

**Weierstraß-Institut**  
**für Angewandte Analysis und Stochastik**  
**Leibniz-Institut im Forschungsverbund Berlin e. V.**

Preprint

ISSN 0946 – 8633

**Adaptive smoothing of multi-shell diffusion-weighted  
magnetic resonance data by msPOAS**

Saskia Becker<sup>1</sup>, Karsten Tabelow<sup>1</sup>, Siawoosh Mohammadi<sup>2</sup>, Nikolaus Weiskopf<sup>2</sup>, Jörg Polzehl<sup>1</sup>

submitted: July 1, 2013

<sup>1</sup> Weierstrass Institute  
Mohrenstr. 39  
10117 Berlin  
Germany

E-Mail: saskia.becker@wias-berlin.de  
karsten.tabelow@wias-berlin.de  
joerg.polzehl@wias-berlin.de

<sup>2</sup> Wellcome Trust Centre for Neuroimaging  
UCL Institute of Neurology  
12 Queen Square  
London WC1N 3BG  
United Kingdom

E-Mail: siawoosh.mohammadi@ucl.ac.uk  
n.weiskopf@ucl.ac.uk

No. 1809  
Berlin 2013



---

2010 *Mathematics Subject Classification.* 62P10, 62G05 .

*Key words and phrases.* Diffusion weighted magnetic resonance imaging, POAS, Structural adaptive smoothing, Parameter choice, Multi-shell.

This work was partially supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin and the Stiftung der Deutschen Wirtschaft (SDW) .

Edited by  
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)  
Leibniz-Institut im Forschungsverbund Berlin e. V.  
Mohrenstraße 39  
10117 Berlin  
Germany

Fax: +49 30 20372-303  
E-Mail: [preprint@wias-berlin.de](mailto:preprint@wias-berlin.de)  
World Wide Web: <http://www.wias-berlin.de/>

**ABSTRACT.** In this article we present a noise reduction method (msPOAS) for multi-shell diffusion-weighted magnetic resonance data. To our knowledge, this is the first smoothing method which allows simultaneous smoothing of all  $q$ -shells. It is applied directly to the diffusion weighted data and consequently allows subsequent analysis by any model. Due to its adaptivity, the procedure avoids blurring of the inherent structures and preserves discontinuities. MsPOAS extends the recently developed position-orientation adaptive smoothing (POAS) procedure to multi-shell experiments. At the same time it considerably simplifies and accelerates the calculations. The behavior of the algorithm msPOAS is evaluated on diffusion-weighted data measured on a single shell and on multiple shells.

## 1. INTRODUCTION

Diffusion-weighted magnetic resonance imaging (dMRI) is a versatile tool for in-vivo imaging of anisotropic tissue structure, especially, but not exclusively in the human brain, see Mori [2007] or Jones [2010] for an introduction. The diffusion-weighted contrast enables various types of analysis to characterize the brain structure for the normal brain, for a broad range of conditions affecting the brain, or for developmental studies [Johansen-Berg and Behrens, 2009]. This is often done by assigning a local anisotropy measure or by determining the main diffusion direction to produce fiber tracks.

dMRI data consists of a series of 3D image volumes acquired by applying diffusion weighting magnetic field gradients in various directions. Depending on the experiment varying gradient strengths or diffusion time, which determine the  $b$ -value of the measurement, are used [Callaghan, 1991]. The diffusion profile obtained from the diffusion images reveals information on the intra-voxel structure, see, e.g., Mitra and Sen [1992]. This explains the broad interest in dMRI as it enables measuring tissue properties that exist at the micron level, whereas the voxel size of a dMRI measurement is only at millimeter level. A wide range of models for the diffusion profile have been developed, such as the diffusion tensor model (DTI) in [Basser et al., 1994a,b], tensor mixture models [Behrens et al., 2003, Assaf and Basser, 2005, Tabelow et al., 2012], the orientation distribution function [Tuch, 2004] and higher order tensor models [Özarslan and Mareci, 2003, Liu et al., 2003], to name only a small selection. Some models like DTI can be evaluated based on measurements on a single  $q$ -shell, i.e., for a single  $b$ -value, others like Diffusion Kurtosis Imaging [Jensen et al., 2005, Tabesh et al., 2011], or methods to estimate the full Diffusion Propagator or its radial part [Özarslan et al., 2006, Aganj et al., 2010, Cheng et al., 2010, Descoteaux et al., 2011] require multi-shell data. It is now generally accepted that models beyond the diffusion tensor should be used to explore white matter microstructure [Jones et al., 2012]. In this context models that require multi-shell dMRI data acquisitions are becoming increasingly popular.

As all imaging modalities dMRI suffers from noise. Due to the signal attenuation the signal-to-noise ratio inherently decreases with increasing  $b$ -value for the diffusion weighted images. Thus, noise hampers modeling for dMRI data in general, but for multi-shell data in particular. Furthermore, high spatial resolution is believed to improve resolution of complex fiber bundles [Heidemann et al., 2010, Zhan et al., 2012, Kleinnijenhuis et al., 2012, Kamali et al., 2013]. As the increase in spatial resolution also reduces the signal-to-noise this further deteriorates of the image quality. In order to reduce noise in dMRI data a number of different approaches have been developed starting from Gaussian filtering [Westin et al., 1999], smoothing procedures in tensor space for DTI [Fletcher, 2004], denoising algorithms based on the partial differential equations [Ding et al., 2005, Parker et al., 2000, Duits and Franken, 2011], and many others.

