UNIVERSITY OF
BATH

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Optimal Control of Spins by Analytical Lie Algebraic Derivatives [⋆]

Mohammadali Foroozandeh [a], Pranav Singh [b],

[a] *Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK*

[b] *Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK*

Abstract

Computation of derivatives (gradient and Hessian) of a fidelity function is one of the most crucial steps in many optimization algorithms. Having access to accurate methods for computing these derivatives is even more desirable where the optimization process requires propagation of these computations over many steps, which is particularly important in optimal control of spin systems. Here we propose a novel numerical approach, ESCALADE (Efficient Spin Control using Analytical Lie Algebraic Derivatives), that offers the exact first and second derivatives of the fidelity function by taking advantage of the properties of the Lie group of 2×2 unitary matrices, SU(2), and its Lie algebra, the Lie algebra of skew-Hermitian matrices, $\mathfrak{su}(2)$. A full mathematical treatment of the proposed method along with some numerical examples are presented.

Key words: Optimal control; Lie algebra; Derivatives; Numerical algorithms; ESCALADE.

1 Introduction

Controlling quantum spin dynamics using time-dependent Hamiltonians via optimal control theory is an increasingly popular approach in many areas of science from spectroscopy and quantum metrology to quantum information processing and computing [1].

Two of the main challenges in the field of optimal control of spin systems are the controllability of the dynamics and the convergence rate of the control process. In principle, three main approaches can be considered when optimal control has been applied to spin systems: 1) derivative-free techniques [2, 3] which are especially important when due to experimental requirements not many iterations or function evaluations by the optimisation protocol can be allowed, 2) gradient-based techniques like GRAPE [4, 5] and KROTOV [6, 7], and 3) Newton–Raphson method [8, 9] where in addition to the gradient (first derivative), the Hessian (second derivative of the objective function with respect to the control parameters) is also utilized. Although the latter approach results in quadratic convergence rate, it suffers

from numerical complexity due to computation and update of a dense Hession matrix in the course of optimization. Additionally, computation of derivatives using finite differences can be expensive, inaccurate and potentially unstable when the objective function involves numerical propagators with limited accuracy [10, Chapter 8]. Therefore having access to the exact form of these derivatives is of great interest, in particular in the optimal control of spin systems, where the optimization process requires propagation of these computations over many steps, and inaccurate estimations of derivatives can result in a large accumulated numerical error.

The objective of this paper is to present a novel approach that facilitates the optimal control of spins using Newton–Raphson utilizing an analytical computation of derivatives. Control of dynamical systems using properties of Lie groups and their algebras covers a surprisingly wide range of applications from controlling of landing a plane, rotations of rigid bodies in robotics and estimation of camera poses in computer vision, to the time evolution of quantum systems [11–19]. In these applications, the underlying geometric structure is described by a Lie group. Finite difference methods suffer from a further disadvantage here since they do not respect the Lie group structure and result in derivatives that do not live in the tangent space (the Lie algebra) [20].

[⋆] Corresponding authors: M. Foroozandeh, P. Singh
   Email addresses:
mohammadali.foroozandeh@chem.ox.ac.uk (Mohammadali Foroozandeh), ps2106@bath.ac.uk (Pranav Singh).

Here we propose ESCALADE (Efficient Spin Control using Analytical Lie Algebraic Derivatives), a novel numerical approach that harnesses the exact first and second derivatives of the fidelity function. These derivatives are computed by exploiting the properties of the Lie group of 2×2 unitary matrices, SU(2), and its Lie algebra – the Lie algebra of skew-Hermitian matrices, $\mathfrak{su}(2)$. Since the Lie groups, SU(2) and SO(3) are closely related (see [21, Chapter 5] and [22, Chapter 6]), there is a close parallel between some of the properties exploited here to the Rodrigues rotation formula [23], which is utilized widely in computer vision and robotics applications for computation of rotation matrices in SO(3) [19, 20, 24, 25].

Although here we present the technique on the optimal control of the dynamic of non-interacting qubits, this is a general approach and can be applied to spin systems with more diverse Hamiltonian structures. It has the potential to find applications in a variety of areas where taking advantage of Lie algebra for efficient optimal control of spins is beneficial. Examples include geometric [11, 26–29] and adiabatic optimal control [30–33] methods.

2 Theory

2.1 Optimal Control of spin-½

The state of a single spin-½ particle is described by the Hermitian density matrix $\rho(t)$, i.e. $\mathrm{i}\rho(t) \in \mathfrak{su}(2)$, and its dynamics are governed by the Liouville–von Neumann equation,

$$\partial_t \rho(t) = -\mathrm{i}[\mathcal{H}(t), \rho(t)], \qquad \rho(0) = \rho_0, \qquad (1)$$

where

$$\mathcal{H}(t) = \boldsymbol{h}(t) \cdot \boldsymbol{\sigma}, \qquad (2)$$

$$\boldsymbol{h}(t) = (f(t), g(t), \Omega)^\top \in \mathbb{R}^3, \quad \boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^\top, \quad (3)$$

and

$$\sigma_x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \frac{1}{2} \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the normalized Pauli matrices. Ω describes the offset frequency of a spin.

Remark 1 *In the case of multiple non-interacting spin-½ particles, the kth spin evolves under the influence of $\mathcal{H}_k(t) = \boldsymbol{h}_k(t) \cdot \boldsymbol{\sigma}$, where the offset Ω_k in $\boldsymbol{h}_k(t) = (f(t), g(t), \Omega_k)$ varies with the particle but $f(t)$ and $g(t)$ are common across all spins.*

In a numerical solution of equation (1), we compute ρ at time intervals $t_0, t_1, \ldots t_N$, with the unitary numerical propagation being described by

$$\rho_n = \mathrm{U}_n \rho_{n-1} \mathrm{U}_n^\dagger, \qquad \mathrm{U}_n = \mathrm{e}^{-\mathrm{i}\boldsymbol{s}_n \cdot \boldsymbol{\sigma}}. \qquad (4)$$

We restrict our attention to a piecewise constant approximation of the pulse, f and g, where $\boldsymbol{s}_n = (\Delta t)(f(t_{n-1}), g(t_{n-1}), \Omega)^\top$.

