

Weierstraß-Institut für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint

ISSN 0946 – 8633

Structural adaptive smoothing: Principles and applications in imaging

Jörg Polzehl, Karsten Tabelow

submitted: April 2nd, 2009

Weierstrass Institute
for Applied Analysis and Stochastics
Mohrenstr. 39
10117 Berlin
Germany
email: polzehl@wias-berlin.de
tabelow@wias-berlin.de

No. 1412

Berlin 2009



2000 *Mathematics Subject Classification.* 68U10, 92C55, 62G08 .

Key words and phrases. Image Enhancement, Functional Magnetic Resonance Imaging, Diffusion Tensor Imaging, Structural Adaptive Smoothing.

K. Tabelow was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

Edited by

Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)

Mohrenstraße 39

10117 Berlin

Germany

Fax: + 49 30 2044975

E-Mail: preprint@wias-berlin.de

World Wide Web: <http://www.wias-berlin.de/>

Abstract

Structural adaptive smoothing provides a new concept of edge-preserving non-parametric smoothing methods. In imaging it employs qualitative assumption on the underlying homogeneity structure of the image. The chapter describes the main principles of the approach and discusses applications ranging from image denoising to the analysis of functional and diffusion weighted Magnetic Resonance experiments.

1 Introduction

Images are often characterized by qualitative properties of their spatial structure, e.g. spatially extended regions of homogeneity that are separated by discontinuities. Images or image data with such a property are the target of the methods considered in this chapter. Alternative geometric characterizations using orientation or channels in feature space are discussed in [22, 21, 26, 19, 28] and could be combined with the approach pursued here.

The methods summarized under the term structural adaptive smoothing try to employ a qualitative assumption on the spatial structure of the data to simultaneously describe the structure and efficiently estimate parameters like image intensities. Structural adaptive smoothing generalizes several concepts in non-parametric regression. These include kernel smoothing and local polynomials, see e.g. [73, 20, 61] or [9], the filter proposed by Lee [42], bilateral smoothing [68] and scale space methods, see e.g. [51, 12]. Relations probably exist to diffusion methods in the Beltrami framework.

Our approach provides an alternative to non-linear diffusion methods, see e.g. [60, 74], and generalizes linear diffusion in a different way. Information on the error distribution and qualitative assumptions on the underlying structure are effectively coded by the definition of statistical penalties. The methods are designed to provide intrinsic balance between variability and bias of the reconstruction results. In contrast to diffusion methods this leads to a meaningful limit for increasing bandwidth, or correspondingly diffusion time. An approach derived from a related idea is described in [59].

A first attempt to use the idea of structural adaptive smoothing was proposed in [52] under the name adaptive weights smoothing. This was generalized and refined especially in [53] providing a theory for the case of one-parameter exponential families. Several

extensions have been made to cover locally smooth images [54], color images [55] and special applications like functional Magnetic Resonance Imaging (fMRI) [66, 56] and Diffusion Tensor Imaging (DTI) [64, 57].

The next section introduces the general approach. We then illustrate how this translates to different imaging modalities and problems like image denoising in 2D and 3D, signal detection in fMRI and smoothing in DTI. Within the summary we give some information on implementations and numerical complexity of the algorithms.

2 Structural adaptive smoothing

Within this chapter we assume the following data structure. We denote by $x_1, \dots, x_n \in \mathcal{X} \subseteq \mathbb{R}^p$ the experimental design. In imaging x_i usually will be a point on a p dimensional grid, although this assumption is not necessary for the approach. At each design point x_i we assume to observe a scalar or vector $Y_i \in \mathcal{Y} \subseteq \mathbb{R}^q$.

We assume that the observed values Y_i follow a probability distribution $P_\theta(x_i)$ from a family $\mathcal{P} = \{\mathcal{P}_\theta; \theta \in \times\}$ and that we are interested in estimating θ or some function $g(\theta)$ as a function of x . Traditional methods in non-parametric regression allow for varying parameters but usually assume that $\theta(x)$ is a smooth function in x . This is violated for image data characterized by strong discontinuities.

Instead we try to describe the image by its local homogeneity structure. We assume that there exists a partitioning

$$\mathcal{X} = \bigcup_{m=1}^M \mathcal{X}_m \tag{1}$$

such that

$$\theta(x) \approx \theta(x_i) \Leftrightarrow \exists m : x \in \mathcal{X}_m \wedge x_i \in \mathcal{X}_m$$

i.e. that θ is approximately constant on each \mathcal{X}_m . This assumption is very weak in the sense that the number M of partitions may be large and that there are no restrictions on the form of the sets \mathcal{X}_m . Nevertheless the assumption will prove helpful if there exists a partitioning with $M \ll n$ and where the \mathcal{X}_m have some spatial extend. This structural assumption is used within an iterative procedure.

We do not directly enforce the partitioning (1). Instead, for each design point x_i , describe a set $U(x_i)$ containing x_i by a weighting scheme

$$W(x_i) = (w_1(x_i), \dots, w_n(x_i)) = (w_{i1} \dots, w_{in}).$$

A positive weight w_{ij} will be assigned if the estimates of θ_j and θ_i are not significantly

different. In this case x_j would be contained in $U(x_i)$. Within the iteration process $U(x_i)$ can be sought of as being a subset or an estimate of a set \mathcal{X}_m containing x_i from the assumed partition. $U(x_i)$ and $U(x_j)$ will usually not coincide at any stage of the process even if the structural assumption is valid exactly and x_i and x_j belong to the same set \mathcal{X}_m .

We try to determine the structure, e.g. sets of similar parameters, and estimate the parameters in an iterative procedure. We start at each design point with an initial estimate, if possible solely obtained from the observation at this point, and initialize a bandwidth h such that a ball of radius h just contains some neighboring points. We will formalize this later. We now alternate the following steps. At all design points x_i and for all design points x_j within a ball of radius h we assign a positive weight w_{ij} if $\hat{\theta}(x_j)$ belongs to a confidence region for $\theta(x_i)$ and a zero weight for all other points, thereby creating a new weighting scheme $W_i = W(x_i)$. This means we employ information from the estimates to learn on the underlying structure. We then use the generated weighting scheme to obtain a new estimate $\hat{\theta}(x_i)$ by weighted local likelihood or minimizing a weighted risk. Before continuing we synchronize and increase the bandwidth h thereby allowing for more positive weights and for a decrease in variability of the estimates. We stop iterating when a prespecified bandwidth, corresponding to a maximal possible variance reduction, is reached.

We now more formally describe how we generate the weighting schemes. Let x_i be fixed and $\hat{\theta}(x_i)$ be obtained employing a weighting scheme $W^{(k-1)}(x_i) = W_i^{(k-1)} = (w_{i1}^{(k-1)}, \dots, w_{in}^{(k-1)})$. We consider all x_j such that $\|x_i - x_j\|^2/h^2 \leq 1$. New weights $w_{ij}^{(k)}$ are then generated as the product of two terms

$$w_{ij}^{(k)} = K_{\text{loc}}(l_{ij})K_{\text{st}}(s_{ij}^{(k-1)})$$

where

$$l_{ij} = \|x_i - x_j\|^2/h^2 \quad \text{and} \quad s_{ij}^{(k-1)} = \frac{N_i}{\lambda} T(\hat{\theta}(x_j), \hat{\theta}(x_i)) \quad \text{with} \quad N_i = \sum_{j=1}^n w_{ij}^{(k-1)}$$

are two penalties measuring the spatial distance between the two design points and the difference between the two estimates. Both terms depend on kernel functions K_{loc} and K_{st} respectively. The second term should reflect both the difference of the estimated parameters and the variability of the parameter estimate at point x_i . Such a statistics $T(\hat{\theta}(x_j), \hat{\theta}(x_i))$ can often be derived as the Kullback-Leibler distance $\mathcal{K}(\hat{\theta}_j, \hat{\theta}_i)$ distance between the probability distributions $P_{\hat{\theta}_i}$ and $P_{\hat{\theta}_j}$. As an alternative, if $\hat{\theta}(x_i)$ is obtained by minimization of a risk $R(Y, W_i; \theta)$ we may define

$$T(\hat{\theta}(x_j), \hat{\theta}(x_i)) = 2(R(Y, W_i; \hat{\theta}(x_j)) - R(Y, W_i; \hat{\theta}(x_i)))$$

which in case of a logarithmic likelihood corresponds to using likelihood profiles.

