The 3d Space-Varying Coefficient Model and its Influence on Diffusion Tensor Estimation and Fiber Tractography

Susanne Heim

Institute of Statistics
Ludwig-Maximilians-Universität München

September 28, 2007

Diffusion - biologically

[image removed]

Mori, van Zijl; NMR Biomed. 2002; **15**: 468–480

Diffusion - geometrically

[image removed]

Diffusion - mathematically

Diffusion process $\{\mathbf{X}(t) : t_0 \leq t \leq T\}$ with state space $\mathbb{R}^3$:

$$d\mathbf{X}(t) = \mathbf{D}^{\frac{1}{2}}(\mathbf{X}(t))d\mathbf{W}(t),$$

where	$\{\mathbf{W}(t), t \geq 0\}$	Wiener process
and	$\mathbf{D}(\mathbf{X}(t))$	local diffusion tensor.

Tracking principle:

[image removed]

Diffusion - technically

Brain: $n_1 \times n_2 \times n_3$ voxels indexed by s

Theory:

$$S_i(s) = S_0(s) \exp\{-b\, \mathbf{g}_i{}'\mathbf{D}(s)\mathbf{g}_i\}$$

Practice: $S_i(s)$ is noisy, $\mathbf{D}(s)$ has to be estimated

$$y_i(s) = -\frac{1}{b}\log\left(\frac{S_i(s)}{S_0(s)}\right) = \mathbf{x}_i{}'\boldsymbol{\beta}(s) + \varepsilon_i(s)$$

with $\boldsymbol{\beta}(s) = (D_1, D_2, D_3, D_4, D_5, D_6)'(s)$

$$\mathbf{x}_i = (g_{1i}^2, g_{2i}^2, g_{3i}^2, 2g_{1i}g_{2i}, 2g_{1i}g_{3i}, 2g_{2i}g_{3i})'$$

and $\varepsilon_i(s) \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

Diffusion tensor estimation

So far: voxelwise regression & post-hoc smoothing

Goal: **one** model for estimation & regularization & interpolation

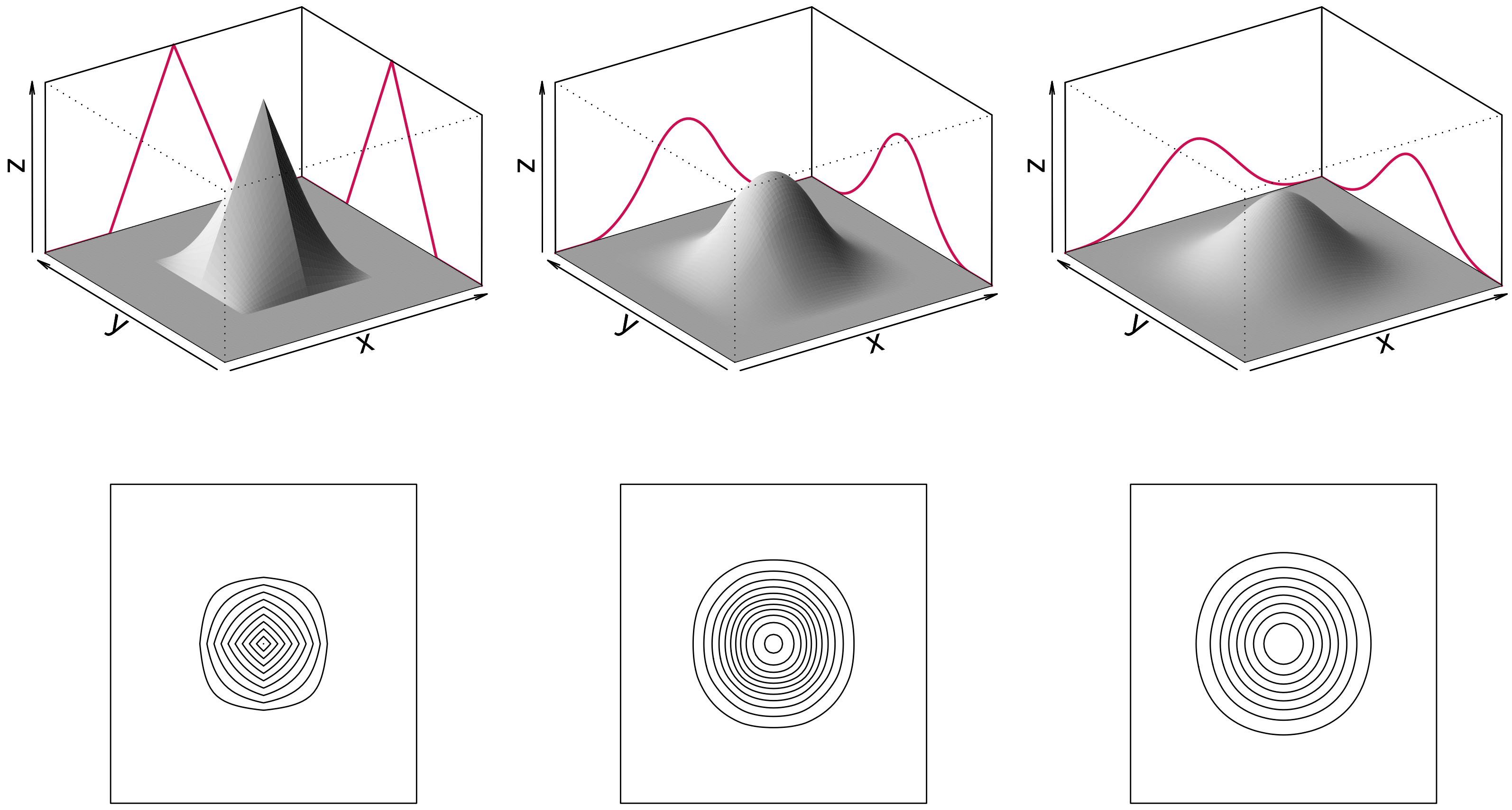
$$\mathbf{y} = \sum_{j=1}^{p} \big(\mathbf{I}_n \otimes \mathbf{X}(\cdot, j)\big)\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_{rn}),$$

using **3d space-varying coefficients** $\boldsymbol{\beta}_j = (\beta_j(1), \ldots, \beta_j(n))'$,

