Chemistry

UNIVERSITY OF Southampton

Optimal Control of Quantum Systems

ss-NMR, Noise Resilience, and a Scalable Newton's Method

D.L. Goodwin, iMR-CDT: 1st year Presentation, University of Nottingham.
3rd December 2014

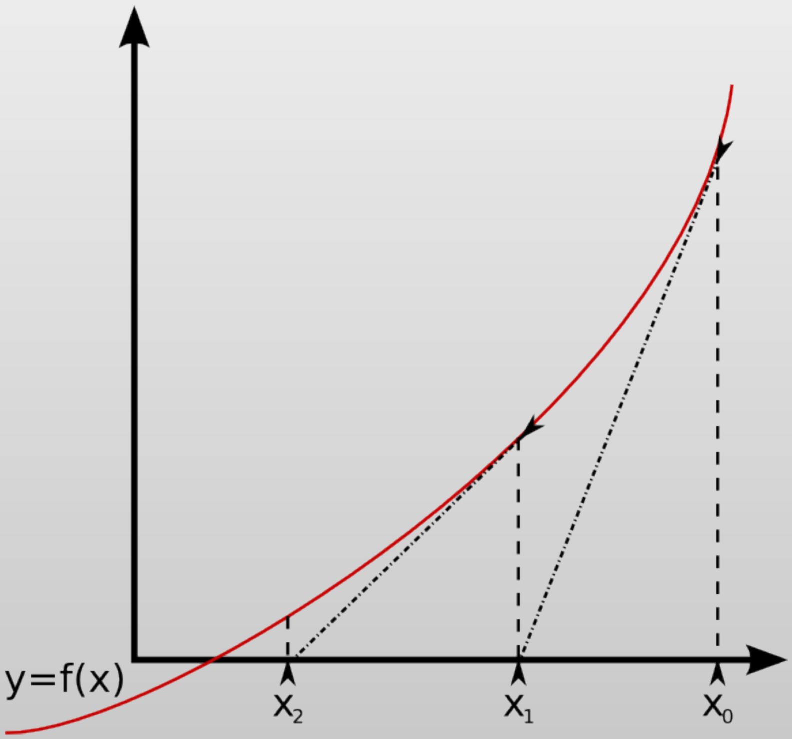
Optimal Control at Southampton

UNIVERSITY OF Southampton

UNIVERSITY OF Southampton

Introduction to Optimal Control

- Optimisation is the process of finding minimum or maximum (extrema) of functions
- Newton-Raphson method uses the gradient at a trial point to, incrementally, approach an extrema of an objective function
- $x_{k+1} = x_k - \frac{f(x)}{\nabla f(x)}$



- Basic method to use a **step length**, α, and a **step direction**, p, to iteratively find and extrema.
- We calculate a step direction, then decide how far to move in that direction:

$$x_{k+1} = x_k + \alpha_k p_k$$

- for minimisation, the step direction should be a descent direction.

$$p_k = -B_k^{-1} \nabla f_k$$

Gradient Descent B_k is the identity matrix

Newton's Method B_k is the exact Hessain matrix $\nabla^2 f(x_k)$

Quasi-Newton Methods B_k is an approximation to the Hessian.

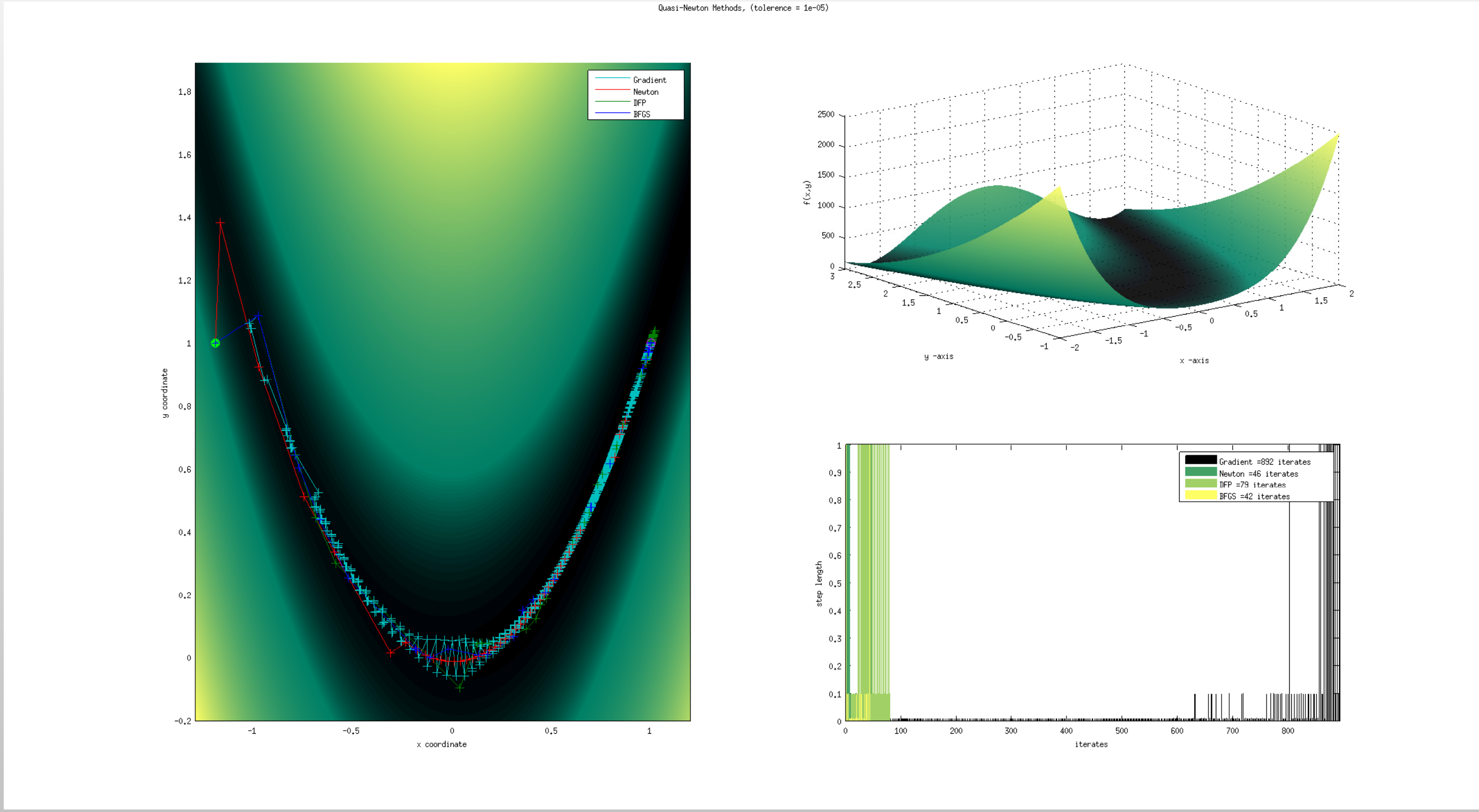
Gradient Descent Cheap, Scalable, Slow Convergence