By using equation (4), one can see that the final density matrix is given by

$$\rho_N = \underbrace{\mathrm{U}_N \mathrm{U}_{N-1} \ldots \mathrm{U}_2 \mathrm{U}_1}_{\mathrm{U}_{\mathrm{tot}}} \rho_0 \underbrace{\mathrm{U}_1^\dagger \mathrm{U}_2^\dagger \ldots \mathrm{U}_{N-1}^\dagger \mathrm{U}_N^\dagger}_{\mathrm{U}_{\mathrm{tot}}^\dagger}. \qquad (5)$$

Typically we want to maximize the fidelity functional,

$$\mathcal{F} = \langle \varrho | \rho_N \rangle := \mathrm{Tr}(\varrho^\dagger \rho_N) \quad \in [0, 1],$$

to have maximum overlap (i.e. $\mathcal{F} = 1$) with the (normalized) target state ϱ, which is also Hermitian.

In a gradient-based optimization scheme one needs to compute the gradient of the fidelity function $\mathcal{F}$,

$$\frac{\partial \mathcal{F}}{\partial \theta_{n,k}} = \mathrm{Tr}\left(\varrho^\dagger \frac{\partial \rho_N}{\partial \theta_{n,k}} \right), \qquad (6)$$

while a Newton–Raphson optimization scheme also requires the Hessian,

$$\frac{\partial^2 \mathcal{F}}{\partial \theta_{m,j} \partial \theta_{n,k}} = \mathrm{Tr}\left(\varrho^\dagger \frac{\partial^2 \rho_N}{\partial \theta_{m,j} \partial \theta_{n,k}} \right). \qquad (7)$$

Here $n, m \in \{1, \ldots, N\}$ and $j, k \in \{1, 2\}$, and

$$\theta_{n,1} = f(t_{n-1}), \qquad \theta_{n,2} = g(t_{n-1})$$

are the control parameters that solely affect the nth propagator, U_n.

In the computation of the gradient of the fidelity function (6), we require the gradient of the final state ρ_N. Since $\theta_{n,k}$ only affects the nth propagator,

$$\frac{\partial \rho_N}{\partial \theta_{n,k}} = 2\mathrm{Re}\left(\mathrm{L}_{n+1} \frac{\partial \mathrm{U}_n}{\partial \theta_{n,k}} \mathrm{R}_{n-1} \rho_0 \mathrm{U}_{\mathrm{tot}}^\dagger \right), \qquad (8)$$

$$\mathrm{L}_n = \mathrm{U}_N \mathrm{U}_{N-1} \ldots \mathrm{U}_n, \qquad (9)$$

$$\mathrm{R}_n = \mathrm{U}_n \mathrm{U}_{n-1} \ldots \mathrm{U}_1, \qquad (10)$$

where L_n and R_n for all $n = 1, \ldots, N$ can be computed in $\mathcal{O}(N)$ time.

Here we present a method for computing the gradient $\partial \mathrm{U}_n / \partial \theta_{n,k}$, and therefore the gradient of the fidelity function, analytically using Lie algebraic techniques. This approach is also extended for computing the Hessian analytically.

2.2 Computation of gradient

In this section we present the analytic approach for computing the derivative of the nth unitary propagator (4), $U_n = \exp(-i\boldsymbol{s}_n(\theta_{n,k}) \cdot \boldsymbol{\sigma})$, with respect to a control parameter $\theta_{n,k}$. Here we write $\boldsymbol{s}_n(\theta_{n,k})$ to highlight the fact that $\boldsymbol{s}_n$ depends on $\theta_{n,k}$.

In general, the derivative of the exponential of $X(\theta)$ with respect to a control parameter θ can be expressed as [34, 35]

$$\frac{\partial}{\partial\theta}\exp(X(\theta)) = \exp(X(\theta))\,\mathrm{dexp}_{X(\theta)}X'(\theta), \qquad (11)$$

where the dexp function,

$$\mathrm{dexp}_X X' = \sum_{p=0}^{\infty} \frac{(-1)^p}{(p+1)!}\mathrm{ad}_X^p(X'), \qquad (12)$$

is expressed as a power series of the adjoint operator, ad. The powers of ad are given by

$$\mathrm{ad}_X^0(X') = X', \qquad \mathrm{ad}_X^{n+1}(X') = [X, \mathrm{ad}_X^n(X')].$$

Equations (11) and (12) allow us to express the derivative of U_n,

$$\begin{aligned}\frac{\partial U_n}{\partial\theta_{n,k}} &= \frac{\partial}{\partial\theta_{n,k}}e^{-i\boldsymbol{s}_n(\theta_{n,k})\cdot\boldsymbol{\sigma}} \\ &= U_n\left(\sum_{p=0}^{\infty}\frac{(-1)^p}{(p+1)!}\mathrm{ad}_{-i\boldsymbol{s}_n\cdot\boldsymbol{\sigma}}^p\right)\left(-i\frac{\partial\boldsymbol{s}_n}{\partial\theta_{n,k}}\cdot\boldsymbol{\sigma}\right).\end{aligned}$$

Remark 2 *For ease of notation, we suppress the dependence of $\boldsymbol{s}$ on the control parameters, θ.*

An explicit formula can be derived for the dexp series when $X(\theta) \in \mathfrak{su}(2)$. To see this, we introduce the map $\sim$ which maps vectors in $\mathbb{R}^3$ to matrices in su(2),

$$\widetilde{\boldsymbol{s}} = -i\boldsymbol{s}\cdot\boldsymbol{\sigma}, \qquad \boldsymbol{s}\in\mathbb{R}^3.$$

It is easy to verify that $\mathrm{ad}_{-i\boldsymbol{s}_n\cdot\boldsymbol{\sigma}}(\widetilde{\boldsymbol{r}}) = [\widetilde{\boldsymbol{s}_n}, \widetilde{\boldsymbol{r}}] = \widetilde{\boldsymbol{S}_n\boldsymbol{r}}$ and

$$\mathrm{ad}_{-i\boldsymbol{s}_n\cdot\boldsymbol{\sigma}}^p(\widetilde{\boldsymbol{r}}) = \widetilde{\boldsymbol{S}_n^p\boldsymbol{r}} = -i(\boldsymbol{S}_n^p\boldsymbol{r})\cdot\boldsymbol{\sigma}, \quad p\geq 0,$$

where $\boldsymbol{S}_n$ is the matrix,

$$\boldsymbol{S}_n = \begin{pmatrix} 0 & -s_{n,z} & s_{n,y} \\ s_{n,z} & 0 & -s_{n,x} \\ -s_{n,y} & s_{n,x} & 0 \end{pmatrix}. \qquad (13)$$