$\mathbf{y} = (y_1(1), \ldots, y_r(1), \ldots, \ldots, y_1(n), \ldots, y_r(n))'$,

r-dimensional regressor $\mathbf{X}(\cdot, j)$.

Multidimensional B-spline basis functions



Direct approach I

Projection of $\boldsymbol{\beta}_j$ onto 3d B-splines:

$$\boldsymbol{\beta}_j(s) = \sum_{v=1}^{KLM} (\mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1)(s,v)\gamma_j(v) = \mathbf{B}(s,\cdot)\boldsymbol{\gamma}_j, \quad j = 1,\ldots,p$$

Estimation of the $(KLMp \times 1)$ - amplitude vector γ by minimizing:

$$Q(\boldsymbol{\gamma},\boldsymbol{\lambda}) = \|\mathbf{y} - \sum_{j=1}^{p} \big(\mathbf{I}_n \otimes \mathbf{X}(\cdot,j)\big)\mathbf{B}\boldsymbol{\gamma}_j\|^2 + \lambda_1 \left\|(\mathbf{I}_L \otimes \mathbf{I}_M \otimes \boldsymbol{\Delta}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma}\right\|^2$$

$$+ \lambda_2 \left\|(\mathbf{I}_M \otimes \boldsymbol{\Delta}_2 \otimes \mathbf{I}_K \otimes \mathbf{I}_p)\boldsymbol{\gamma}\right\|^2 + \lambda_3 \left\|(\boldsymbol{\Delta}_3 \otimes \mathbf{I}_K \otimes \mathbf{I}_L \otimes \mathbf{I}_p)\boldsymbol{\gamma}\right\|^2$$

$$= \left\|\mathbf{y} - (\mathbf{B} \otimes \mathbf{X})\boldsymbol{\gamma}\right\|^2 + \lambda_1 \left\|\mathbf{P}_1\boldsymbol{\gamma}\right\|^2 + \lambda_2 \left\|\mathbf{P}_2\boldsymbol{\gamma}\right\|^2 + \lambda_3 \left\|\mathbf{P}_3\boldsymbol{\gamma}\right\|^2$$

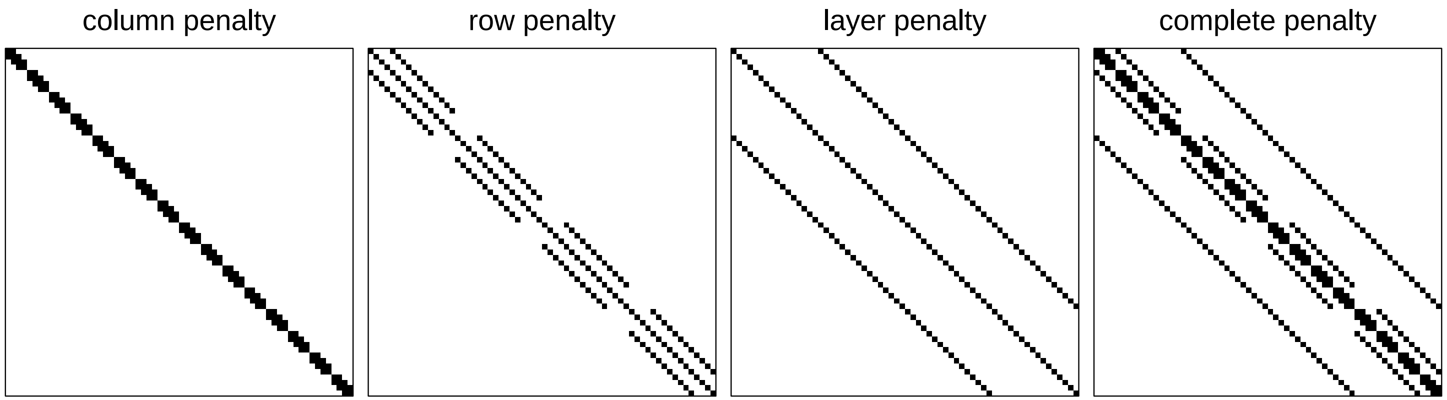
Direct approach II

Solution:

$$\hat{\gamma} = (\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}\mathbf{U}'\mathbf{y},$$

where $\mathbf{U} = \mathbf{B} \otimes \mathbf{X}$ is $(rn \times pKLM)$-dimensional,

and $\mathbf{P} = \lambda_1 \mathbf{P}_1' \mathbf{P}_1 + \lambda_2 \mathbf{P}_2' \mathbf{P}_2 + \lambda_3 \mathbf{P}_3' \mathbf{P}_3$.



