Chemistry

Optimal Control of Spin Systems:

A Modified Newton-Raphson Method (David Goodwin , Ilya Kuprov)

UNIVERSITY OF Southampton

Optimal Control - Gradient Ascent Pulse Engineering (GRAPE)

The evolution of a quantum system can be characterised by a Liouville equation with a vector representation of the density operator, $\frac{\partial}{\partial t}|\hat{\rho}(t)\rangle = -i\hat{\hat{\mathcal{L}}}(t)\,|\hat{\rho}(t)\rangle$, having the general solution,

$$|\hat{\rho}(t)\rangle = \exp_{(0)}\left(-i\int_0^t \hat{\hat{\mathcal{L}}}(t)\,\mathrm{d}t\right)|\hat{\rho}(t)\rangle$$

A practical Magnetic Resonance system can be split into two parts; that is beyond experimental control, $\hat{\hat{\mathcal{L}}}_0$, and a set of controllable electromagnetic pulses $\left\{\hat{\hat{\mathcal{L}}}_k\right\}$:

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

The problem is simplified if the control sequences $\left\{c^{(k)}(t)\right\}$ are assumed to be piecewise constant. For a piecewise constant Hamiltonian, we sequentially multiply each of the, discrete in time, propagators into the initial conditions:

$$\hat{\hat{\mathcal{P}}}_n = \exp\left[-i\left(\hat{\hat{\mathcal{L}}}_0 + \sum_k \underbrace{c^{(k)}(t_n)}_{\substack{\| \\ c_n^{(k)}}}\hat{\hat{\mathcal{L}}}_k\right)\Delta t\right]$$

The expression for fidelity, the overlap between the current state of the system ρ_0 and the target state σ, is

$$J = \mathrm{Re}\,\langle\sigma|\,\hat{\hat{\mathcal{P}}}_N\hat{\hat{\mathcal{P}}}_{N-1}\ldots\hat{\hat{\mathcal{P}}}_2\hat{\hat{\mathcal{P}}}_1\,|\rho_0\rangle$$

Since $\left\{c_n^{(k)}\right\}$ are vectors of finite dimension, we can use the standard non-linear numerical optimisation to find the maximum of J in their space (it is useful to note that the maximum of J is the same as the minimum of $1-J$).

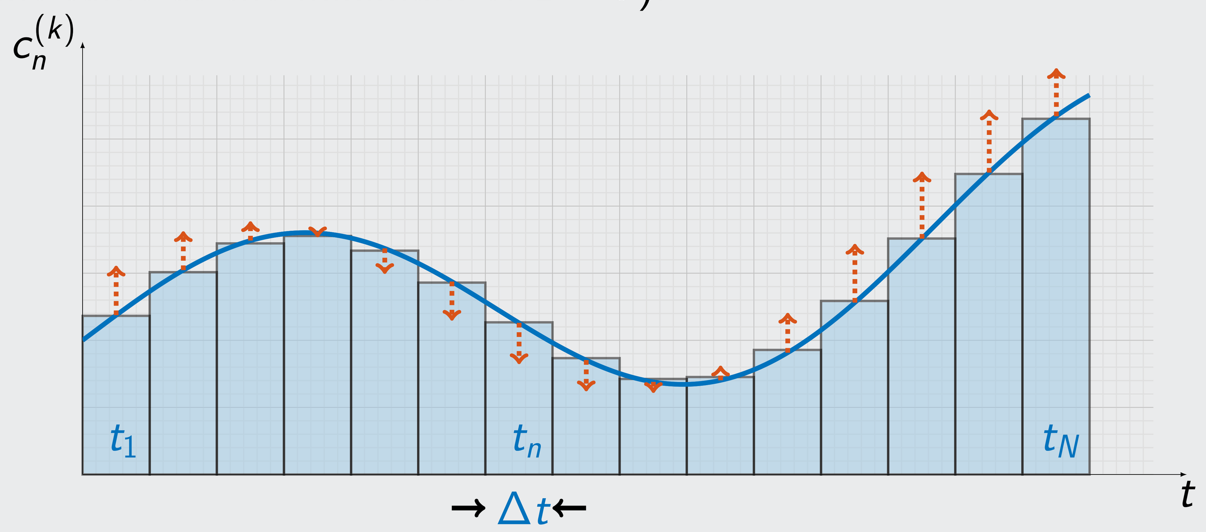
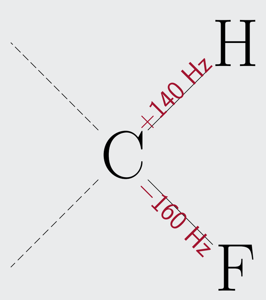