Recently, we developed a position-orientation adaptive smoothing (POAS) algorithm [Becker et al., 2012] based on the Propagation-Separation approach [Polzehl and Spokoiny, 2006, Becker and Mathé, 2013]. The method directly smoothes diffusion weighted images measured on a single  $q$ -shell: It is applied to the dMRI data prior to any modeling. Hence, it does not introduce a model-specific bias into the data and any model can be used after smoothing. The algorithm is suitable even for very low signal-to-noise-ratio due to the gained efficiency by considering the geometric properties of the measurement space:

Each signal value in a diffusion weighted image is associated to its position in voxel space and the (gradient) orientation. Here,  $\mathbb{R}^3$  stands for the 3D voxel position space and  $\mathbb{S}^2$  for the unit sphere, where the diffusion gradient orientations are distributed. The measurement space can thus be described by a combined space  $\mathbb{R}^3 \times \mathbb{S}^2$ . In [Becker et al., 2012], POAS has been proven to be able to reduce noise in dMRI data without blurring the structural borders in the images which constitutes its adaptive properties.

This article aims to extend and substantiate the original POAS proposal in several very important directions. We will start with the consideration of *multi-shell data* as required by higher order models. Basically, the method POAS could be applied separately for each  $b$ -value. However, this is not efficient as we would ignore helpful information for adaptation by considering only data from one shell. We will illustrate this in the paper. We present an algorithm allowing for simultaneous smoothing of all  $q$ -shells. We denote this generalization multi-shell POAS (msPOAS). To our best knowledge this is the first smoothing algorithm for multi-shell dMRI data, that uses the geometry of the measurement space and a vector structure of the data from the different shells. Second, we will make significant simplifications to the algorithm in comparison to the single-shell method to accelerate it for practical purposes and enable easier theoretical accessibility. Third, we will present some heuristics which suggest the generalizability of the theoretical properties in Becker and Mathé [2013] to msPOAS. A theoretical proof is hampered by the lack of an explicit formula for the Kullback-Leibler divergence in case of dMRI data. The corresponding simulations show that msPOAS behaves as expected in typical situations. Finally, we evaluate msPOAS considering real multi-shell and single-shell dMRI data and give recommendations how to use the method.

## 2. THEORY

When performing a dMRI-scan, we do not directly observe the image, which we are interested in. The scanner yields complex-valued data in k-space, which relates to the signal attenuation due to water diffusion. For Cartesian acquisitions the data is transformed via Fourier transformation to a diffusion weighted image [Callaghan, 1991]. In case of parallel imaging with an NMR phased array of  $L$  receiver coils [Roemer et al., 1990], this image is reconstructed from the data of all coils. The details depend on the acquisition method among which the dominant ones are SENSE [Pruessmann et al., 1999] and GRAPPA [Griswold et al., 2002]. The complex signal in image space is typically transformed to a real positive number by considering a magnitude image and thus neglecting the phase component. The data is generally pre-processed to compensate for artifacts due to motion [Storey et al., 2007, Mohammadi et al., 2013], magnetic field inhomogeneities [Andersson et al., 2003, Mohammadi et al., 2012b, Ruthotto et al., 2012], eddy currents [Jezzard et al., 1998, Andersson and Skare, 2002, Mohammadi et al., 2010] or noise [Kristoffersen, 2012, Mohammadi et al., 2012a]. The corresponding methods are applied at different steps of the processing pipeline. Our method for multi-shell position-orientation adaptive smoothing will be directly applied to the reconstructed diffusion weighted images to improve and stabilize the subsequent modeling and the analysis of the data. Ideally, it should be executed after all other pre-processing steps have been done.

After a short summary of the original method POAS [Becker et al., 2012], we discuss necessary extensions for the generalization of this algorithm to multi-shell dMRI data. First, we derive a specific perspective on dMRI multi-shell data based on the measurement process that allows a description by a quite sophisticated function  $S$ , see Equations (8) and (10). Second, we reconstruct missing data for this function by spherical interpolation. Third, we study the probability distribution of the measurement error. Then, our new algorithm for msPOAS is presented in Section 2.4. Afterwards, we discuss further modifications that improve, simplify and accelerate the method. Specifically, we introduce an advanced parameter choice strategy for the crucial parameter of the algorithm and a simplified discrepancy for the measurement space that relates to the associated geodesics and exhibits the properties of a pseudo-metric. Additionally, we justify a simplified approximation for the statistical penalty which determines the

adaptivity of the procedure. In the first instance some of the parameters are introduced only on a superficial level to improve readability. For details we refer the reader to their specific discussion in Section 2.4 and to Becker et al. [2012].

**2.1. Background: Position-orientation adaptive smoothing (POAS).** In this section, we summarize the main ideas of the original method POAS [Becker et al., 2012]. This takes advantage of the following interpretation of the measurement space.

In diffusion weighted imaging, one measures integral values on a regular grid of voxel  $\vec{v} \in V \subseteq \mathbb{R}^3$  to infer on the structure of the analyzed tissue from the properties of the diffusion of water therein. Measurements may be performed for a single or for varying  $b$ -values ( $b > 0$ ). For each  $b$ -value, data is acquired with varying diffusion gradients  $\vec{g} \in G$ . These can be identified with elements of the 2-sphere such that  $G \subseteq \mathbb{S}^2 := \{\vec{g} \in \mathbb{R}^3 : \|\vec{g}\| = 1\}$ . The method POAS is limited to dMRI-measurements with a single  $b$ -value allowing a description of the measured signal by the following function,

$$(1) \quad S : V \times G \ni (\vec{v}, \vec{g}) \mapsto S(\vec{v}, \vec{g}) \in \mathbb{R}, \quad V \times G \subseteq \mathbb{R}^3 \times \mathbb{S}^2.$$

Here, the signal depends on the voxel position  $\vec{v} \in V$  and the diffusion gradient direction  $\vec{g} \in G$ . A generalized version for multiple  $b$ -values is given in Section 2.2. Additionally, the full dataset contains at least one non-diffusion weighted image  $S_0$  that is acquired without applying any magnetic field gradient. In case of several  $S_0$ -images, we consider the corresponding mean image  $\hat{S}_0$ . However, for better readability we will use  $S_0$  instead of  $\hat{S}_0$  for notation. The signal attenuation  $S(\vec{v}, \vec{g})/S_0(\vec{v})$  can then be related to the diffusion properties.