Given a weighting scheme W_i we estimate parameters by either weighted (local) likelihood

$$\hat{\theta}(x_i) = \arg \max_{\theta} l(Y, W_i; \theta) = \arg \max_{\theta} \sum_{j=1}^n w_{ij} p(Y_j; \theta)$$

or weighted (local) risk minimization, e.g. least squares,

$$\hat{\theta}(x_i) = \arg \min_{\theta} R(Y, W_i; \theta) = \arg \min_{\theta} \sum_{j=1}^n w_{ij} \|Y_j - f(\theta)\|^2$$

where $f : \Theta \mapsto \mathbb{R}^q$ is a suitable function on the parameter space. A formal description of the algorithm is then given as

- Initialization: Set $k = 0$, $W_i^{(0)}$ such that $w_{ij}^{(0)} = \delta_{ij}$, $\hat{\theta}^{(0)}(x_i)$ defined as a weighted likelihood or least squares estimate, $h^{(0)} = 1$.
- Adaptation: $\forall i, j$ define

$$w_{ij}^{(k)} = K_{loc}(l_{ij}^{(k)}) K_{st}(s_{ij}^{(k-1)})$$

- Estimation: $\forall i$ define

$$\hat{\theta}^{(k)}(x_i) = \arg \max_{\theta} l(Y, W_i^{(k)}; \theta) \quad (\text{or} \quad \arg \min_{\theta} R(Y, W_i^{(k)}; \theta))$$

- Iterate: Stop if $k \geq k^*$, else select $h^{(k+1)}$ such that $\sum_j K_{loc}(l_{ij}^{(k+1)}) = c_h \sum_j K_{loc}(l_{ij}^{(k)})$ ($c_h = 1.25$), set $k := k + 1$, and continue with adaptation.

The proposed procedure involves several parameters. The most important one is the scale parameter λ in the statistical penalty s_{ij} . The special case $\lambda = \infty$ simply leads to a kernel estimate with bandwidth $h_{\max} = h^{(k^*)}$. We propose to choose λ as the smallest value satisfying a propagation condition 2 [53]. This condition requires that, if the local assumption is valid globally, i.e. $\theta(x) \equiv \theta$ does not depend on x , then with high probability and for all k the estimate coincides at every point with the nonadaptive estimate. More formally we request that in this case for each iteration k

$$\mathbf{E} \sum_{i=1}^n |\hat{\theta}^{(k)}(x_i) - \check{\theta}^{(k)}(x_i)| < \alpha \mathbf{E} \sum_{i=1}^n |\check{\theta}^{(k)}(x_i) - \theta| \quad (2)$$

for a specified constant $\alpha > 0$. Here

$$\check{\theta}^{(k)}(x_i) = \sum_j K_{loc}(l_{ij}^{(k)}) Y_j / \sum_j K_{loc}(l_{ij}^{(k)})$$

denotes the non-adaptive kernel estimate employing the bandwidth $h^{(k)}$ from step k . The value λ provided by this condition usually does not depend on the unknown model parameter θ and can therefore be found by simulation in a global parametric situation. This enables us to select default values for λ depending on the specified family of the probability distribution $\mathcal{P} = (P_\theta, \theta \in \Theta)$ and the chosen statistics T . Default values for λ in the examples below are selected for a value of $\alpha = 0.2$.

The second parameter of interest is the maximal bandwidth h_{\max} which controls both numerical complexity of the algorithm and smoothness within homogeneous regions.

Additionally we specify a number of parameters and kernel functions that have less influence on the resulting estimates. As a default the kernel functions are chosen as $K_{\text{loc}}(x) = (1 - x^2)_+$ and $K_{\text{st}}(x) = \min(1, 2(1 - x))_+$. If the design is on a grid, e.g. for images, the initial bandwidth $h^{(0)}$ is chosen as the distance between neighboring pixel.

Applications usually require an appropriate description of the statistical model, the structural assumption and a corresponding definition of the statistical penalty s_{ij} . Table 1 provides an overview of currently implemented models and the corresponding software packages for the R environment for statistical computing [58].

3 Image denoising

The algorithm described in the last section is essentially dimension free. It can be easily applied to reconstruct 2D and 3D images. We illustrate this using a 3D-MR image of a head.

We apply adaptive weights smoothing assuming a model with additive Gaussian errors

$$Y_i = \theta(x_i) + \varepsilon_i,$$

to describe the gray value in voxel x_i . The statistical penalty used is $s_{ij} = \frac{N_i}{2\lambda\sigma^2}(\hat{\theta}_i - \hat{\theta}_j)^2$. The error variance σ^2 is estimated from the image. We employ a maximal bandwidth $h_{\max} = 6$. The value of $\lambda = 3.45$ fulfills the propagation condition for $\alpha = 0.1$. Special interest in this example is in detection and/or enhancement of tissue borders. We illustrate the results in Fig. 1. Additionally to the image we provide the results in terms of absolute values of a Laplacian filter which illustrates the gain in edge detection.

Within this example we essentially assumed that the image intensity is locally constant. This assumption may be too rigid and can be replaced, at the cost of sensitivity to disconti-

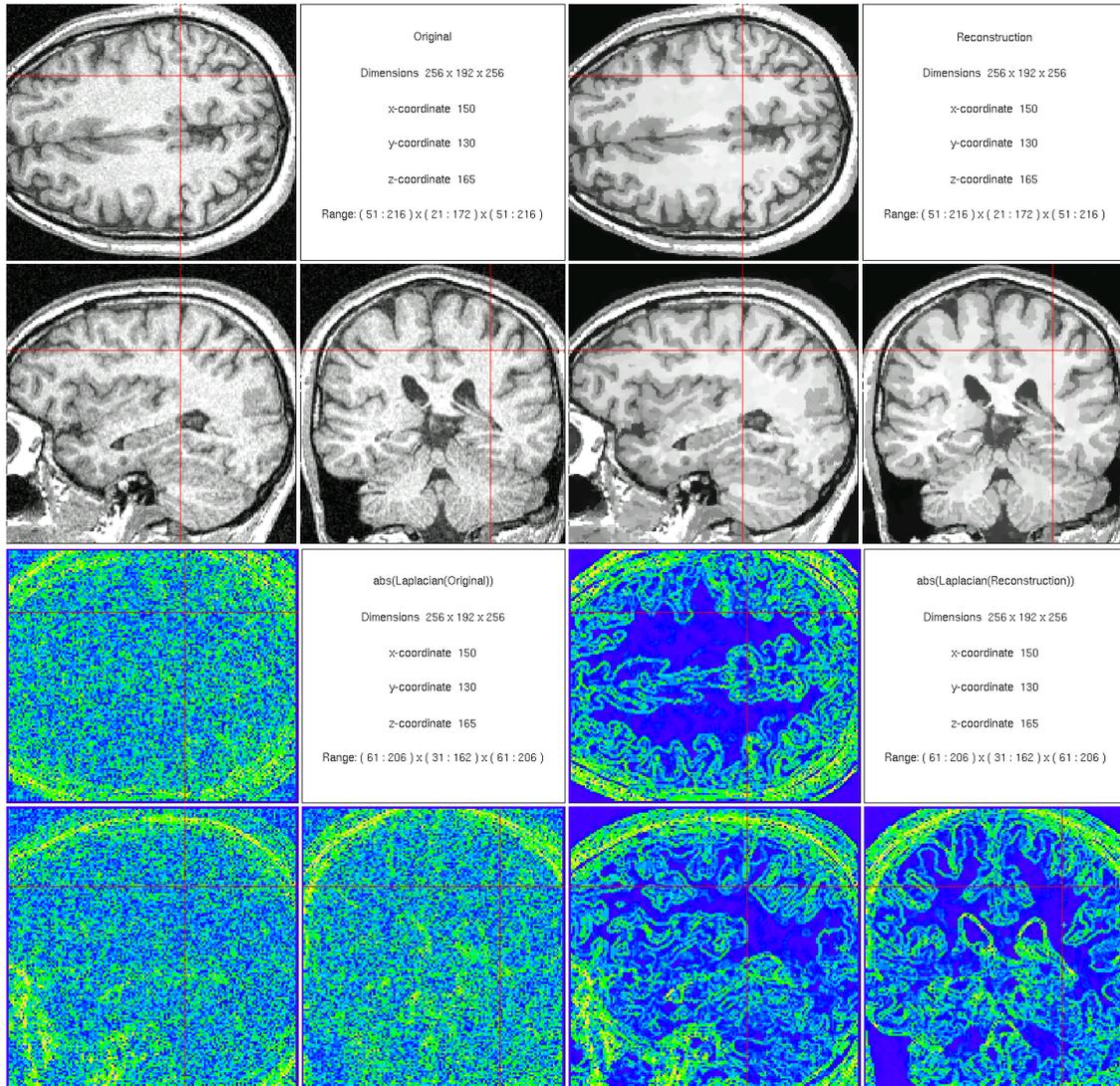


Figure 1: Top: Original (left) and reconstruction (right) of an 3D-MR-image. Bottom: Corresponding absolute values of a Laplacian edge filter.