Figure: Piecewise constant approximation in a GRAPE simulation

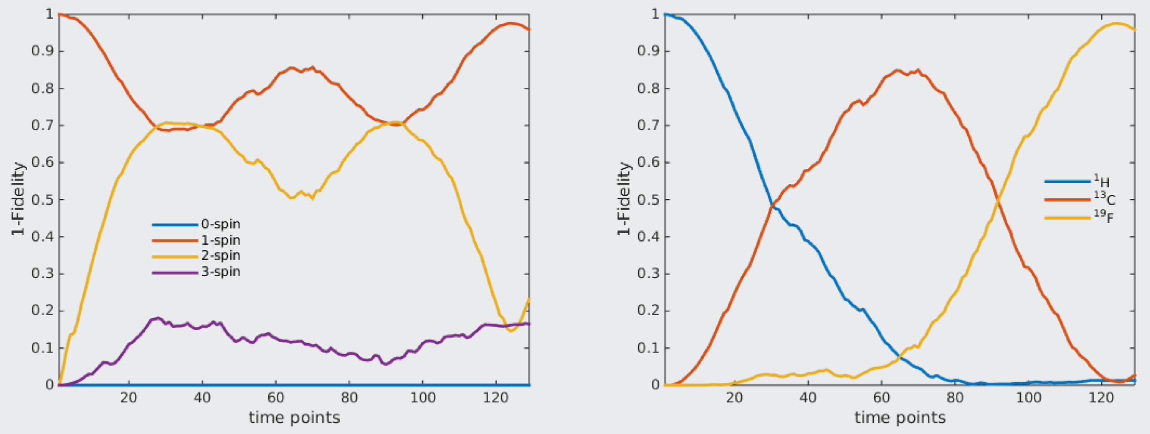
Simulation System

We prepare the *Spinach* software to find a pulse set to transfer magnetisation from ^{1}H atom to ^{19}F atom with the following molecule:

Interaction parameters of a molecular group used in state transfer simulations on a system characterising the fragment of a fluorohydrocarbon molecule (magnetic induction = 9.4 Tesla). In this case the set of control channels operators are $\left\{\hat{L}_x^{(H)}, \hat{L}_y^{(H)}, \hat{L}_x^{(C)}, \hat{L}_y^{(C)}, \hat{L}_x^{(F)}, \hat{L}_y^{(F)}\right\}$.



Visualise the set of optimal pulses by the population of their correlation subspaces and population of coherence local at each spin:



Regularise the Hessian, so it is non-singular; take the eigendecomposition the Hessian matrix and add a multiple of the identity:

$$[\nabla^2 J] = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1},\ \ \sigma = \max(0, \delta - \min(\mathbf{\Lambda}_{ii}))$$
$$[\nabla^2 J]_{\mathrm{reg}} = \mathbf{Q}(\mathbf{\Lambda} + \sigma\mathbf{1})\mathbf{Q}^{-1}$$

where δ is chosen to give a sufficiently positive definite Hessian. To condition the Hessian, we proceed as before, except the shifting is applied to an augmented Hessian:

$$[\nabla^2 J]^{\mathrm{aug}} = \begin{pmatrix}\alpha^2\nabla^2 J & \alpha\nabla J\\ \alpha\nabla J^T & 0\end{pmatrix} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1},\ \ \sigma = \max(0, -\min(\mathbf{\Lambda}_{ii}))$$
$$[\nabla^2 J]^{\mathrm{aug}}_{\mathrm{reg}} = \frac{1}{\alpha^2}\mathbf{Q}(\mathbf{\Lambda} + \sigma\mathbf{1})\mathbf{Q}^{-1}$$

where the scaling constant α is reduced until the condition number becomes acceptable, for example:

$$\alpha_{r+1} = \phi\alpha_r \text{ while } \frac{\min(\mathbf{\Lambda}_{ii})}{\max(\mathbf{\Lambda}_{ii})} > \frac{1}{\sqrt{\varepsilon}}$$

where ε is machine precision and $\alpha_0 = 1$. The factor $0 < \phi < 1$ is used to iteratively decrease the condition number of the Hessian

