Chemistry, Aarhus University, Aarhus, DK-8000, Denmark; david.goodwin@partner.kit.edu

†msv@cfin.au.dk

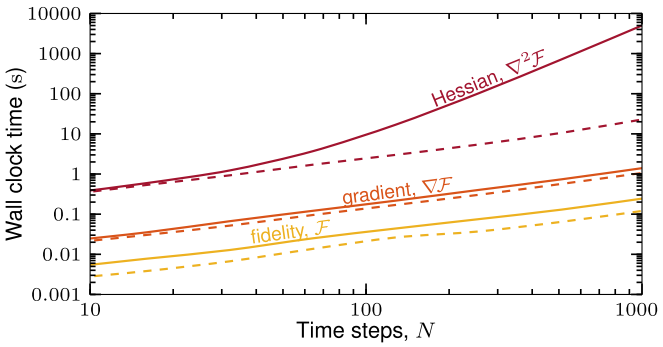

FIG. 1. Average wall-clock time of fidelity, gradient, and Hessian for increasing N. Solid lines show the Newton-Raphson GRAPE method [32] and dashed lines show the proposed *accelerated* Newton-Raphson GRAPE method. Both methods use the auxiliary matrix method [48] to calculate propagator derivatives.

exact Newton-Raphson GRAPE methods in the irreducible spherical-tensor basis of a Liouville space and show an alternative method to calculate the Hessian matrix with $\mathcal{O}(N)$ scaling, compared to the previous $\mathcal{O}(N^2)$ scaling (see Fig. 1). Results show the comparative speedup of this method to the original in the context of state-to-state MRI problems, including a generalization of the ESCALADE method with calculations of optimal z controls, desired by a practical implementation in modern optimal MRI [42,49,50], and with formulas for three-level systems.

II. EXACT NEWTON-RAPHSON OPTIMAL CONTROL

A. Optimal control in Liouville space

A quantum system can be described by a density operator $\hat{\rho}(t)$, a time-dependent system state. The evolution of this state is dictated by the Liouville–von Neumann equation,

$$\frac{\partial}{\partial t}\hat{\rho}(t) = -i[\hat{H}(t), \hat{\rho}(t)], \tag{1}$$

where $\hat{H}(t)$ is a time-dependent Hamiltonian, an operator in a Hilbert space with a spectrum of allowed energy levels. The usual factor of $\hbar$ is dropped here, resulting in the eigenspectrum of $\hat{H}$ expressed in angular frequency units. The methods presented in this Letter are particular to a Liouville space, also named the adjoint representation [31,32,37], although it is only the matrix-vector mathematics of Liouville space that is required. Other matrix-vector OC problems include optimizing pure states in a Hilbert space [5] and magnetization vectors in the real $3d$ space of the Bloch equation [15]. A system state in a Liouville space is represented by a vector $|\hat{\rho}\rangle$, obtained by stacking columns of the density operator $\hat{\rho}$ with Eq. (1) becoming

$$\frac{\partial}{\partial t}|\hat{\rho}(t)\rangle = -i\hat{\hat{L}}(t)|\hat{\rho}(t)\rangle, \quad \hat{\hat{L}} \triangleq \mathbb{1} \otimes \hat{H} - \hat{H}^{\intercal} \otimes \mathbb{1}, \tag{2}$$

where $\hat{\hat{L}}$ is a Liouvillian, and the identity matrix $\mathbb{1}$ and the Hamiltonian have the same dimension.

A bilinear control problem splits that which is controllable, *the controls*, from that which is not, *the drift*. The Liouvillian for a control problem with x, y, and z controls on a two-level system can be written as

$$\hat{\hat{L}}(t) = \underbrace{\omega\hat{\hat{L}}_{\mathrm{z}}}_{\text{drift}} + \underbrace{c_{\mathrm{x}}(t)\hat{\hat{L}}_{\mathrm{x}} + c_{\mathrm{y}}(t)\hat{\hat{L}}_{\mathrm{y}} + c_{\mathrm{z}}(t)\hat{\hat{L}}_{\mathrm{z}}}_{\text{controls}}, \tag{3}$$

where the angular frequency ω is the time-independent resonant frequency offset, $c_{\mathrm{x,y,z}}(t)$ are time-dependent control amplitudes, and $\hat{\hat{L}}_{\mathrm{x,y,z}}$ are Pauli matrices of a Liouville space.

The OC method of GRAPE [13] uses piecewise constant control pulses, where control pulses are constant over a small time interval Δt [29,30],

$$c_k(t) \to [c_{k,1} \quad c_{k,2} \quad \cdots \quad c_{k,N}], \quad k \in \{\mathrm{x}, \mathrm{y}, \mathrm{z}\}, \tag{4}$$

using the notation $c_{k,n} \equiv c_k(t_n)$ for convenience, and $t_N = N\Delta t$. This discrete formulation allows a numerical solution to Eq. (2), given an initial state of the system $|\hat{\rho}_0\rangle$, through time-ordered propagation

$$|\hat{\rho}_N\rangle = \hat{\hat{P}}_N\hat{\hat{P}}_{N-1}\cdots\hat{\hat{P}}_2\hat{\hat{P}}_1|\hat{\rho}_0\rangle, \tag{5}$$

where $\hat{\hat{P}}_n$ are time propagators of an isolated time slice and are defined through the exponential map

$$|\hat{\rho}_n\rangle = \hat{\hat{P}}_n|\hat{\rho}_{n-1}\rangle, \quad \hat{\hat{P}}_n = e^{-i\hat{\hat{L}}_n\Delta t}. \tag{6}$$

The matrix exponential of Eq. (6) is usually calculated with the Padé approximant, Taylor series, or Krylov propagation [37].