POAS is based on the Propagation-Separation approach for pointwise estimation of piecewise smooth functions [Polzehl and Spokoiny, 2006]. Here, we assume the simplest version, i.e., a local constant model, to accelerate and stabilize the estimation procedure. In particular, we assume that the measurement space can be partitioned into regions with almost constant values separated by discontinuities. This assumption seems to be reasonable for dMRI data, see Becker et al. [2012]. No assumption is made on the size and shape of the homogeneous regions. In comparison, the application of the common Gaussian filter to the diffusion-weighted images actually relies on the assumption of a globally smooth image intensity value. The obvious violation of this assumption in dMRI data manifests itself as the blurring effect at borders. In Sections 2.4 and 5, we discuss possible consequences of a violated structural assumption for POAS and msPOAS.

The Propagation-Separation approach is an iterative procedure that uses a weighted mean of the observations to infer on the true function value via the expectation of the respective observation. The adaptive weights depend on a location and on an adaptation kernel, denoted by  $K_{\text{loc}}$  and  $K_{\text{ad}}$ . The former determines the local neighborhood under consideration and the latter tests for local homogeneity. Specifically, it uses the information of the previous iteration step to improve the new estimation by restricting the estimator as far as possible to observations of the same homogeneous region. This avoids blurring of the observed structure in dMRI data. During iteration the local neighborhood increases while the adaptation becomes more restrictive.

In POAS, the location kernel determines a local neighborhood of the whole measurement space of dMRI,  $V \times G \subseteq \mathbb{R}^3 \times \mathbb{S}^2$ , using the discrepancy

$$(2) \quad \Delta_\kappa(m_1, m_2) := \left\| \left( \mathbf{R}_{\vec{g}_2}^{-1}(\vec{v}_1 - \vec{v}_2), \mathbf{R}_{\vec{g}_2}^{-1} \mathbf{R}_{\vec{g}_1} \right) \right\|_{R, \kappa}, \quad m_1, m_2 \in V \times G, \quad m_i = (\vec{v}_i, \vec{g}_i),$$

where  $\|\cdot\|_{R, \kappa}$  denotes the  $\kappa$ -weighted Riemannian 2-norm on the special Euclidean motion group  $\text{SE}(3)$  and  $\mathbf{R}_{\vec{g}_i} \in \text{SO}(3)$  is any rotation with  $\mathbf{R}_{\vec{g}_i} \mathbf{e}_z = \vec{g}_i$ , see Becker et al. [2012, Section 2.2 and Appendix A.3]. For a given sequence of increasing bandwidths  $\{h^{(k)}\}_{k=0}^{k^*}$  with  $h^{(0)} > 0$ , this allows to define the

following non-adaptive estimator

$$(3) \quad \bar{S}^{(k)}(m) = \sqrt{\sum_{n \in V \times G} \bar{w}_{mn}^{(k)} S^2(n) / \bar{N}_m^{(k)}}$$

with non-adaptive weights

$$(4) \quad \bar{w}_{mn}^{(k)} := K_{\text{loc}} \left( \Delta_{\kappa}(m, n) / h^{(k)} \right) \quad \text{and} \quad \bar{N}_m^{(k)} := \sum_{n \in V \times G} \bar{w}_{mn}^{(k)}.$$

Here, we consider a weighted quadratic mean instead of a weighted arithmetic mean to ensure that the estimator remains in the same class of probability distributions as the observations, which we assumed to be non-central  $\chi$ -distributed. If this distribution belongs to an exponential family this estimator is related to local maximum likelihood estimation, see Polzehl and Spokoyny [2006]. The non-adaptive estimator for  $k = 0$  with a sufficiently small location bandwidth  $h^{(0)}$  is used to initialize POAS.

A slight modification of the weights leads to the adaptive estimator

$$(5) \quad \tilde{S}^{(k)}(m) = \sqrt{\sum_{n \in V \times G} \tilde{w}_{mn}^{(k)} S^2(n) / \tilde{N}_m^{(k)}},$$

as it is used in POAS. Here, we set

$$(6) \quad \tilde{w}_{mn}^{(k)} := \bar{w}_{mn}^{(k)} \cdot K_{\text{ad}} \left( s_{mn}^{(k)} / \lambda \right) \quad \text{and} \quad \tilde{N}_m^{(k)} := \sum_{n \in V \times G} \tilde{w}_{mn}^{(k)},$$

where  $\lambda > 0$  denotes the adaptation bandwidth and the statistical penalty

$$(7) \quad s_{mn}^{(k)} := \tilde{N}_m^{(k-1)} \mathcal{KL} \left( \mathbb{P}_{\tilde{S}^{(k-1)}(m)/\hat{\sigma}}, \mathbb{P}_{\tilde{S}^{(k-1)}(n)/\hat{\sigma}} \right)$$

bases upon the Kullback-Leibler divergence between the two standardized non-central  $\chi^2$ -distributions with  $2L$  degrees of freedom and non-centrality parameters

$$\left( (\tilde{S}^{(k-1)}(m))^2 / \hat{\sigma}^2 - 2L \right)_+ \quad \text{and} \quad \left( (\tilde{S}^{(k-1)}(n))^2 / \hat{\sigma}^2 - 2L \right)_+,$$

where  $(x)_+$  denotes the maximum of  $x$  and 0. More details concerning the probability distribution of the observations and the approximation of the Kullback-Leibler divergence can be found in Sections 2.3 and 2.7.

After initialization we iterate computation of the adaptive weights in Equation (6), to infer on the homogeneous regions, and computation of the adaptive estimator (5), to approximate the corresponding expected value. Increasing  $h^{(k)}$  with each iteration step  $k$ , the procedure stops after a prespecified number  $k^*$  of steps. Since POAS provides a certain stability of estimates,  $k^*$  should be chosen to provide a reasonable computation time and still a sufficient quality of estimates within homogeneous regions. The specific choice of parameters for POAS and the appropriate smoothing of the  $S_0$ -image are discussed in Becker et al. [2012, Section 2.5].

