On the design of regularized Newton-type methods for all-at-once reconstructions from HARDI datasets

Christoph Rügge

Institute for Numerical and Applied Mathematics
University of Göttingen

24 November 2011

Outline

Introduction

Different approaches to reconstruct information on water diffusion from DW images:

- Series of several independent reconstruction steps, like:
  1. k-space Fourier inversion
  2. Voxel-wise q-space reconstruction
  3. Smoothing
- "All-at-once" approach: Formulate the entire reconstruction process as a non-linar inverse problem.
  - Advantages: Handling incomplete data, no propagation of errors between individual steps

Introduction

Different approaches to reconstruct information on water diffusion from DW images:

- Series of several independent reconstruction steps, like:
  1. k-space Fourier inversion
  2. Voxel-wise q-space reconstruction
  3. Smoothing
- "All-at-once" approach: Formulate the entire reconstruction process as a non-linar inverse problem.
  - Advantages: Handling incomplete data, no propagation of errors between individual steps
- This talk focusses on combining voxel-wise reconstruction and smoothing, i.e. spatial regularization.
- Fourier inversion may be included based on

M. Uecker, A. Karaus, J. Frahm, Inverse Reconstruction Method for Segmented Multi-shot Diffusion-weighted MRI With Multiple Coils, Magnetic Resonance in Medicine, 2009

Introduction

- Given noisy DW images $S(x, q)$ for $x \in \mathbb{R}^3$ and $q \in \mathcal{Q} \subset \cup_{i=1}^{N} q_{0,i} S^2$, solve

$$F(\theta) = S,$$

where θ contains information on the directional structure of water diffusion.

- F is the model operator. Possible choices include Mono-exponential decay, PAS-MRI, Spherical Deconvolution ...

I. Aganj, C. Lenglet, G. Sapiro, ODF reconstruction in Q-ball imaging with solid angle consideration, Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging, 2009

J. Tournier et al., Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution, NeuroImage 23, 2004

K. Jansons, D. Alexander, Persistent Angular Structure: New Insights from Diffusion MRI Data, Information Processing in Medical Imaging, 2003

The Model

- Spherical Deconvolution:

$$S(x,q) = \int_{S^2} k(q \cdot u)\theta(u)\, du$$

- Multi-Tensor:

$$S(x,q) = \sum_{i=1}^{N(x)} \lambda_i(x) \exp\left(-q^T D_i(x) q\right)$$

The Model

- For $x \in \mathbb{R}^3$ and $u \in S^2$, define $D(x,u)$ as a rotationally symmetric tensor with main eigenvector u:

$$D(x,u) = \alpha(x,u)\mathbb{1} + \beta(x,u)uu^T$$

- Parametrise data as a mixture of Gaussians, one for each direction:

$$S(x,q) = \int_{S^2} \lambda(x,u) \exp\left(-q^T D(x,u) q\right)\, du$$

- Forward-Operator:

$$F(\lambda,\alpha,\beta)(x,q) := \int_{S^2} \lambda(x,u) \exp\left(-\alpha(x,u)\,\|q\|^2 - \beta(x,u)(q\cdot u)^2\right)\, du$$

- Reconstruction of λ and α requires measurements on at least 2 q-shells.

Regularization

Iteratively Regularized Gauss-Newton Method:

$$\theta_{n+1} = \underset{\lambda,\beta\geq 0,\alpha\geq\alpha_0}{\operatorname{argmin}} \left[\frac{1}{2} \left\| F(\theta_n) + F'|_{\theta_n}(\theta - \theta_n) - S \right\|^2 + \frac{\gamma_n}{2} \|\theta\|^2 + \frac{\gamma_n}{2} \|\nabla_{\mathbb{R}^3 \times S^2} \theta\|^2 \right],$$

where $\theta \equiv (\lambda, \alpha, \beta)$ and

$$F'|_\theta(h_\lambda, h_\alpha, h_\beta)(x, q) = \int_{S^2} \exp\left(-\alpha(x,u)\|q\|^2 - \beta(x,u)(q\cdot u)^2\right) \left(h_\lambda(x,u) - \lambda(x,u)(h_\alpha(x,u)q^2 + h_\beta(x,u)(q\cdot u)^2)\right) du$$

is the Fréchet derivative.

