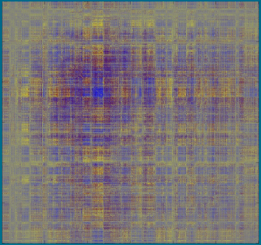
Chemistry

UNIVERSITY OF Southampton

Methods of Optimal Control in Magnetic Resonance

Hessian Calculations, Newton-GRAPE & some Applications



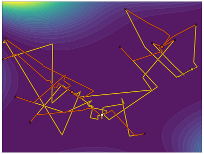
David L. Goodwin &
Ilya Kuprov
d.goodwin@soton.ac.uk

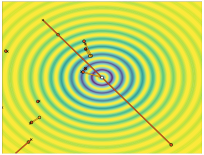
iMR-CDT@Southampton, Wednesday 23rd March 2016

Spinach

Newton methods

What should I minimise?





- We always minimise a number (or minimise the negative of a number - maximising).
- This number can be created in any way - it is only a metric (although it can have a physical significance).
- Minimising many variables can be equivalent to minimising the sum of those elements (usually with normalisation i.e an average).
- Usually, we minimise this metric as the output of a function.
- These functions can be “physics” or even “a black-box machine”.
- Essentially - we can treat any well behaved functions as a “black-box” - with only inputs and outputs seen by the optimisation algorithm.

- A condition for minimisation is the $\Delta J = J(\mathbf{c}_{s+1}) - J(\mathbf{c}_s) > 0$
- Taylor series approximated to second order[1].
  - If J is continuously differentiable

$$\Delta J = J(\mathbf{c}_{s+1}) - J(\mathbf{c}_s) \approx \langle \nabla J_s | \mathbf{c}_s \rangle$$

  - If J is twice continuously differentiable

$$\Delta J = J(\mathbf{c}_{s+1}) - J(\mathbf{c}_s) \approx \langle \nabla J_s | \mathbf{c}_s \rangle + \frac{1}{2} \langle \mathbf{c}_s | \nabla^2 J_s | \mathbf{c}_s \rangle$$

- 1st order necessary condition: $\nabla J_s \Big|_{\text{min}} = 0$
- 2nd order necessary condition: $\nabla^2 J_s \Big|_{\text{min}}$ is positive semidefinite

[1] I. Newton. Methodus fluxionum et serierum infinitarum, 1671, B. Taylor. Methodus Incrementorum Directa & Inversa, 1715, J. Nocedal and S. J. Wright. Numerical optimization. Springer, 2006.

Optimal Control

Gradient assisted pulse engineering (GRAPE)

- Although some of quantum physics uses linear control theory (Quantum Kalman filter), bilinear control is more suited to that dealing with Lie algebras.
- Split Hamiltonian (or Liouvillian) to controllable and uncontrollable parts[2],

$$\hat{\hat{L}}(t) = \hat{\hat{L}}_0 + \sum_k c^{(k)}(t)\hat{\hat{L}}_k$$

- Maximise the fidelity measure, $J = \Re e \langle \hat{\sigma} | \exp_{(O)} \left[-i \int_0^T \hat{\hat{L}}(t) dt \right] | \hat{\rho}(0) \rangle$
- A gradient calculation is reduced to:

$$J = \langle \sigma | \hat{\hat{\mathcal{P}}}_N \hat{\hat{\mathcal{P}}}_{N-1} \hat{\hat{\mathcal{P}}}_{N-2} \hat{\hat{\mathcal{P}}}_{N-3} \underbrace{\hat{\hat{\mathcal{P}}}_{N-4} \ldots \hat{\hat{\mathcal{P}}}_3 \hat{\hat{\mathcal{P}}}_2 \hat{\hat{\mathcal{P}}}_1 | \rho_0 \rangle}_{\text{(I) propagate forwards from source}}$$

$$\frac{\partial}{\partial c_{N-3}^{(k)}} \hat{\hat{\mathcal{P}}}_{N-3}$$

(III) compute expectation of the derivative

$$J = \overbrace{\langle \sigma | \hat{\hat{\mathcal{P}}}_N \hat{\hat{\mathcal{P}}}_{N-1} \hat{\hat{\mathcal{P}}}_{N-2}}^{\text{(II) propagate backwards from target}} \hat{\hat{\mathcal{P}}}_{N-3} \hat{\hat{\mathcal{P}}}_{N-4} \ldots \hat{\hat{\mathcal{P}}}_3 \hat{\hat{\mathcal{P}}}_2 \hat{\hat{\mathcal{P}}}_1 | \rho_0 \rangle$$