Newton's Method Can be Expensive, Good Convergence.

BFGS Requires line search, No matrix inversion, Each update in $O(n^2)$.

l-BFGS Requires line search, No matrix inversion, Each update in $O(n)$.

Numerical Optimisation
Example - Rosenbruck Function



Optimal Control Theory

UNIVERSITY OF Southampton

- The idea of optimal control is to use all controllable parts of the full Hamiltonian to steer an initial quantum state into a target state.
- A measure of "success" for this control can be fidelity; the overlap between the controlled, final state and the target state (1 begin identical states, 0 being orthogonal states).
- The control parts of the Magnetic Resonance Hamiltonians are rf pulses for NMR, and mw pulses for ESR.
- The optimal solutions to the stated optimisation problem are a set of discrete-time control pulses.

Optimal Control Theory

- Separate Hamiltonian into the drift, $\hat{\mathcal{H}}_0$, and the parts we can control, $\hat{\hat{\mathcal{H}}}^{(k)}$.

$$\hat{\mathcal{H}}(t) = \hat{\mathcal{H}}_0 + \sum_k c^{(k)}(t)\hat{\hat{\mathcal{H}}}^{(k)}$$

- The control operators (e.g. Electromagnetic fields), each have a coefficient, $c^{(k)}(t)$, dictating the controllable time-evolution.
- Find $\{c^{(k)}(t)\}$ to take our system from an initial state, $|\psi_0\rangle$, to a target state, $|\sigma\rangle$

Optimal Control Theory

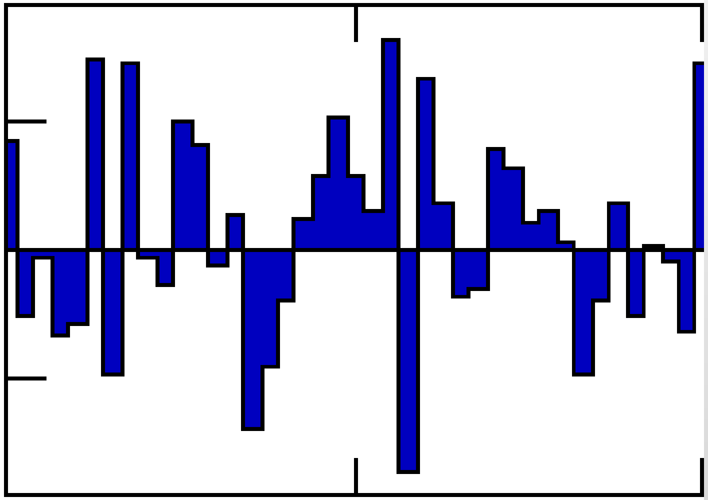
- Use fidelity as a metric to measure the overlap between target state and final state (1 = total overlap, 0 = no overlap)

$$J\{c^{(k)}(t)\} = \mathrm{Re}\,\langle\sigma|\,\exp_{(0)}\int_0^T -i\left(\hat{\mathcal{H}}_0 + \sum_k c^{(k)}(t)\hat{\hat{\mathcal{H}}}^{(k)}\right)dt\,|\psi_0\rangle$$

- The problem is to find the maximum of the fidelity (or the minimum of 1−fidelity).