Regularization

Iteratively Regularized Gauss-Newton Method:

$$\theta_{n+1} = \underset{\lambda,\beta\geq 0,\alpha\geq\alpha_0}{\operatorname{argmin}} \left[\frac{1}{2} \left\| F(\theta_n) + F'|_{\theta_n}(\theta - \theta_n) - S \right\|^2 + \frac{\gamma_n}{2} \|\theta\|^2 + \frac{\gamma_n}{2} \|\nabla_{\mathbb{R}^3\times S^2}\theta\|^2 \right],$$

where $\theta \equiv (\lambda, \alpha, \beta)$.

- Smoothness information: Tensors at nearby voxels in similar directions should have similar eigenvalues ($\nabla_{\mathbb{R}^3\times S^2}\alpha$, $\nabla_{\mathbb{R}^3\times S^2}\beta$).
- Positivity: $\lambda \geq 0$ (no fibers have negative volumen fraction), $\alpha \geq \alpha_0 > 0$ (diffusion tensor is positive), $\beta \geq \beta_0 > 0$ (u is eigenvector to *largest* eigenvalue of $D(x, u)$)
- More on the $\nabla_{\mathbb{R}^3\times S^2}\lambda$-term later.

Constrained Minimization Algorithm

- Schematically, we try to solve

$$\text{Minimize } \varphi(\theta) := \frac{1}{2}\|T\theta - g\|^2 + \frac{\gamma}{2}\|\theta\|^2 \quad \text{s.th. } \theta \geq \underline{\theta}, \qquad (1)$$

  where T contains both the Fréchet derivative and the $\nabla_{\mathbb{R}^3 \times S^2}$-terms.
- KKT conditions lead to

$$(T^*T + \gamma)\theta - \lambda = T^*g, \qquad \lambda = \max(0, \lambda - k(\theta - \underline{\theta}))$$

  (where $k > 0$).
- Define the *active set* $\mathcal{A} := \{\lambda - k(\theta - \underline{\theta}) > 0\}$. Then

$$(T^*T + \gamma)\theta - \lambda = T^*g, \qquad \theta = \underline{\theta} \text{ in } \mathcal{A}, \qquad \lambda = 0 \text{ in } \mathcal{A}^c.$$

Constrained Minimization Algorithm

- Semismooth Newton Algorithm: Take initial guesses θ_0, λ_0 and repeat:
  - Set $\mathcal{A}_n = \{\lambda_n - k(\theta_n - \underline{\theta}) > 0\}$.
  - Solve

$$P(T^*T + \gamma)P\theta = P(T^*g - (T^*T + \gamma)(1 - P)\underline{\theta}),$$

    for $P\theta$, where $P = P_{\mathcal{A}_n}$ is the orthogonal projector onto $\mathcal{A}_n^c$.
  - Set $\theta_{n+1} = P\theta + (1 - P)\underline{\theta}$.
  - Set $\lambda_{n+1} = (T^*T + \gamma)\theta_{n+1} - T^*g$.
- The solution is found if the active sets $\mathcal{A}_n$ do not change anymore.
- The algorithm has been shown to converge superlinearly.

Hintermüller, Ito, Kunisch, The Primal-Dual Active Set Strategy as a Semismooth Newton Method, SIAM J. on Optimization, 2002

Duality Gap

A dual problem to (1) is given by:

$$\text{Maximize } \psi(p) := \langle p, g\rangle - \frac{1}{2}\|p\|^2 + \frac{\gamma}{2}\left\|P_C(\gamma^{-1}T^*p)\right\|^2 - \langle T^*p, P_C(\gamma^{-1}T^*p)\rangle,$$

where P_C is the L^2-projector onto $C := \{\theta : \theta \geq \underline{\theta}\}$.

- Strong duality: $\max_p \psi(p) = \min_{\theta\in C}\varphi(\theta)$. The extrema p^*, θ^* fulfill $p^* = g - T\theta^*$.
- After n-th Semismooth Newton step, let $\tilde{\theta}_n = P_C\theta_n$ and $\tilde{p}_n = g - T^*\tilde{\theta}_n$.
- Stop the iterations if $\varphi(\tilde{\theta}_n) - \psi(\tilde{p}_n)$ is small enough and return $\tilde{\theta}_n$.