---

[2] N. Khaneja et al. In: *Journal of Magnetic Resonance* 172.2 (2005), pp. 296 –305.

- Propagator over a time slice:

$$\hat{\hat{\mathcal{P}}}_n = \exp\left[-i\left(\hat{\hat{H}}_0 + \sum_k c_n^{(k)}\hat{\hat{H}}_k\right)\Delta t\right]$$

- Optimality conditions, $\frac{\partial J}{\partial c_k(t)} = 0$ at a minimum, and the Hessian matrix should be positive definite

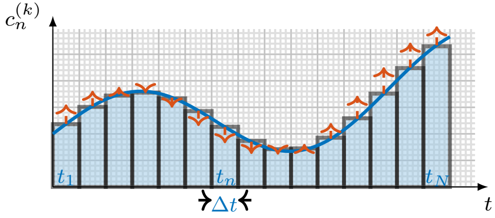
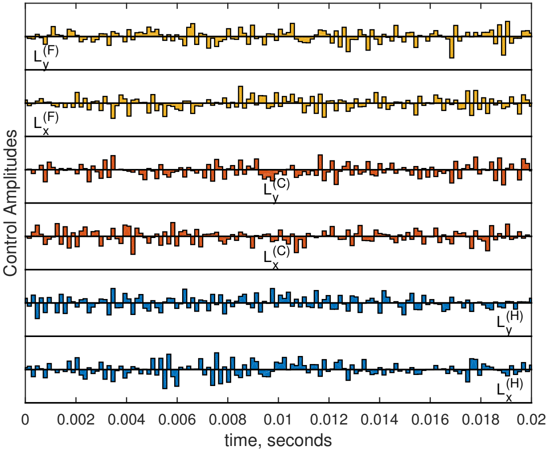


Figure: Piecewise constant approximation in a GRAPE simulation

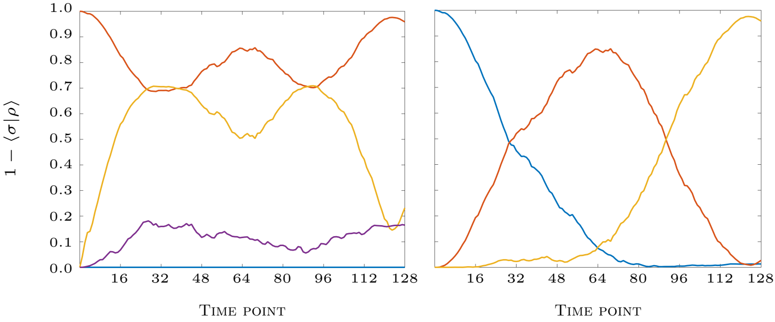
NMR state transfer

Control Pulses



NMR state transfer

Control Pulses



Population of correlation order subspaces (left)

0-spin 2-spin
1-spin 3-spin

Coherence orders local to each spin (right)