GRAPE

- Assumption: control pulse sequence is piecewise constant.
- Time-ordered exponential now a sequential multiplication of propagators.



$$J\{c^{(k)}(t)\} = \mathrm{Re}\left\langle \sigma \middle| \hat{\mathcal{U}}_N \hat{\mathcal{U}}_{N-1} \cdots \hat{\mathcal{U}}_2 \mathcal{U}_1 \middle| \psi_0 \right\rangle$$

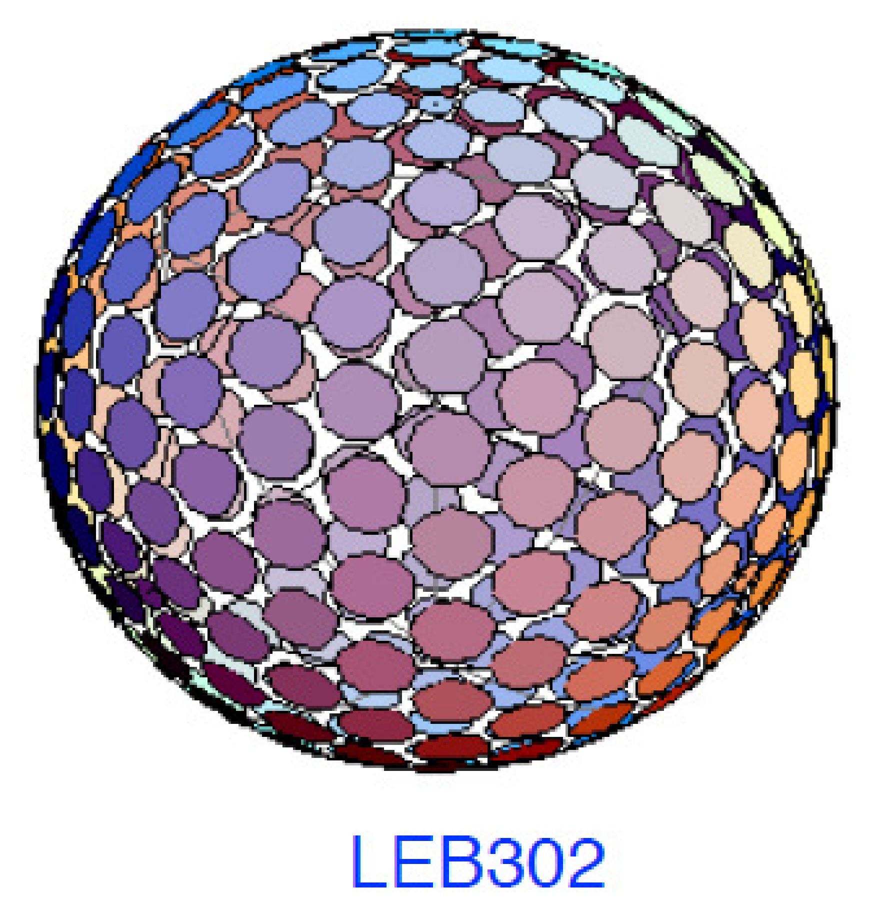
$$\hat{\mathcal{U}}_n = \exp_{(0)}\left[-i \left(\hat{\mathcal{H}}_0 + \sum_k c^{(k)}(t_n) \hat{\hat{\mathcal{H}}}^{(k)} \right) \Delta t \right]$$