Consequently,

$$\frac{\partial U_n}{\partial\theta_{n,k}} = -iU_n\left[\underbrace{\left(\sum_{p=0}^{\infty}\frac{(-\boldsymbol{S}_n)^p}{(p+1)!}\right)}_{3\times 3}\underbrace{\left(\frac{\partial\boldsymbol{s}_n}{\partial\theta_{n,k}}\right)}_{3\times 1}\right]\cdot\boldsymbol{\sigma}.$$

Using $\boldsymbol{S}_n^3 = -\|\boldsymbol{s}_n\|^2\,\boldsymbol{S}_n$, We can further simplify the dexp series as

$$\boldsymbol{D}_n = \sum_{p=0}^{\infty}\frac{(-\boldsymbol{S}_n)^p}{(p+1)!} = I + c_1\boldsymbol{S}_n + c_2\boldsymbol{S}_n^2, \qquad (14)$$

$$c_1 = \frac{\cos(\|\boldsymbol{s}_n\|)-1}{\|\boldsymbol{s}_n\|^2}, \quad c_2 = \frac{\|\boldsymbol{s}_n\|-\sin(\|\boldsymbol{s}_n\|)}{\|\boldsymbol{s}_n\|^3}. \qquad (15)$$

For a piece-wise constant pulse,

$$\boldsymbol{s}_n = (\Delta t)\,(f(t_{n-1}), g(t_{n-1}), \Omega)^\top,$$

and the control parameters are $\theta_{n,1} = f(t_{n-1})$ and $\theta_{n,2} = g(t_{n-1})$. Consequently,

$$\frac{\partial\boldsymbol{s}_n}{\partial\theta_{n,1}} = (\Delta t)\,(1,0,0)^\top, \quad \frac{\partial\boldsymbol{s}_n}{\partial\theta_{n,2}} = (\Delta t)\,(0,1,0)^\top. \quad (16)$$

To summarise, the analytic derivative of U_n is given by

$$\frac{\partial U_n}{\partial\theta_{n,k}} = -iU_n\left(\left[\boldsymbol{D}_n\frac{\partial\boldsymbol{s}_n}{\partial\theta_{n,k}}\right]\cdot\boldsymbol{\sigma}\right), \qquad (17)$$

and the derivative of U_tot by

$$\frac{\partial U_\mathrm{tot}}{\partial\theta_{n,k}} = -iL_n\left(\left[\boldsymbol{D}_n\frac{\partial\boldsymbol{s}_n}{\partial\theta_{n,k}}\right]\cdot\boldsymbol{\sigma}\right)R_{n-1}. \qquad (18)$$

Using the definition of L_n and R_n in equations (9) and (10) it is evident that $L_nR_{n-1} = U_\mathrm{tot}$ and consequently we may write

$$L_n = U_\mathrm{tot}R_{n-1}^\dagger, \qquad R_{n-1} = L_n^\dagger U_\mathrm{tot}. \qquad (19)$$

Therefore (18) can be re-written as:

$$\frac{\partial U_\mathrm{tot}}{\partial\theta_{n,k}} = -i\mathcal{L}_{n,k}U_\mathrm{tot}, \qquad (20)$$

where for any pulse segment n and any control parameter k:

$$\mathcal{L}_{n,k} = L_n\left(\left[\boldsymbol{D}_n\frac{\partial\boldsymbol{s}_n}{\partial\theta_{n,k}}\right]\cdot\boldsymbol{\sigma}\right)L_n^\dagger, \qquad (21)$$

In a practical implementation, L_n is given by equations (9), and we compute $\boldsymbol{D}_n$ using equation (14).

Combining equations (6), (8) and (20), the final form of the analytical gradient of the fidelity is:

$$\frac{\partial \mathcal{F}}{\partial \theta_{n,k}} = 2\mathrm{Im}\mathrm{Tr}(\mathcal{L}_{n,k}\rho_N \varrho^\dagger). \tag{22}$$

This completes the description of the analytic gradients. Note that ρ_N is required in computation of the fidelity $\mathcal{F}$ and is assumed to be available already. The 2N matrices $\mathcal{L}_{n,k}$ are computed along with L_n and $\boldsymbol{D}_n$ in $\mathcal{O}(N)$ time, following which the full gradient involves 2N computations of (22), bringing overall cost to $\mathcal{O}(N)$.

2.3 Computation of Hessian

In the computation of the Hessian of the fidelity function (7) we require the Hessian of the final state,

$$\frac{\partial^2 \rho_N}{\partial \theta_{m,j}\partial \theta_{n,k}} = 2\mathrm{Re}\left(\frac{\partial^2 \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{m,j}\partial \theta_{n,k}}\rho_0 \mathrm{U}_{\mathrm{tot}}^\dagger \right.$$
$$\left. + \frac{\partial \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{n,k}}\rho_0 \frac{\partial \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{m,j}}^\dagger\right). \tag{23}$$

An analytic form for the gradient of $\mathrm{U}_{\mathrm{tot}}$ with respect to control parameters $\theta_{n,k}$ and $\theta_{m,j}$ has already been obtained in equation (20). In this section, we derive an analytic form for $\partial^2 \mathrm{U}_{\mathrm{tot}}/\partial \theta_{m,j}\partial \theta_{n,k}$.

2.3.1 Off-diagonal entries ($n > m$) of the Hessian

When $n > m$, the Hessian is typically computed as

$$\frac{\partial^2 \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{m,j}\partial \theta_{n,k}} = \mathrm{L}_{n+1}\frac{\partial \mathrm{U}_n}{\partial \theta_{n,k}}\mathrm{M}_{n-1,m+1}\frac{\partial \mathrm{U}_m}{\partial \theta_{m,j}}\mathrm{R}_{m-1}.$$
$$\mathrm{M}_{n,m} = \mathrm{U}_n \mathrm{U}_{n-1}\ldots \mathrm{U}_{m-1}\mathrm{U}_m. \tag{24}$$

Similarly, we can derive the corresponding expression for $m > n$. Overall, since n and m range between 1 and N, the various values of $\mathrm{M}_{n,m}$ are typically computed in $\mathcal{O}(N^2)$ time in such a procedure.