Alternating Projections

- Unfortunately, solving the above problem in the whole volume is very slow (large number of CG steps for the operator T^*T).
- Instead, use a splitting approach. Separate:

$$\theta_{n+1} = \operatorname*{argmin}_{\theta} \Bigg[\underbrace{\frac{1}{2} \left\| F(\theta_n) + F'|_{\theta_n}(\theta - \theta_n) - S \right\|^2 + \frac{\gamma_n}{2} \|\theta\|^2 + \iota_C(\theta)}_{f_{\text{vox}}(\theta)} + \underbrace{\frac{\gamma_n}{2} \|\nabla_{\mathbb{R}^3 \times S^2} \theta\|^2}_{f_{\text{smooth}}(\theta)} \Bigg],$$

- Iteration:

$$\theta_{n+1}^{k+1} = \operatorname{prox}_{f_{\text{vox}}} \operatorname{prox}_{f_{\text{smooth}}} \theta_{n+1}^{k}$$

where

$$\operatorname{prox}_f u = \operatorname*{argmin}_x \left(f(x) + \frac{1}{2} \|x - u\|^2 \right)$$

Alternating Projections

- $\text{prox}_{f_{\text{vox}}}$ is Tikhonov-regularized, constrained data fitting in each voxel (can be solved efficiently by the primal-dual method above).
- $\text{prox}_{f_{\text{smooth}}}\ u$ (roughly) corresponds to isotropic diffusion-reaction on $\mathbb{R}^3 \times S^2$:

$$\partial_t \theta = \gamma_n \Delta_{\mathbb{R}^3 \times S^2} \theta + (u - \theta)$$

Alternating Projections

- $\text{prox}_{f_{\text{vox}}}$ is Tikhonov-regularized, constrained data fitting in each voxel (can be solved efficiently by the primal-dual method above).
- $\text{prox}_{f_{\text{smooth}}}\ u$ (roughly) corresponds to isotropic diffusion-reaction on $\mathbb{R}^3 \times S^2$:

$$\partial_t \theta = \gamma_n \Delta_{\mathbb{R}^3 \times S^2} \theta + (u - \theta)$$

- This open possibilities to include smoothing that is better adapted to the structure of HARDI reconstructions.

Smoothing step for λ

- Pre-Smoothing: Solve

$$\tilde{\lambda}_n^k = \operatorname*{argmin}_f \frac{1}{2} \left\| f - \lambda_n^k \right\|^2 + \frac{\nu}{2} \left\| \nabla_{\mathbb{R}^3 \times S^2} f \right\|^2 .$$

- Determine a diffusion tensor D_n^k from $\tilde{\lambda}_n^k$ using the $\mathrm{SE}(3)$ formalism using the left-invariant Hessian of $\tilde{\lambda}_n^k$.