**2.2. Description of the data.** Using the notation from Section 2.1 for the voxel set  $V$  and the gradient set  $G$ , we generalize the description of the signal to multi-shell dMRI data. The signal value in Equation (1) then additionally depends on the  $b$ -value  $b \in B \subseteq (0, \infty)$ . We denote the number of different shells or applied  $b$ -values by  $\mathfrak{B} := |B| \in \mathbb{N}$ . Furthermore, we interpret the non-weighted data  $S_0$  as a shell with  $b = 0$ . In other words, the data are sampled from  $\mathfrak{B} + 1$  functions

$$(8) \quad S_b : V \times G_b \ni (\vec{v}, \vec{g}) \mapsto S_b(\vec{v}, \vec{g}) \in \mathbb{R}, \quad b \in B_0 := B \cup \{0\},$$

where  $G_b$  is the set of gradients that have been measured on the shell with  $b$ -value  $b$ . As in Section 2.1, we consider the mean of all measured non-diffusion weighted images and denote it by  $S_0$ , too. Furthermore,  $S_0(\vec{v}, \vec{0}) := S_0(\vec{v})$  and  $G_0 = \{\vec{0}\}$ .

If the gradient scheme  $G_b$  is identical on all shells, the measured data can be easily arranged in a vector of length  $\mathfrak{B} + 1$  such that

$$(9) \quad (S_0(\vec{v}), S_{b_1}(\vec{v}, \vec{g}), \dots, S_{b_{\mathfrak{B}}}(\vec{v}, \vec{g}))^T \in \mathbb{R}^{\mathfrak{B}+1}.$$

This will be advantageous for our new method msPOAS as it enables the use of the whole information of all  $q$ -shells for adaptation. Else, that is, if the gradient schemes  $G_b \subseteq \mathbb{S}^2$  do not coincide for the measured  $b$ -values  $b \in B$  the vector in (9) will not be complete. We use the following interpolation method to replace the missing values  $S_b(\vec{v}, \vec{g})$ , for every  $b \in B$  with  $\vec{g} \notin G_b$ . The data structure is visualized in Figure 1.

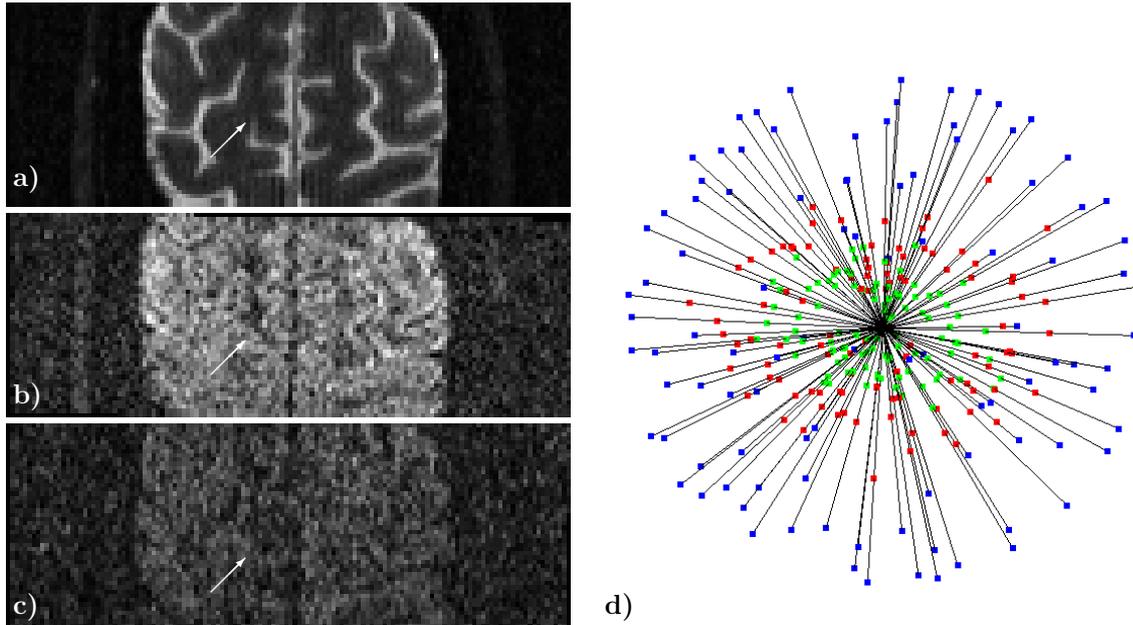


FIGURE 1. Diffusion weighted data  $S$  for an arbitrarily selected slice of the multi-shell data described in Section 3. a) Slice of the non-diffusion weighted data. b) Same slice taken with diffusion-weighting gradient  $\vec{g} \in \mathbb{S}^2$  and  $b$ -value  $b = 800\text{s/mm}^2$ . c) Same slice taken with the same gradient  $\vec{g}$  and  $b$ -value  $b = 2000\text{s/mm}^2$ . The intensity of the non-diffusion weighted image in a) has been down-weighted to make all three images visually feasible at once. d) shows the data within a single voxel (see arrow) as a 3D plot for all measured diffusion gradients in red and green at  $b = 800\text{s/mm}^2$  and  $b = 2000\text{s/mm}^2$ , respectively. For comparison the non-diffusion-weighted value is repeatedly shown as blue point for each gradient. The distance of the points to the center of the sphere is the corresponding signal value. Each 3D diffusion weighted image (a-c) is fully described by the set  $\{S_b(\vec{v}, \vec{g})\}_{\vec{v} \in V} \subseteq \mathbb{R}$  of signals in voxel space for a fixed diffusion gradient direction  $\vec{g} \in G$  (if  $b \neq 0$ ) and the  $b$ -value  $b \in B_0$ . Conversely, the data in a single voxel  $\vec{v} \in V$  equals the set  $\{S(\vec{v}, \vec{g})\}_{\vec{g} \in G} \subseteq \mathbb{R}^{\mathfrak{B}+1}$  of vectors, see (d).