Here we introduce an alternative approach for computing $\partial^2 \mathrm{U}_{\mathrm{tot}}/\partial \theta_{m,j}\partial \theta_{n,k}$ that does not require the computation of $\mathrm{M}_{n,m}$. Since L_{n+1} and R_{m-1} are unitary,

$$\mathrm{L}_{n+1}^\dagger \mathrm{L}_{n+1} = I, \qquad \mathrm{R}_{m-1}^\dagger \mathrm{R}_{m-1} = I, \tag{25}$$

we can express $\mathrm{M}_{n,m}$ as

$$\mathrm{M}_{n,m} = \mathrm{L}_{n+1}^\dagger \mathrm{U}_{\mathrm{tot}} \mathrm{R}_{m-1}^\dagger.$$

Thus, $\mathrm{M}_{n,m}$ can be replaced in the computation of the Hessian and equation (24) can be written in the form

$$\frac{\partial^2 \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{m,j}\partial \theta_{n,k}} = \mathrm{L}_{n+1}\frac{\partial \mathrm{U}_n}{\partial \theta_{n,k}}\mathrm{L}_n^\dagger \mathrm{U}_{\mathrm{tot}}\mathrm{R}_m^\dagger \frac{\partial \mathrm{U}_m}{\partial \theta_{m,j}}\mathrm{R}_{m-1}. \tag{26}$$

By substituting (17) in (26), combining it with (21) and (19) we have,

$$\frac{\partial^2 \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{m,j}\partial \theta_{n,k}} = -\mathcal{L}_{n,k}\mathcal{L}_{m,j}\mathrm{U}_{\mathrm{tot}}. \tag{27}$$

2.3.2 Diagonal entries ($m = n$) of the Hessian

For the case $m = n$ we have,

$$\frac{\partial^2 \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{n,j}\partial \theta_{n,k}} = \mathrm{L}_{n+1}\frac{\partial^2 \mathrm{U}_n}{\partial \theta_{n,j}\partial \theta_{n,k}}\mathrm{R}_{n-1}. \tag{28}$$

Differentiating equation (17) with respect to $\theta_{n,j}$,

$$\frac{\partial^2 \mathrm{U}_n}{\partial \theta_{n,j}\partial \theta_{n,k}} =$$
$$- \mathrm{U}_n \left\{ \left[\left(\boldsymbol{D}_n \frac{\partial \boldsymbol{s}_n}{\partial \theta_{n,j}}\right)\cdot \boldsymbol{\sigma}\right]\left[\left(\boldsymbol{D}_n \frac{\partial \boldsymbol{s}_n}{\partial \theta_{n,k}}\right)\cdot \boldsymbol{\sigma}\right] \right.$$
$$\left. +\mathrm{i}\left(\frac{\partial \boldsymbol{D}_n}{\partial \theta_{n,j}}\frac{\partial \boldsymbol{s}_n}{\partial \theta_{n,k}} + \boldsymbol{D}_n \frac{\partial^2 \boldsymbol{s}_n}{\partial \theta_{n,j}\partial \theta_{n,k}}\right)\cdot \boldsymbol{\sigma}\right\}, \tag{29}$$

where $\partial^2 \boldsymbol{s}_n/\partial \theta_{n,j}\partial \theta_{n,k}$ vanishes due to (16). The derivative of $\boldsymbol{D}_n$ (14) can be computed explicitly,

$$\frac{\partial \boldsymbol{D}_n}{\partial \theta_{n,j}} = c_1'\frac{\partial \|\boldsymbol{s}_n\|}{\partial \theta_{n,j}}\boldsymbol{S}_n + c_1\frac{\partial \boldsymbol{S}_n}{\partial \theta_{n,j}} \tag{30}$$
$$+ c_2'\frac{\partial \|\boldsymbol{s}_n\|}{\partial \theta_{n,j}}\boldsymbol{S}_n^2 + c_2\left(\boldsymbol{S}_n \frac{\partial \boldsymbol{S}_n}{\partial \theta_{n,j}} + \frac{\partial \boldsymbol{S}_n}{\partial \theta_{n,j}}\boldsymbol{S}_n\right),$$

where

$$c_1' = \frac{-2\cos(\|\boldsymbol{s}_n\|) - \|\boldsymbol{s}_n\|\sin(\|\boldsymbol{s}_n\|) + 2}{\|\boldsymbol{s}_n\|^3}, \tag{31}$$

$$c_2' = \frac{3\sin(\|\boldsymbol{s}_n\|) - \|\boldsymbol{s}_n\|\cos(\|\boldsymbol{s}_n\|) - 2\|\boldsymbol{s}_n\|}{\|\boldsymbol{s}_n\|^4}, \tag{32}$$

$$\frac{\partial \|\boldsymbol{s}_n\|}{\partial \theta_{n,j}} = \frac{\boldsymbol{s}_n \cdot \frac{\partial \boldsymbol{s}_n}{\partial \theta_{n,j}}}{\|\boldsymbol{s}_n\|}, \tag{33}$$

and $\partial \boldsymbol{S}_n/\partial \theta_{n,j}$ is obtained directly by differentiation of the matrix $\boldsymbol{S}$ in (13).

With a similar approach as in 2.3.1, using (19) and (21) we can rewrite (28) as

$$\frac{\partial^2 \mathrm{U}_{\mathrm{tot}}}{\partial \theta_{n,j}\partial \theta_{n,k}} = -\left(\mathcal{L}_{n,j}\mathcal{L}_{n,k} + \mathrm{i}\mathcal{D}_{n,j,k}\right)\mathrm{U}_{\mathrm{tot}}, \tag{34}$$

where

$$\mathcal{D}_{n,j,k} = \mathrm{L}_n \left[\left(\frac{\partial \boldsymbol{D}_n}{\partial \theta_{n,j}}\frac{\partial \boldsymbol{s}_n}{\partial \theta_{n,k}}\right)\cdot \boldsymbol{\sigma}\right]\mathrm{L}_n^\dagger. \tag{35}$$