Table 1: Statistical model, corresponding statistical penalty and R-package.

Model	penalty s_{ij}	R-package
1D-, 2D- and 3D- regression models	$\frac{N_i}{2\lambda\sigma^2}(\hat{\theta}_i - \hat{\theta}_j)^2$	aws / adimpro
1D-, 2D- and 3D- exponential families	$\frac{N_i}{\lambda}\mathcal{K}(\hat{\theta}_j, \hat{\theta}_i)$	aws
1D-, 2D- local polynomial regression	$\frac{N_i}{\lambda}(R(Y, W_i; \hat{\theta}_j) - R(Y, W_i; \hat{\theta}_i))$	aws / adimpro
1D-, 2D- and 3D- Gaussian models with parametric mean-variance model $g(x, \theta, \eta)$ (η -global parameter)	$\frac{N_i}{2\lambda\hat{\sigma}^2(x_i)}(\hat{\theta}_i - \hat{\theta}_j)^2$ $\sigma(x_i) = g(x_i, \theta_i, \eta)$	aws
Color images with constant and linear parametric mean-variance model and spatial correlation	$\frac{N_i}{\lambda C(h, g)}(\hat{\theta}_i - \hat{\theta}_j)^T \hat{\Sigma}_i^{-1}(\hat{\theta}_i - \hat{\theta}_j)$	adimpro
functional MR (smoothing of SPM's)	$\lambda^{-1}(\text{Var } \hat{\theta}_i)^{-1}(\hat{\theta}_i - \hat{\theta}_j)^2$	fmri
Diffusion tensor imaging (DTI)	$\frac{N_i}{\lambda}[R(Y, W_i; \hat{\theta}_j) - R(Y, W_i; \hat{\theta}_i)]$	dti

nities, by assuming the image to consist of locally smooth regions. The Propagation-Separation approach from [53] assumes that within a homogeneous region containing $x_i = (i_h, i_v)$, i.e. for $x_j \in U(x_i)$, the gray value or color Y_{j_h, j_v} can be modelled as

$$Y_{j_h, j_v} = \theta(x_i)^\top \Psi(j_h - i_h, j_v - i_v) + \varepsilon_{j_h, j_v},$$

where the components of $\Psi(\delta_h, \delta_v)$ contain values of basis functions

$$\psi_{m_1, m_2}(\delta_h, \delta_v) = (\delta_h)^{m_1} (\delta_v)^{m_2}$$

for integers $m_1, m_2 \geq 0$, $m_1 + m_2 \leq p$ and some polynomial order p . For a given local model $W(x_i)$ estimates of $\theta(x_i)$ are obtained by local Least Squares as

$$\tilde{\theta}(x_i) = B_i^{-1} \sum_j w_{ij} \Psi(j_h - i_h, j_v - i_v) Y_{j_h, j_v},$$

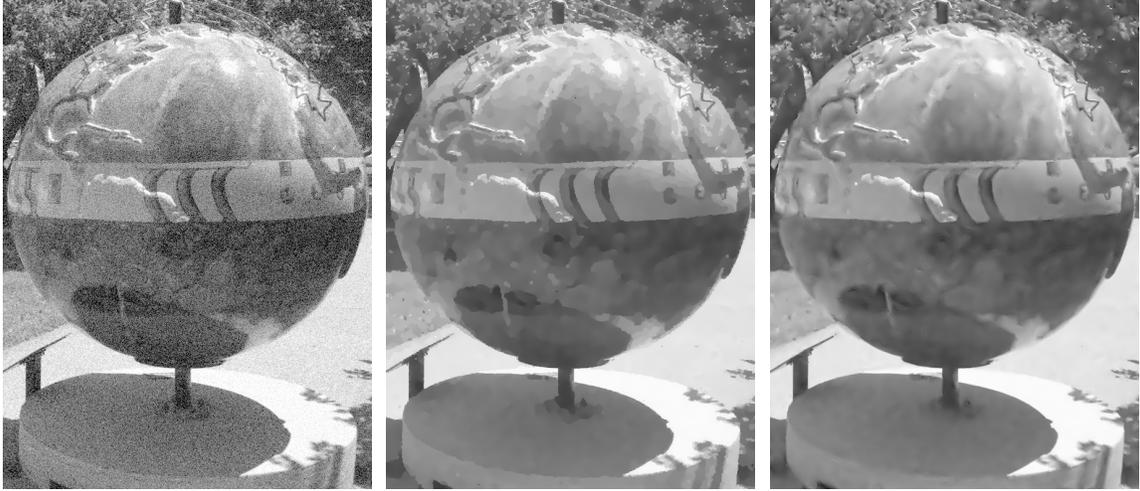


Figure 2: From left to right: noisy image, local constant reconstruction and local quadratic reconstruction ($h_{max} = 12$)

with

$$B_i = \sum_j w_{ij} \Psi(j_h - i_h, j_v - i_v) \Psi(j_h - i_h, j_v - i_v)^\top.$$

The parameters $\theta(x_i)$ are defined with respect to a system of basis functions centered in x_i . Parameter estimates $\hat{\theta}(x_j, x_i)$ employing the local model $W(x_j)$ with basis functions centered at x_i can be obtained by a linear transformation from $\hat{\theta}(x_j)$. In iteration k a statistical penalty can now be defined as

$$s_{ij}^{(k)} = \frac{1}{\lambda 2\sigma^2} (\hat{\theta}^{(k-1)}(x_i) - \hat{\theta}^{(k-1)}(x_j, x_i))^\top B_i (\hat{\theta}^{(k-1)}(x_i) - \hat{\theta}^{(k-1)}(x_j, x_i)).$$

For a more detailed description and discussion of the resulting algorithm see [54].

Figure 2 illustrates results obtained by local constant and a quadratic structural adaptive smoothing for a piecewise smooth image. The local constant reconstruction gives a cartoonlike impression which is due to the use of an, for this image, inappropriate structural assumption.

In digital color images the information in each pixel consists of a vector of three values. Each value is a intensity in one channel of a three dimensional color space, usually the RGB space.

If the image was recorded under bad light conditions, employing a high sensitivity of the sensor, such images can carry a substantial noise. This noise is usually spatially correlated, i.e. colored. Additionally we observe a correlation between the noise components in the three RGB channels.