As a notational convenience for what follows, the following *effective propagators* are defined as the effect of the pulse between time slices m and n,

$$\mathbf{U}_m^n \triangleq \hat{\hat{P}}_n\hat{\hat{P}}_{n-1}\cdots\hat{\hat{P}}_{m+1}\hat{\hat{P}}_m, \quad \forall(1 \leqslant m \leqslant n \leqslant N), \tag{7}$$

where a backward (time-reversed) propagation can be denoted by $\mathbf{U}_n^m = \mathbf{U}_m^{n\,\dagger}$.

Optimal control requires a metric to optimize, in this case termed *the fidelity* [13,51] $\mathcal{F}$, a measure of how well the pulses perform a desired control task. The task of the OC problem is to find a set of control amplitudes $c_{\mathrm{x,y,z}}(t)$ that maximize the fidelity, e.g., the real part of an inner product:

$$\max_{c_{\mathrm{x,y,z}}(t)}(\mathcal{F}) = \max_{c_{\mathrm{x,y,z}}(t)}\left(\mathrm{Re}\langle\hat{\sigma}|\mathbf{U}_1^N|\hat{\rho}_0\rangle\right). \tag{8}$$

This form of the fidelity metric is defined in terms of state-to-state problems, where a system is in a defined initial state, $|\hat{\rho}_0\rangle$, and the control task is to take this state to a desired one, $|\hat{\sigma}\rangle$. With the notation introduced in Eq. (7), $\mathbf{U}_1^N$ is interpreted as the effective propagator over the shaped pulse. In addition to Eq. (5), the system is propagated backwards from the desired target state,

$$|\hat{\chi}_n\rangle = \hat{\hat{P}}_n^{\dagger}\hat{\hat{P}}_{n+1}^{\dagger}\cdots\hat{\hat{P}}_{N-1}^{\dagger}\hat{\hat{P}}_N^{\dagger}|\hat{\sigma}\rangle, \tag{9}$$

which is termed the *adjoint state* of the control problem [37].

GRAPE is a gradient-following numerical optimization method and requires derivatives of the fidelity with respect to the controls. In turn, this requires the directional propagator derivatives $D_{k,n}^a$, with the subscripts k denoting the derivative in the direction of $\hat{\hat{L}}_k$ and n denoting the derivative operating on the time propagator $\hat{\hat{P}}_n$. For each time slice n, and for each

control direction $\hat{\hat{L}}_k \in \{\hat{\hat{L}}_x, \hat{\hat{L}}_y, \hat{\hat{L}}_z\}$, the ath-order derivative takes the form

$$\nabla^a \mathcal{F}(c_{k,n}) = \mathrm{Re}\, \underbrace{\langle \hat{\chi}_{n+1}}_{N \text{ times}} | \overbrace{D^a_{k,n} \hat{\rho}_{n-1}\rangle}^{N \text{ times}}, \tag{10}$$

where bra-ket notation explicitly shows vector structures, i.e., $|D^a_{k,n}\hat{\rho}_{n-1}\rangle = D^a_{k,n}|\hat{\rho}_{n-1}\rangle$. The number of forward and backward propagations is indicated for each control channel to produce a gradient vector $\nabla\mathcal{F}$ or the diagonal elements of a Hessian matrix $\nabla^2\mathcal{F}$.

This is sufficient for a fidelity gradient, scaling linearly with N, but a fidelity Hessian also requires mixed second-order derivatives [32], where the off-diagonal Hessian elements are

$$\nabla^2 \mathcal{F}(c_{k,n}, c_{j,m}) = \mathrm{Re}\, \underbrace{\langle \hat{\chi}_{n+1} D_{k,n}}_{N \text{ times}} | \overbrace{\mathbf{U}^{n-1}_{m+1} |D_{j,m}\hat{\rho}_{m-1}\rangle}^{\frac{1}{2}N(N-1) \text{ times}}. \tag{11}$$

Clearly, the form of Eq. (11) has a central propagator that cannot be absorbed into the bra or ket because it depends on both t_n and t_m, and therefore the computation scales polynomially with $\frac{1}{2}N(N-1)$ (the factor $\frac{1}{2}$ comes from the symmetric property of a Hessian [37]). This is known within the OC community [47] and is highlighted in Fig. 1. The linear plots of the fidelity and gradient, on log-log axes, show these calculations are efficient with increasing N, whereas the Hessian calculation time is not linear on these log-log axes. The subject of this Letter is to mitigate this undesirable scaling, resulting in a linearly scaling Hessian calculation (dashed lines in Fig. 1), after the following section outlines the calculation of directional propagator derivatives.