Let  $b \in B$  be a fixed  $b$ -value and consider a gradient direction  $\vec{g} \notin G_b$ , that has not been included in the measurement on the corresponding shell. Then, we search for a triple of measured gradients  $\{\mathfrak{g}_{(b, \vec{g})}^{(l)}\}_{l=1}^3 \subseteq G_b$ , such that the gradient direction  $\vec{g}$  lies within the resulting spherical triangle

$$\vec{g} \in \Delta(\mathfrak{g}_{(b, \vec{g})}^{(1)}, \mathfrak{g}_{(b, \vec{g})}^{(2)}, \mathfrak{g}_{(b, \vec{g})}^{(3)})$$

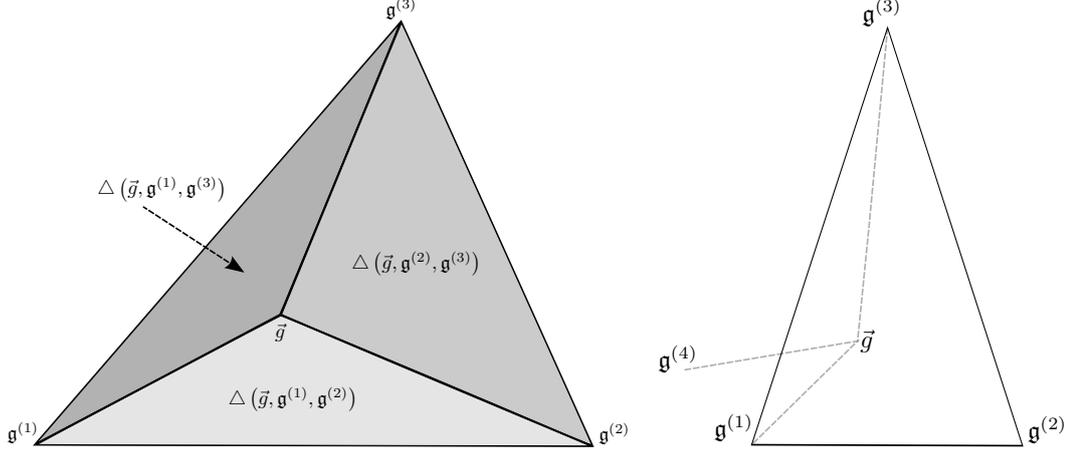


FIGURE 2. Spherical interpolation: The gradient  $\vec{g} \notin G_b$  lies within the spherical triangle which is formed by  $\mathbf{g}_{(b,\vec{g})}^{(1)}, \mathbf{g}_{(b,\vec{g})}^{(2)}, \mathbf{g}_{(b,\vec{g})}^{(3)} \in G_b$ . The corresponding weights  $c_{(b,\vec{g})}^{(l)}$ ,  $l = 1, 2, 3$ , depend on the respective proportion of the surface area of the partial triangles  $\Delta(\vec{g}, \mathbf{g}_{(b,\vec{g})}^{(l_1)}, \mathbf{g}_{(b,\vec{g})}^{(l_2)})$  with  $l_1, l_2 \neq l$ . On the right, we illustrate an example, where minimization of the total angular distance leads to a triangle which does not contain the gradient  $\vec{g}$ . Here, the angular distances satisfy  $\arccos(\langle \vec{g}, \mathbf{g}_{(b,\vec{g})}^{(4)} \rangle) < \arccos(\langle \vec{g}, \mathbf{g}_{(b,\vec{g})}^{(3)} \rangle)$ , but  $\vec{g} \notin \Delta(\mathbf{g}^{(4)}, \mathbf{g}_{(b,\vec{g})}^{(1)}, \mathbf{g}_{(b,\vec{g})}^{(2)})$ . The same holds with  $\mathbf{g}_{(b,\vec{g})}^{(2)}$  in place of  $\mathbf{g}_{(b,\vec{g})}^{(3)}$ . For simplicity, the figure shows two-dimensional triangles instead of spherical triangles.

and the corresponding total angular distance  $\sum_{l=1}^3 \arccos(\langle \vec{g}, \mathbf{g}_{(b,\vec{g})}^{(l)} \rangle)$  is minimal. Here,  $\langle \cdot, \cdot \rangle$  denotes the Euclidean scalar product of two vectors. The missing value  $S_b(\vec{v}, \vec{g})$  is then generated by linear spherical interpolation

$$S_b(\vec{v}, \vec{g}) := \sum_{l=1}^3 c_{(b,\vec{g})}^{(l)} S_b(\vec{v}, \mathbf{g}_{(b,\vec{g})}^{(l)})$$

with weights  $c_{(b,\vec{g})}^{(l)}$  defined as the spherical Bary-coordinates [Carfora, 2007, 3.1(c)] of  $\vec{g}$  within the spherical triangle  $\Delta(\mathbf{g}_{(b,\vec{g})}^{(1)}, \mathbf{g}_{(b,\vec{g})}^{(2)}, \mathbf{g}_{(b,\vec{g})}^{(3)})$ . For  $l \in \{1, 2, 3\}$  we set

$$c_{(b,\vec{g})}^{(l)} := \frac{\text{area}(\Delta(\vec{g}, \mathbf{g}_{(b,\vec{g})}^{(l_1)}, \mathbf{g}_{(b,\vec{g})}^{(l_2)}))}{\text{area}(\Delta(\mathbf{g}_{(b,\vec{g})}^{(1)}, \mathbf{g}_{(b,\vec{g})}^{(2)}, \mathbf{g}_{(b,\vec{g})}^{(3)}))}, \quad \text{where } l_1, l_2 \in \{1, 2, 3\} \setminus \{l\}, l_1 \neq l_2.$$

This interpolation is illustrated in Figure 2.