^{1}H ^{13}C ^{19}F

Work done and work in progress

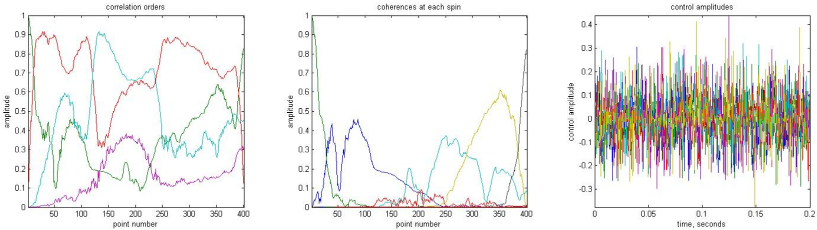
UNIVERSITY OF Southampton

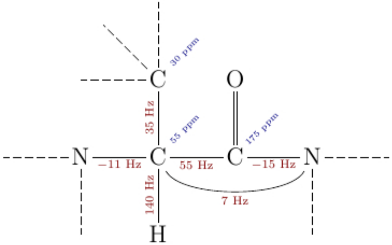
- Simple systems with many controls
- Singlet to multiplet optimisation
- Powder averaged optimisation for solid state NMR
- Controlling systems with noise
- Long range state-transfer (?)
- Newton-GRAPE method
- Conditioned BFGS-Krotov method
- Augmented exponentials

- Optimal control of MRI pulses (project due to start this summer, with Swansea)
- EPR pulse optimisation (treating the "machine" as the objective function - optimise the integral of an echo or the shape of a signal compared to a reference (project with Oxford)
- Solid state - making a general pulse set tolerant to a NQI range, different MAS rates, different power levels; with a small total pulse duration (experiments to be performed soon)
- Secret maths... (with Munich)
- Penalty functions
- Optimal control review

Larger systems

Simulation

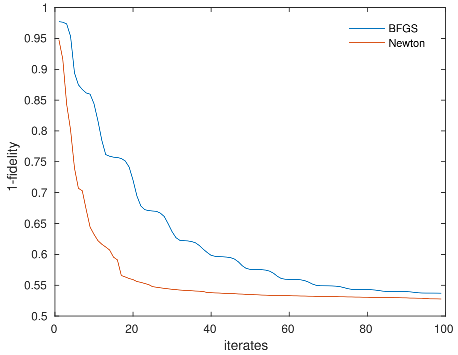


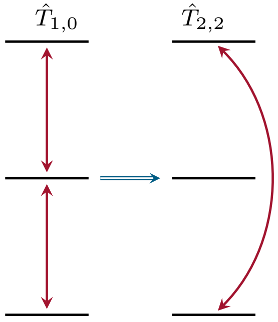


Overtone excitation

Simulation

- Excite ^{14}N from a state $\hat{T}_{1,0} \rightarrow \hat{T}_{2,2}$.
- Solid state powder average, with objective functional weighted over the crystalline orientations (rank 17 Lebedev grid - 110 points).
- Nuclear quadrupolar interaction.
- 400 time points for total pulse duration of $40\mu s$





"*nineteen dubious way to compute the exponential of a matrix*"[3].

Taylor series - concerning the error of truncating the arithmetic from roundoff error of floating point arithmetic, eps; increasing as $t\|A\|$ increases.

Padé approximation - Again, roundoff error can make this method unreliable. Also, unreliability occurs when A has widely spread eigenvalues.

Scaling and squaring - Exploits the relationship $\mathrm{e}^{A} = \left(\mathrm{e}^{A/m}\right)^{m}$ choosing m to be the smallest power of 2 for which $\|A\|/m \leq 1$. Can be used together with the Padé approximation (or Taylor series method).

Chebyshev rational approximation - $c_{qq}(x)$ is the ratio of two polynomials of degree q. A minimizer for $\|c_{qq}(A) - \mathrm{e}^{A}\|$ can be found from coefficients given in[4]. This is a useful method for sparse matrices.

Other methods Ordinary differential equation methods; Polynomial methods; Matrix decomposition methods... Particularly interesting in[5], investigating integrals and differentials involving the matrix exponential.

---

[3] C. Moler and C. Van Loan. In: *SIAM review* 20.4 (1978), pp. 801–836.

[4] W. Cody, G Meinardus and R. Varga. In: *Journal of Approximation Theory* 2.1 (1969), pp. 50–65.

[5] C. Kallstrom. In: *Lund Instit. Technol., Div. Automatic Control, Rep* 7309 (1973), C. F. Van Loan. In: *Automatic Control, IEEE Transactions on* 23.3 (1978), pp. 395–404.