'The devil in disguise'

Realistic scenario: $64 \times 64 \times 24$ voxels with $50 \times 50 \times 19$ knots

➡ $4.6 \cdot 10^9$ elements in $\mathbf{B}$, i. e. $\mathbf{B} \approx 37$ Gb

➡ $50 \times 50 \times 19 \times 6 = 285 \cdot 10^3$ normal equations

➡ regression diagnostics hard to obtain

⇨ software libraries for sparse matrices

⇨ R package **svcm**

Sequential approach

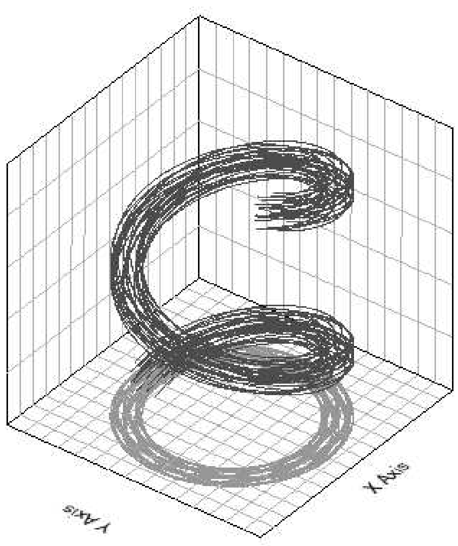
Minimize:

$$Q(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \|\mathbf{y} - (\mathbf{B} \otimes \mathbf{X})\boldsymbol{\gamma}\|^2 + \mathrm{Pen}(\boldsymbol{\gamma}, \boldsymbol{\lambda}),$$

where the penalty term consists of 15 term of Kronecker products.

- approximation to the direct tensor product approach
- efficient successive computation by exploiting the array structure
- easy computation of the hat matrix

Simulation study



$N = 100$ runs corrupted by Gaussian noise

$$\mathrm{VMSE}_j(s) = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\beta}_j(s) - \hat{\boldsymbol{\beta}}_j^{(i)}(s)\right)^2$$

$$\log\left(\mathrm{VMSE}_{\text{method A}}/\mathrm{VMSE}_{\text{method B}}\right) \in (-\infty, +\infty)$$

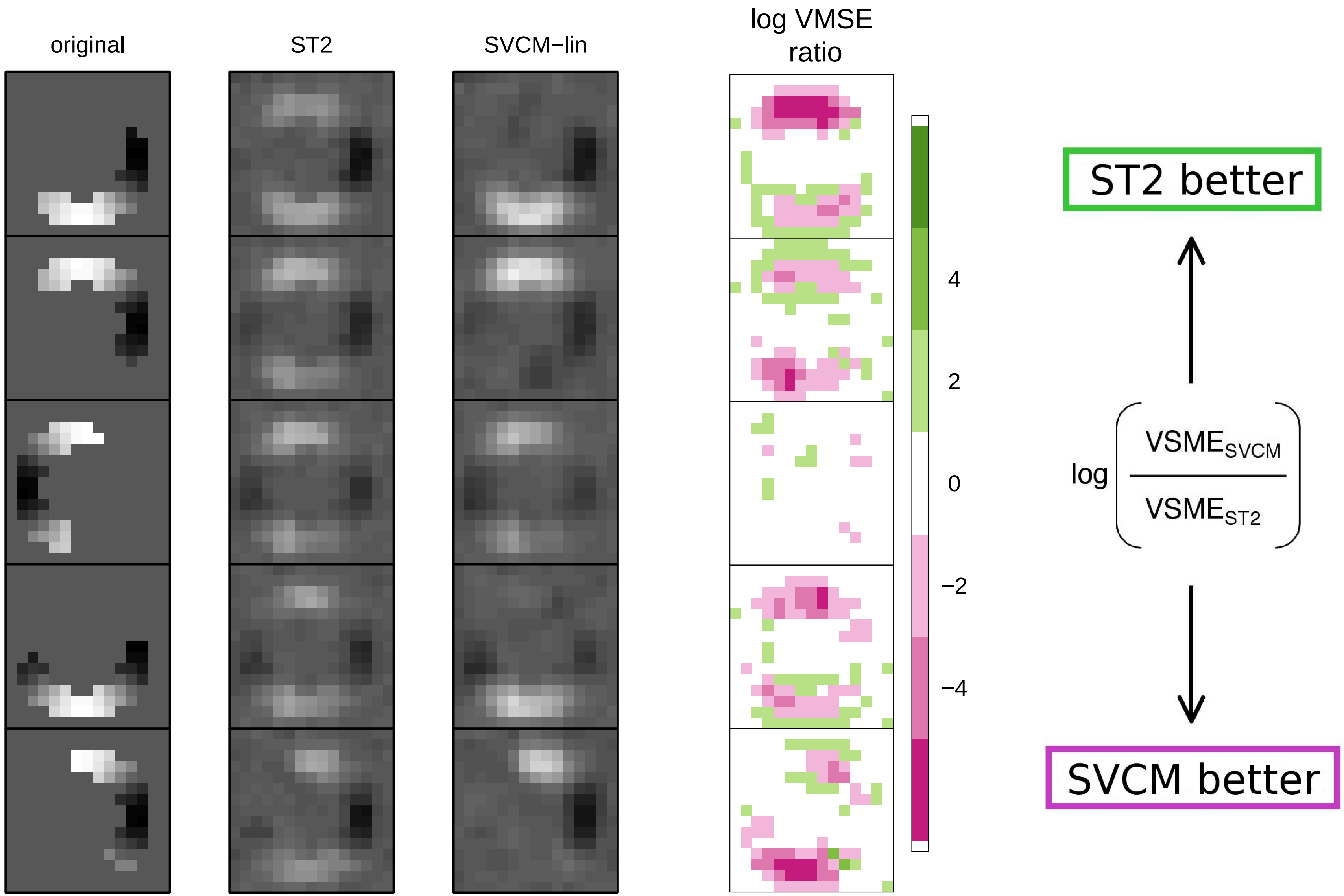
Methods:

– voxelwise regression + Gaussian kernel (+ linear interpolation)

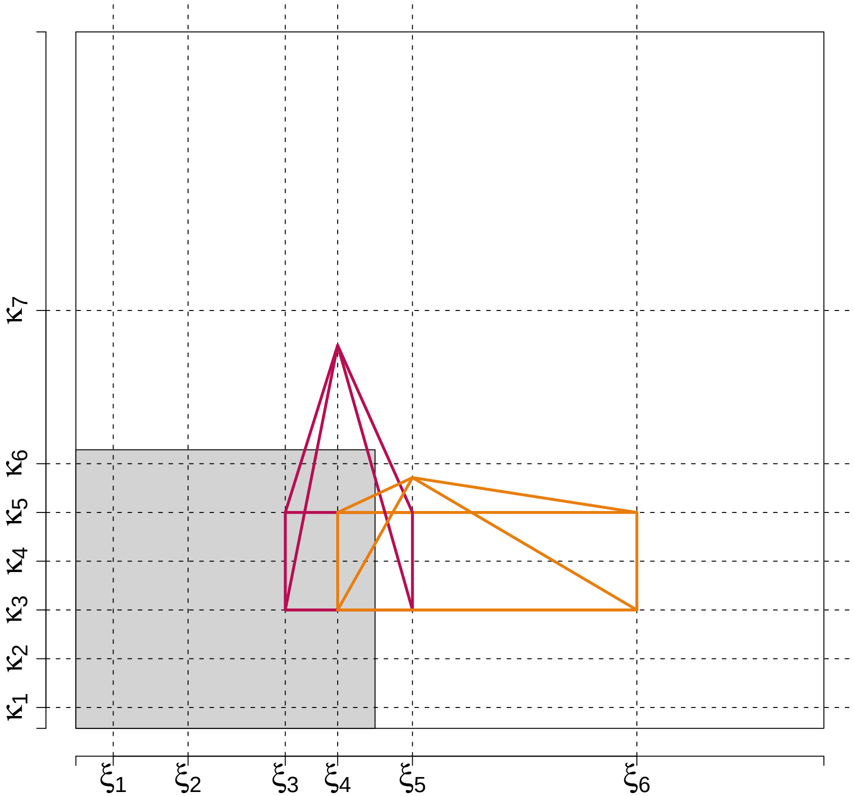
– B-splines of degree 1 and 2 + first order difference penalties:

direct and successive approach respectively

Error ratio map of one diagonal element



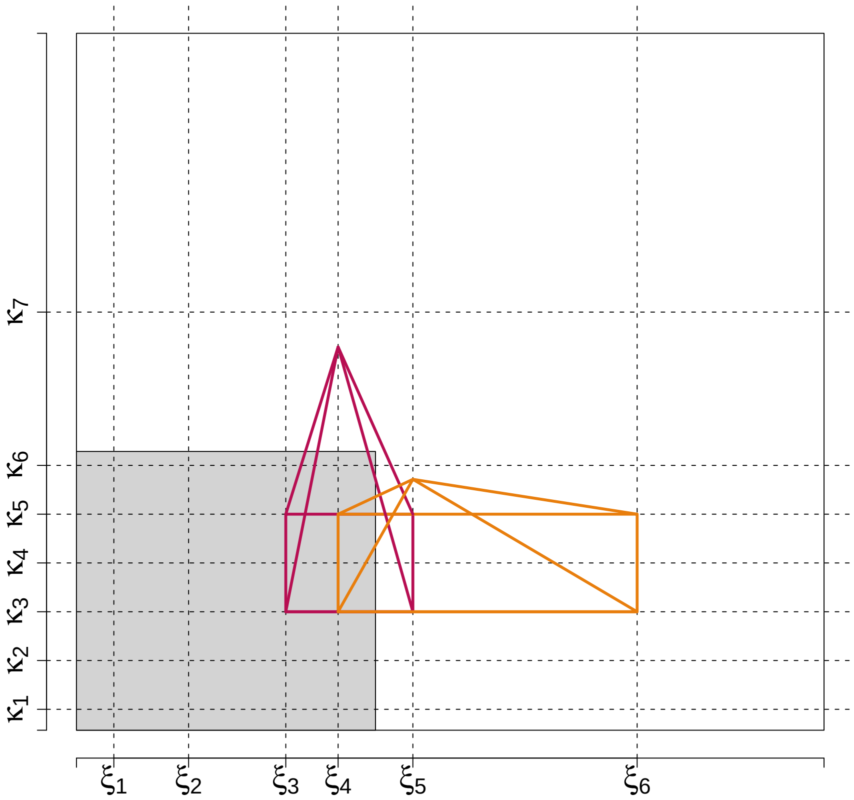
Interim result

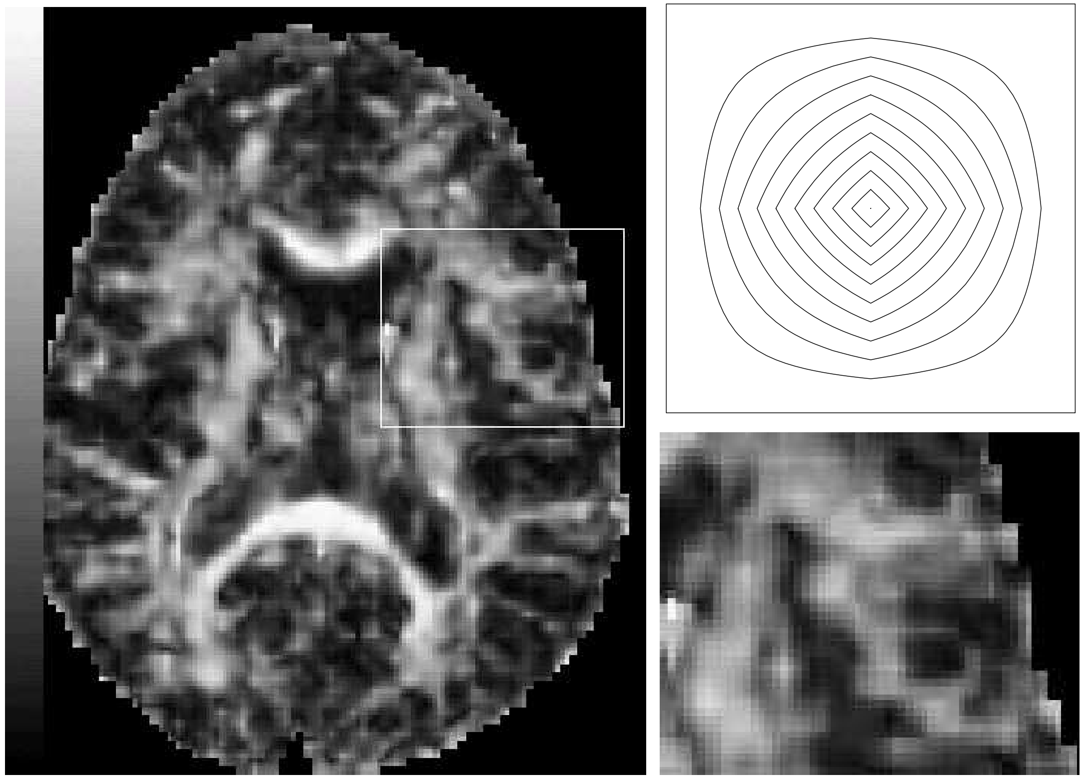


lacking local adaptivity

➡ weighted penalization