4

2.3.3 General description of the full Hessian

The complete description of the Hessian of the fidelity is obtained by combining equations (7), (23), (27), and (34). The general form of the diagonal elements of the Hessian matrix will be:

$$\frac{\partial^2 \mathcal{F}}{\partial \theta_{n,j} \partial \theta_{n,k}} = 2\mathrm{ReTr}(\mathcal{L}_{n,k} \rho_N \mathcal{L}_{n,j}^\dagger \varrho^\dagger \\ - (\mathcal{L}_{n,j} \mathcal{L}_{n,k} + \mathrm{i}\mathcal{D}_{n,j,k}) \rho_N \varrho^\dagger), \quad (36)$$

and the general form of the upper-diagonal elements of the Hessian matrix can be written as

$$\frac{\partial^2 \mathcal{F}}{\partial \theta_{m,j} \partial \theta_{n,k}} = 2\mathrm{ReTr}(\mathcal{L}_{n,k} \rho_N \mathcal{L}_{m,j}^\dagger \varrho^\dagger - \mathcal{L}_{n,k} \mathcal{L}_{m,j} \rho_N \varrho^\dagger). \tag{37}$$

$\mathcal{L}$ and $\mathcal{D}$ can be precomputed in $\mathcal{O}(N)$ time along with L. The factorization (21) reduces the computational effort by a factor of three since only two matrix multiplications are required for each entry of the Hessian. Note that the lower-diagonal elements ($n < m$) can be easily obtained using the symmetry of the Hessian matrix and do not need to be computed separately. Equation (37) for these entries can be written as

$$\frac{\partial^2 \mathcal{F}}{\partial \theta_{m,j} \partial \theta_{n,k}} = 2\mathrm{ReTr}(\mathcal{L}_{n,k} \rho_N \mathcal{L}_{m,j}^\dagger \varrho^\dagger - \mathcal{L}_{m,j} \mathcal{L}_{n,k} \rho_N \varrho^\dagger). \tag{38}$$

Finally, using equations (36), (37) and (38), a general form of all Hessian entries can be expressed as a single equation,

$$\frac{\partial^2 \mathcal{F}}{\partial \theta_{m,j} \partial \theta_{n,k}} = 2\mathrm{ReTr}(\mathcal{L}_{n,k} \rho_N \mathcal{L}_{m,j}^\dagger \varrho^\dagger \\ - (\underbrace{\mathcal{L}_{n,k} \mathcal{L}_{m,j}}_{\substack{\mathcal{L}_{m,j} \mathcal{L}_{n,k} \\ \text{if } m > n}} + \underbrace{\mathrm{i}\mathcal{D}_{n,j,k}}_{\substack{0 \\ \text{if } n \neq m}}) \rho_N \varrho^\dagger). \quad (39)$$

3 Numerical demonstrations

3.1 Optimization routines

The proposed approach is very flexible, and once the analytic gradients (22) and Hessians (39) are available, they can be used in a variety of optimization routines. For instance, the gradients can be used in the context of gradient descent or quasi–Newton optimization techniques such as BFGS, while the Hessians can be used in any Newton type methods.

A numerical implementation of the proposed method in MATLAB along with additional functions for optimization and visualization of the performance are freely available via the following DOI: `10.17632/8zz84359m5.1`.

In our implementation, we have used MATLAB's constrained optimization function `fmincon` with the default algorithm `interior-point` for gradient-only optimization, and the `trust-region-reflective` algorithm when computing both gradients and Hessians. These can equally be used in the context of unconstrained optimization such as MATLAB's `fminunc`, and regularization using penalty functions can also be easily incorporated.

Similarly, the method can be combined, straightforwardly, with a range of algorithms in PYTHON, e.g. via `scipy.optimize.minimize` including BFGS, L-BFGS-B, and variants of `trust-region` algorithm (`dogleg`, `trust-ncg`, `trust-krylov`, `trust-exact`, and `trust-constr`).

3.2 Comparison with finite difference method

Finite difference approximation of the gradient of numerical propagators U_n requires computing U_n for multiple values of $\theta_{n,k}$ differing by a 'finite difference step'. Figure 1 demonstrates that while the finite difference step size must be kept sufficiently small for accuracy, the approximations become unstable for very small steps. Thus the suitability of a finite difference step may prove difficult to asses a-priori. This balance between accuracy and stability becomes more precarious when (i) the time step of the numerical propagator (Δt) is large, (ii) the numerical propagator is of limited accuracy or (iii) higher derivatives are required.

In addition to respecting the Lie algebraic structure and being relatively inexpensive, the proposed approach for computing analytic derivatives does not suffer from such instability. It should be noted that the accuracy of the computed derivatives using the proposed approach is entirely independent of size of the time step Δt.

3.3 Example for pulse design in magnetic resonance

In NMR and ESR, designing radiofrequency (RF) and microwave (MW) pulses for robust excitation of signals over a very wide range of frequencies and reduced sensitivity to instrumental imperfections is still among the most challenging areas of method design and is of great interest. The development of methods for pulse design in these applications generally follows one or more of three distinct routes: composite pulse design [36–41], evolutionary numerical methods like optimal control theory (OCT) [4, 42–48], and design of swept-frequency pulses [30, 49–58].

Here we demonstrate one of the applications of the proposed method for the design of broadband excitation pulses in NMR spectroscopy. The simplest case would be control of an ensemble of non-interacting spin-½ particles. Conventional instantaneous radio-frequency pulses have limited bandwidth due to high power requirements
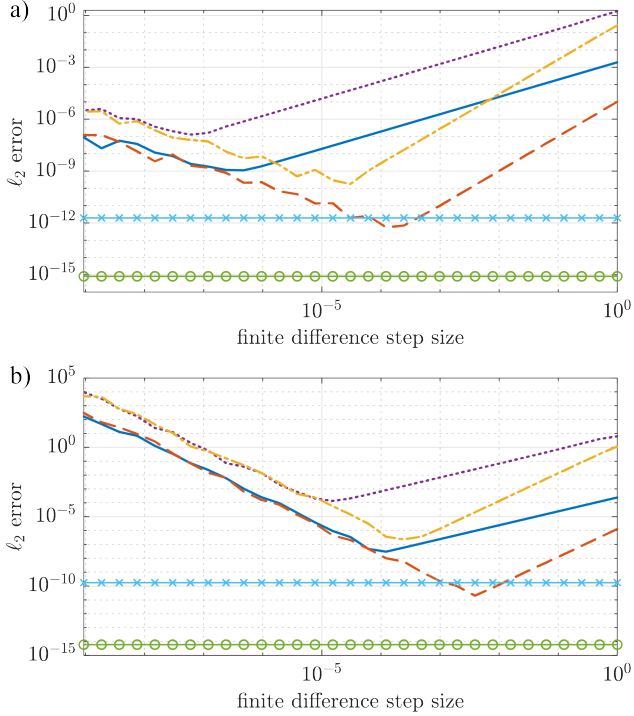
Fig. 1. Accuracy and stability of forward differences and central differences approximation to (a) the gradient $(\partial U_n / \partial \theta_{n,k})$, and (b) the Hessian $(\partial^2 U_n / \partial \theta_{n,k}^2)$ of a single step propagator U_n. The ℓ_2 error in approximation of derivatives is shown for two different propagation time steps (distinct from the finite difference step size): $\Delta t = 10^{-6}$ (forward difference [solid blue], central difference [dotted purple], analytic [solid green with circles]) and $\Delta t = 10^{-4}$ (forward difference [dashed orange], central difference [densely dashed yellow], analytic [solid light blue with crosses]).