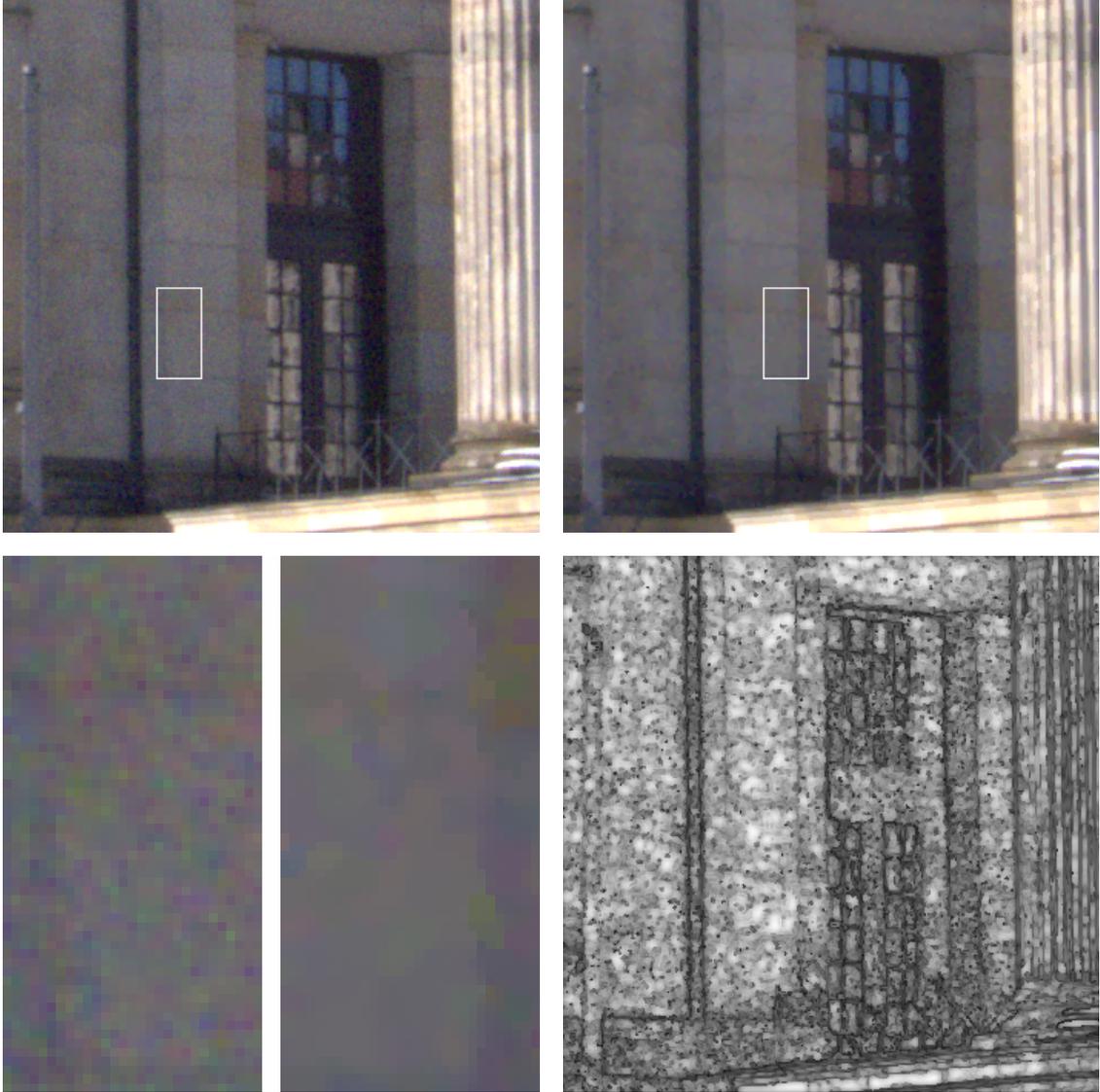


Figure 3: Noisy image (upper left) and reconstruction by the proposed algorithm (bandwidth $h_{max} = 6$, upper right). The bottom row shows details from both images and the sum of weights N_i in each voxel.

An appropriate model to describe such a situation is given by

$$Y_{i_h, i_v} = \theta(x_i) + \varepsilon_{i_h, i_v},$$

where the components of $x_i = (i_h, i_v)$ are the horizontal and vertical image coordinates. Y_{i_h, i_v} , $\theta(x_i)$ and ε_{i_h, i_v} take values in R^3 . The errors follow a distribution with $\mathbf{E}\varepsilon_{i_h, i_v} = 0$, $\text{Var}\varepsilon_{i_h, i_v} = \Sigma$ and $\mathbf{E}\varepsilon_{i_h, i_v}^c \varepsilon_{i_h+1, i_v}^c = \mathbf{E}\varepsilon_{i_h, i_v}^c \varepsilon_{i_h, i_v+1}^c = \rho$ for each color channel c . The covariance matrix Σ may vary with the value of θ_{i_h, i_v} .

Structural adaptive smoothing can be applied in this situation with a statistical penalty

$$s_{ij}^{(k)} = \frac{N_i^{(k-1)}}{2\lambda C(g, h)} (\hat{\theta}_i^{(k-1)} - \hat{\theta}_j^{(k-1)})^\top \Sigma^{-1} (\hat{\theta}_i^{(k-1)} - \hat{\theta}_j^{(k-1)})$$

where $C(g, h)$ is a correction term for spatial correlation, see [55].

Figure 3 illustrates the effect of structural adaptive smoothing for color images. The lower row provides details for the regions marked in the original and the reconstructed image as well as an image of sum of weights N_i that illustrates the adaptivity of the approach.

4 Signal detection in functional MR

Functional Magnetic Resonance Imaging (fMRI) is nowadays a standard tool for in-vivo examination of human brain function with plenty of applications both in research as well as in clinical practice such as diagnosis and treatment of brain lesions. Data obtained in human fMRI consists of time series of three dimensional datasets of the brain. The interscan interval is usually in the order of seconds, while the spatial resolution is commonly in the millimeter range [38, 39] with recent studies entering the submillimeter domain [13, 37, 36]. An interesting fact about fMRI is, that the blood oxygenation serves as a natural contrast making the method non-invasive [46, 47]. This effect is known as the BOLD-effect and can be used for example to localize cognitive functions within the brain. When performing a cognitive task, the MR signal in some voxels is increased due to the higher oxygenation level at the active site. Other voxels remain in their resting state. The increase of the signal can be described by the hemodynamic response function, which has been extensively studied in the past years. The BOLD-effect leads to the creation of various typical experimental designs, mainly block- or event-related. In recent years it has also been proposed that even the resting state pattern of the brain contains valuable information about the working brain and hence has attracted much interest [44].

Many different methods exist to analyze fMRI data depending on the experiment or the focus of the scientific questions, ranging from correlation analysis, ICA, spatio-temporal to fully Bayesian models, see [40] for an overview. Since from the knowledge of the design of an experiment the expected BOLD response is known, the linear model for fMRI data has perhaps become the most widely used approach.

Any analysis method however has to consider the fact, that fMRI data suffers from significant noise. Signal detection in fMRI data inherently involves a severe multiple test problem. In common experiments decisions have to be made at more than 100 000 voxel, leading either to high thresholds (low sensitivity) or a high number of false positives (low

specificity). Analysis methods in fMRI use the fact, that activated areas have a spatial extend of several voxel. Spatial correlation, as introduced by smoothing, significantly reduces the number of independent tests. Under the hypothesis of no activation and in a linear model based analysis spatial smoothing of fMRI data results in statistical parametric maps (SPM) that form random t - or F -fields. Results on excursion sets of random fields [2, 77] can be therefore be used to define suitable thresholds for signal detection, see e.g. [79, 78]. However, common non-adaptive smoothing methods involve a significant loss of information on spatial structure and shape of activated areas. Several algorithms based on different methodology, from noise reduction with anisotropic diffusion processes, Bayesian approaches using spatial priors, region growing methods as established for image segmentation, and others, have been suggested to circumvent this.

Recently we proposed the use of structural adaptive smoothing [66], to avoid the loss of spatial information. The algorithm was developed in the context of the linear model for the BOLD-fMRI data [29, 76] but can easily be translated to other contexts. Our approach for smoothing fMRI data is mainly based on the observation that the structures of interest are defined by areas in which the parameter values corresponding to the BOLD signal are similar and differ significantly from zero. The common non-adaptive filtering approaches smooth the data cube at each time step separately, without making use of the information contained in the time series. We therefore suggested to first evaluate the linear model

$$Y_i = X\beta_i + \varepsilon_i$$

for the time series $Y_i = (Y_{it})_{t=1\dots T}$ at each voxel i . The design matrix X contains the expected BOLD response evaluated at scan acquisition times and nuisance parameters such as a slowly varying drift. After performing some appropriate prewhitening procedure, the error vector $\varepsilon_i = (\varepsilon_{it})_{t=1\dots T}$ can be assumed to have zero expectation and to be approximately uncorrelated in time.

We obtain fields of least squares estimates $\hat{\beta}_i$ for the parameter value β_i and, what is most important, its error variance $\text{Var} \hat{\beta}_i$. Equipped with these the development of a specific structural adaptive smoothing algorithm as outlined in previous sections is canonical. First, we define a structural assumption of spatial homogeneity, which should be valid for the field of the true parameter β_i . In non-activated areas the parameter value is assumed to be zero. This serves as the null hypothesis and allows to again use Random Field Theory [2, 77] for signal detection, see [66]. In areas which are activated during the scan the parameter values differ from zero and are similar, provided that the BOLD %-changes are similar. Hence, our structural assumption is a local constant model for the BOLD-parameter. Activated areas may consist of more than one region with similar parameters. Based on this assumption,

we use an iterative smoothing algorithm for the statistical parametric map (SPM) that is based on pairwise tests of homogeneity. The result is a smoothed SPM where the shape and borders of the activation structure are preserved. As a consequence, in contrast to other non-adaptive smoothing methods, the procedure does reduce noise while preserving the resolution of the scan as required by many modern applications.