B. Directional propagator derivatives with auxiliary matrices

As has been published previously [32,37,48], exact propagator derivatives required by Eqs. (10) and (11) can be calculated by exponentiating an auxiliary matrix (AUXMAT), resulting in an upper triangular block matrix [52] with a time propagator on the block diagonal, and with a directional derivative of that propagator in the upper triangular block. The propagator and propagator derivatives are extracted from

$$\begin{bmatrix} \hat{\hat{P}}_n & D_{j,n} & \frac{1}{2}D^2_{jk,n} \\ \mathbf{0} & \hat{\hat{P}}_n & D_{k,n} \\ \mathbf{0} & \mathbf{0} & \hat{\hat{P}}_n \end{bmatrix} = \exp\left(\begin{bmatrix} \mathbf{A}_n & \mathbf{C}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_n & \mathbf{C}_k \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_n \end{bmatrix} \right), \tag{12}$$

where the block matrix is formed from $\mathbf{A}_n = -i\hat{\hat{L}}_n \Delta t$, as in Eq. (6), and $\mathbf{C}_k = -i\hat{\hat{L}}_k \Delta t$ are functions of the control operators with $j, k \in \{x, y, z\}$.

C. Directional propagator derivatives with ESCALADE

An additional method to calculate the time propagators of a two-level system is by explicitly calculating elements of the matrix. In a spherical-tensor basis this is the Wigner $\mathcal{D}$ matrix [53],

$$\hat{\hat{P}}_n = \mathcal{D}^{(1)} = \begin{bmatrix} \alpha^2 & \sqrt{2}\alpha\beta & \beta^2 \\ -\sqrt{2}\alpha\beta^* & \alpha\alpha^* - \beta\beta^* & \sqrt{2}\alpha^*\beta \\ \beta^{*2} & -\sqrt{2}\alpha^*\beta^* & \alpha^{*2} \end{bmatrix}, \tag{13}$$

which is formulated in terms of the complex elements

$$\alpha = \cos\phi - i\frac{z}{r}\sin\phi, \quad \beta = -\frac{y}{r}\sin\phi - i\frac{x}{r}\sin\phi, \tag{14}$$

where the convenient and compact notation $x = c_{x,n}$, $y = c_{y,n}$, and $z = c_{z,n} + \omega$ are used, $\phi = \frac{1}{2}r\Delta t$ is a polar angle of rotation, and $r = \sqrt{x^2 + y^2 + z^2}$ is the polar radius.

Whereas Eq. (12) is exact, the matrix exponential is expensive. Foroozandeh and Singh recently derived a method that is free from this expensive matrix exponential calculation, in the propagator derivative calculation method of efficient spin control using analytical Lie algebraic derivatives (ESCALADE) [46]. This method can also be extended with interaction propagator splitting [54].

The calculation of directional derivatives in Eqs. (10) and (11), in this single-spin model, can proceed with ESCALADE by constructing a matrix with rows containing all elements needed to construct the propagator derivatives,

$$\begin{bmatrix} \mathbf{\Theta}_x \\ \mathbf{\Theta}_y \\ \mathbf{\Theta}_z \end{bmatrix} = \mathrm{vec}[\mathbb{1}] + \frac{\sin^2\phi}{\phi r}\mathrm{vec}[\mathbf{S}] + \frac{2\phi - \sin 2\phi}{2\phi r^2}\mathrm{vec}[\mathbf{S}^2], \tag{15}$$

where vec[$\mathbb{1}$] is a vectorized identity matrix, with a vectorization operation on a matrix $\mathbf{A}$ such that $\mathbf{A} = \mathrm{vec}^{-1}[\mathrm{vec}[\mathbf{A}]]$. The skew-symmetric matrix $\mathbf{S}$ is

$$\mathbf{S} = \begin{bmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{bmatrix}, \tag{16}$$

and the symmetric matrix $\mathbf{S}^2$ is calculated algebraically. In turn, the directional propagator derivatives can be written as

$$D_{k,n} = \hat{\hat{P}}_n[\mathrm{vec}^{-1}[\mathbf{\Sigma\Theta}_k]], \tag{17}$$

$$D^2_{jk,n} = \hat{\hat{P}}_n[\mathrm{vec}^{-1}[\mathbf{\Sigma\Theta}_{jk}]] + \hat{\hat{P}}_n[\mathrm{vec}^{-1}[\mathbf{\Sigma\Theta}_j]][\mathrm{vec}^{-1}[\mathbf{\Sigma\Theta}_k]], \tag{18}$$

where $\mathbf{\Sigma} = [\mathrm{vec}[\mathbf{C}_x] \quad \mathrm{vec}[\mathbf{C}_y] \quad \mathrm{vec}[\mathbf{C}_z]]$ is a three-column matrix to be multiplied with the three-row matrix of $\mathbf{\Theta}_k$. Using notation similar to Eq. (15), the $\mathbf{\Theta}_{jk}$ matrices required by the second-order derivatives in Eq. (18) are