- Find the derivative of the control pulse at a specific time point
- set

$$\int_0^1 e^{\mathbf{A}(1-s)}\mathbf{B}e^{\mathbf{A}s}ds = D_{c_n}(t)\exp\left(-i\hat{\hat{L}}\Delta t\right) \Rightarrow \mathbf{B} = -i\hat{\hat{H}}_n^{(k)}\Delta t$$

- leading to an efficient calculation of the gradient element

$$\exp\begin{pmatrix} -i\hat{\hat{L}}\Delta t & -i\hat{\hat{H}}_n^{(k)}\Delta t \\ \mathbf{0} & -i\hat{\hat{L}}\Delta t \end{pmatrix} = \begin{pmatrix} e^{-i\hat{\hat{L}}\Delta t} & \frac{\partial}{\partial c_n^{(k)}} e^{-i\hat{\hat{L}}\Delta t} \\ \mathbf{0} & e^{-i\hat{\hat{L}}\Delta t} \end{pmatrix}$$

- Furthermore, we can use an $n+1 \times n+1$ block matrix to find the n propagator derivatives or higher order derivatives in one exponential operation[6].

---

[6] D. L. Goodwin and I. Kuprov. In: *The Journal of Chemical Physics* 143.8 (2015), p. 084113.

- Can use augmented exponential with Krylov propagation

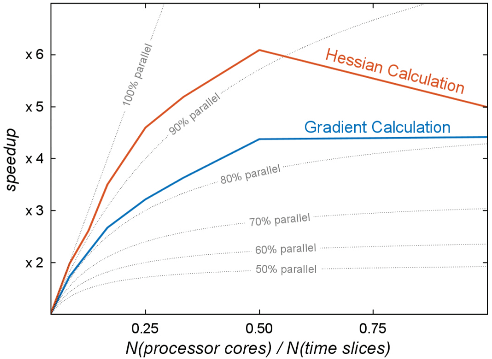
$$\exp\begin{bmatrix} -i\hat{\hat{L}}\Delta t & -i\hat{\hat{H}}_n^{(k)}\Delta t \\ \mathbf{0} & -i\hat{\hat{L}}\Delta t \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \rho_n \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{\partial}{\partial c_n^{(k)}} e^{-i\hat{\hat{L}}\Delta t} \\ \mathbf{0} \end{bmatrix}$$

- We need n of these Krylov propagations $+2 \times n$ Krylov propagations to (forwards and backwards to each n) to calculate the gradient element

$$\nabla J_n = \begin{bmatrix} \rho_{n+1} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \frac{\partial}{\partial c_n^{(k)}} e^{-i\hat{\hat{L}}\Delta t} \\ \mathbf{0} \end{bmatrix}$$

- If Krylov propagation is not appropriate, Spinach will calculate the explicit propagators with the matrix exponential of our auxiliary matrix.
- Particularly, we use the Taylor method (with gpu calculation if available), and preserve sparsity with a matrix cleanup during the iterative algorithm (remove any elements $< 10^{-14}$)

Parallelisation



- [7] Amdahl's law: parallelisation efficiency analysis for the Hessian calculation compared to the gradient calculation within Spinach implementation of GRAPE[8]. The optimal control problem involves 24 time slices and 6 control channels, yielding a fidelity functional gradient with 144 elements and a 144 × 144 Hessian.

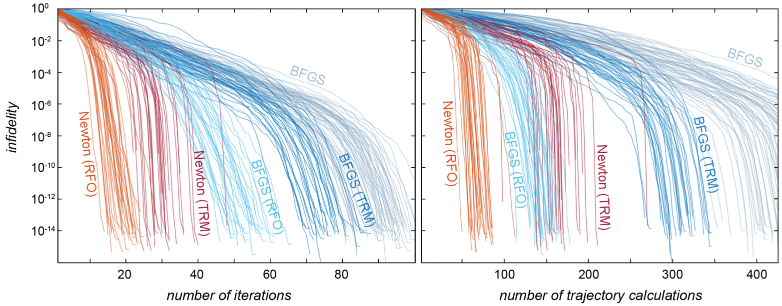
[7] D. L. Goodwin and I. Kuprov. In: *arXiv:1510.02420v4* (2016).