that cannot be afforded on most instruments; therefore they can only satisfy the desired state manipulation in a rather limited range of frequencies close to the transmitter offset of the pulse, i.e. they are only effective for spins with relatively small frequency offsets; additionally, the performance of these pulses can be considerably affected by instrumental imperfections or instabilities. The goal here is to circumvent these problems by designing a pulse propagator that satisfies certain objectives for all spins within the desired frequency range, with a robust performance that does not depend on frequency offset of spins or instrumental imperfections.

The example here demonstrates an excitation pulse designed using the proposed method to bring all spins in the ensemble from z to y. Figure 2 (a) shows the final state of spin across the frequency range of interest (50 kHz here), and figure 2 (b) shows variations of one of the components, y, for five different offset frequencies during the 200 μs pulse. Additionally, we can incorporate an additional optimization step that significantly
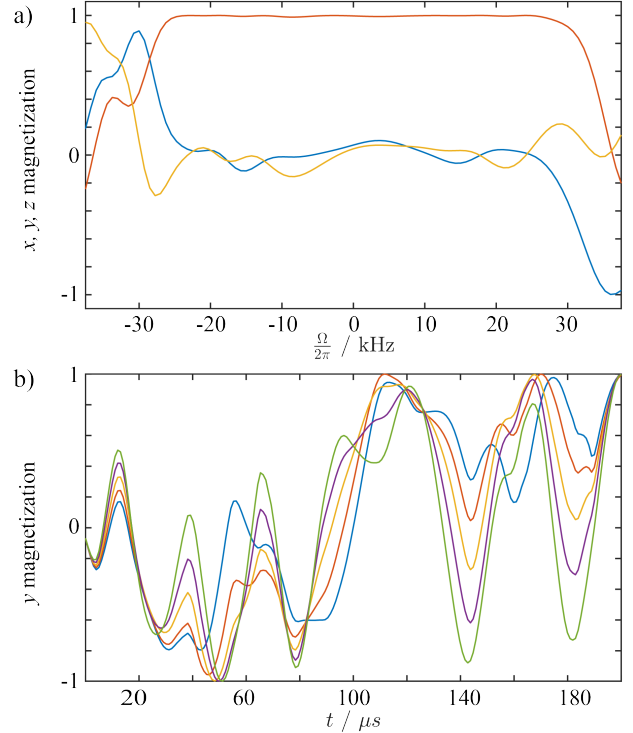

Fig. 2. (a) calculated excitation profiles for different component of density operator, x (blue), y (red), and z (orange) at $\omega_1 = \omega_1^0$ (20 kHz) for a 200 μs pulse acting on 101 non-interacting spin-½ over a 50 kHz frequency range; (b) variations of y component of spin trajectories during the pulse for five different frequencies.

reduces the sensitivity of the pulse to instrumental imperfections. Here we consider reducing the sensitivity of the pulse performance to unknown variations of the RF amplitude. Figure 3 (a) shows corresponding graphs for the variations of the target state, y, versus RF field, B_1. One common example of such imperfection is the position-dependent B_1 field across an RF coil used to generate the pulse. These variations introduce position-dependent phase of the signal across the ensemble of spins and therefore results in significant signal loss and non-uniform excitation profile of the pulse. Here an additional objective is to minimize the variation of signal phase with respect to the variation of B_1 field $\left(\frac{\mathrm{d}\phi}{\mathrm{d}B_1}\right)$, figure 3 (b) shows that for a given nominal RF amplitude with $\pm 20\%$ variations in the amplitude of B_1 field, $\frac{\mathrm{d}\phi}{\mathrm{d}B_1}$ is zero for all frequencies in the desired range.

4 Conclusions

In the present work, we have introduced a new approach, ESCALADE, for computation of derivatives of the cost function in optimal control of spin systems. We demonstrated that using the proposed mathematical framework, derivatives (gradient and Hessian) can be computed analytically using Lie algebraic techniques.
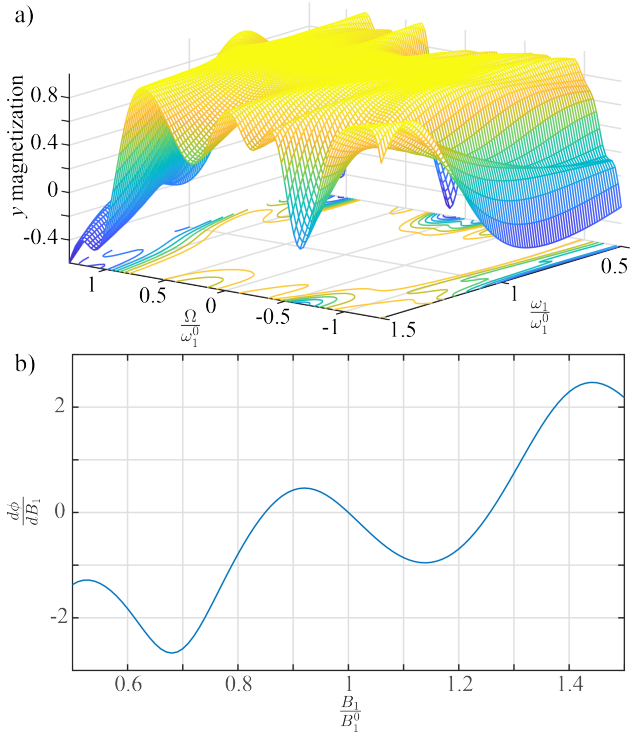
a)

b)

Fig. 3. (a) 3D plot and projection showing the y-magnetization excited as a function of relative resonance offset $(\frac{\Omega}{\omega_1^0})$ and relative RF amplitude $\frac{\omega_1}{\omega_1^0}$; (b) Sensitivity of signal phase to field strength $(\frac{\mathrm{d}\phi}{\mathrm{d}B_1})$ as a function of relative field strength $(\frac{B_1}{B_1^0})$.

