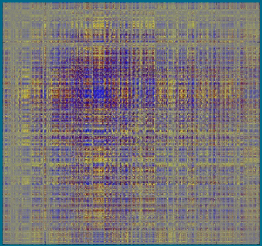
Chemistry

Quadratically Converging Control Sequence Optimisation Algorithms

Newton-GRAPE; Hessian Calculations & Regularisation



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

QUAINT@Swansea, Wednesday 1st July 2015

Two equations...

$$\frac{\partial^2 J}{\partial c_{n_2}^{(k)} \partial c_{n_1}^{(k)}} = \langle \sigma | \hat{\hat{\mathcal{P}}}_N \hat{\hat{\mathcal{P}}}_{N-1} \cdots \frac{\partial}{\partial c_{n_2}^{(k)}} \hat{\hat{\mathcal{P}}}_{n_2} \cdots \frac{\partial}{\partial c_{n_1}^{(k)}} \hat{\hat{\mathcal{P}}}_{n_1} \cdots \hat{\hat{\mathcal{P}}}_2 \hat{\hat{\mathcal{P}}}_1 | \psi_0 \rangle \qquad (1)$$

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

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

Abstract

We report an efficient algorithm for computing the full Hessian of the optimal control objective function in a large class of practically relevant cases, and use it to analyze the convergence properties of Newton-Raphson optimization methods using the full Hessian matrix.

Newton-Raphson method

- Taylor series approximated to second order[1].
  - If f is continuously differentiable

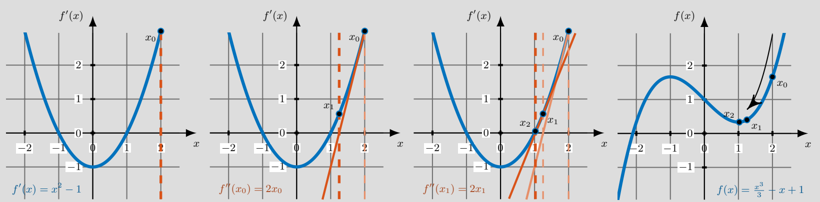
$$f(x+p) = f(x) + \nabla f(x+tp)^T p$$

  - If f is twice continuously differentiable

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+\alpha p) p$$

- 1st order necessary condition: $\nabla f(x^*) = 0$
- 2nd order necessary condition: $\nabla^2 f(x^*)$ is positive semidefinite

Finding a minimum to the function using tangents of the gradient



[1] B. Taylor. Inny, 1717, J. Nocedal and S. J. Wright. 1999.

The Hessian Matrix

The Newton step: $\quad p_k^N = -\mathbf{H}_k^{-1}\nabla f_k$

- $\nabla^2 f_k = \mathbf{H}_k$ is the Hessian matrix, one of second order partial derivatives[2]:

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

- The steepest descent method results from the same equation when we set $\mathbf{H}$ to the identity matrix.
- Quasi-Newton methods initialise $\mathbf{H}$ to the identity matrix, then to approximate it from an update formula using a gradient history.
- The Hessian proper must be positive definite (an quite well conditioned) to make an inverse; an indefinte Hessian results in non-descent search directions.

[2] L. O. Hesse. Chelsea, 1972.

GRAPE; gradient and Hessian

GRAPE gradient

- Split Liouvillian to controllable and uncontrollable parts[3], $\hat{\hat{L}}(t) = \hat{\hat{H}}_0 + \sum_k c^{(k)}(t)\hat{\hat{H}}_k$
- Maximise the fidelity measure, $J = \Re e \langle \hat{\sigma} | \exp_{(0)} \left[-i \int_0^T \hat{\hat{L}}(t) dt \right] | \hat{\rho}(0) \rangle$
- Optimality conditions, $\frac{\partial J}{\partial c_k(t)} = 0$ at a minimum, and the Hessian matrix should be positive definite
- Gradient found from forward and backward propagation