Interim result





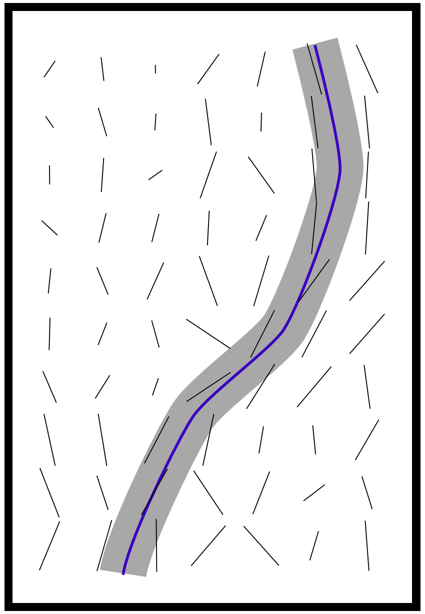
lacking local adaptivity

➡ weighted penalization

'blips' when interpolated

➡ basis exchange

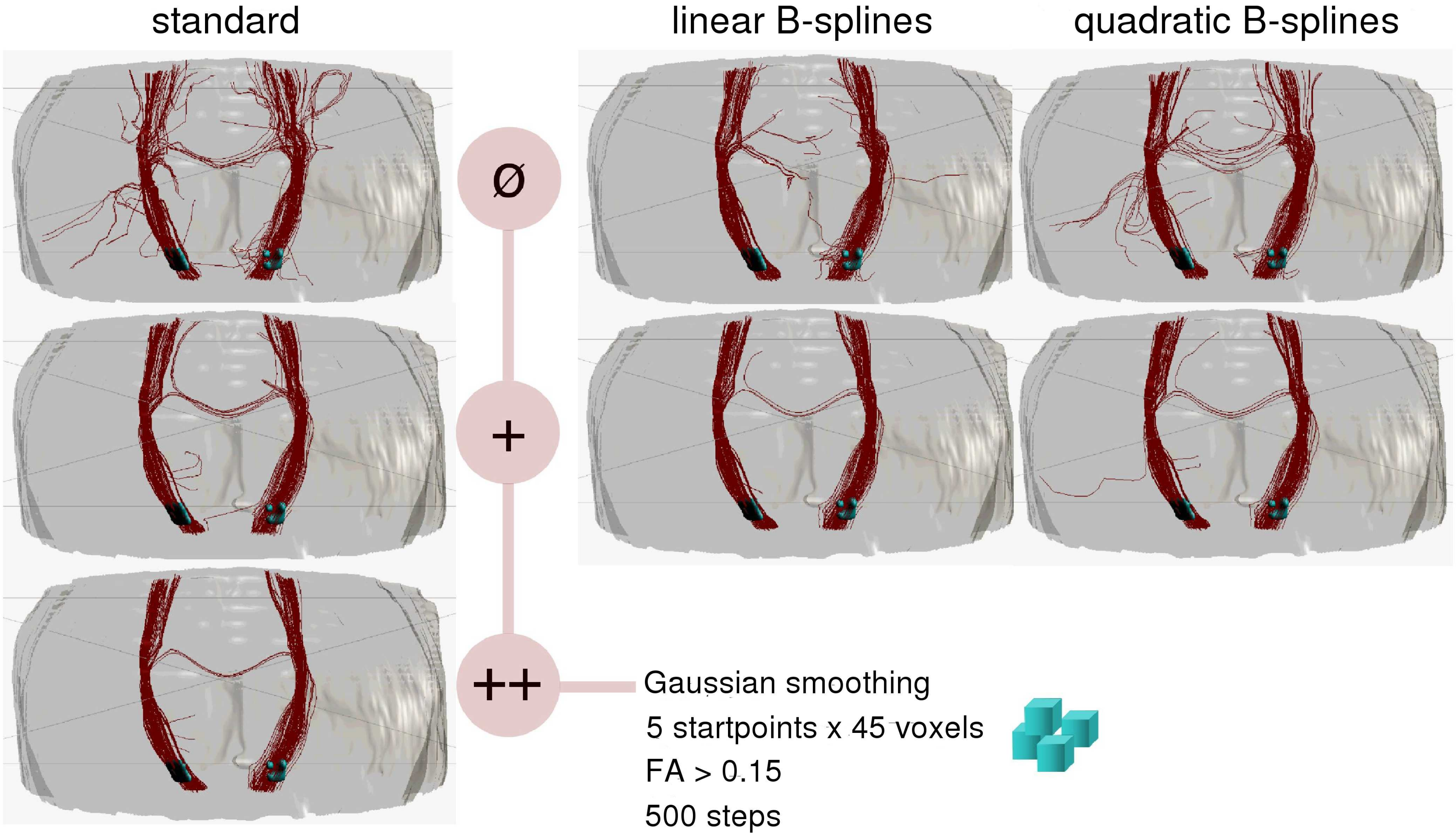
How relevant are the tensor elements?