Simulation Results

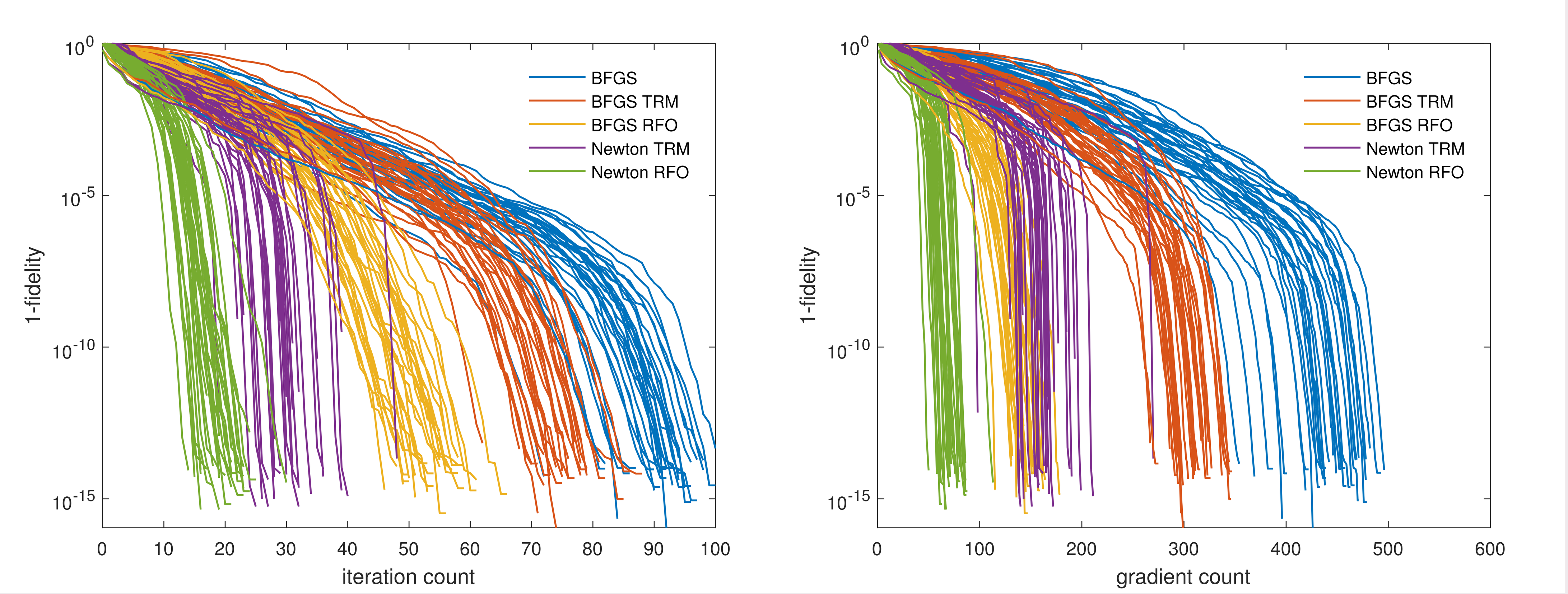
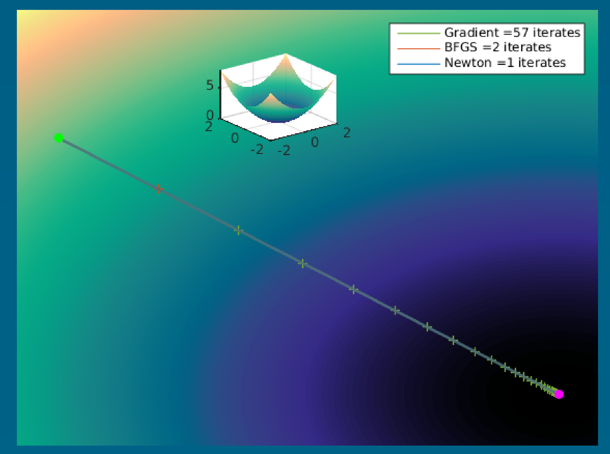
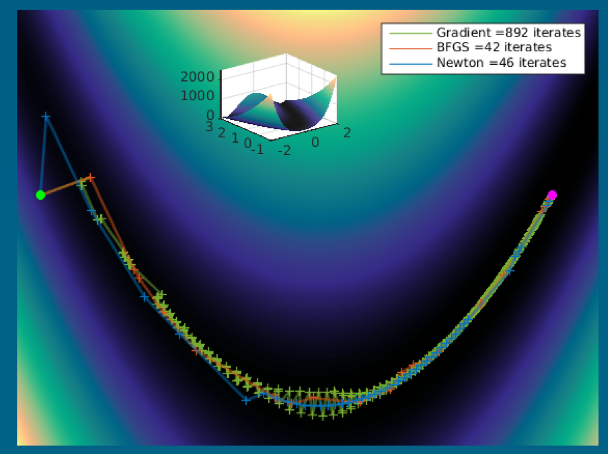