$$\begin{aligned} \begin{bmatrix} \mathbf{\Theta}_{jx} \\ \mathbf{\Theta}_{jy} \\ \mathbf{\Theta}_{jz} \end{bmatrix} &= \frac{\cos(2\phi) - 1}{r^2}\mathrm{vec}\left[\mathbf{S}\frac{d\mathbf{S}}{dj} + \frac{d\mathbf{S}}{dj}\mathbf{S}\right] \\ &+ \frac{2\phi - \sin(2\phi)}{r^3}\mathrm{vec}\left[\frac{d\mathbf{S}}{dj}\right] \\ &+ \frac{2k(1 - \cos(2\phi) - \phi\sin(2\phi))}{r^3}\mathrm{vec}[\mathbf{S}] \\ &+ \frac{3k\sin(2\phi) - 2k\phi(2 + \cos(2\phi))}{r^4}\mathrm{vec}[\mathbf{S}^2], \end{aligned} \tag{19}$$

where three such equations are required for $j \in \{x, y, z\}$ and the derivatives of $\mathbf{S}$ are derived algebraically from Eq. (16).

Since the directional derivatives of Eqs. (17) and (18), together with the time propagators of Eq. (13), do not involve any matrix operations, other than a few trivial multiplications, ESCALADE offers substantial computational gains relative to the AUXMAT method in Eq. (12).

D. Propagators of ESCALADE for higher-level systems

Only two-level systems are considered so far in ESCALADE [46] and its extension to interacting two-level systems [54]. One route to using these methods for higher-level systems is to consider the control with the offset part of the Liouvillian, as in this Letter in Eq. (3), and then to split the interaction part of the Liouvillian, e.g., for a spin-1, three-level system this would correspond to quadrupolar terms. This is detailed in the quantum optimal control by adaptive low-cost algorithm (QOALA) paper [54] and not reproduced here. In terms of the complex elements in Eq. (14), propagators can be formulated in the spherical-tensor basis with Wigner $\mathcal{D}$ matrices [55]; for a three-level system (ignoring those quadrupolar terms) this is

$$\hat{\hat{P}}_n = \begin{bmatrix} \mathcal{D}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathcal{D}^{(2)} \end{bmatrix}, \tag{20}$$

which is formulated in terms of the Wigner $\mathcal{D}$ matrix in Eq. (13) and a second-rank Wigner matrix. It should be emphasized that the quadrupolar part of the Liouvillian needs to be included and that this should use propagator splitting methods recently presented by Goodwin *et al.* [54]. The elements of the second-rank Wigner $\mathcal{D}$ matrix can be formulated as

$$\mathcal{D}^{(2)} = \begin{bmatrix} \alpha^4 & 2\alpha^3\beta & \sqrt{6}\alpha^2\beta^2 & 2\alpha\beta^3 & \beta^4 \\ -2\alpha^3\beta^* & \alpha^2(2(\alpha\alpha^* - \beta\beta^*) - 1) & \sqrt{6}\alpha\beta(\alpha\alpha^* - \beta\beta^*) & \beta^2(2(\alpha\alpha^* - \beta\beta^*) + 1) & 2\alpha^*\beta^3 \\ \sqrt{6}\alpha^2\beta^{*2} & -\sqrt{6}\alpha\beta^*(\alpha\alpha^* - \beta\beta^*) & \frac{1}{2}(3(\alpha\alpha^* - \beta\beta^*)^2 - 1) & \sqrt{6}\alpha^*\beta(\alpha\alpha^* - \beta\beta^*) & \sqrt{6}\alpha^{*2}\beta^2 \\ -2\alpha\beta^{*3} & \beta^{*2}(2(\alpha\alpha^* - \beta\beta^*) + 1) & -\sqrt{6}\alpha^*\beta^*(\alpha\alpha^* - \beta\beta^*) & \alpha^{*2}(2(\alpha\alpha^* - \beta\beta^*) - 1) & 2\alpha^{*3}\beta \\ \beta^{*4} & -2\alpha^*\beta^{*3} & \sqrt{6}\alpha^{*2}\beta^{*2} & -2\alpha^{*3}\beta^* & \alpha^{*4} \end{bmatrix}. \tag{21}$$

III. ACCELERATED NEWTON-RAPHSON OPTIMAL CONTROL

Moving away from a chosen calculation method of directional propagator derivatives, the remaining bottleneck of the Newton-Raphson methods presented above is the off-diagonal Hessian elements, the mixed derivatives of Eq. (11).

As a starting point, the ESCALADE paper [46] includes an additional efficiency of Newton-Raphson GRAPE where the central effective propagator in Eq. (11) can be split into $\mathbf{U}_{m+1}^{n-1} = \mathbf{U}_{m+1}^{n-1}\mathbf{U}_1^m\mathbf{U}_m^1 = \mathbf{U}_1^{n-1}\mathbf{U}_m^1$ so Eq. (11) becomes

$$\nabla^2\mathcal{F}(c_{k,n}, c_{j,m}) = \mathrm{Re}\langle\hat{\chi}_{n+1}D_{k,n}|\mathbf{U}_1^{n-1}\mathbf{U}_m^1|D_{j,m}\hat{\rho}_{m-1}\rangle. \tag{22}$$

An interpretation of these two central effective propagators is as follows: The right-hand side, $\langle\hat{\chi}_{n+1}D_{k,n}|$, is multiplied by the effective propagator $\mathbf{U}_1^{n-1}$, presenting the directional derivative at t_n to now be evaluated at t_0; the left-hand side, $|D_{j,m}\hat{\rho}_{m-1}\rangle$, is multiplied by the time-reversed effective propagator $\mathbf{U}_m^1$, presenting the directional derivative at t_m to also now be evaluated at t_0.