Solid State Nuclear Magnetic Resonance

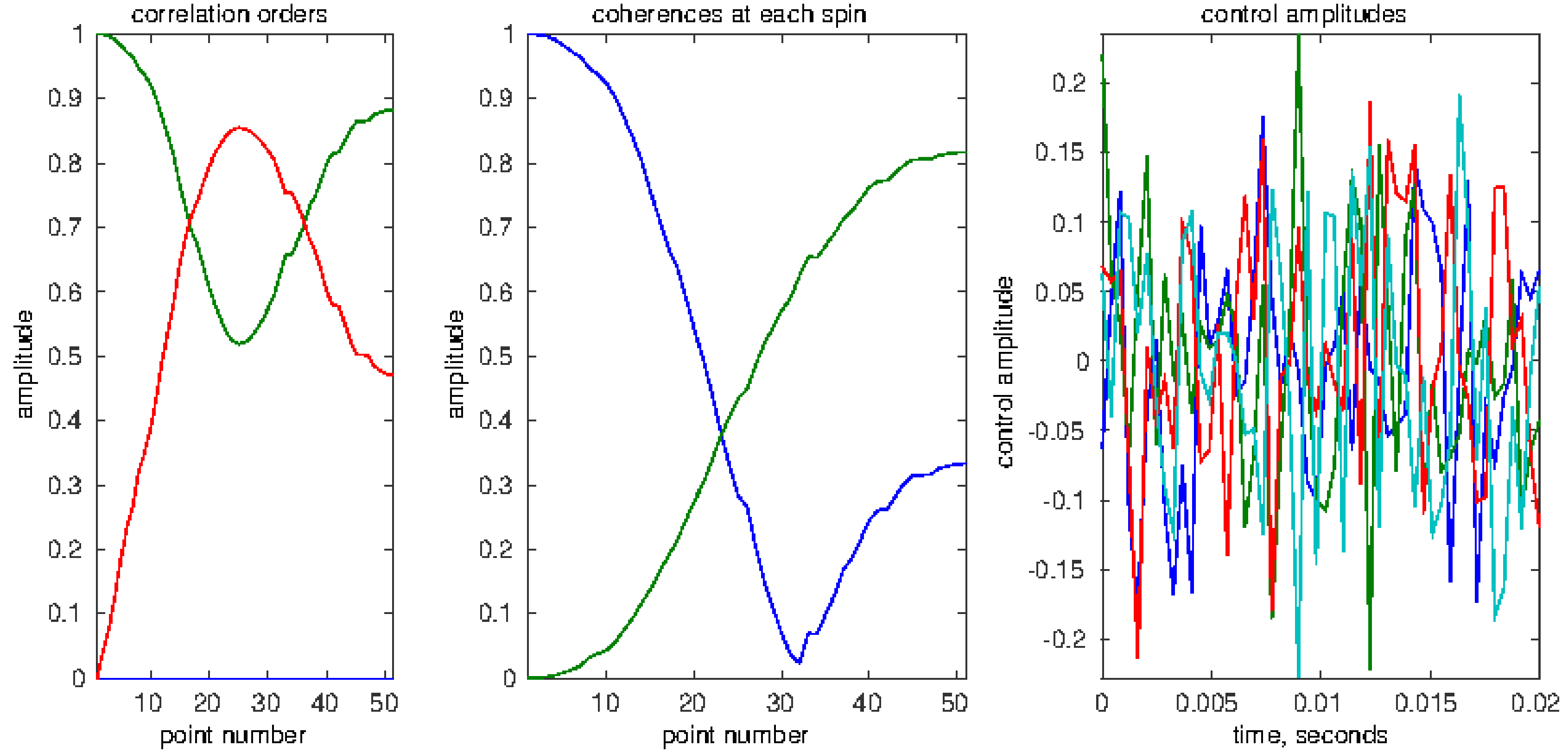
Ensemble Average

Crystallite orientation, Lebedev grid, Floquet space

- Take an ensemble average over the crystallite orientations.
- Average over a Lebedev grid.
- Initialise each of the ensemble to the same random guess.
- Calculate the Liouvillian in a Floquet basis at each of the grid weights.
- find the local cost and its gradient.
- Take the weighted average over the cost and gradient.
- Feed into a numerical optimisation through GRAPE.



Ensemble Average



UNIVERSITY OF Southampton

Noise Resilience of Optimal Control Pulses

HCF simulations

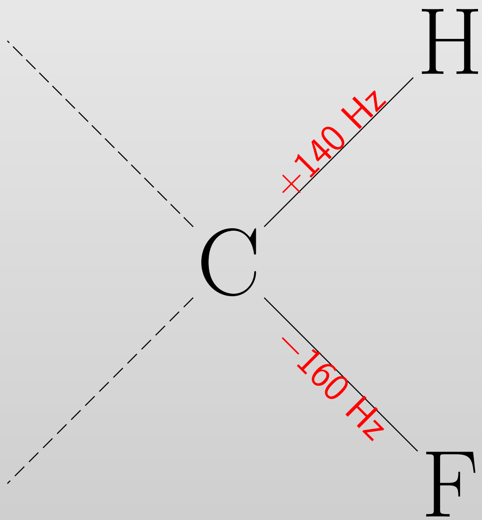


Figure: Interaction parameters of a molecular group used in HCF small system state transfer simulations, with a magnetic induction of 9.4 Tesla.