R. Duits, Perceptual Organization in Image Analysis, PhD thesis, 2005


E. Franken, Enhancement of Crossing Elongated Structures in Images, PhD thesis, 2009

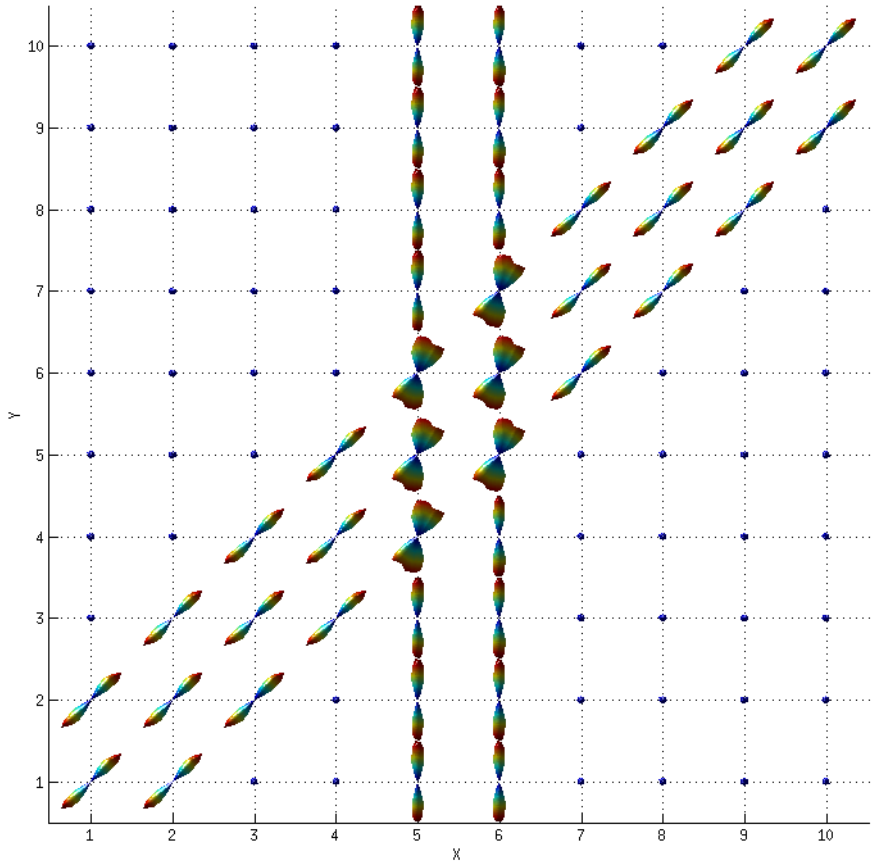
- Solve

$$\lambda_n^{k+1} = \operatorname*{argmin}_f \frac{1}{2} \left\| f - \lambda_n^k \right\|^2 + \frac{\gamma_n}{2} \langle \nabla_{\mathbb{R}^3 \times S^2} f , D_n^k \nabla_{\mathbb{R}^3 \times S^2} f \rangle .$$

Outline

Numerical Examples: Linear Deconvolution

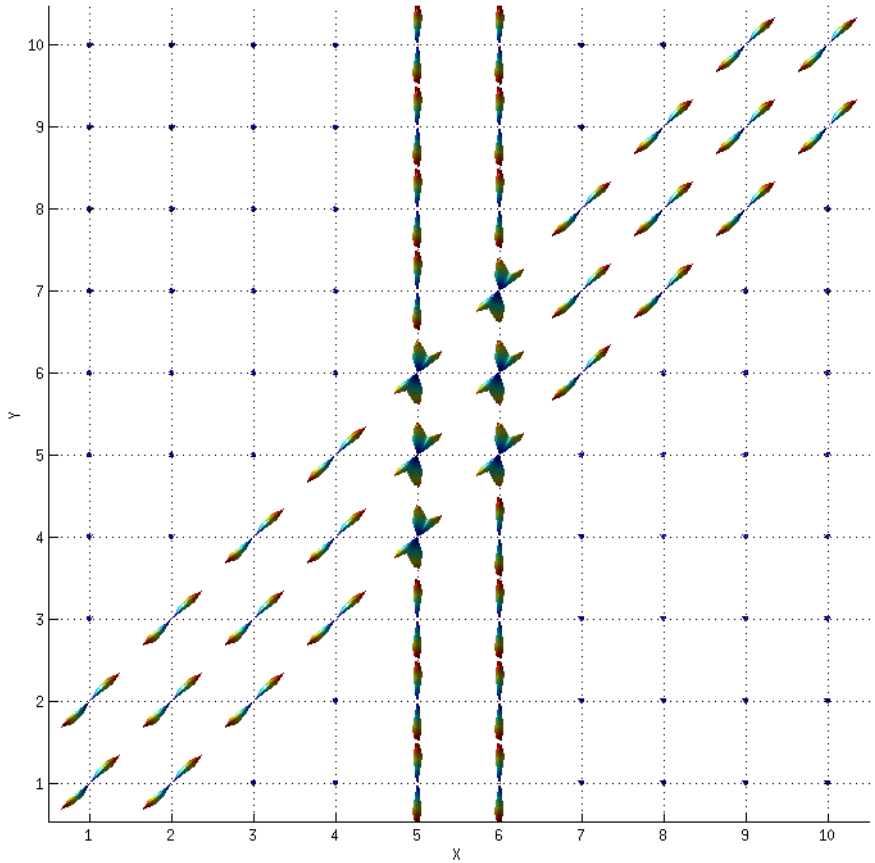
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. Exact data. Reconstruction on 128 points / voxel.