To outline how the bottleneck of Newton-Raphson GRAPE can be avoided, a representation of a *trajectory* is introduced,

$$[\boldsymbol{\rho}]_m^n \triangleq [|\hat{\rho}_n\rangle \quad |\hat{\rho}_{n-1}\rangle \quad \cdots \quad |\hat{\rho}_{m+1}\rangle \quad |\hat{\rho}_m\rangle], \tag{23}$$

which is an array of a column vector. The whole trajectory from the dynamics in Eq. (3) is contained in $[\boldsymbol{\rho}]_0^N$ and its analysis is useful for visualizing pulse dynamics [56]. Taking this concept a step further, a *directional derivative trajectory*, evaluated at t_0 as with the right-hand side of Eq. (22), is defined as

$$\left[\partial_j^{[0]}\boldsymbol{\rho}\right]_0^n \triangleq \left[\left|\mathbf{U}_n^1 D_{j,n}\hat{\rho}_{n-1}\right\rangle \quad \cdots \quad \left|\mathbf{U}_2^1 D_{j,2}\hat{\rho}_1\right\rangle \quad \left|\mathbf{U}_1^\dagger D_{j,1}\hat{\rho}_0\right\rangle\right], \tag{24}$$

where the superscript [0] is used to indicate evaluation at t_0.

With the realization that a *directional derivative trajectory* in Eq. (24) is a matrix in itself, n Hessian elements can be calculated with a one matrix-vector product with

$$\begin{bmatrix} \nabla^2\mathcal{F}(c_{k,n}, c_{j,n}) \\ \vdots \\ \nabla^2\mathcal{F}(c_{k,n}, c_{j,2}) \\ \nabla^2\mathcal{F}(c_{k,n}, c_{j,1}) \end{bmatrix}^\mathsf{T} = \mathrm{Re}\underbrace{\langle\hat{\chi}_{n+1}D_{k,n}\mathbf{U}_1^{n-1}}_{N\text{times backward}}\overbrace{|\left[\partial_j^{[0]}\boldsymbol{\rho}\right]_0^n}^{N\text{times forward}}. \tag{25}$$

Given that the total effective propagator $\mathbf{U}_1^N$ can be calculated with forward propagation, $\mathbf{U}_1^{n-1}$ can be updated during subsequent backward propagation with $\mathbf{U}_1^{n-1} = \hat{\hat{P}}_{n-1}^\dagger\mathbf{U}_1^n$. The single vector-matrix product of Eq. (25), per time slice, is expected to be much more efficient than the $n-1$ vector-vector products, per time slice, in Eq. (11).

In contrast to Eq. (11), using an arrangement for calculating the mixed derivatives in Eq. (25) allows the central propagators to be split and absorbed into the bra and ket involving each of the directional derivatives, which reduces the computational complexity to that of the gradient calculations and nonmixed second-order directional derivatives of Eq. (10). The reason this can be done is because one of the directional derivatives is evaluated at t_0 using Eq. (22).

IV. ACCELERATION OF A BROADBAND MRI EXAMPLE

The authors choose to set the test of the method presented above in the context of MRI because there is a very real need to have computationally fast OC methods to run *on the fly* with a need for patient-specific solutions. Often, OC-facilitated MRI exploits simultaneous/parallel control systems, e.g., eight radio-frequency channels (x and y controls) [15,49] and three magnetic field gradients and/or numerous local shims (z controls) [42,49,50].