For completion and for the estimation of denoised  $S_0$ -images in msPOAS, we additionally create a vector  $(9)$  for  $\vec{g} \in G_0$ , i.e., for  $\vec{g} = \vec{0}$ , using the mean value of the signals on the corresponding shell

$$S_b(\vec{v}, \vec{0}) := |G_b|^{-1} \sum_{\vec{g} \in G_b} S_b(\vec{v}, \vec{g}),$$

where  $|G_b| \in \mathbb{N}$  is the number of gradients measured on the shell with  $b$ -value  $b$ .

Finally, we get the desired vector function defined on the measurement space  $V \times G \subseteq \mathbb{R}^3 \times \mathbb{S}^2$  by

$$(10) \quad \mathcal{S} : V \times G \ni (\vec{v}, \vec{g}) \mapsto (S_0(\vec{v}), S_{b_1}(\vec{v}, \vec{g}), \dots, S_{b_{\mathbb{N}}}(\vec{v}, \vec{g}))^T \in \mathbb{R}^{2\mathbb{N}+1},$$

where  $G := \bigcup_{b \in B_0} G_b \subseteq \mathbb{S}^2 \cup \{\vec{0}\}$  is the union of all gradient directions measured on any shell including the artificial direction  $\vec{0}$ . Here, we distinguish for  $b$ -values  $b > 0$  between values  $\mathcal{S}_b(\vec{v}, \vec{g})$  that have been measured ( $\vec{g} \in G_b$ ), values that have been interpolated on some shell as the gradient direction has been used on a different shell ( $\vec{g} \in G_{b'} \setminus G_b, b' > 0$ ) and a mean signal for  $\vec{g} \in G_0$ , such that

$$(11) \quad \mathcal{S}_b(\vec{v}, \vec{g}) = \begin{cases} S_b(\vec{v}, \vec{g}), & \text{if } \vec{g} \in G_b \\ \sum_{l=1}^3 c_{(b, \vec{g})}^{(l)} S_b(\vec{v}, \mathbf{g}_{(b, \vec{g})}^{(l)}), & \text{if } \vec{g} \in G_{b'} \setminus G_b, b' > 0 \\ |G_b|^{-1} \sum_{\vec{g} \in G_b} S_b(\vec{v}, \vec{g}) & \text{if } \vec{g} \in G_0. \end{cases}$$

Due to the interpolation we lose the statistical independence of the “observations”  $\{\mathcal{S}_b(\vec{v}, \vec{g})\}_{\vec{g} \in G}$ . Estimates will be constructed as weighted averages of the observed values  $\{\mathcal{S}_b(\vec{v}, \vec{g})\}_{\vec{g} \in G_b}$ . The interpolated observations  $\{\mathcal{S}_b(\vec{v}, \vec{g})\}_{\vec{g} \in G \setminus G_b}$  are solely used to determine the adaptive weights, see Section 2.4.

**2.3. Probability distribution of the observations.** The statistical penalty in Equation (7) is based on the Kullback-Leibler divergence. This requires knowledge of the probability distribution of the observations, up to some parameter  $\theta$ . For the generalization to multi-shell data we consider the joint probability distribution, see Equation (14). Additionally, we discuss the probability distribution of the single-shell data in more detail. In [Becker et al., 2012, Section 2.3], we treated this topic very briefly. This time, we want to provide a better understanding for the relationship of the unknown parameter  $\theta$  and the expected value, which we estimate by msPOAS. In particular, this sheds light on the impact of the variance  $\sigma$ , i.e., the amount of noise in the data, and the number of coils  $L$ .

For a fixed  $b$ -value, the measurement error of a complex dMRI-signal is usually assumed to follow a Gaussian distribution. This is motivated by the central limit theorem valid for the sum of many independent random errors summing up to the final measurement error in general experimental setups. Assuming the same variance  $\sigma^2$  for the real and the imaginary part of a signal, this leads with  $L = 1$  for the standardized magnitude image  $S_b(\vec{v}, \vec{g})/\sigma$  to a non-central  $\chi$ -distribution with two degrees of freedom and non-centrality parameter  $\theta = \sqrt{\mathbb{E}[S_b^2(\vec{v}, \vec{g})] / \sigma^2 - 2}$ .

For  $L > 1$ , the probability distribution of the reconstructed signals  $S$  depends on the applied reconstruction algorithm. Let  $S_{b,c}(\vec{v}, \vec{g})$  denote the absolute value of the Fourier transformed signal measured on receiver coil  $c$  for the voxel  $\vec{v} \in V$  with gradient  $\vec{g} \in G \cup \{\vec{0}\}$  and  $b$ -value  $b \in B_0$ . For the Sum-of-Squares reconstruction algorithm the diffusion weighted image is reconstructed as the square-root of the quadratic sum of the single-coil signals

$$S_b(\vec{v}, \vec{g}) := \sqrt{\sum_{c=1}^L S_{b,c}^2(\vec{v}, \vec{g})},$$

see e.g. Roemer et al. [1990], Aja-Fernández et al. [2009]. With a homogeneous variance  $\sigma^2$  for all receiver coils, we get

$$(12) \quad S_b(\vec{v}, \vec{g})/\sigma \sim \chi_{2L}(\theta) \quad \text{with} \quad \theta = \sqrt{\mathbb{E}[S_b^2(\vec{v}, \vec{g})] / \sigma^2 - 2L}$$

The method msPOAS estimates the true expected value  $\mathbb{E}[S_b(\vec{v}, \vec{g})]$  of the observations which relates to the parameter  $\theta$  via the function

$$\begin{aligned}
(13) \quad \mu(\theta) &:= \sqrt{\frac{\pi}{2}} \frac{\Gamma(L + \frac{1}{2})}{\Gamma(\frac{3}{2})\Gamma(L)} {}_1F_1\left(-\frac{1}{2}, L, -\frac{\theta^2}{2}\right) \\
&= \sqrt{\frac{\pi}{2}} \mathbb{L}_{1/2}^{(L-1)}\left(-\frac{\theta^2}{2}\right) \\
&= \mathbb{E}[S_b(\vec{v}, \vec{g})] / \sigma,
\end{aligned}$$

where  $\mathbb{L}_{1/2}^{(L-1)}$  denotes the generalized Laguerre polynomial,  $\Gamma$  is the Gamma function, and  ${}_1F_1$  the confluent hypergeometric function. For their relation used in Equation (13) see e.g. El-Sayed [2000].