protein backbone simulations

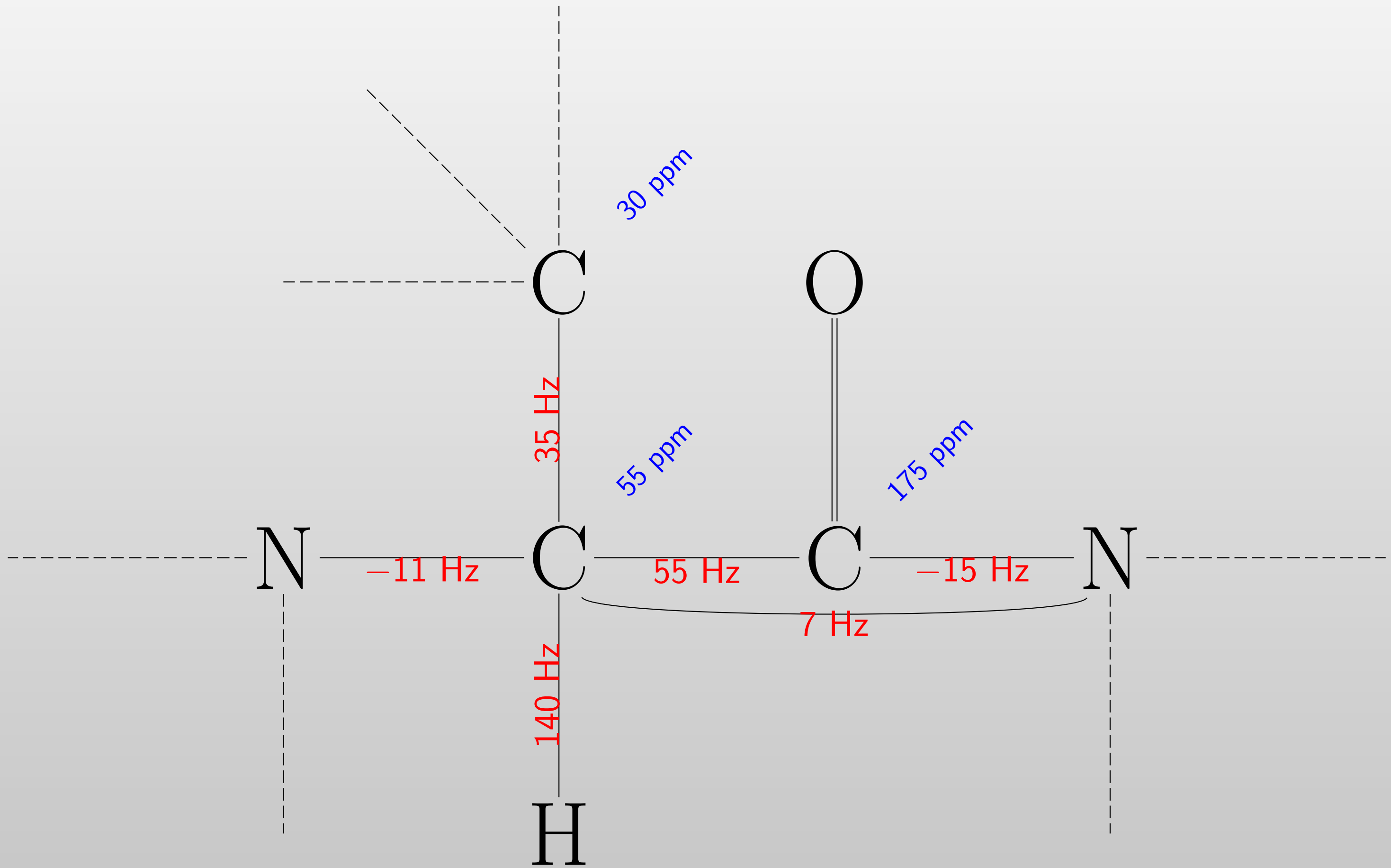


Figure: Interaction parameters of a typical protein backbone used in large system state transfer simulations, with a magnetic induction of 9.4 Tesla.

Ensemble Average

UNIVERSITY OF Southampton

- In real experimental apparatus, the control channels will include some level of noise as most electrical systems do
- Just as with the ssNMR optimisation, we average over an ensemble of systems.
- each system has the same initial random guess
- In simulating this noise, the solution is to create an ensemble of systems, each with their own instance of noise affecting the control channels.
- Noise is additive Gaussian, and is defined at the start of the simulation for each noise instance.
- We then optimise over the ensemble.