Iteration 1

Numerical Examples: Linear Deconvolution

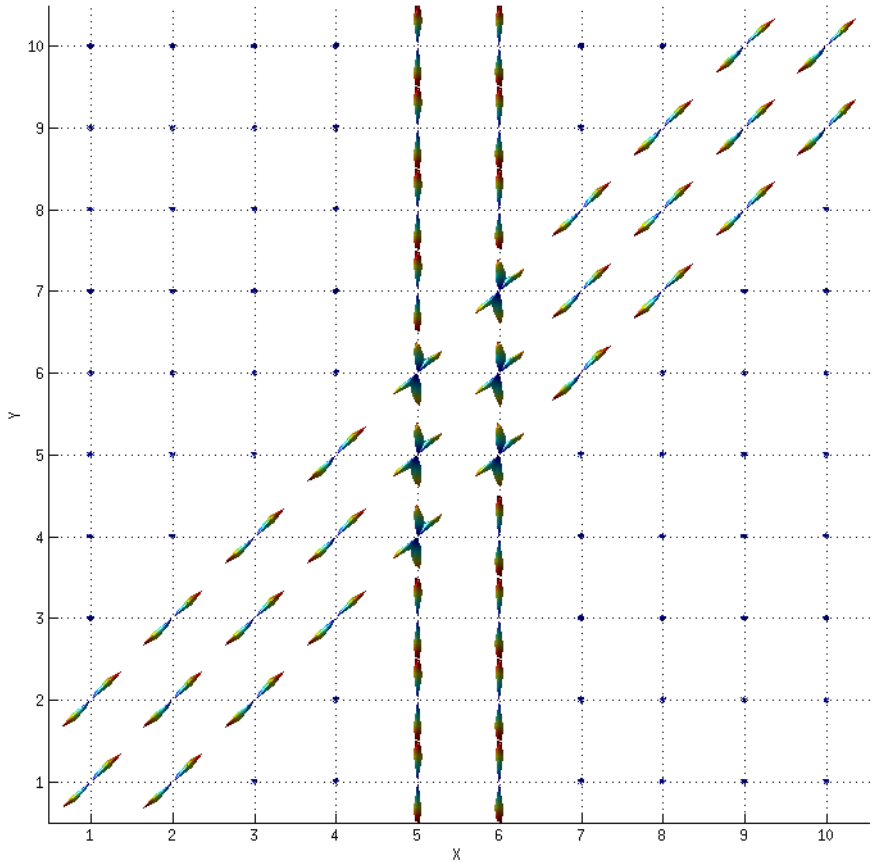
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. Exact data. Reconstruction on 128 points / voxel.



Iteration 5

Numerical Examples: Linear Deconvolution

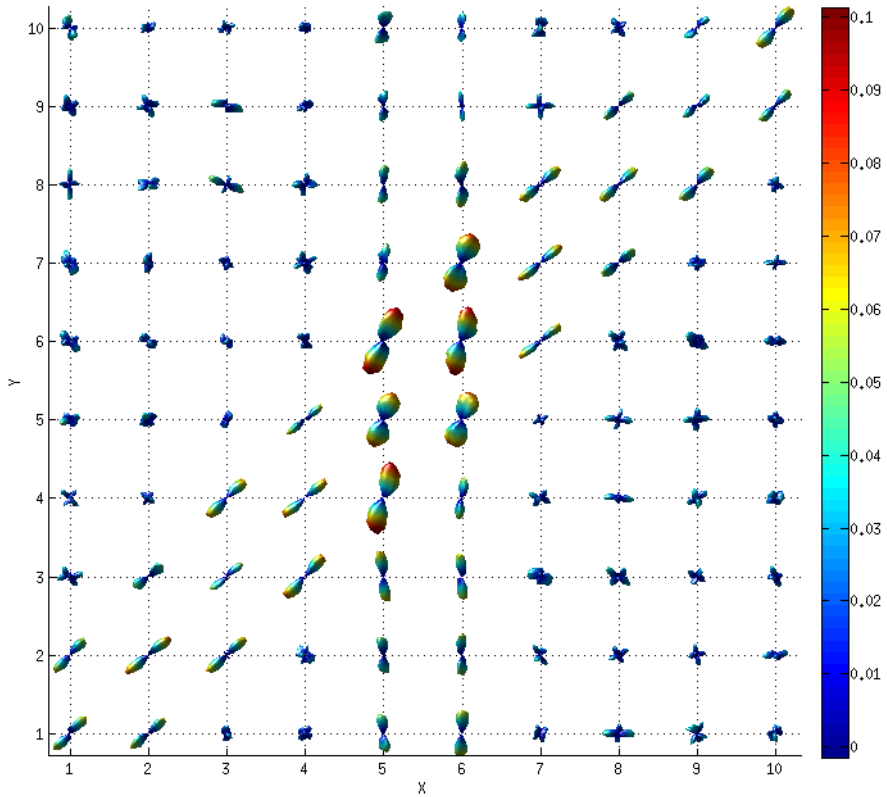
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. Exact data. Reconstruction on 128 points / voxel.