The non-central  $\chi$ -distribution remains a valid model, with lower number of degrees of freedom, if combination weights are either 0 or 1, i.e., location dependent subsets of images are combined. Generally, the sum-of-squares formula is the simplest approximation of the reconstruction of the NMR phased array signal, that does not require any assumption on the magnetic field of the different receiver coils [Roemer et al., 1990]. More sophisticated parallel acquisition methods like SENSE and GRAPPA, lead to heteroscedastic variance and correlated noise with slight changes in the distribution [Thunberg and Zetterberg, 2007]. However, it has been shown [Dietrich et al., 2008] that the non-central  $\chi$ -distribution is also approximately valid for these reconstruction methods. For a detailed discussion concerning GRAPPA we refer to Aja-Fernández et al. [2011].

For the sake of simplicity, we concentrate on the case of a fixed number of degrees of freedom  $2L' \leq 2L$ . Additionally, we assume the variance  $\sigma^2$  to be known or an appropriate approximation  $\hat{\sigma}^2$  to be achievable. Generally, variance estimation for dMRI data is a challenging problem. A survey of estimation procedures can be found in Aja-Fernández et al. [2009]. In Becker et al. [2012, Appendix C], we proposed a new estimation procedure that is also feasible for low signal-to-noise-ratio and if no background can be defined in the image. Similar to POAS and msPOAS is based on the Propagation-Separation approach [Polzehl and Spokoiny, 2006]. The achieved results could be improved by the new approximation of the Kullback-Leibler divergence that we discuss in Section 2.7. In any case, the variance estimation should be performed before pre-processing steps such as registration-based motion and eddy-current correction. The interpolation which is part of the registration process, leads to a reduced variability and to spatial correlation of the data, while the bias  $\theta - \mathbb{E}S$  between the true parameter and the expectation of the signal remains unchanged. This would in turn bias the variance estimates.

The measurement errors for different  $b$ -values are statistically independent since each  $q$ -shell is measured on its own. Therefore, the joint probability density  $f_{S/\sigma}$  of the random variable  $\mathcal{S}(\vec{v}, \vec{g}) \in \mathbb{R}^{\mathfrak{B}+1}$  in Equation (10) equals the product of the corresponding marginal densities  $f_{S_b/\sigma}$ ,  $b \in B_0$ , i.e.

$$(14) \quad f_{S/\sigma}(u) = \prod_{l=0}^{\mathfrak{B}} f_{S_{b_l}/\sigma}(u_l), \quad u = (u_0, \dots, u_{\mathfrak{B}})^T \in \mathbb{R}^{\mathfrak{B}+1}, \quad b_0 := 0.$$

For simplicity, we do not distinguish the notation of the signals  $S/\sigma$  and the interpolated signals  $\mathcal{S}/\sigma$  and the corresponding random variables  $S/\sigma \sim f_{S/\sigma}$  and  $\mathcal{S}/\sigma \sim f_{\mathcal{S}/\sigma}$ , respectively.

**2.4. Multi-shell position-orientation adaptive smoothing (msPOAS).** The method msPOAS can be applied to multi-shell data with fixed or varying gradient schemes. In the following, we modify the non-adaptive estimator (3) which we used for initialization of POAS. In particular, we discuss the choices of the location bandwidths  $\{h^{(k)}\}_{k=0}^{k^*}$  and of the balancing parameter  $\varkappa$  for the case of multi-shell dMRI data. Afterwards, we introduce a generalized statistical penalty which uses the information of all  $q$ -shells for adaptation. This is based on the interpolated signals which we described by a vector function in Equation (10). However, the interpolated values will be used for adaptation only and not in the estimator

in order to ensure statistical independence of the observations that are included into the estimator. Finally, we present the algorithm msPOAS.

As for POAS, we fix two non-increasing kernel functions  $K_{\text{loc}}$  and  $K_{\text{ad}}$  with compact support  $[0, 1]$ . The effect of different choices for the kernels is negligible, see e.g. Scott [1992, Section 6.2.3]. For the implementation of the algorithm we set

$$(15) \quad K_{\text{loc}}(x) := (1 - x^2)_+ \quad \text{and} \quad K_{\text{ad}}(x) := \begin{cases} 1 & \text{for } 0 \leq x < 0.5, \\ (2 - 2x)_+ & \text{for } 0.5 \leq x. \end{cases}$$

Furthermore, let  $\delta_\kappa(\cdot, \cdot)$  denote some appropriate discrepancy on  $\mathbb{R}^3 \times \mathbb{S}^2$ . For instance, we may use  $\delta_\kappa(g_1, g_2) := \Delta_\kappa(g_1, g_2)$ , see Equation (2), as in the original work or the simplified metric in Section 2.6, Equation (23). For simplicity, we do in the following no longer distinguish between the original data  $S$  and the interpolated data  $\tilde{S}$  denoting both by  $S$ . Additionally, we apply the interpolation in Equation (10) to the adaptive estimates  $\tilde{S}_b^{(k)}$  defined below, replacing  $S_b$  by  $\tilde{S}_b^{(k)}$  in the respective formulas. Once again, we denote the resulting values  $\tilde{S}_b^{(k)}$  by  $\tilde{S}_b^{(k)}$ .