Ensemble Average

UNIVERSITY OF Southampton

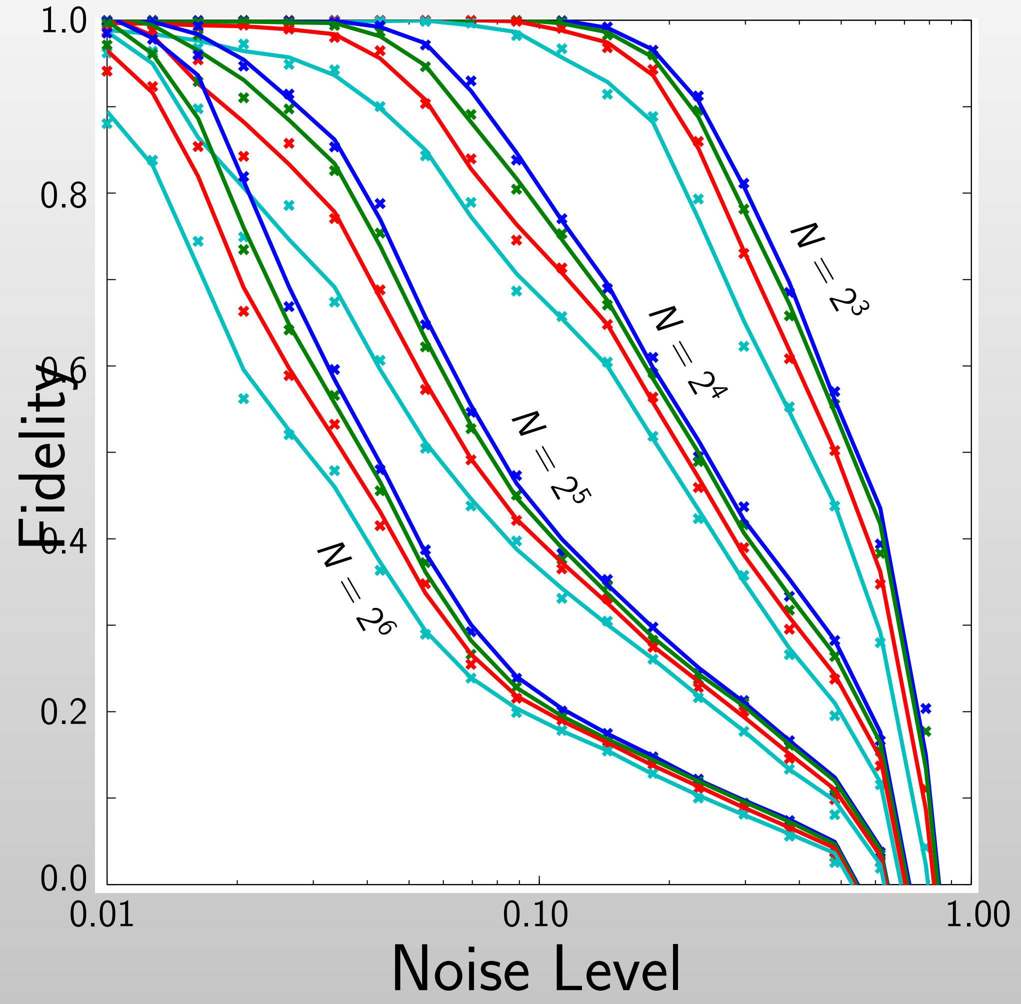


Figure: Convergence of the noise simulations to a number of noise instances

Ensemble Average

UNIVERSITY OF Southampton

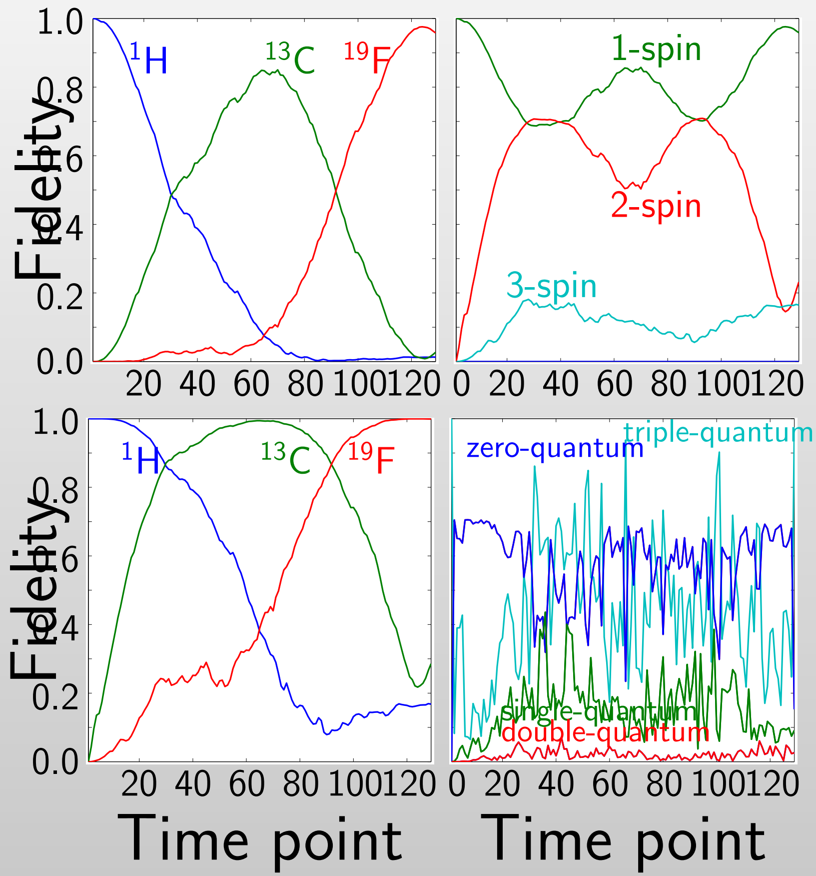


Figure: Population of states local at each spin of HCF (upper-left), and total population of each state (lower-left).

Total populations of correlations in the system (upper-right), and coherence orders (lower-right).

a direct product basis set the full state space, $\mathfrak{L}$, is a direct sum of correlation order subspaces:

$$\mathfrak{L} = \mathfrak{L}_0 \oplus \mathfrak{L}_1 \oplus \mathfrak{L}_2 \oplus \mathfrak{L}_3$$

where the population of a correlation order, p_k, is given by the projection onto its subspace

$$p_k = \left\| \hat{\hat{P}}_{\mathfrak{L}_k} \left| \hat{\rho} \right\rangle \right\|$$

Ensemble Average

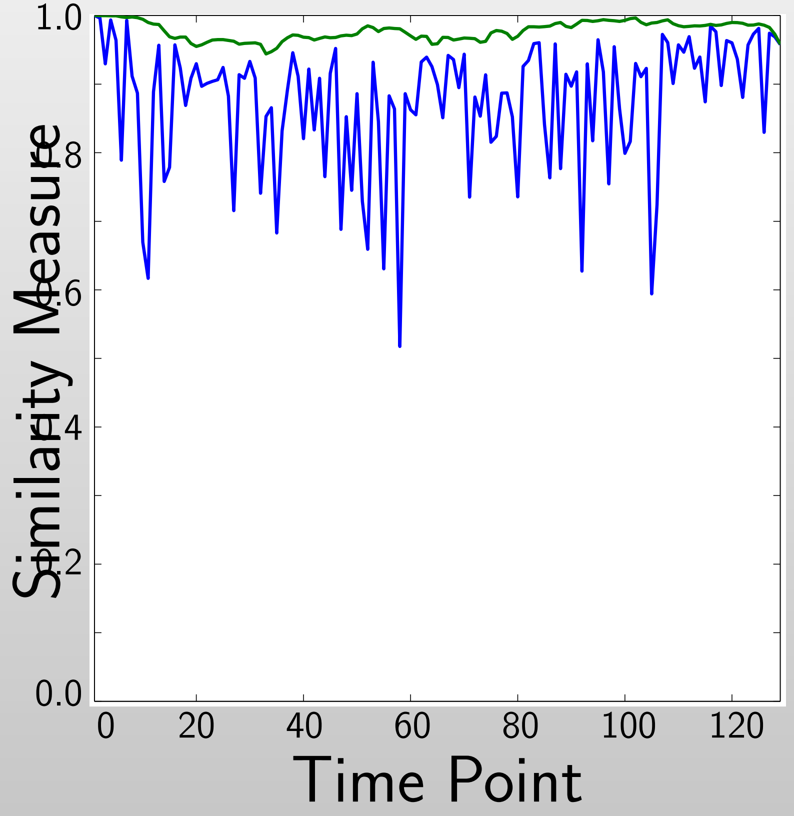


Figure: Measures of similarity between the trajectory of a noisy system compared to that of a similar one without noise

The Newton-Raphson Method

Objective function and Gradient Calculation

- The problem of second order methods, that require a gradient calculation, is reduced to:
  1. Propagate forwards from the source
  2. Propagate backwards from the target
  3. Compute the expectation of the derivative

$$\left\langle \sigma \middle| \hat{\mathcal{U}}_N \hat{\mathcal{U}}_{N-1} \cdots \frac{\partial}{\partial c_{n=t}^{(k)}} \hat{\mathcal{U}}_{n=t} \cdots \hat{\mathcal{U}}_2 \mathcal{U}_1 \middle| \psi_0 \right\rangle$$