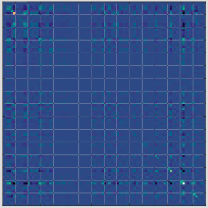
$$\frac{\partial J}{\partial c_n^{(k)}} = \langle \sigma | \hat{\hat{\mathcal{P}}}_N \hat{\hat{\mathcal{P}}}_{N-1} \cdots \frac{\partial}{\partial c_n^{(k)}} \hat{\hat{\mathcal{P}}}_n \cdots \hat{\hat{\mathcal{P}}}_2 \hat{\hat{\mathcal{P}}}_1 | \psi_0 \rangle$$

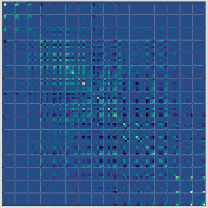
- Propagator over a time slice:

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

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

GRAPE Hessian





- (block) diagonal elements

$$\frac{\partial^2 J}{\partial {c_n^{(k)}}^2} = \langle\sigma|\,\hat{\hat{\mathcal{P}}}_N\,\hat{\hat{\mathcal{P}}}_{N-1}\cdots\frac{\partial^2}{\partial {c_n^{(k)}}^2}\hat{\hat{\mathcal{P}}}_n\cdots\hat{\hat{\mathcal{P}}}_2\,\hat{\hat{\mathcal{P}}}_1\,|\psi_0\rangle$$

- non-diagonal elements

$$\frac{\partial^2 J}{\partial c_{n_2}^{(k)}\partial c_{n_1}^{(k)}} = \langle\sigma|\,\hat{\hat{\mathcal{P}}}_N\,\hat{\hat{\mathcal{P}}}_{N-1}\cdots\frac{\partial}{\partial c_{n_2}^{(k)}}\hat{\hat{\mathcal{P}}}_{n_2}\cdots\frac{\partial}{\partial c_{n_1}^{(k)}}\hat{\hat{\mathcal{P}}}_{n_1}\cdots\hat{\hat{\mathcal{P}}}_2\,\hat{\hat{\mathcal{P}}}_1\,|\psi_0\rangle$$

- All propagators of the non-diagonal blocks have been calculated within a gradient calculation, and can be reused. Only need to find the diagonal blocks.

Van Loan's Augmented Exponentials

- Among the many complicated functions encountered in magnetic resonance simulation context, chained exponential integrals involving square matrices A_k and B_k occur particularly often:

$$\int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-2}} dt_{n-1} \left\{ \mathrm{e}^{A_1(t-t_1)}\, B_1\, \mathrm{e}^{A_2(t_1-t_2)}\, B_2 \cdots \mathrm{e}^{A_1(t-t_1)}\, B_{n-1}\, \mathrm{e}^{A_n t_{n-1}} \right\}$$

- A method for computing some of the integrals of the general type shown in Equation of this type was proposed by Van Loan in 1978[4] (pointed out by Sophie Schirmer[5])
- Using this augmented exponential technique, we can write an upper-triangular block matrix exponential as

$$\exp\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{A} \end{pmatrix} = \begin{pmatrix} e^{\mathbf{A}t} & \int_0^t e^{\mathbf{A}(t-s)}\mathbf{B}e^{\mathbf{A}s}ds \\ \mathbf{0} & e^{\mathbf{A}t} \end{pmatrix} = \begin{pmatrix} e^{\mathbf{A}} & \int_0^1 e^{\mathbf{A}(1-s)}\mathbf{B}e^{\mathbf{A}s}ds \\ \mathbf{0} & e^{\mathbf{A}} \end{pmatrix}$$

[4] C. F. Van Loan. In: *Automatic Control, IEEE Transactions on* 23.3 (1978), pp. 395–404.
[5] F. F. Floether, P. de Fouquieres and S. G. Schirmer. In: *New Journal of Physics* 14.7 (2012), p. 073023.