Iteration 10

Numerical Examples: Linear Deconvolution

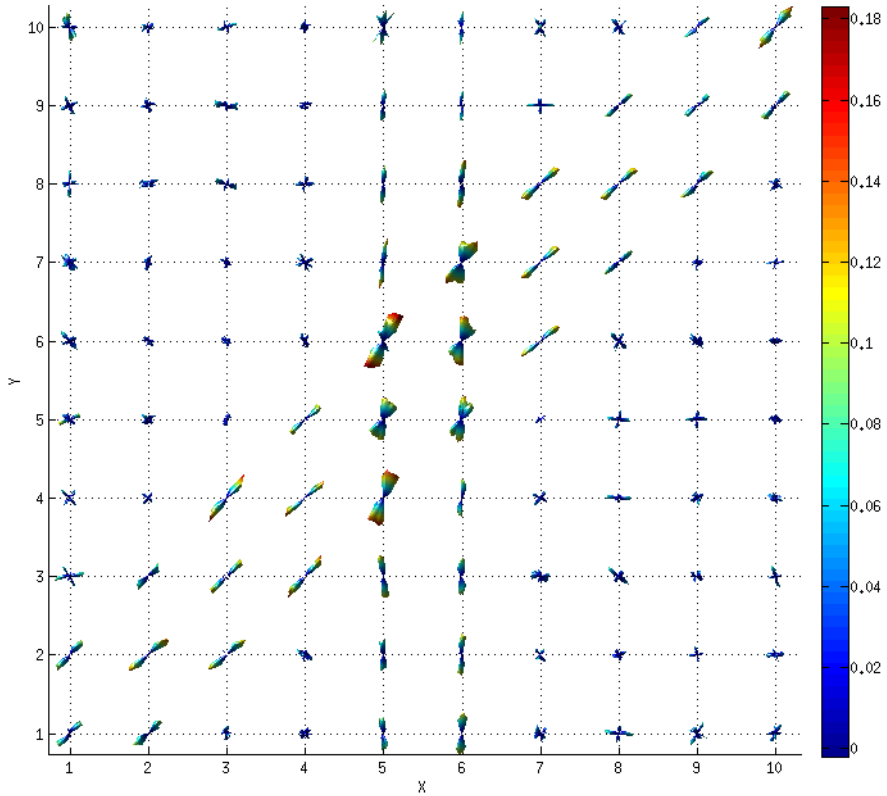
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. 5 % Gaussian noise. Reconstruction on 128 points / voxel.



Iteration 1

Numerical Examples: Linear Deconvolution

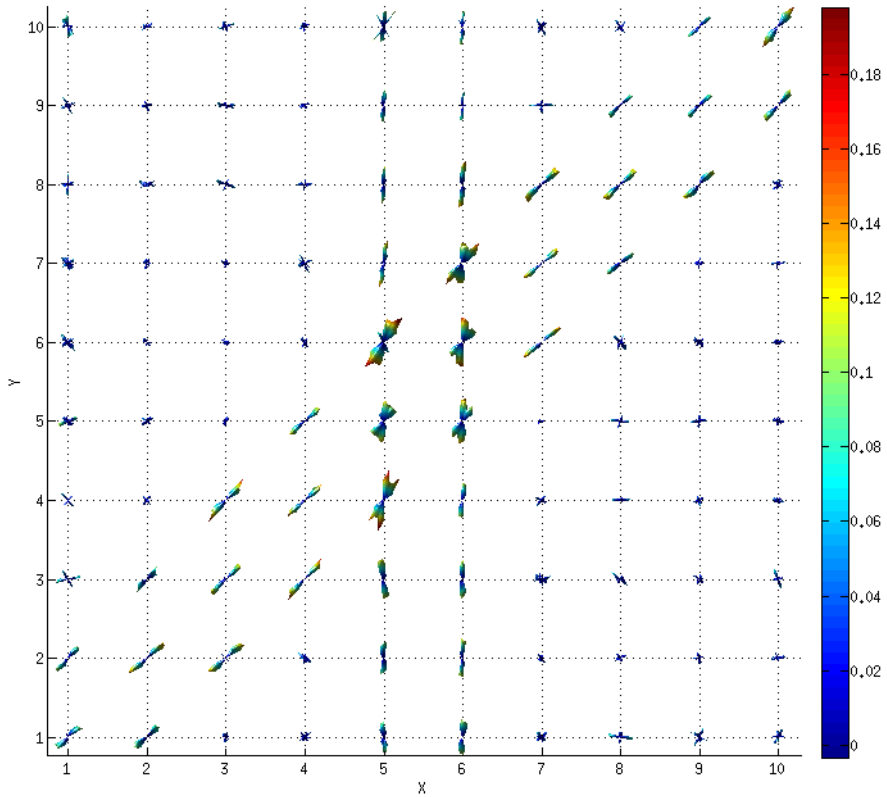
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. 5 % Gaussian noise. Reconstruction on 128 points / voxel.