[8] P. de Fouquieres et al. In: *Journal of Magnetic Resonance* 212.2 (2011), pp. 412 –417, H. Hogben et al. In: *Journal of Magnetic Resonance* 208.2 (2011), pp. 179 –194.

State transfer

Comparison of BFGS and Newton-Raphson I

UNIVERSITY OF Southampton

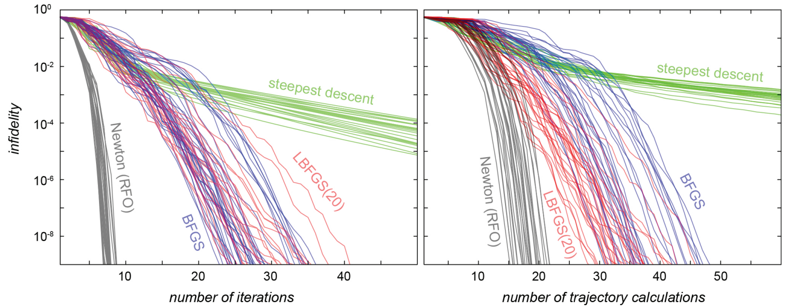


- [9] State transfer in an H–C–F group in a 9.4 Tesla magnet with ^{1}H isotope for hydrogen, ^{13}C isotope for carbon and ^{19}F isotope for fluorine, with the ^{1}H-^{13}C J-coupling of 140 Hz, ^{13}C-^{19}F J-coupling of –160 Hz and all three signals assumed to be on resonance with the transmitters on the corresponding NMR spectrometer channels. A six-channel $\{H_X, H_Y, C_X, C_Y, F_X, F_Y\}$ shaped pulse with a duration of 100 ms, a quadratic penalty for excursions outside the 10 kHz power envelope and 50 time discretisation points was optimized to perform longitudinal magnetization transfer from ^{1}H to ^{19}F.

[9] D. L. Goodwin and I. Kuprov. In: *arXiv:1510.02420v4* (2016).

State transfer

Comparison of BFGS and Newton-Raphson II



- [10] State transfer from longitudinal polarization into a two-spin singlet state, while allowing for up to 20% miscalibration of the control channel power level. The spin system contains two ^{13}C spins in a 14.1 Tesla magnet with chemical shifts of 0.00 and 0.25 ppm and a J-coupling of 60 Hz. The system is prepared in the $C_Z^{(1)} + C_Z^{(2)}$ state and a two-channel control sequence on $C_X^{(1)} + C_X^{(2)}$ and $C_X^{(1)} + C_X^{(2)}$ control operators with 50 time discretization points, the nominal power of 60 Hz and the duration of 50 milliseconds is optimized simultaneously for ten different power levels spaced equally between 80

[10] D. L. Goodwin and I. Kuprov. In: *arXiv:1510.02420v4* (2016).

Acknowledgements

UNIVERSITY OF Southampton

- Sophie Schirmer (University of Swansea)
- Thomas Schulte-Herbrüggen (TUM, Munich)
- Ville Bergholm (TUM, Munich)
- Steffan Glaser (TUM, Munich)
- Shai Machines (Wisemann institute, Israel)
- David Tannor (Wisemann institute, Israel)
- Willaim Myers (Oxford University)
- Christiane Timmel (Oxford University)
- Arzhang Ardavan (Oxford University)
- Marina Caravetta (University of Southampton)
- Phil Williamson (University of Southampton)

This work was made possible by EPSRC through a grant to IK group (EP/H003789/1) and a CDT studentship to DLG. The European Commission has facilitated this project through a coordination action grant (297861/QUAINT). EPR is additionally supported by EPSRC through a grant to Centre for Advanced Spin Resonance (EP/L011972/1)