Focus on fiber reconstruction

- regularization and interpolation
- choice of starting points
- determination of stopping criteria

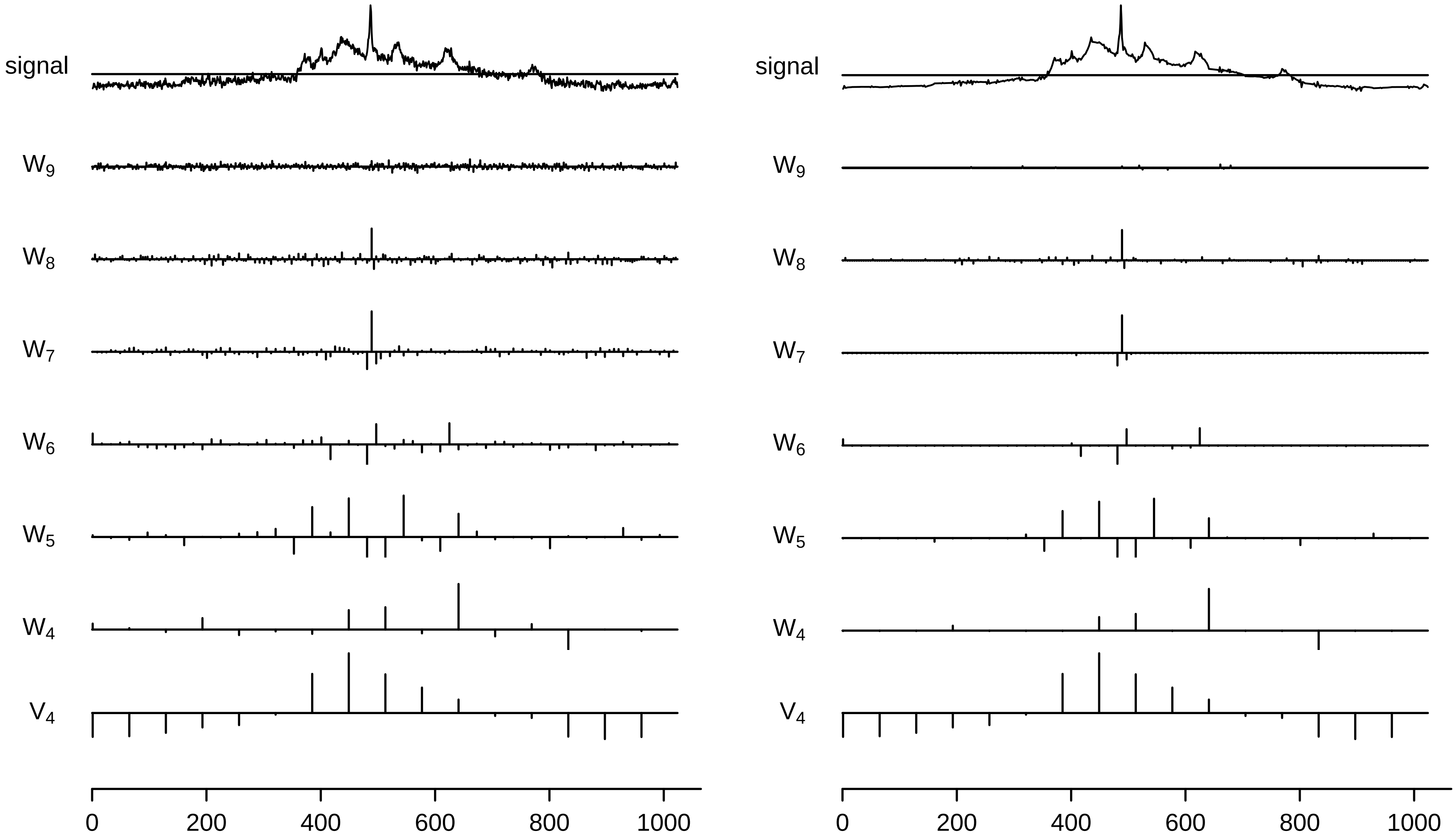
Fiber reconstruction



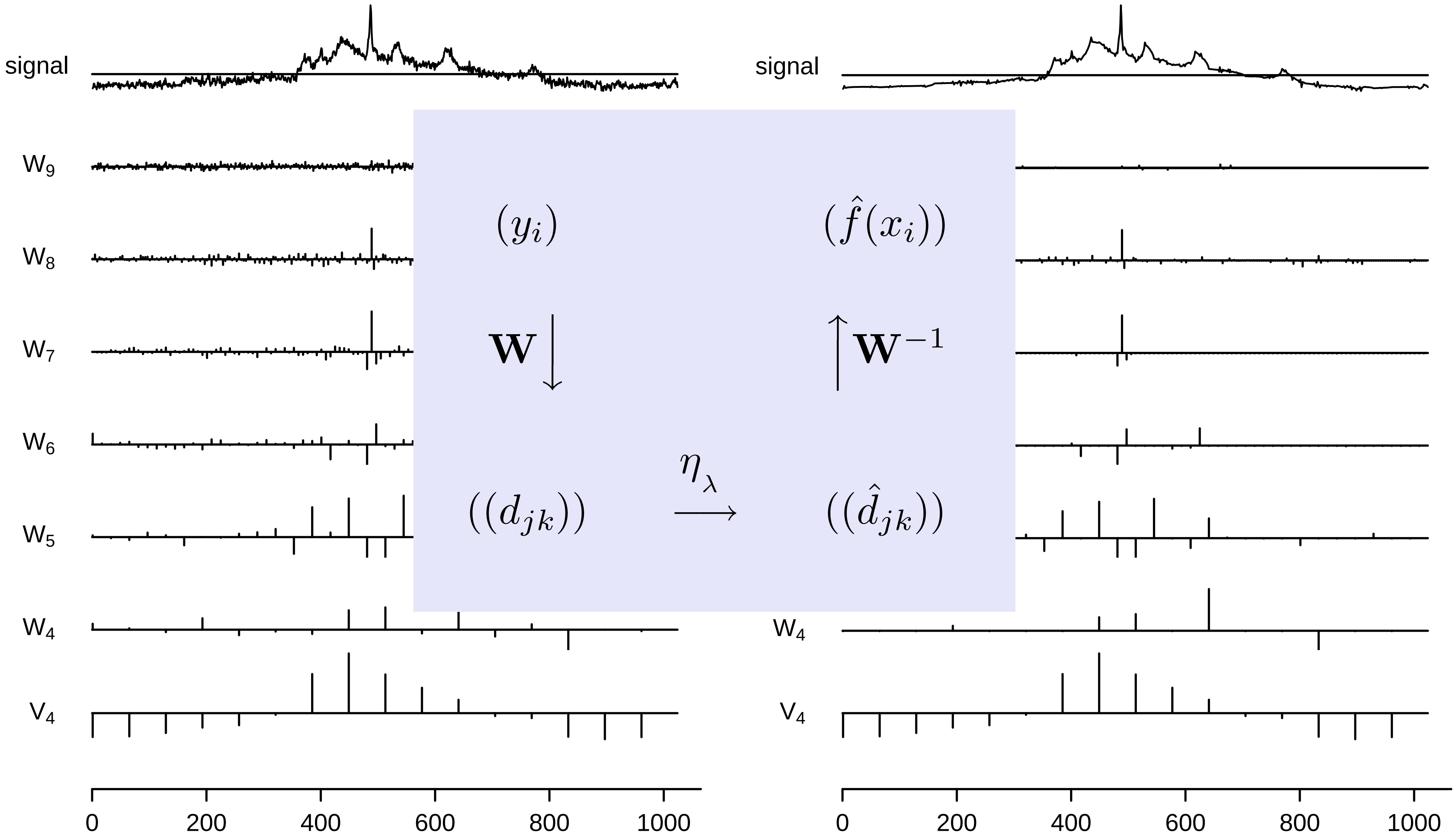
Regression model with 3d space-varying coefficients

Results	**Extensions**
published *Heim, Fahrmeir, Eilers, Marx (2007, CSDA)*	basis exchange to wavelets
implemented *R package 'svcm'*	confidence intervals
evaluated using DTI data	more complex simulation model
reintegrated into fiber reconstruction	similarity measures, Brownian bridge

Wavelet approach: Motivation



Wavelet approach: Motivation



$$(y_i) \qquad (\hat{f}(x_i))$$

$$\mathbf{W}\downarrow \qquad \uparrow\mathbf{W}^{-1}$$

$$((d_{jk})) \quad \xrightarrow{\eta_\lambda} \quad ((\hat{d}_{jk}))$$

2d decomposition example

[image removed]

Wavelet regularization

"Hard" thresholding is "keep or kill" strategy:

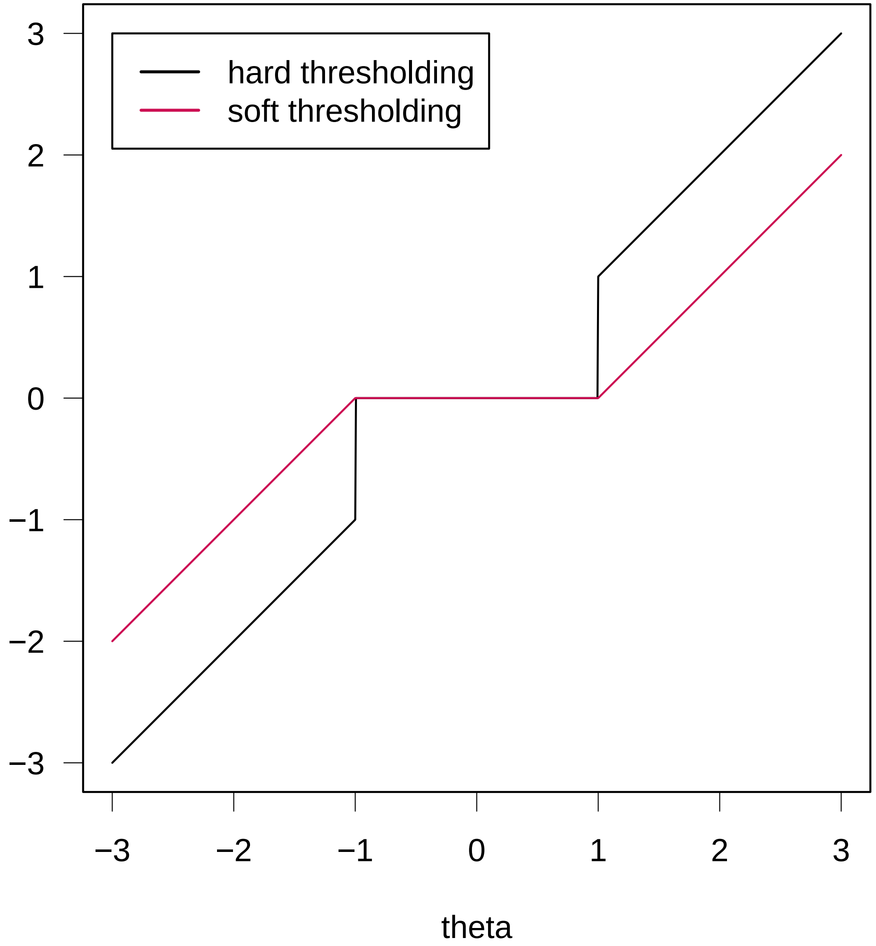
$$\hat{\theta}_j = \eta_\lambda^H(\hat{\theta}_j^{LS}) = \begin{cases} \hat{\theta}_j^{LS} & |\hat{\theta}_j^{LS}| \geq \lambda \\ 0 & \text{otherwise} \end{cases}$$

"Soft" thresholding is shrinkage:

$$\|\mathbf{y} - \mathbf{W}\boldsymbol{\theta}\|_2^2 + 2\lambda\|\boldsymbol{\theta}\|_1 \longrightarrow \min_{\boldsymbol{\theta}}$$

with solution

$$\hat{\theta}_j = \eta_\lambda^S(\hat{\theta}_j^{LS}) = \text{sign}(\hat{\theta}_j^{LS})(|\hat{\theta}_j^{LS}| - \lambda)_+$$



Application to DTI data: A proposal

Recall joint regression model

$$\mathbf{y} = \sum_{j=1}^{p} \big(\mathbf{X}(\cdot, j) \otimes \mathbf{I}_n\big)\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_{rn}).$$

Project each surface $\boldsymbol{\beta}_j$ onto 3d tensor product wavelets:

$$\boldsymbol{\beta}_j = \mathbf{W}\boldsymbol{\gamma}_j, \quad j = 1, \ldots, p$$

Restate the SVCM:

$$\mathbf{y} = (\mathbf{X} \otimes \mathbf{W})\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

with $\mathbf{X}$ of dimension $(r \times p)$, $\mathbf{W}$ of dimension $(n \times n)$, and $\boldsymbol{\gamma}$ of dimension $(pn \times 1)$.

Transfer to the wavelet domain

Initial least squares estimation:

$$\begin{aligned}\hat{\gamma}^{LS} &= \big((\mathbf{X}\otimes\mathbf{W})'(\mathbf{X}\otimes\mathbf{W})\big)^{-1}(\mathbf{X}\otimes\mathbf{W})'\mathbf{y} \\ &= \left(\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{p\times r}\otimes\mathbf{I}_n\right)\begin{pmatrix}\mathbf{W}'\mathbf{y}_1\\ \vdots \\ \mathbf{W}'\mathbf{y}_r\end{pmatrix}_{nr\times 1}\end{aligned}$$

Alternative formulation of the SVCM:

$$\mathbf{y}_{\mathrm{W}} = (\mathbf{X}\otimes\mathbf{I}_n)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_{\mathrm{W}},$$

with $\quad \mathbf{y}_{\mathrm{W}} = (\mathbf{I}_r\otimes\mathbf{W}')\mathbf{y}$

and $\quad \mathrm{Var}(\boldsymbol{\varepsilon}_{\mathrm{W}}) = \mathrm{Var}\big((\mathbf{I}_r\otimes\mathbf{W}')\boldsymbol{\varepsilon}\big) = \sigma^2\mathbf{I}_{rn}.$

Regularization and backsubstitution

Independent thresholding:

$$\hat{\gamma}_{\lambda_j,j} \quad = \quad \delta_{\lambda_j}\left(\hat{\gamma}_j^{LS}\right), \qquad j = 1, \ldots, p$$

Final smooth estimates of the coefficient surfaces:

$$\hat{\boldsymbol{\beta}}_j \quad = \quad \mathbf{W}\hat{\gamma}_{\lambda_j,j}$$

➡ What is the most suitable wavelet family for DTI data?

➡ Which shrinkage rule is appropriate?

➡ How can the resolution be increased?