UNIVERSITY OF Southampton

- 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

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

Efficient Hessian Calculation

Augmented Exponentials

- second order derivatives can be calculated with a 3×3 augmented exponential[6]
- set

$$\int_0^1 \int_0^s e^{\mathbf{A}(1-s)} \mathbf{B}_{n_1} e^{\mathbf{A}(s-r)} \mathbf{B}_{n_2} e^{\mathbf{A}r} dr ds = D^2_{c_{n_1} c_{n_2}}(t) \exp\left(-i\hat{\hat{L}}\Delta t\right) \Rightarrow \mathbf{B}_n = -i\hat{\hat{H}}_n^{(k)}\Delta t$$

- Giving the efficient Hessian element calculation

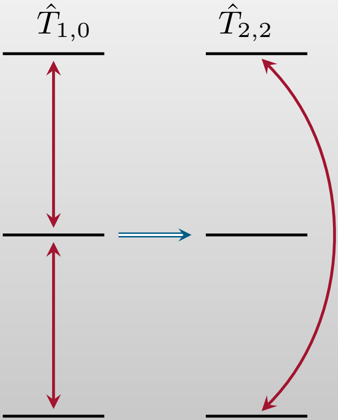
$$\exp\begin{pmatrix} -i\hat{\hat{L}}\Delta t & -i\hat{\hat{H}}_{n_1}^{(k_1)}\Delta t & 0 \\ 0 & -i\hat{\hat{L}}\Delta t & -i\hat{\hat{H}}_{n_2}^{(k_2)}\Delta t \\ 0 & 0 & -i\hat{\hat{L}}\Delta t \end{pmatrix} =$$

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

---

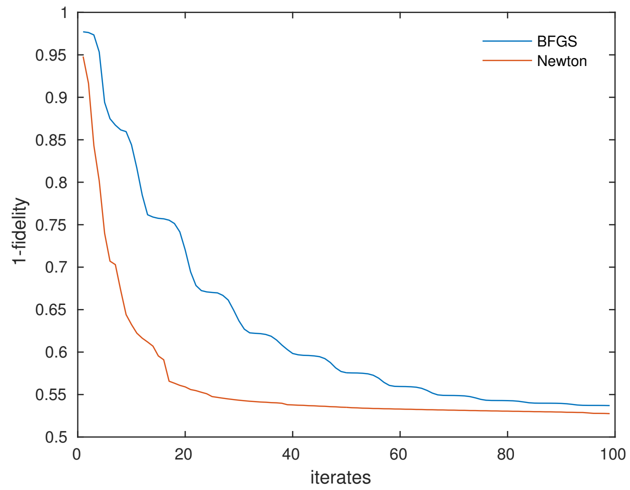
[6] T. F. Havel, I Najfeld and J. X. Yang. In: *Proceedings of the National Academy of Sciences* 91.17 (1994), pp. 7962–7966, I. Najfeld and T. Havel. In: *Advances in Applied Mathematics* 16.3 (1995), pp. 321 –375.

- 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 40μs



Overtone excitation

Comparison of BFGS and Newton-Raphson



Regularisation

- BFGS (using the DFP formula) is guaranteed to produce a positive definite Hessian update
- The Newton-Raphson method does not:

$$p_k^N = -\mathbf{H}_k^{-1} \nabla f_k$$

- Properties of the Hessian matrix:
  1. Must be symmetric: $\frac{\partial^2}{\partial c^{(i)} \partial c^{(j)}} = \frac{\partial^2}{\partial c^{(j)} \partial c^{(i)}}$; not if control operators commute
  2. Must be sufficiently positive definite; non-singular; invertible.
  3. The Hessian is diagonally dominant.

- Common when we have negative eigenvalues, regularise the Hessian to be positive definite[7].
- Check eigenvalues performing an eigendecomposition of the Hessian matrix:

$$\mathbf{H} = Q\Lambda Q^{-1}$$

- Add the smallest negative eigenvalue to all eigenvalues, then reform the Hessian with initial eigenvectors:

$$\lambda_{\text{min}} = \max\left[0, -\min(\Lambda)\right]$$
$$\mathbf{H}_{\text{reg}} = Q(\Lambda + \lambda_{\text{min}}\hat{I})Q^{-1}$$

TRM Introduce a constant δ; region of a radius we trust to give a sufficiently positive definite Hessian.

$$\mathbf{H}_{\text{reg}} = Q(\Lambda + \delta\lambda_{\text{min}}\hat{I})Q^{-1}$$

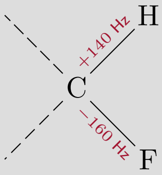
However, if δ is too small, the Hessian will become ill-conditioned.