The proposed method is general and can be adapted to and used in many potential applications where efficient optimal control of spin systems is required. These include high-resolution magnetic resonance spectroscopy and imaging [4], terahertz technologies [59, 60], and control of trapped ions [61], cold atoms [62] and NV-centers in diamond [63, 64].

Acknowledgements

References

[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(12), 2015.

[2] C. Cartis, J. Fiala, B. Marteau, and L. Roberts. Improving the flexibility and robustness of model-based derivative-free optimization solvers. *ACM Trans. Math. Softw.*, 45(3):1–41, 2019.

[3] D. L. Goodwin, W. K. Myers, C. R. Timmel, and I. Kuprov. Feedback control optimisation of esr experiments. *J. Magn. Reson.*, 297:9–16, 2018.

[4] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbruggen, and S. J. Glaser. Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms. *J. Magn. Reson.*, 172(2):296–305, 2005.

[5] D. Lu, K. Li, J. Li, H. Katiyar, A. J. Park, G. Feng, T. Xin, H. Li, G. Long, A. Brodutch, J. Baugh, B. Zeng, and R. Laflamme. Enhancing quantum control by bootstrapping a quantum processor of 12 qubits. *Npj Quantum Inf.*, 3(1):45, 2017.

[6] R. Eitan, M. Mundt, and D. J. Tannor. Optimal control with accelerated convergence: Combining the krotov and quasi-newton methods. *Phys. Rev. A*, 83(5), 2011.

[7] S. G. Schirmer and P. de Fouquieres. Efficient algorithms for optimal control of quantum dynamics: the krotov method unencumbered. *New J. Phys.*, 13(7), 2011.

[8] P. de Fouquieres, S. G. Schirmer, S. J. Glaser, and I. Kuprov. Second order gradient ascent pulse engineering. *J. Magn. Reson.*, 212(2):412–7, 2011.

[9] D. L. Goodwin and I. Kuprov. Modified newton-raphson grape methods for optimal control of spin systems. *J. Chem. Phys.*, 144(20):204107, 2016.

[10] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.

[11] V. Jurdjevic. Optimal control on lie groups and integrable hamiltonian systems. *Regul. Chaotic Dyn.*, 16(5):514–535, 2011.

[12] G. C. Walsh, R. Montgomery, and S. S. Sastry. Optimal path planning on matrix lie groups. In *Proc. IEEE Conf. Decis. Control*, volume 2, pages 1258–1263, 1994.

[13] S. Amari. Natural gradient works efficiently in learning. *Neural Comput.*, 10(2):251–276, 1998.

[14] Y. Nishimori. Learning algorithm for independent component analysis by geodesic flows on orthogonal group. In *Proc. Int. Jt. Conf. Neural Netw.*, volume 2, pages 933–938, 1999.

[15] M. D. Plumbley. Lie group methods for optimization with orthogonality constraints. In C. G. Puntonet and Alberto Prieto, editors, *Independent Component Analysis and Blind Signal Separation*, pages 1245–1252. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[16] G. Dirr, U. Helmke, K. Hüper, M. Kleinsteuber, and Y. Liu. Spin dynamics: A paradigm for time optimal control on compact lie groups. *J. Global Optim.*, 35(3):443–474, 2006.

[17] K. Spindler. Optimal control on lie groups: Theory and applications. *WSEAS Trans. Math.*, 12(5):531–

542, 2013.

[18] T. Drummond and R. Cipolla. Application of lie algebras to visual servoing. *Int. J. Comput. Vis.*, 37(1):21–41, 2000.

[19] O. Tuzel, R. Subbarao, and P. Meer. Simultaneous multiple 3d motion estimation via mode finding on lie groups. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 18–25 Vol. 1, 2005.

[20] C. J. Taylor and D. J. Kriegman. Minimization on the lie group so(3) and related manifolds. Technical Report 9405, Yale University, 1994.

[21] Y. Kosmann-Schwarzbach. *Groups and Symmetries*. Springer, 2010. ISBN 978-0-387-78865-4. doi: 10.1007/978-0-387-78866-1.

[22] P. Woit. *Quantum Theory, Groups and Representations*. Springer, 2017. ISBN 978-3-319-87835-5. doi: 10.1007/978-3-319-64612-1.

[23] Rodrigues. Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *J. Math. Pures Appl.*, pages 380–440, 1840.

[24] G. Gallego and A. Yezzi. A compact formula for the derivative of a 3-d rotation in exponential coordinates. *J. Math. Imaging Vis.*, 51:378–384, 2015.

[25] G. Terzakis, M. Lourakis, and D. Ait-Boudaoud. Modified rodrigues parameters: An efficient representation of orientation in 3d vision and graphics. *J. Math. Imaging Vis.*, 60(3):422–442, 2018.

[26] N. Khaneja, S. J. Glaser, and R. Brockett. Subriemannian geometry and time optimal control of three spin systems: Quantum gates and coherence transfer. *Phys. Rev. A*, 65(3):1–11, 2002.

[27] N. Khaneja, S. Glaser, and R. Brockett. Subriemannian geodesics and optimal control of spin systems. *Proc. Am. Control Conf.*, 4:2806–2811, 2002.

[28] B. Bonnard, S. J. Glaser, and D. Sugny. A review of geometric optimal control for quantum systems in nuclear magnetic resonance. *Adv. Math. Phys.*, 2012:1–29, 2012.

[29] B. Bonnard, O. Cots, S. J. Glaser, M. Lapert, D. Sugny, and Zhang Yun. Geometric optimal control of the contrast imaging problem in nuclear magnetic resonance. *IEEE Trans. Autom. Control*, 57 (8):1957–1969, 2012.

[30] T. Kato. On the adiabatic theorem of quantum mechanics. *J. Phys. Soc. Jpn.*, 5(6):435–439, 1950.