The main parameter of our procedure is the maximum achievable variance reduction or equivalently a maximum achievable smoothness. Both can be specified by selecting a maximum bandwidth. Oversmoothing is avoided in the algorithm by construction as long as differences between the parameter values of two homogeneity regions are statistically significant. The largest homogeneous region is expected to be the non-activation area, where parameter values do not significantly differ from zero. Therefore we can choose the maximum bandwidth larger than in non-adaptive smoothing and achieve a larger amount of variance reduction without blurring. This has the effect of lowering the thresholds for signal detection, since the smoothness in non-activation areas, which determine the threshold under the hypothesis of no signal, is directly proportional to the bandwidth.

As statistical penalty we use

$$s_{ij}^{(k)} = \frac{1}{\lambda \text{Var } \hat{\beta}_i^{(k-1)}} (\hat{\beta}_i^{(k-1)} - \hat{\beta}_j^{(k-1)})^2$$

where $\hat{\beta}_i^{(k-1)}$ is the estimated BOLD-parameter from the previous iteration step. Its variance $\text{Var } \hat{\beta}_i^{(k-1)}$ is estimated from the spatially smoothed residuals of the time series. From the final estimates for $k = k^*$ a random t -field $\hat{\beta}_i^{(k^*)} / (\text{Var } \hat{\beta}_i^{(k^*)})^{1/2}$ can be constructed such that again Random Field Theory can be applied for signal detection [66].

We now consider an application of this algorithm and compare it with the signal detection using no smoothing and Gaussian filtering. Experiments were performed on healthy volunteers and approved by the Institutional Review Board of Weill Cornell Medical College. Data was acquired on a 3.0 T General Electric (Milwaukee, WI) Signa Excite MRI scanner, using two-dimensional gradient echo echo planar imaging pulse sequences (GE-EPI) on an eight-channel head receive-only coil. A somatosensory motor task was performed by one male subject. For functional MRI, a GE-EPI sequence with TE/TR = 40/2000 ms was used and 20 axial slices of 4 mm thickness were acquired. We used a field-of-view of 24 cm with a matrix size of 128 × 128, yielding voxel dimensions of 1.88 mm, respectively. A task was performed in three blocks of 60 s duration; each block consisted of 30 s task and 30 s rest. The first 4 scans before these block were discarded, yielding in total 105 scans. The task consisted of bimanual tapping of the thumb against all fingers of the same hand, one by one and in quick succession. In Fig. 4 the effect of smoothing and structural adaptive smoothing in particular is demonstrated. While without smoothing only very few active

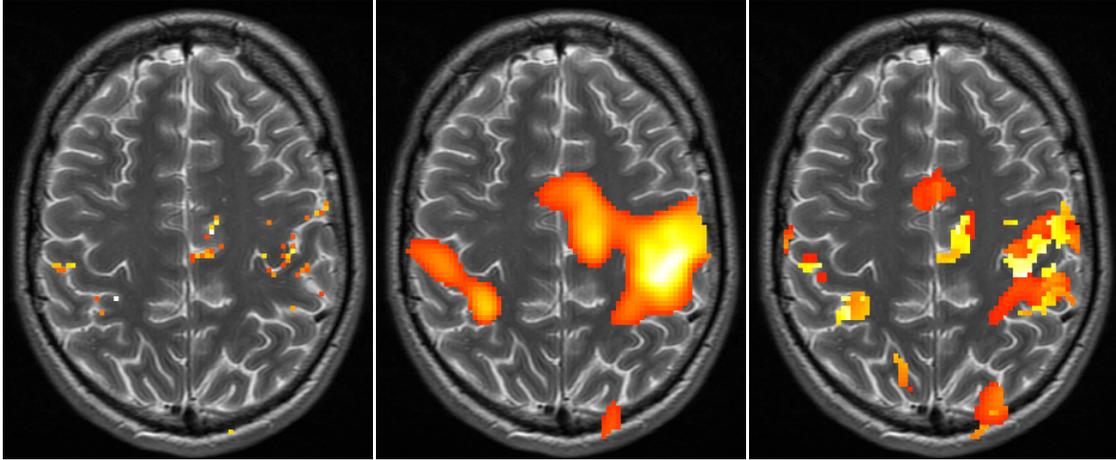


Figure 4: Signal detection in an fMRI experiment: without smoothing (left), with Gaussian filtering (center) and structural adaptive smoothing (right).

voxels can be detected, smoothing in general leads to better detection results. However the significant and inherent blurring with Gaussian filtering can be avoided using structural adaptive smoothing. It has been recently shown, that this procedure is especially helpful in high resolution scans [63]. Structural adaptive smoothing in fMRI is capable to fully use high resolution, to correctly locate activation at tumor borders for pre-surgical planning, and to extract information on spatial structure and shape of the activation areas. For successful applications see e.g. [65, 72, 63].

5 Diffusion Tensor Imaging

Since the early times of nuclear magnetic resonance, it has been known that this phenomenon is sensitive to, and thus can be used to measure, diffusion of molecules in complex systems [11]. The basic principles of magnetic resonance diffusion weighted imaging (DWI) were introduced in the 1980's [8, 45, 67]. Since then, DWI has evolved into a versatile tool for in-vivo examination of tissues in the human brain and spinal cord, leading to a plethora of clinical and neuroscience applications. The broad interest in this technique grows from the fact that DWI probes microscopic structures well beyond typical image resolutions through water molecule displacement, which can be used in particular to characterize the integrity of neuronal tissue in the central nervous system.

Diffusion in neuronal tissue is usually not isotropic but depends on the particular microscopic structure of the tissue. Different diffusion directions can be probed by application of corresponding bipolar magnetic field diffusion gradients [62]. Compared to the non-

diffusion weighted images S_0 the diffusion weighted images $S_{\vec{b}}$ for gradients in direction \vec{b} are exponentially attenuated

$$S_{\vec{b}} = S_0 \exp(-bD(\vec{b})). \quad (3)$$

with the apparent diffusion coefficient $D(\vec{b})$ depending on the tested direction \vec{b} and the “b-value” b depending on the magnetic field gradient parameters. In DTI [4, 5] this information is reduced to a three dimensional Gaussian distribution model for diffusion. Within this model, diffusion is completely characterized by the diffusion tensor \mathbf{D} , a symmetric positive definite 3×3 matrix with six independent components. Eq. 3 thus generalizes to

$$S_{\vec{b}} = S_0 \exp(-b \cdot \vec{b}^T \mathbf{D} \vec{b}).$$

The tensor itself is not invariant against rotations of the observation frame. Hence, only rotationally invariant quantities derived from the tensor contain physically meaningful measures. They are mainly based on the eigenvalues μ_i ($i = 1, 2, 3$) of the tensor \mathbf{D} with $\mu_i > 0$ for positive definite tensors. Mainly used are the trace, corresponding to the mean diffusivity $\langle \mu \rangle$, and the fractional (FA) or geodetic anisotropy (GA) measuring the anisotropy of the tensor:

$$\begin{aligned} \langle \mu \rangle &= \frac{1}{3} \sum_{i=1}^3 \mu_i. \\ FA &= \sqrt{\frac{3}{2}} \sqrt{\frac{\sum_{i=1}^3 (\mu_i - \langle \mu \rangle)^2}{\sum_{i=1}^3 \mu_i^2}} \\ GA &= \sqrt{\frac{\sum_{i=1}^3 (\log(\mu_i) - \frac{1}{3} \sum_{i=1}^3 \log(\mu_i))^2}{\sum_{i=1}^3 \log^2(\mu_i)}} \end{aligned}$$

The diffusion tensor can be visualized as an ellipsoid with the length of main axis corresponding to the eigenvalues and the eigenvectors to the direction in space. Furthermore, the eigenvector for the largest eigenvalue directs in the main fiber direction. Evaluating the components of the vector as three components in a color space like RGB lead to color-coded directional maps with high diagnostic value due to its high contrast for interesting structures.

The diffusion tensor model describes diffusion completely if the microscopic diffusion properties within a voxel are homogeneous. In the presence of partial volume effects, like crossing or bifurcating fibers, the Gaussian model is only an approximation. Such effects are addressed in High Angular Resolution Diffusion Imaging (HARDI) [71, 27], see also [26, 18] within this book. For HARDI more sophisticated models exist, e.g. Q-ball [69, 16], higher order tensors [48], multi-tensor models [70] and tensor distribution functions [34, 43]. In

this chapter, we restrict ourselves to the Gaussian diffusion tensor model for anisotropic diffusion, as used in DTI.