Iteration 10

Numerical Examples: Linear Deconvolution

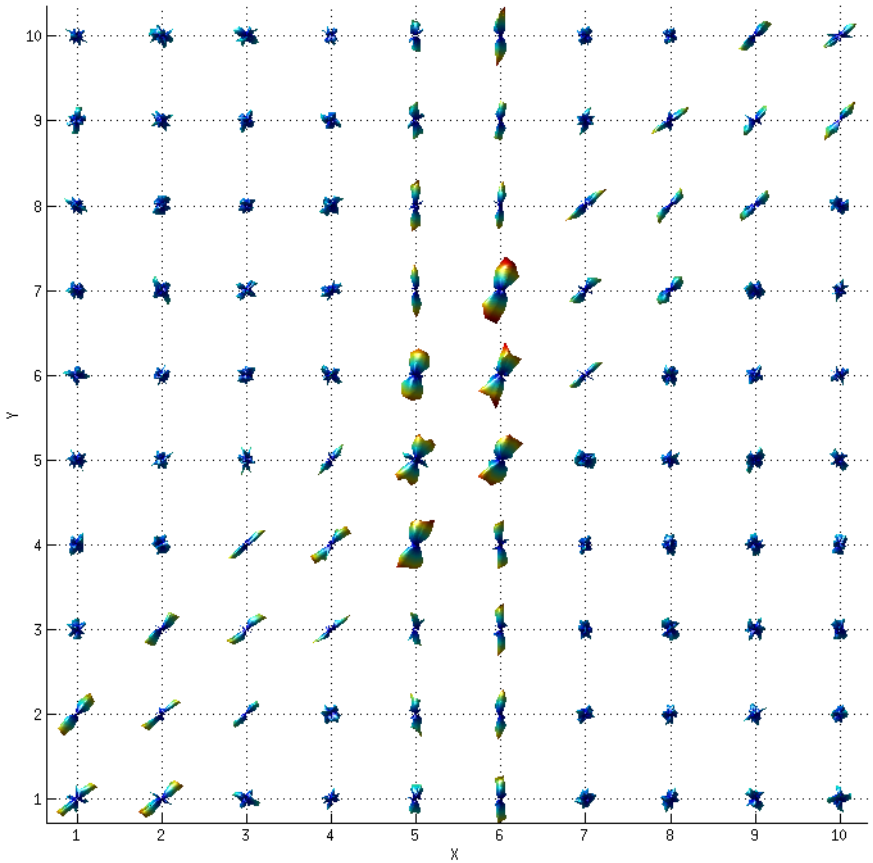
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. 5 % Gaussian noise. Reconstruction on 128 points / voxel.



Iteration 20

Numerical Examples: Non-linear Deconvolution

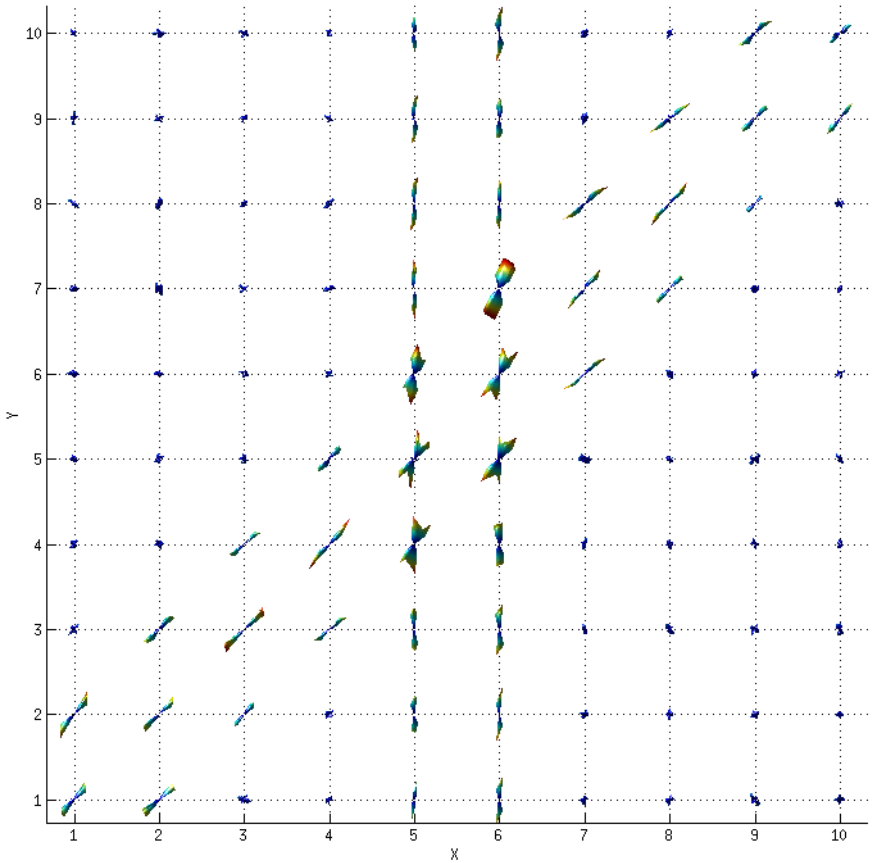
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. 5 % Gaussian noise. Reconstruction on 50 points / voxel. Initial guess $\lambda = 0$, $\alpha = \alpha_0$ and β 3 times as large as the exact value.