Figure: Newton-Raphson methods and quasi-Newton methods with regularisation/conditioning. GRAPE optimal control for state transfer on a fragment of a fluorohydrocarbon molecule. Simulation was run multiple times from different initial guesses of control pulses. *Left*, Iterate count. *Right*, Calculation count





Hessians are normally so expensive that a significant body of work exists on the subject of avoiding their calculation and recovering second derivative information in an approximate way from the gradient history. The recent BFGS-GRAPE algorithm is an example of such approach. The fact that the Hessian is cheap suggests that Newton-Raphson type algorithms with the control sequence update rule at step s

$$\mathbf{c}_{s+1} = \mathbf{c}_s - \left[\nabla^2 J_s\right]^{-1}\nabla J_s$$

formulated in terms of the gradient ∇J and the Hessian $\nabla^2 J$ of the fidelity functional $J(\mathbf{c})$ with a suitable line search procedure are a natural next step. Newton-Raphson and quasi-Newton methods rely on the necessary conditions for Taylor's theorem and use a local quadratic approximation:

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

The first order necessary condition requires any minimiser $\tilde{\mathbf{c}}$ to be a stationary point

$$\nabla J(\tilde{\mathbf{c}}) = 0$$

Imposing this condition gives the control sequence update rule. The second order necessary condition is

$$\langle\mathbf{c}|\,\nabla^2 J(\tilde{\mathbf{c}})\,|\mathbf{c}\rangle > \varepsilon\,\langle\mathbf{c}|\mathbf{c}\rangle\ \ \forall\mathbf{c}\in\mathbb{R}^{KN}$$

where ε is a positive scalar, i.e. the Hessian $\nabla^2 J(\tilde{\mathbf{c}})$ must be positive definite at $\tilde{\mathbf{c}}$. This is evident above, in which a negative Hessian eigenvalue would result in a step being performed up, rather than down, the corresponding gradient direction. A significant problem is that far away from a minimiser, the Hessian is not actually expected to be positive definite. Small Hessian eigenvalues are also problematic because they result in overly long steps that can be detrimental because most fidelity functionals are not actually quadratic.

Efficient Propagator Derivative Calculations

The numerical optimisation method simulated in the results section above require a gradient calculation. This 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

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

The total cost of the gradient of J is therefore one forward simulation, one backward simulation and (n steps) $\times$ (k controls) derivatives of matrix exponentials with respect to scalar parameters. The expectation of first order derivatives is

$$\left\langle\frac{\partial J}{\partial c_{n=t}^{(k)}}\right\rangle = \langle\sigma|\,\hat{\hat{\mathcal{P}}}_N\hat{\hat{\mathcal{P}}}_{N-1}\cdots\frac{\partial}{\partial c_{n=t}^{(k)}}\hat{\hat{\mathcal{P}}}_{n=t}\cdots\hat{\hat{\mathcal{P}}}_2\hat{\hat{\mathcal{P}}}_1\,|\rho_0\rangle$$