RFO The method proceeds to construct an augmented Hessian matrix

$$\mathbf{H}_{\text{aug}} = \begin{bmatrix} \delta^2\mathbf{H} & \delta\vec{g} \\ \delta\vec{g} & \mathbf{0} \end{bmatrix} = Q\Lambda Q^{-1}$$

[7] X. P. Resina. PhD thesis. Universitat Autónoma de Barcelona, 2004.

- Controls $:= \left\{ \hat{\hat{L}}_x^{(H)}, \hat{\hat{L}}_y^{(H)}, \hat{\hat{L}}_x^{(C)}, \hat{\hat{L}}_y^{(C)}, \hat{\hat{L}}_x^{(F)}, \hat{\hat{L}}_y^{(F)} \right\}$
- valid vs. invalid parametrisation of Lie groups.

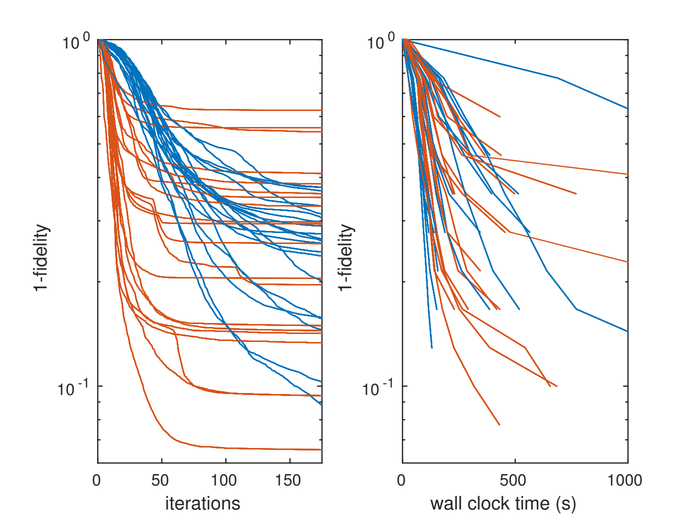


Interaction parameters of a hydroflourocarbon molecular group used in state transfer simulations, with a magnetic induction of 9.4 Tesla

State transfer

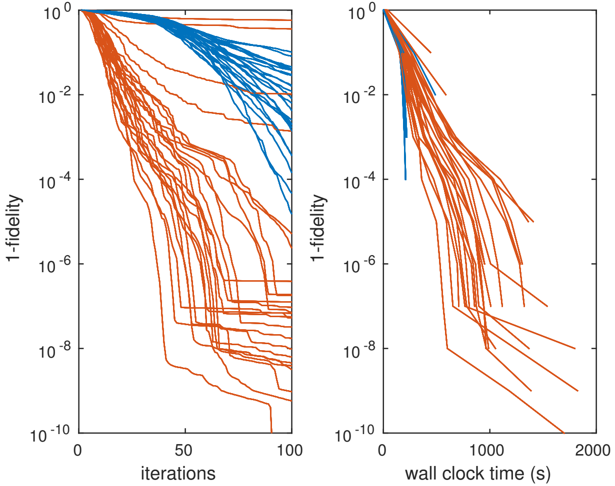
Comparison of BFGS and Newton-Raphson I

UNIVERSITY OF Southampton



State transfer

Comparison of BFGS and Newton-Raphson II



Closing Remarks

- Hessian calculation that scales with $O(n)$ computations.
- Efficient directional derivative calculation with augmented exponentials
- better regularisation and conditioning
- infeasible start points?
- Different line search methods?
- forced symmetry of a Hessian