Outer Iteration 1

Numerical Examples: Non-linear Deconvolution

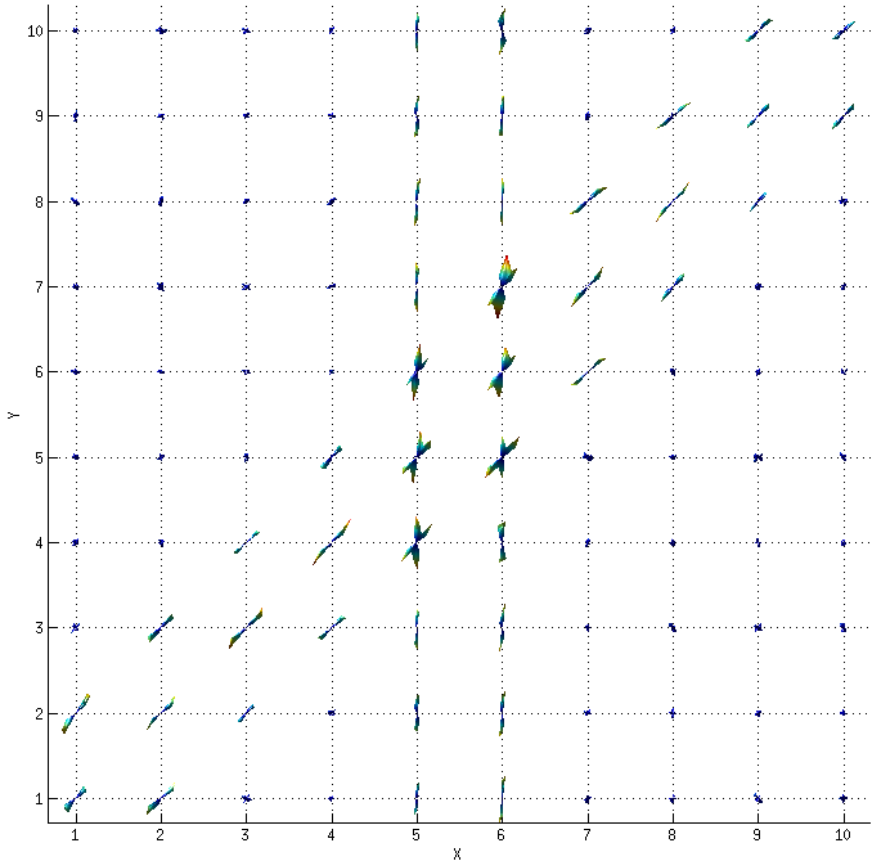
Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. 5 % Gaussian noise. Reconstruction on 50 points / voxel. Initial guess $\lambda = 0$, $\alpha = \alpha_0$ and β 3 times as large as the exact value.



Outer Iteration 5

Numerical Examples: Non-linear Deconvolution

Data simulated on 2 shells with 12 and 32 points, 10^3 voxels. 5 % Gaussian noise. Reconstruction on 50 points / voxel. Initial guess $\lambda = 0$, $\alpha = \alpha_0$ and β 3 times as large as the exact value.



Outer Iteration 10

Conclusions

- Image quality can benefit from including post-processing steps into the reconstruction.
- The method can handle non-linear models and missing data in q-space and (hopefully) k-space.
- Next steps will be
  - analyzing the dependece on the various parameters,
  - improving the efficiency of the implementation,
  - including the Fourier inversion,
  - real data.

Thank you for your attention.