Efficient calculation of the expectation of first order derivatives can be made utilising the work of C.Van Loan; using an augmented exponential in the following form

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

extracting the derivative from the upper right block. In practice, the exponential is calculated using a two-point finite difference stencil with Krylov propagation. The Newton-Raphson method is a second order method, additionally requiring the explicit calculation of the Hessian matrix. This requires the expectation of the second order derivatives:

$$\left\langle\frac{\partial^2 J}{\partial c_n^2}\right\rangle = \langle\sigma|\,\hat{\hat{\mathcal{P}}}_N\cdots\hat{\hat{\mathcal{P}}}_{n+1}\frac{\partial^2\hat{\hat{\mathcal{P}}}_n}{\partial c_n^2}\hat{\hat{\mathcal{P}}}_{n-1}\cdots\hat{\hat{\mathcal{P}}}_1\,|\rho_0\rangle$$

$$\left\langle\frac{\partial^2 J}{\partial c_m\partial c_n}\right\rangle = \langle\sigma|\,\hat{\hat{\mathcal{P}}}_N\cdots\hat{\hat{\mathcal{P}}}_{n+1}\frac{\partial\hat{\hat{\mathcal{P}}}_n}{\partial c_n}\hat{\hat{\mathcal{P}}}_{n-1}\cdots\hat{\hat{\mathcal{P}}}_{m+1}\frac{\partial\hat{\hat{\mathcal{P}}}_m}{\partial c_m}\hat{\hat{\mathcal{P}}}_{m-1}\cdots\hat{\hat{\mathcal{P}}}_1\,|\rho_0\rangle$$

- Computation to scale with $O(n\times k)$ by storing propagators from gradient calculation.
- Problem now reduces to finding $n\times k$ second-order derivatives on the block diagonal of the Hessian with a 3×3 augmented exponential:

$$\exp\begin{pmatrix}-i\hat{\hat{\mathcal{L}}}\Delta t & -i\hat{\hat{\mathcal{L}}}_n^{(k)}\Delta t & \mathbf{0}\\ \mathbf{0} & -i\hat{\hat{\mathcal{L}}}\Delta t & -i\hat{\hat{\mathcal{L}}}_m^{(k)}\Delta t\\ \mathbf{0} & \mathbf{0} & -i\hat{\hat{\mathcal{L}}}\Delta t\end{pmatrix} = \begin{pmatrix}e^{-i\hat{\hat{\mathcal{L}}}\Delta t} & \frac{\partial}{\partial c_n^{(k)}}e^{-i\hat{\hat{\mathcal{L}}}\Delta t} & \frac{1}{2}\frac{\partial^2}{\partial c_n^{(k)}\partial c_m^{(k)}}e^{-i\hat{\hat{\mathcal{L}}}\Delta t}\\ \mathbf{0} & e^{-i\hat{\hat{\mathcal{L}}}\Delta t} & \frac{\partial}{\partial c_m^{(k)}}e^{-i\hat{\hat{\mathcal{L}}}\Delta t}\\ \mathbf{0} & \mathbf{0} & e^{-i\hat{\hat{\mathcal{L}}}\Delta t}\end{pmatrix}$$

Selected Bibliography

- *Goodwin, Kuprov; J. Chem. Phys. 143(8):084113, (2015).*
- *Floether, de Fouquieres, Schirmer,; New J. Phys. 14, 073023, (2012).*
- *de Fouquieres, Schirmer, Glaser, Kuprov; J. Mag. Res. 212, 412-417, (2011).*
- *Hogben, Krzystyniak, Charnock, Hore, Kuprov; J. Mag. Res. 208, 179-194, (2011).*
- *Khaneja, Reiss, Kehlet, Schulte-Herbruggen, Glaser; J. Mag. Res. 172, 296-305, (2005).*
- *Nocedal, Wright; Numerical optimization, (1999).*
- *Najfeld, Havel; Adv. App. Math. 16, 321-375, (1995).*
- *Van Loan; IEEE Trans. 23(3), 395–404, (1978).*