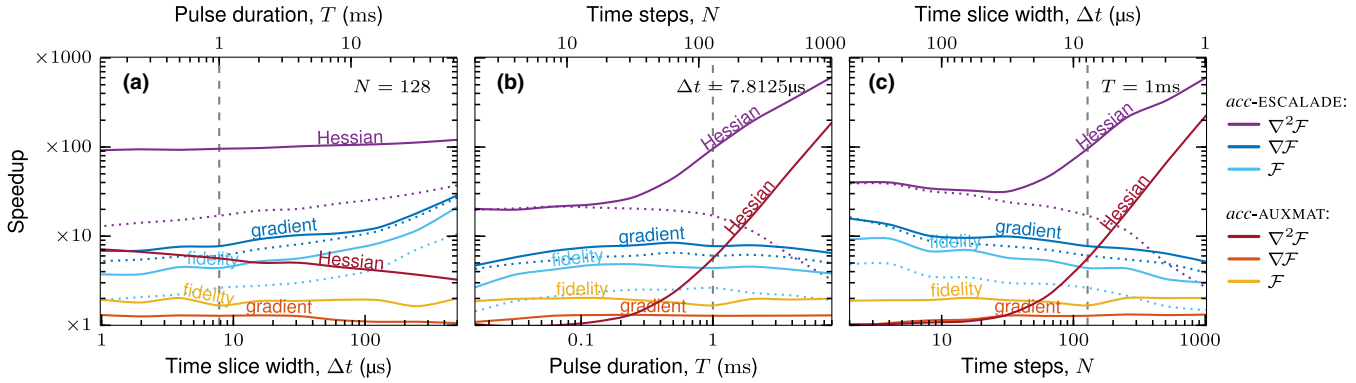

FIG. 2. The average speedup of the accelerated ESCALADE method (*acc*-ESCALADE) and the accelerated auxiliary matrix method (*acc*-AUXMAT), compared to the auxiliary matrix method. The number of time steps N, the time-slice width Δt, and the pulse duration T are parameters affecting computation time. In each of the three plots one parameter is kept constant: (a) $N = 128$, (b) $\Delta t = 7.8125$ μs, and (c) $T = 1$ ms. The dotted lines in each plot show the speedup of *acc*-ESCALADE relative to *acc*-AUXMAT.

The optimization is set as robust for an ensemble of 101 spin-$\frac{1}{2}$ systems, with an offset bandwidth $\frac{\omega}{2\pi} \in [-\frac{1}{2}, +\frac{1}{2}]$ kHz, simulated using a single block-diagonal Liouvillian with each block operating on a single ensemble member [39,43]. Although this work does not present pulses robust to control field imperfections, this optimization can be simulated in the same way as the previous control field robustness [36] where fidelity and fidelity derivatives are averaged from a distribution of control amplitudes [44]. This part of the OC algorithm can be computed efficiently in parallel.

The time slice Δt is often fixed to the hardware digitalization dwell time, typically a few μs. Modern MRI systems handle $N > 1000$, but the pulse duration, $T = N\Delta t$, is usually kept around a few ms to avoid degradation of pulse performance due to transverse relaxation. As a set of speed tests, these three variables are incremented and set as a range of OC problems and are shown in Fig. 2. The vertical dashed lines indicate a set of variables physically relevant to MRI.

Figure 2 shows the average speedup of the accelerated auxiliary matrix method (*acc*-AUXMAT) and the accelerated ESCALADE method (*acc*-ESCALADE), both relative to the standard auxiliary matrix method. A separate fidelity, gradient, and Hessian calculation is performed for each method from a random control pulse shape, with c_x, c_y, and c_z controls in the order of $2\pi \times 10^3$ rads^{-1}. The average is over 84 different random control pulses, run in parallel on 28 CPU cores. The parallel computation is the same form as that which would be used for control field robustness.

Figure 2(a) shows that the effect of increasing T while also increasing Δt is, approximately, a constant speedup of *acc*-AUXMAT and *acc*-ESCALADE Hessians. This can be attributed to the effective propagator splitting of Eq. (22). The small speedup of *acc*-AUXMAT fidelity and gradient calculations is due to an algorithmic efficiency from coding in a way that lends itself to the accelerated Hessian methods.

The main result of this Letter is the speedup of Hessian calculations of the *acc*-AUXMAT and *acc*-ESCALADE methods, with speedup increasing as N increases, in Figs. 2(b) and 2(c), to over ×100 when a large N is used. Furthermore, this trend does not appear to dampen, indicating a further speedup when $N \gg 1000$.

From the evidence in Figs. 2(b) and 2(c), the accelerated Hessian calculations of Eq. (25) do indeed remove the $\mathcal{O}(N^2)$ scaling of the original Newton-Raphson method [32], reducing to a linear scaling $\mathcal{O}(N)$.

V. CONCLUSION

An alternative mathematical formulation of the Newton-Raphson GRAPE method [32] has been presented in Liouville space and applies to OC problems with unitary evolution. The ESCALADE method [46] recast the cumbersome problem of finding derivatives—using trigonometric evaluations of vectorized arrays rather than matrix-matrix products together with computationally expensive matrix exponentials—and factorized the central propagator to avoid $\mathcal{O}(N^2)$ scaling, now reduced to $\mathcal{O}(N)$, in the computation of the Hessian. The ESCALADE method is shown in a different light of Liouville space, with the additional derivation of z controls, which are important to MRI, and also extends ESCALADE and its related method [46,54] to systems with more than two levels.

The main result is a different formulation of the Newton-Raphson GRAPE method in Liouville space, being an optimization method with quadratic convergence to an optimal solution, reducing the expensive polynomial scaling of the control problem to a linear scaling when increasing the number of piecewise-constant pulses in an optimal solution. Speedup increases as the number of time slices N increases: ×4 to ×200, for $N = 100$ and $N = 1000$, respectively. Furthermore, employing ESCALADE within this Hessian calculation method shows a further speedup of ×70 to ×600, for $N = 100$ and $N = 1000$, respectively, when compared to the original Newton-Raphson GRAPE method.