Augmented Exponential
Gradient Calculation

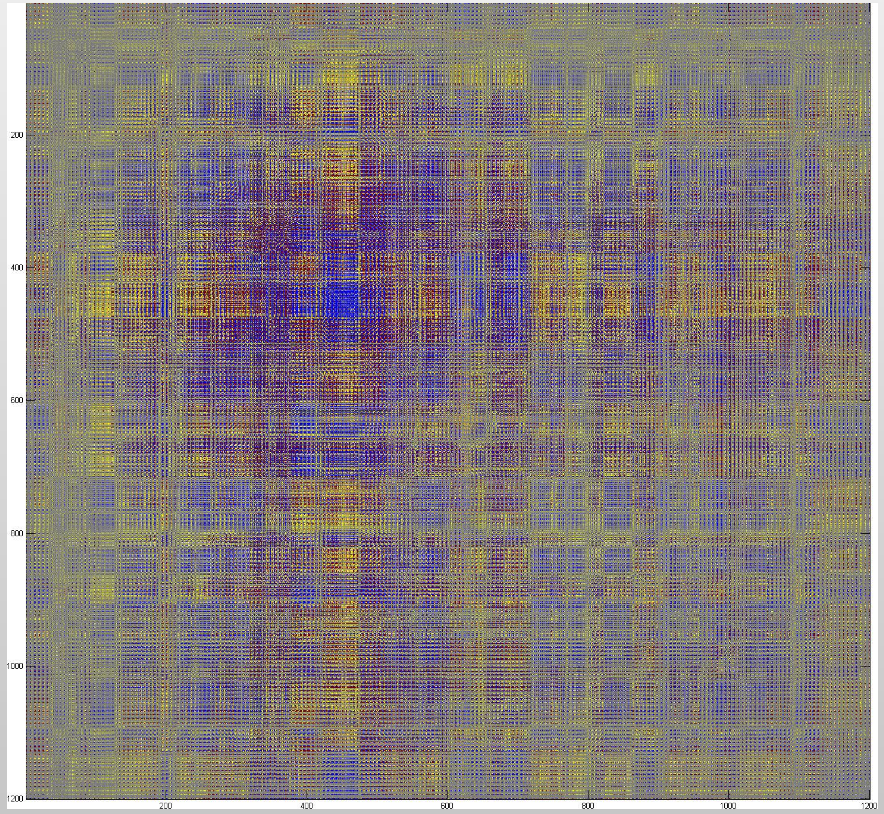
- We can use a the work of C. Van Loan (1970's) and Sophie Schirmer to find the derivative of the control pulse at a specific time point

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

Hessian Matrix
Avoiding Singularities

UNIVERSITY OF Southampton

- Symmetric
- Non-singular
- Size is time-points multiplied by control channels.
- Diagonally dominant.



Hessian Matrix
Calculating the Hessian elements

- Calculation of the Hessian elements requires the second order derivatives
- Scales to $O(n^2)$ calculations (compared with $O(n)$ for a gradient calculation).
- second order derivatives can be calculated with a 3×3 augmented exponential.

$$\exp\begin{pmatrix} -i\hat{\hat{L}}\Delta t & -i\hat{\hat{H}}_{n_1}^{(k_1)}\Delta t & 0 \\ 0 & -i\hat{\hat{L}}\Delta t & -i\hat{\hat{H}}_{n_2}^{(k_2)}\Delta t \\ 0 & 0 & -i\hat{\hat{L}}\Delta t \end{pmatrix} =$$

$$\begin{pmatrix} e^{-i\hat{\hat{L}}\Delta t} & \frac{\partial}{\partial c_{n_1}^{(k_1)}} e^{-i\hat{\hat{L}}\Delta t} & \frac{\partial^2}{\partial c_{n_1}^{(k_1)} \partial c_{n_2}^{(k_2)}} e^{-i\hat{\hat{L}}\Delta t} \\ 0 & e^{-i\hat{\hat{L}}\Delta t} & \frac{\partial}{\partial c_{n_2}^{(k_2)}} e^{-i\hat{\hat{L}}\Delta t} \\ 0 & 0 & e^{-i\hat{\hat{L}}\Delta t} \end{pmatrix}$$

Hessian Calculation
Expectation of second order derivatives

- We can reduce the computation to scale with $O(n)$ by realising that we can store propagators from gradient calculation, and then perform an extra set of 3×3 augmented exponentials for the (block) diagonal elements.

$$\left\langle \sigma \left| \hat{\mathcal{U}}_N \hat{\mathcal{U}}_{N-1} \cdots \frac{\partial^2}{\partial {c_{n=t}^{(k)}}^2} \hat{\mathcal{U}}_{n=t} \cdots \hat{\mathcal{U}}_2 \mathcal{U}_1 \right| \psi_0 \right\rangle$$

$$\left\langle \sigma \left| \hat{\mathcal{U}}_N \hat{\mathcal{U}}_{N-1} \cdots \frac{\partial}{\partial c_{n=t}^{(k)}} \hat{\mathcal{U}}_{n=t+1} \frac{\partial}{\partial c_{n=t}^{(k)}} \hat{\mathcal{U}}_{n=t} \cdots \hat{\mathcal{U}}_2 \mathcal{U}_1 \right| \psi_0 \right\rangle$$

$$\left\langle \sigma \left| \hat{\mathcal{U}}_N \hat{\mathcal{U}}_{N-1} \cdots \frac{\partial}{\partial c_{n=t}^{(k)}} \hat{\mathcal{U}}_{n=t_2} \cdots \frac{\partial}{\partial c_{n=t}^{(k)}} \hat{\mathcal{U}}_{n=t_1} \cdots \hat{\mathcal{U}}_2 \mathcal{U}_1 \right| \psi_0 \right\rangle$$