The diffusion tensor can be estimated by non-linear regression minimizing the risk:

$$\mathbf{R}(S_{:,i}, \theta, \mathbf{D}) = \sum_{\vec{b}} \frac{(S_{\vec{b},i} - \theta \exp(-\vec{b} \cdot \vec{b}^\top \mathbf{D} \vec{b}))^2}{\sigma_{\vec{b},i}^2} \quad (4)$$

with respect to the non-diffusion weighted parameter θ and the diffusion tensor \mathbf{D} with the variability $\sigma_{\vec{b},i}^2$ of the diffusion weighted images.

DTI suffers from significant noise which may render subsequent analysis or medical decisions more difficult. This is especially important in low signal-to-noise applications, such as high-resolution DTI or DTI with high b -values [15, 80, 35]. It has been shown that noise may induce a systematically biased assessment of features. For example, a well known phenomenon is the biased estimation of anisotropy indices in the presence of noise [6, 32]. At high noise levels, in addition to the common random errors, the order of the diffusion eigenvectors is subject to a sorting bias. Noise reduction is therefore essential. Several approaches have been proposed for smoothing diffusion tensor data. They include common methods such as Gaussian smoothing [75], anisotropic kernel estimates [41], and methods based on non-linear diffusion [51, 74, 49, 17] or splines [33].

Procedures proposed within this book include coherence enhancing diffusion for matrix fields [10] and tensor regularization methods [31]. Smoothing of tensor data requires to choose a Riemannian [24, 50, 81, 82, 25] or log-Euclidian metric in the tensor space [3, 23]. We see some conceptual advantages in smoothing the diffusion weighted images instead of the tensor estimates. Estimating the tensor by Eq. (4) from noisy data leads, with a certain probability, to results outside the tensor space. This requires some kind of regularization. Reducing the noise level in the diffusion weighted images allows for a reduction of this probability in case of an underlying non-degenerate tensor. In case of high noise level in the diffusion weighted images, both the Rician distribution and the non-linearity of Eq. (3) lead to a bias in the tensor estimate. This bias can be reduced by smoothing the diffusion weighted images, but is not addressed if smoothing is performed in the tensor space itself. A correction for Rician bias [30, 7, 57] can be incorporated.

We therefore developed a structural adaptive smoothing algorithm for DWI data in the context of the diffusion tensor model [64] with extensions to include Rician bias correction and non-linear tensor estimation. Our underlying structural assumption is that for every voxel there is a neighborhood of this voxel in which the diffusion tensor is nearly constant. This assumption reflects the fact that the structures of interest are regions with a homogeneous fractional anisotropy, a homogeneous diffusivity, and a locally constant direction

field. The shape of this neighborhood can be quite different for different voxels and cannot be described by few simple characteristics like bandwidth or principal directions.

The algorithm involves the statistical penalty

$$s_{ij}^{(k)} = \frac{N_i^{(k-1)}}{\lambda} \left[\mathbf{R} \left(\hat{\zeta}_{\cdot,i}^{(k-1)}, \hat{\theta}_{0,j}^{(k-1)}, \hat{\mathbf{D}}_j^{(k-1)} \right) - \mathbf{R} \left(\hat{\zeta}_{\cdot,i}^{(k-1)}, \hat{\theta}_{0,i}^{(k-1)}, \hat{\mathbf{D}}_i^{(k-1)} \right) \right]$$

based on previous estimates for the diffusion tensor and its variability. The corresponding weighting schemes are then directly applied to the diffusion weighted images, from which new estimates for the tensors with lower variability can be estimated.

In contrast to non-linear diffusion methods [81, 82, 50, 10] or non-adaptive smoothing [25, 3, 23] this algorithm takes the variability of the tensor estimates into account and effectively uses the estimated underlying local structure to restrict the averaging process.

In Fig. 5 we demonstrate the effect of this procedure on experimental data. We use a DWI data set [14] made available by the NIH/NCCR Center for Integrative Biomedical Computing, P41-RR12553. This data set contains twelve diffusion weighted volumes and one non-diffusion-weighted ($b = 0$) reference volume. The data has a spatial resolution of 1.5 mm on each axis. The front of the head is at the top of the image. The scan goes from the top of the head down to about the middle of the brain, below the corpus callosum, but above the eyes.

The DTI data was collected on a 3 Tesla MRI scanner in the W.M. Keck Laboratory for Functional Brain Imaging and Behavior by Dr. Andrew Alexander, Departments of Medical Physics and Psychiatry, University of Wisconsin, Madison, funding: NIH RO1 EB002012.

6 Implementations

The structural adaptive smoothing procedures described in this chapter have been mainly implemented as packages for the R environment for statistical computing [58].

Basic algorithms for denoising 1D- 2D and 3D structures using one parameter exponential family models and local polynomial models are implemented in the package *aws*.

The image processing of two dimensional color images has become omnipresent in the past years due to the availability of digital cameras. With the package *adimpro* we provide basic image processing functions, from reading/writing and color space transformations to structural adaptive smoothing.

The package *fmri* provides functions for reading and writing medical imaging formats and

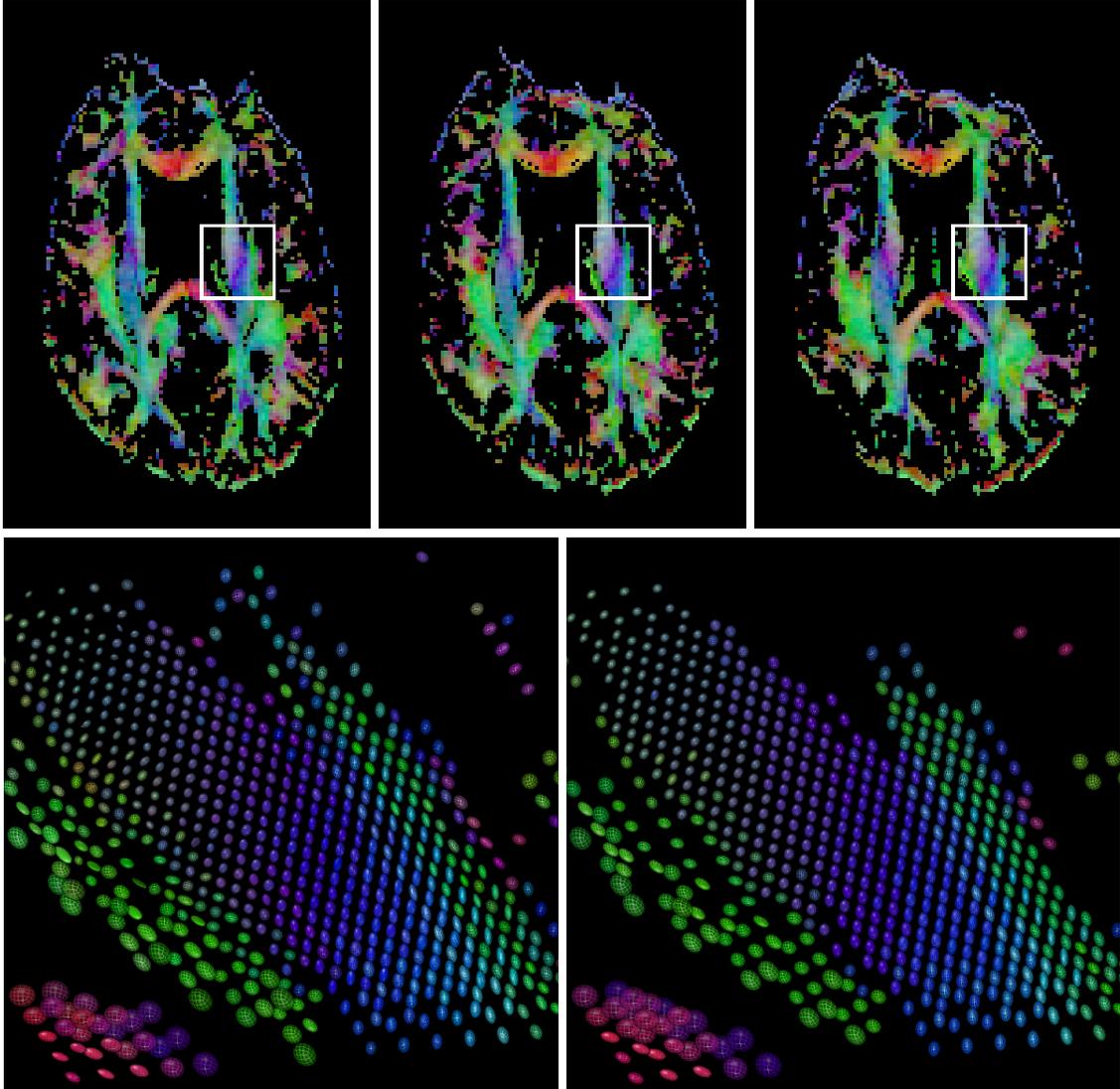


Figure 5: Real DWI data example: The upper row shows the estimated color-coded directional map weighted with FA for the slices 22-24 of the CIBC-dataset [14]. White square marks the extent of the region specified for the lower row. There, the noisy (left) and smoothed (bandwidth 4, right) tensors are shown. The structural adaptive smoothing apparently leads to a homogenization of the regions without blurring the structural borders.

performing an analysis of BOLD-fMRI on the basis of the linear model. This includes the structural adaptive smoothing procedure described in the previous sections as well as signal detection based on Random Field Theory.