ACKNOWLEDGMENTS

D.L.G. thanks B. Luy for his continued support, P. Singh and M. Foroozandeh for discussions and their basic MATLAB code on the ESCALADE method, and M. Goodwin for proofreading. M.S.V. would like to thank VILLUM Fonden, Eva og Henry Fraenkels Mindefond, Harboefonden, and Kong Christian Den Tiendes Fond.

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

[2] A. Acín, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm, The Quantum Technologies Roadmap: A European community view, New J. Phys. **20**, 080201 (2018).

[3] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. J. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, Quantum optimal control in quantum technologies. Strategic report on current status, visions and goals for research in Europe, EPJ Quantum Technol. **9**, 19 (2022).

[4] J. C. Saywell, I. Kuprov, D. Goodwin, M. Carey, and T. Freegarde, Optimal control of mirror pulses for cold-atom interferometry, Phys. Rev. A **98**, 023625 (2018).

[5] J. C. Saywell, Optimal control of cold atoms for ultra-precise quantum sensors, Ph.D. thesis, University of Southampton, 2020.

[6] J. Saywell, M. Carey, M. Belal, I. Kuprov, and T. Freegarde, Optimal control of Raman pulse sequences for atom interferometry, J. Phys. B: At., Mol. Opt. Phys. **53**, 085006 (2020).

[7] J. Saywell, M. Carey, N. Dedes, I. Kuprov, and T. Freegarde, Can optimised pulses improve the sensitivity of atom interferometers? in *Quantum Technology: Driving Commercialisation of an Enabling Science II*, edited by M. J. Padgett, K. Bongs, A. Fedrizzi, and A. Politi, International Society for Optics and Photonics, Vol. 11881 (SPIE, Bellingham, WA, 2021), pp. 83–92.

[8] J. Zhang, R. Laflamme, and D. Suter, Experimental Implementation of Encoded Logical Qubit Operations in a Perfect Quantum Error Correcting Code, Phys. Rev. Lett. **109**, 100503 (2012).

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

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

[11] N. Khaneja, R. Brockett, and S. J. Glaser, Time optimal control in spin systems, Phys. Rev. A **63**, 032308 (2001).

[12] T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, and S. J. Glaser, Application of optimal control theory to the design of broadband excitation pulses for high-resolution NMR, J. Magn. Reson. **163**, 8 (2003).

[13] 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).

[14] Z. Tošner, S. J. Glaser, N. Khaneja, and N. C. Nielsen, Effective Hamiltonians by optimal control: Solid-state NMR double-quantum planar and isotropic dipolar recoupling, J. Chem. Phys. **125**, 184502 (2006).

[15] M. S. Vinding, D. Brenner, D. H. Y. Tse, S. Vellmer, T. Vosegaard, D. Suter, T. Stöcker, and I. I. Maximov, Application of the limited-memory quasi-Newton algorithm for multi-dimensional, large flip-angle RF pulses at 7T, Magn. Reson. Mater. Phys. **30**, 29 (2017).

[16] E. V. Reeth, H. Ratiney, K. Tse Ve Koon, M. Tesch, D. Grenier, O. Beuf, S. J. Glaser, and D. Sugny, A simplified framework to optimize MRI contrast preparation, Magn. Reson. Med. **81**, 424 (2019).

[17] J. Somlói, V. A. Kazakov, and D. J. Tannor, Controlled dissociation of I_2 via optical transitions between the X and B electronic states, Chem. Phys. **172**, 85 (1993).

[18] W. Zhu, J. Botina, and H. Rabitz, Rapidly convergent iteration methods for quantum optimal control of population, J. Chem. Phys. **108**, 1953 (1998).

[19] Y. Maday and G. Turinici, New formulations of monotonically convergent quantum control algorithms, J. Chem. Phys. **118**, 8191 (2003).

[20] R. Eitan, M. Mundt, and D. J. Tannor, Optimal control with accelerated convergence: Combining the Krotov and quasi-Newton methods, Phys. Rev. A **83**, 053426 (2011).

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

[22] N. Rach, M. M. Müller, T. Calarco, and S. Montangero, Dressing the chopped-random-basis optimization: A bandwidth-limited access to the trap-free landscape, Phys. Rev. A **92**, 062343 (2015).

[23] G. Ciaramella, A. Borzì, G. Dirr, and D. Wachsmuth, Newton methods for the optimal control of closed quantum spin systems, SIAM J. Sci. Comput. **37**, A319 (2015).

[24] 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).

[25] D. Lucarelli, Quantum optimal control via gradient ascent in function space and the time-bandwidth quantum speed limit, Phys. Rev. A **97**, 062346 (2018).

[26] D. Quiñones Valles, S. Dolgov, and D. Savostyanov, Tensor product approach to quantum control, in *Integral Methods in Science and Engineering: Analytic Treatment and Numerical Approximations*, edited by C. Constanda and P. Harris (Springer, Berlin, 2019), pp. 367–379.

[27] S. Conolly, D. Nishimura, and A. Macovski, Optimal control solutions to the magnetic resonance selective excitation problem, IEEE Trans. Med. Imaging **5**, 106 (1986).