[31] C Brif, M. D. Grace, M. Sarovar, and K. C. Young. Exploring adiabatic quantum trajectories via optimal control. *New J. Phys.*, 16(6):065013, 2014.

[32] S. Meister, J. T. Stockburger, R. Schmidt, and J. Ankerhold. Optimal control theory with arbitrary superpositions of waveforms. *J. Phys. A*, 47 (49), 2014.

[33] N. Augier, U. Boscain, and M. Sigalotti. Adiabatic ensemble control of a continuum of quantum systems. *SIAM J. Control Optim.*, 56(6):4045–4068, 2018.

[34] F. Schur. Zur theorie der endlichen transformationsgruppen. *Abh. Math. Sem. Univ. Hamburg*, 4: 15–32, 1891.

[35] W. Rossmann. *Lie Groups: An Introduction Through Linear Groups*. Oxford University Press, 2006.

[36] M. H. Levitt and R. R. Ernst. Composite pulses constructed by a recursive expansion procedure. *J. Magn. Reson.*, 55(2):247–254, 1983.

[37] A. J. Shaka and R. Freeman. Composite pulses with dual compensation. *J. Magn. Reson.*, 55(3): 487–493, 1983.

[38] M. H. Levitt, D. Suter, and R. R. Ernst. Composite pulse excitation in three-level systems. *J. Chem. Phys.*, 80(7):3064–3068, 1984.

[39] R. Tycko, H. M. Cho, E. Schneider, and A. Pines. Composite pulses without phase distortion. *J. Magn. Reson.*, 61(1):90–101, 1985.

[40] M. H. Levitt. Composite pulses. *Prog. NMR Spectrosc.*, 18(2):61–122, 1986.

[41] A. J. Shaka and A. Pines. Symmetric phase-alternating composite pulses. *J. Magn. Reson.*, 71 (3):495–503, 1987.

[42] N. Khaneja, R. Brockett, and S. J. Glaser. Time optimal control in spin systems. *Phys. Rev. A*, 63 (3):1–13, 2001.

[43] N. Khaneja, T. Reiss, B. Luy, and S. J. Glaser. Optimal control of spin dynamics in the presence of relaxation. *J. Magn. Reson.*, 162(2):311–319, 2003.

[44] K. Kobzar, B. Luy, N. Khaneja, and S. J. Glaser. Pattern pulses: design of arbitrary excitation profiles as a function of pulse amplitude and offset. *J. Magn. Reson.*, 173(2):229–35, 2005.

[45] Z. Tosner, T. Vosegaard, C. Kehlet, N. Khaneja, S. J. Glaser, and N. C. Nielsen. Optimal control in nmr spectroscopy: numerical implementation in simpson. *J. Magn. Reson.*, 197(2):120–34, 2009.

[46] K. Kobzar, S. Ehni, T. E. Skinner, S. J. Glaser, and B. Luy. Exploring the limits of broadband 90 degrees and 180 degrees universal rotation pulses. *J. Magn. Reson.*, 225:142–60, 2012.

[47] I. Kuprov. Spin system trajectory analysis under optimal control pulses. *J. Magn. Reson.*, 233:107–12, 2013.

[48] E. Van Reeth, H. Rafiney, M. Tesch, S. J. Glaser, and D. Sugny. Optimizing mri contrast with b1 pulses using optimal control theory. In *Proc. IEEE Int. Symp. Biomed. Imaging*, pages 310–313, 2016.

[49] J. Baum, R. Tycko, and A. Pines. Broadband population inversion by phase modulated pulses. *J. Chem. Phys.*, 79(9):4643–4644, 1983.

[50] J. Baum, R. Tycko, and A. Pines. Broadband and adiabatic inversion of a two-level system by phase-modulated pulses. *Phys. Rev. A*, 32(6):3435–3447, 1985.

[51] D. Kunz. Use of frequency-modulated radiofrequency pulses in mr imaging experiments. *Magn. Reson. Med.*, 3(3):377–84, 1986.

[52] T. Fujiwara and K. Nagayama. Optimized frequency/phase-modulated broadband inversion pulses. *J. Magn. Reson.*, 86(3):584–592, 1990.

[53] R. Fu and G. Bodenhausen. Broadband decoupling in nmr with frequency-modulated 'chirp' pulses. *Chem. Phys. Lett.*, 245(4-5):415–420, 1995.

[54] Ē. Kupče and R. Freeman. Adiabatic pulses for wideband inversion and broadband decoupling. *J. Magn. Reson. A*, 115(2):273–276, 1995.

[55] A. Doll and G. Jeschke. Fourier-transform electron spin resonance with bandwidth-compensated chirp pulses. *J. Magn. Reson.*, 246:18–26, 2014.

[56] J. E. Power, M. Foroozandeh, R. W. Adams, M. Nilsson, S. R. Coombes, A. R. Phillips, and G. A. Morris. Increasing the quantitative bandwidth of nmr measurements. *Chem. Commun.*, 52(14):2916–9, 2016.

[57] N. Khaneja. Chirp excitation. *J. Magn. Reson.*, 282:32–36, 2017.

[58] M. Foroozandeh, M. Nilsson, and G. A. Morris. Improved ultra-broadband chirp excitation. *J. Magn. Reson.*, 302:28–33, 2019.

[59] E. Rasanen, A. Castro, J. Werschnik, A. Rubio, and E. K. Gross. Optimal control of quantum rings by terahertz laser pulses. *Phys. Rev. Lett.*, 98(15): 157404, 2007.

[60] L. H. Coudert. Optimal control of the orientation and alignment of an asymmetric-top molecule with terahertz and laser pulses. *J. Chem. Phys.*, 148(9), 2018.

[61] M. Zhao and D. Babikov. Coherent and optimal control of adiabatic motion of ions in a trap. *Phys. Rev. A*, 77(1), 2008.

[62] J. C. Saywell, I. Kuprov, D. L. Goodwin, M. Carey, and T. Freegarde. Optimal control of mirror pulses for cold-atom interferometry. *Phys. Rev. A*, 98(2), 2018.

[63] Y. Chou, S. Huang, and H. 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(5), 2015.

[64] J. Tian, T. Du, Y. Liu, H. Liu, F. Jin, R. S. Said, and J. Cai. Optimal quantum optical control of spin in diamond. *Phys. Rev. A*, 100(1), 2019.