Hessian Regularisation
Avoiding Singularities

- A singular matrix is one that is not invertable.
- Common when we have negative eigenvalues.
- We can regularise the Hessian matrix:

TRM - We can add a multiple of the identity to the Hessian

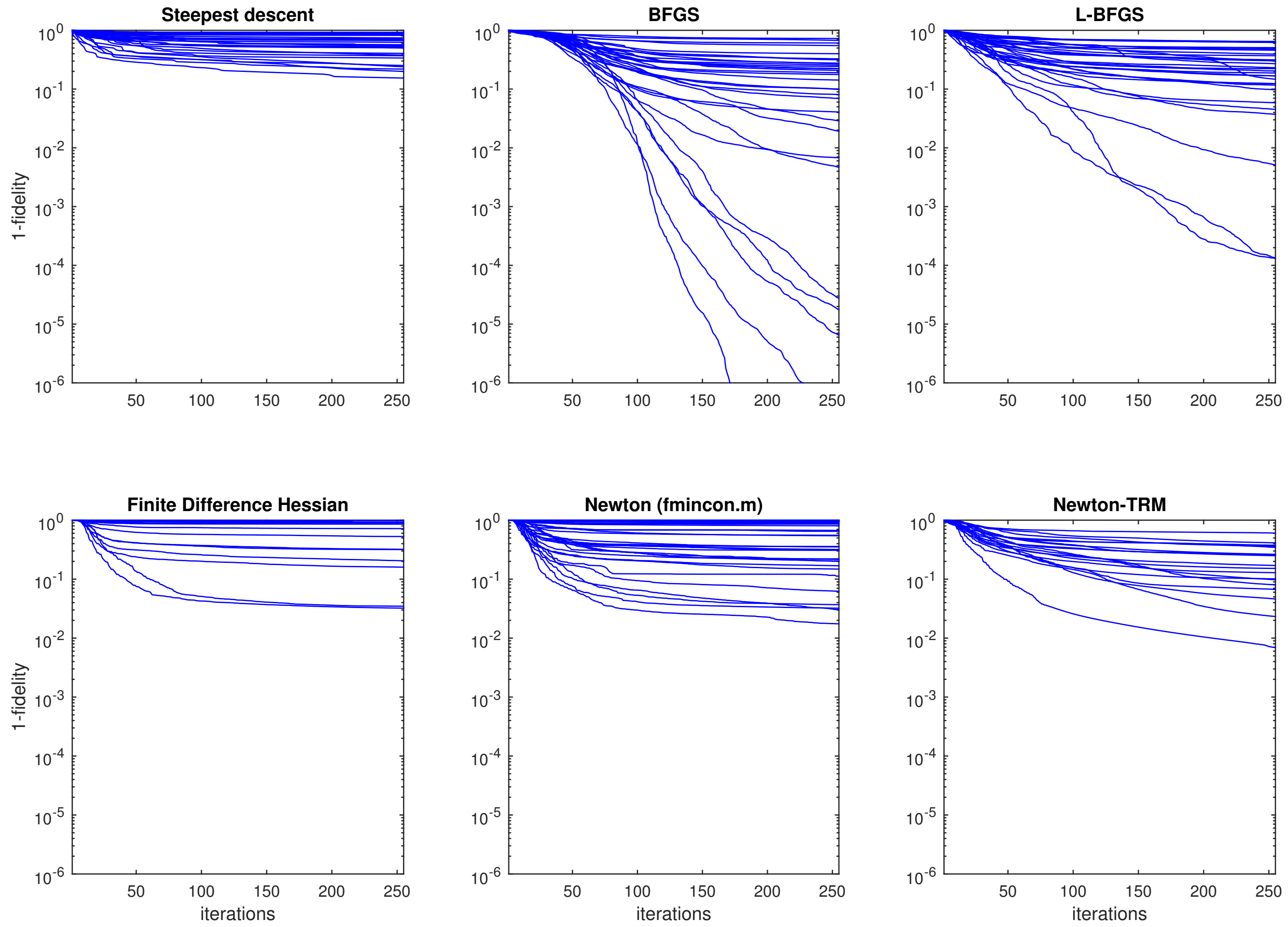
$$H_{reg} = H + \lambda I$$

where $\lambda > max\{0, -min[eig(H)]\}$

RFO - Prevents the algorithm taking large steps....

Results

Comparing the Newton method



Acknowledgement

Acknowledgement

Marina Caravetta University of Southampton

Sophie Schirmer Swansea University

Frank Langbein Cardiff University

Gavin Morley University of Warwick

Anton Tcholakov University of Warwick

Sean Wright University of Southampton

Spinach

Ilya Kuprov University of Southampton

Hannah Hogben University of Oxford

Luke Edwards University of Oxford

Matthew Krzystyniak Oxford e-Research Centre

Peter Hore University of Oxford

Gareth Charnock Oxford e-Research Centre

Dmitry Savostyanov University of Southampton

Sergey Dolgov Max-Planck Institute for Mathematics in the Sciences

Zenawi Welderufael University of Southampton