[28] J. Mao, T. H. Mareci, K. N. Scott, and E. R. Andrew, Selective inversion radiofrequency pulses by optimal control, J. Magn. Reson. **70**, 310 (1986).

[29] L. Viola and S. Lloyd, Dynamical suppression of decoherence in two-state quantum systems, Phys. Rev. A **58**, 2733 (1998).

[30] L. Viola, E. Knill, and S. Lloyd, Dynamical Decoupling of Open Quantum Systems, Phys. Rev. Lett. **82**, 2417 (1999).

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

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

[33] N. I. Gershenzon, T. E. Skinner, B. Brutscher, N. Khaneja, M. Nimbalkar, B. Luy, and S. J. Glaser, Linear phase slope in pulse design: Application to coherence transfer, J. Magn. Reson. **192**, 235 (2008).

[34] A. N. Pechen and D. J. Tannor, Are there Traps in Quantum Control Landscapes? Phys. Rev. Lett. **106**, 120402 (2011).

[35] K. Kobzar, T. E. Skinner, N. Khaneja, S. J. Glaser, and B. Luy, Exploring the limits of broadband excitation and inversion pulses, J. Magn. Reson. **170**, 236 (2004).

[36] K. Kobzar, T. E. Skinner, N. Khaneja, S. J. Glaser, and B. Luy, Exploring the limits of broadband excitation and inversion: II. Rf-power optimized pulses, J. Magn. Reson. **194**, 58 (2008).

[37] D. L. Goodwin, Advanced optimal control methods for spin systems, Ph.D. thesis, University of Southampton, 2017.

[38] Z. Tošner, M. J. Brandl, J. Blahut, S. J. Glaser, and B. Reif, Maximizing efficiency of dipolar recoupling in solid-state NMR using optimal control sequences, Sci. Adv. **7**, eabj5913 (2021).

[39] I. Kuprov, Fokker-Planck formalism in magnetic resonance simulations, J. Magn. Reson. **270**, 124 (2016).

[40] M. S. Vinding, B. Skyum, R. Sangill, and T. E. Lund, Ultrafast (milliseconds), multidimensional RF pulse design with deep learning, Magn. Reson. Med. **82**, 586 (2019).

[41] M. S. Vinding, C. S. Aigner, S. Schmitter, and T. E. Lund, DeepControl: 2DRF pulses facilitating B_1^+ inhomogeneity and B_0 off-resonance compensation in vivo at 7 T, Magn. Reson. Med. **85**, 3308 (2021).

[42] M. S. Vinding, D. L. Goodwin, I. Kuprov, and T. E. Lund, Optimal control gradient precision trade-offs: Application to fast generation of DeepControl libraries for MRI, J. Magn. Reson. **333**, 107094 (2021).

[43] D. L. Goodwin, M. R. M. Koos, and B. Luy, Second order phase dispersion by optimised rotation pulses, Phys. Rev. Res. **2**, 033157 (2020).

[44] J. D. Haller, D. L. Goodwin, and B. Luy, SORDOR pulses: Expansion of the Böhlen–Bodenhausen scheme for low-power broadband magnetic resonance, Magn. Reson. **3**, 53 (2022).

[45] M. F. Dempsey, B. Condon, and D. M. Hadley, MRI safety review, Semin. Ultrasound, CT MRI **23**, 392 (2002).

[46] M. Foroozandeh and P. Singh, Optimal control of spins by analytical Lie algebraic derivatives, Automatica **129**, 109611 (2021).

[47] W. Kallies, Concurrent optimization of robust refocused pulse sequences for magnetic resonance spectroscopy, Ph.D. thesis, Technische Universität München, 2018.

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

[49] M. Vinding, B. Gurin, T. Vosegaard, and N. Nielsen, Local SAR, global SAR, and power-constrained large-flip-angle pulses with optimal control and virtual observation points, Magn. Reson. Med. **77**, 374 (2017).

[50] J. P. Stockmann and L. L. Wald, In vivo B_0 field shimming methods for MRI at 7 T, NeuroImage **168**, 71 (2018).

[51] S. J. Glaser, T. Schulte-Herbrüggen, M. Sieveking, O. Schedletzky, N. C. Nielsen, O. W. Sørensen, and C. Griesinger, Unitary control in quantum ensembles: Maximizing signal intensity in coherent spectroscopy, Science **280**, 421 (1998).

[52] C. F. Van Loan, Computing integrals involving the matrix exponential, IEEE Trans. Autom. Control **23**, 395 (1978).

[53] D. J. Siminovitch, Rotations in NMR: Part I. Euler-Rodrigues parameters and quaternions, Concepts Magn. Reson. **9**, 149 (1997).

[54] D. L. Goodwin, P. Singh, and M. Foroozandeh, Adaptive optimal control of entangled qubits, Sci. Adv. **8**, eabq4244 (2022).

[55] R. M. Lynden-Bell and A. J. Stone, Reorientational correlation functions, quaternions and Wigner rotation matrices, Mol. Simul. **3**, 271 (1989).

[56] I. Kuprov, Spin system trajectory analysis under optimal control pulses, J. Magn. Reson. **233**, 107 (2013).