The package *dti* implements the structural adaptive smoothing in the context of the diffusion tensor model for diffusion weighted data. Table 2 compares the computing times and

Dimensions	$101 \times 146 \times 38$	$72 \times 100 \times 50$	$146 \times 193 \times 47$
# gradients	13	201	150
# voxel in mask	211235	255257	863134
CPU-time for sdpar	41 s	27 s	126 s
CPU-time for dtiTensor	16 s	118 s	156 s
CPU-time for dti.smooth ($h_{max} = 2$)	220 s	667 s	893 s
CPU-time for dti.smooth ($h_{max} = 4$)	415 s	1256 s	2140 s
CPU-time for dtiIndices	3.1 s	3.7 s	11 s
Mean N_i ($h_{max} = 2/4$)	3.67/16.5	1.20/1.87	1.55/4.32

Table 2: Computing time for three DWI data sets. The first column corresponds to the CIBC-dataset [14], the second to a data set kindly made available by A. Anwander, and a third data set kindly made available by H.U. Voss.

mean sum of weights for different datasets. The package also features extended visualization functions in two and three dimensions based on the R-interface to OpenGL provided by the R-package *rgl* [1].

7 Summary

Structural adaptive smoothing as described in this chapter can be very helpful in many contexts to remove noise without blurring interesting structures. While the main idea of structural adaptation in the Propagation-Separation approach is common for all applications, the resulting algorithms may be quite different and depend on the specific properties of the data and the application.

Acknowledgments

This work is supported by the DFG Research Center MATHEON. It was made possible in part by software and a dataset from the NIH/NCRR Center for Integrative Biomedical Computing, P41-RR12553. The authors would like to thank A. Anwander at the Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig, Germany) and H.U. Voss at the Citigroup Biomedical Imaging Center, Weill Cornell Medical College for providing diffusion-weighted and functional MR datasets. Furthermore, the authors would like to thank H.U. Voss for numerous intense and helpful discussions on Magnetic Resonance Imaging and related issues.

References

- [1] D. Adler, O. Nenadic, and W. Zucchini. Rgl: An r-library for 3d visualization with opengl. In *Proceedings of the 35th Symposium of the Interface: Computing Science and Statistics*, Salt Lake City, 2003.
- [2] R.J. Adler. On excursion sets, tube formulae, and maxima of random fields, (special invited paper). *Annals of Applied Probability*, 10:1–74, 2000.
- [3] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Res. Med.*, 56(2):411–421, August 2006. PMID: 16788917.
- [4] P.J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance*, B(103):247–254, 1994.
- [5] P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
- [6] P.J. Basser and S. Pajevic. Statistical artefacts in diffusion tensor MRI (DT-MRI) caused by background noise. *Magnetic Resonance Medicine*, 44(1):41–50, 2000.
- [7] Saurav Basu, Thomas Fletcher, and Ross Whitaker. Rician noise removal in diffusion tensor MRI. In *Medical image computing and computer-assisted intervention*, LNCS 4190, pages 117–125. MICCAI, 2006.
- [8] D. Le Bihan and E. Breton. Imagerie de diffusion in vivo par résonance magnétique nucléaire. *CR Acad Sci Paris*, 301:1109–1112, 1985.
- [9] A.W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford University Press, Oxford, 1997.
- [10] Bernhard Burgeth, Luis Pizarro, Stephan Didas, and Joachim Weickert. *Locally Adaptive Filtering in Signal and Image Processing*, chapter Coherence-Enhancing Diffusion for Matrix Fields. Springer, 2009.
- [11] H. Y. Carr and E. M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94:630–638, 1954.
- [12] Probal Chaudhuri and James Steven Marron. Scale space view of curve estimation. *Annals of Statistics*, 28:408–428, 2000.

- [13] K. Cheng, R.A. Waggoner, and K. Tanaka. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron*, 32:359–374, 2001.
- [14] CIBC. Data sets: Ncrr center for integrative biomedical computing (cibc) data set archive., 2008.
- [15] Chris A. Clark and Denis Le Bihan. Water diffusion compartmentation and anisotropy at high b values in the human brain. *Magnetic Resonance in Medicine*, 44:852–859, 2000.
- [16] Maxime Descoteaux, Elaine Angelino, Shaun Fitzgibbons, and Rachid Deriche. Regularized, fast and robust analytical q-ball imaging. *MRM*, 58:497–512, 2007.
- [17] Z. Ding, J.C. Gore, and A.W. Anderson. Reduction of noise in diffusion tensor images using anisotropic smoothing. *Magnetic Resonance Medicine*, 53(2):485–490, 2005.
- [18] Remco Duits and Eric Franken. *Locally Adaptive Filtering in Signal and Image Processing*, chapter Left-invariant Diffusions on SE(3) and their application to crossing-preserving smoothing of HARDI-images. Springer, 2009.
- [19] Remco Duits, Hartmut Fuehr, and Bart Janssen. *Locally Adaptive Filtering in Signal and Image Processing*, chapter Left Invariant Evolutions on Gabor Transforms. Springer, 2009.
- [20] J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications*. Chapman & Hall, London, 1996.
- [21] M. Felsberg, P.-E. Forssén, and H. Schar. Channel smoothing: Efficient robust smoothing of low-level signal features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:209–222, 2006.
- [22] Michael Felsberg. *Locally Adaptive Filtering in Signal and Image Processing*, chapter Adaptive Filtering using Channel Representations. Springer, 2009.
- [23] Pierre Fillard, Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Clinical DT-MRI estimation, smoothing and fiber tracking with Log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, 26:1472–1482, 2007.
- [24] P. Thomas Fletcher. *Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI*. PhD thesis, University of North Carolina at Chapel Hill, 2004.
- [25] Thomas Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87:250–262, 2007.

- [26] Luc Florack. *Locally Adaptive Filtering in Signal and Image Processing*, chapter Scale Space Representations Locally Adapted to the Geometry of Base and Target Manifold. Springer, 2009.
- [27] L.R. Frank. Anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine*, 45:935–939, 2001.
- [28] Erik M. Franken. *Enhancement of Crossing Elongated Structure in Images*. PhD thesis, Eindhoven University of Technology, Dept. of Biomedical Engineering, 2008.
- [29] K.J. Friston, A.P. Holmes, K. Worsley, J-B. Poline, C.D. Frith, and R.S.J Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.
- [30] Hakon Gudbjartsson and Samuel Patz. The Rician distribution of noisy MRI data. *Magn. Res. Med.*, 34:910–914, 1995.
- [31] Yaniv Gur, Ofer Pasternak, and Nir Sochen. *Locally Adaptive Filtering in Signal and Image Processing*, chapter Fast $GL(n)$ -Invariant framework for Tensors Regularization. Springer, 2009.
- [32] K. Hahn, S. Prigarin, S. Heim, and K. Hasan. Random noise in diffusion tensor imaging, its destructive impact and some corrections. In J. Weickert and H. Hagen, editors, *Visualization and Image Processing of Tensor Fields*, Mathematics and Visualization. Springer, Berlin, 2006.
- [33] S. Heim, L. Fahrmeir, P.H.C. Eilers, and B.D. Marx. 3D space-varying coefficient models with application to diffusion tensor imaging. *Computational Statistics and Data Analysis*, 51:6212–6228, 2007.
- [34] Bing Jian, Baba C. Vemuri, Evren Ozarslan, Paul R. Carney, and Thomas H. Mareci. A novel tensor distribution model for the diffusion-weighted MR signal. *NeuroImage*, 37:164–176, 2007.
- [35] D. K. Jones and P. J. Basser. "Squashing peanuts and smashing pumpkins": How noise distorts diffusion-weighted MR data. *Magn. Res. Med.*, 52:979–993, 2004.
- [36] Andreas Kleinschmidt. Different analysis solutions for different spatial resolutions? moving towards a mesoscopic mapping of functional architecture in the human brain. *NeuroImage*, 38:663–665, 2007.
- [37] Nikolaus Kriegeskorte and Peter Bandettini. Analyzing for information, not activation, to exploit high-resolution fmri. *NeuroImage*, 38:649–662, 2007.

- [38] N. Lange. Statistical approaches to human brain mapping by functional magnetic resonance imaging. *Stat. in Medicine*, 15:389–428, 1996.
- [39] N. Lange and S. L. Zeger. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. 46:1–29, 1997.
- [40] Nicole A. Lazar. *The Statistical Analysis of Functional MRI Data*. Statistics for Biology and Health. Springer, 2008.
- [41] J.H. Lee, M.K. Chung, T.E. Oakes, and A.L. Alexander. Anisotropic gaussian kernel smoothing of DTI data. In *Proceedings of the 13th Annual Meeting of ISMRM*, page 2253, 2005.
- [42] J.S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 165–168, 1980.
- [43] A. Leow, S. Zhu, K. McMahon, G. de Zubicaray, M. Meredith, M. Wright, and P. Thompson. The tensor distribution function. *Magnetic Resonance in Medicine*, 61:205–214, 2009. University of California, Los Angeles, CA, USA.
- [44] D. Mantini, M. G. Perrucci, C. Del Gratta, G. L. Romani, and M. Corbetta. Electrophysiological signatures of resting state networks in the human brain. *PNAS*, 104(32):13170–13175, 2007.
- [45] K.D. Merboldt, W. Hanicke, and J. Frahm. Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance*, 64:479–486, 1985.
- [46] S. Ogawa, T.M. Lee, A.R. Kay, and D.W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA*, 87:9868–9872, 1990.
- [47] S Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA*, 89:5951–5955, 1992.
- [48] Evren Özarslan and Thomas H. Mareci. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magnetic Resonance in Medicine*, 50:955–965, 2003.
- [49] G.J. Parker, J.A. Schnabel, M.R. Symms, D.J. Werring, and G.J. Barker. Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 11:702–710, 2000.

- [50] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *Intern. J. Computer Vision*, 66(1):41–66, January 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.
- [51] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Pattern Anal Machine Intell*, 12:629–639, 1990.
- [52] J. Polzehl and V. Spokoiny. Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society B*, 62:335–354, 2000.
- [53] J. Polzehl and V. Spokoiny. Propagation-separation approach for local likelihood estimation. *Probab. Theory and Relat. Fields*, 135:335–362, 2006.
- [54] J. Polzehl and V. Spokoiny. *Handbook of Data Visualization*, chapter Structural adaptive smoothing by propagation-separation methods, pages 471–492. Series Handbooks of Computational Statistics. Springer-Verlag, 2008.
- [55] J. Polzehl and K. Tabelow. Adaptive smoothing of digital images: The R package adimpro. *J. Statist. Software*, 19:1–17, 2007.
- [56] J. Polzehl and K. Tabelow. fmri: A package for analyzing fmri data. *RNews*, 7(2):13–17, 2007.
- [57] J. Polzehl and K. Tabelow. Structure adaptive smoothing diffusion tensor imaging data: the R package dti. Preprint 1382, WIAS, 2008.
- [58] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0.
- [59] Yves Rozenholc, Markus Reiss, and Charles A. Cuenod. *Locally Adaptive Filtering in Signal and Image Processing*, chapter Preserving time structures while denoising dynamical image Application to Dynamic Contrast Enhanced-CT. Springer, 2009.
- [60] Hanno Schar and Kai Krajsek. *Locally Adaptive Filtering in Signal and Image Processing*, chapter A Short Introduction to Diffusion-like Methods. Springer, 2009.
- [61] J. Simonoff. *Smoothing Methods in Statistics*. Springer, New York, 1996.
- [62] E.O. Stejskal and J.E. Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chem Phys*, 42:288–292, 1965.
- [63] K. Tabelow, V. Piëch, J. Polzehl, and H. U. Voss. High-resolution fmri: Overcoming the signal-to-noise problem. *Journal of Neuroscience Methods*, 178:357–365, 2009.

- [64] K. Tabelow, J. Polzehl, V. Spokoiny, and H.U. Voss. Diffusion tensor imaging: Structural adaptive smoothing. *NeuroImage*, 39:1763–1773, 2008.
- [65] K. Tabelow, J. Polzehl, A.M. Ulug, J.P. Dyke, L.A. Heier, and H.U. Voss. Accurate localization of functional brain activity using structure adaptive smoothing. *IEEE Trans. Med. Imaging*, 27(4):531–537, 2008.
- [66] K. Tabelow, J. Polzehl, H.U. Voss, and V. Spokoiny. Analyzing fMRI experiments with structural adaptive smoothing procedures. *NeuroImage*, 33(1):55–62, 2006.
- [67] D. G. Taylor and M. C. Bushell. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Phys. Med. Biol.*, 30:345–349, 1985.
- [68] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India.*, pages 839–846, 1998.
- [69] David S. Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52:1358–1372, 2004.
- [70] David S. Tuch, Timothy G. Reese, Mette R. Wiegell, and Van J. Wedeen. Diffusion MRI of complex neural architecture. *Neuron*, 40:885–895, 2003.
- [71] D.S. Tuch, R.M. Weisskoff, J.W. Belliveau, and V.J. Wedeen. High angular resolution diffusion imaging of the human brain. In *Proceedings of the 7th Annual Meeting of ISMRM*, page 321, 1999.
- [72] H.U. Voss, K. Tabelow, J. Polzehl, O. Tchernichovski, K.K. Maul, D. Salgado-Commissariat, D. Ballon, and S.A. Helekar. Functional MRI of the zebra finch brain during song stimulation suggests a lateralized response topography. *PNAS*, 104:10667–10672, 2007.
- [73] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall, London, 1995.
- [74] J. Weickert. *Anisotropic Diffusion in Image Processing*. ECMI. Teubner-Verlag, Stuttgart, 1998.
- [75] C.-F. Westin, S. E. Maier, B. Khidhir, P. Everett, F. A. Jolesz, and R. Kikinis. Image processing for diffusion tensor magnetic resonance imaging. In *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science, pages 441–452, September 19–22 1999.
- [76] K. Worsley and K.J. Friston. Analysis of fMRI time series revisited - again. *NeuroImage*, 2:173–181, 1995.

- [77] K.J. Worsley. Local maxima and the expected Euler characteristic of excursion sets of χ^2 , f and t fields. *Advances in Applied Probability*, 26:13–42, 1994.
- [78] K.J. Worsley. Detecting activation in fMRI data. *Statistical Methods in Medical Research*, 12:401–418, 2003.
- [79] K.J. Worsley, C. Liao, J. A. D. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A general statistical analysis for fMRI data. *NeuroImage*, 15:1–15, 2002.
- [80] Takashi Yoshiura, Futoshi Mihara, Atsuo Tanaka, Koji Ogomori, Yasumasa Ohyagi and Takayuki Taniwaki, Takeshi Yamada and Takao Yamasaki, Atsushi Ichimiya, Naoko Kinukawa, Yasuo Kuwabara, and Hiroshi Honda. High b value diffusion-weighted imaging is more sensitive to white matter degeneration in Alzheimer’s disease. *NeuroImage*, 20:413–419, 2003.
- [81] Fan Zhang and Edwin R. Hancock. Riemannian graph diffusion for dt-mri regularization. *MICCAI*, 2006.
- [82] Fan Zhang and Edwin R. Hancock. Smoothing tensor-valued images using anisotropic geodesic diffusion. In D.-Y. Yeung, J.T. Kwok, F. Roli, A. Fred, and D. de Ridder, editors, *Advances in Statistical, Structural and Syntactical Pattern Recognition*, volume 4109 of *Lecture Notes in Computer Science*, pages 83–91. Springer, 2006.