For initialization of the algorithm we introduce a non-adaptive estimator which is similar to Equation (3). There, the estimator was defined as a weighted quadratic mean of the observations in order to ensure that it belongs to the same class of probability distributions as the observations. Here, we follow a slightly different strategy. In Section 2.7, we justify an approximation of the probability distribution of the observations by a Gaussian distribution with appropriate parameters. As the weighted mean of independent Gaussian distributed observations is again Gaussian, this motivates to replace the weighted quadratic mean by a weighted arithmetic mean. Thus, we define for each  $b \in B_0$  and all measurement points  $m \in V \times G_b$  the non-adaptive estimator

$$(16) \quad \bar{S}_b^{(k)}(m) = \sum_{n \in V \times G_b} \bar{w}_{mn}^{(k)} S_b(n) / \bar{N}_{m,b}^{(k)} \in \mathbb{R}$$

with

$$\bar{w}_{mn}^{(k)} := K_{\text{loc}}\left(\delta_\kappa(m, n)/h^{(k)}\right) \quad \text{and} \quad \bar{N}_{m,b}^{(k)} := \sum_{n \in V \times G_b} \bar{w}_{mn}^{(k)},$$

which is similar to Equation (4). For  $m, n \in V \times G_0$  we set  $\bar{w}_{mn}^{(k)} := \bar{w}_{m_{\vec{g}} n_{\vec{g}}}^{(k)}$ , where  $m = (\vec{v}_m, \vec{0})$  implies  $m_{\vec{g}} = (\vec{v}_m, \vec{g})$  and  $n = (\vec{v}_n, \vec{0})$  yields  $n_{\vec{g}} = (\vec{v}_n, \vec{g})$  for some arbitrary but fixed gradient  $\vec{g} \in G$ . This reduces the  $\mathbb{R}^3 \times \mathbb{S}^2$ -discrepancy  $\delta_\kappa$  for  $S_0$ -images to the Euclidean metric in voxel space and hence enables its application to the  $S_0$ -image.

The non-adaptive weights  $\bar{w}_{mn}^{(k)}$  depend on the respective  $b$ -value  $b \in B_0$  via the location bandwidth  $h^{(k)}$ . The bandwidths are measured in units of voxel counts with respect to the smallest of the three voxel extensions. We chose  $h^{(0)} = 1$ , i.e., one unit of the shortest edge of the voxel. The pre-specified sequence of bandwidths  $\{h^{(k)}\}_{k=0}^{k^*}$  should be increasing with  $h^{(0)} > 0$ . However, it seems to be recommendable to ensure a constant variance reduction of the non-adaptive estimates during iteration. Hence, in our implementation, we chose the bandwidths  $\{h^{(k)}\}_{k=0}^{k^*}$  such that the variance of the non-adaptive estimate is reduced by 25%, up to boundary effects in voxel space. As the number of applied gradients may vary for the distinct  $q$ -shells, this leads to a choice of  $\{h^{(k)}\}_{k=0}^{k^*}$  which depends via the respective gradient scheme  $G_b \ni \vec{g}$  on the considered  $b$ -value  $b \in B_0$ . Additionally, we get a dependence on the gradient  $\vec{g}$  itself by compensating possible inhomogeneities of the gradient scheme. More precisely, we fix some voxel  $\vec{v}_m$  close to the center of the voxel space  $V$  and calculate for all  $b \in B_0$  and every  $\vec{g}_m \in G_b$  the sequence of location bandwidths  $\{h^{(k)}(b, \vec{g}_m)\}_{k=1}^{k^*}$  such that

$$\sum_{n \in V \times G_b} \frac{(\bar{w}_{mn}^{(k)})^2}{(\bar{N}_{m,b}^{(k)})^2} = 1.25^{-1} \sum_{n \in V \times G_b} \frac{(\bar{w}_{mn}^{(k-1)})^2}{(\bar{N}_{m,b}^{(k-1)})^2},$$

where  $m = (\vec{v}_m, \vec{g}_m) \in V \times G_b$ . The resulting bandwidths  $\{h^{(k)}(b, \vec{g}_m)\}_{k=1}^{k^*}$  are then used for all voxels  $\vec{v} \in V$ .

The parameter  $\kappa$  balances between spatial and spherical smoothing in  $V \subseteq \mathbb{R}^3$  and  $G \subseteq \mathbb{S}^2$ . One possible choice of  $\varkappa > 0$  has been discussed in Becker et al. [2012, Section 2.5], where we set  $\varkappa^{(k)}(\vec{g}) := \varkappa_0/h^{(k)}(\vec{g})$ . For multi-shell data, this leads again to a dependence on  $b \in B_0$  and  $\vec{g} \in G_b$ . The precise choice of  $\varkappa_0$  depends on the data at hand. More precisely,  $\varkappa_0$  balances between variance reduction and possible bias by smoothing on the sphere. In the initial step it determines spherical resolution. For a given total number of applied gradient directions  $N_g$  (summed over all shells) the quantity  $N_g(1 - \cos(\varkappa_0))$  is the mean number of neighboring gradient directions with positive weights, if  $h^{(0)} = 1$ . We suggest to select  $\varkappa_0$  such that this number is between 5 and 10, i.e.,  $\varkappa_0$  depends on the number of measured diffusion-weighting gradients. The value of  $\varkappa_0$  should be the larger the lower the signal-to-noise-ratio in order to stabilize the estimates in the first iteration steps. However, larger values may introduce a bias. This bias will be effectively reduced during iteration due to the restrictions imposed by the statistical penalty.

For the adaptive estimator, we take advantage of the whole information in the data vector (10). This requires an appropriate modification of the statistical penalty in Equation (7). With a slight abuse of notation we introduce an abbreviatory notion for the Kullback-Leibler divergence of two non-central  $\chi$ -distributions with expectation  $\zeta_1$  and  $\zeta_2$ , respectively, setting

$$\mathcal{KL}(\zeta_1, \zeta_2) := \mathcal{KL}(\mathbb{P}_{(\mu^{-1}(\zeta_1))}, \mathbb{P}_{(\mu^{-1}(\zeta_2))}),$$

where  $\mathbb{P}_\theta = \chi_{2L'}(\theta)$  and the function  $\mu^{-1}$  is defined via Equation (13) for all  $\zeta > \sqrt{2}\Gamma(L + 1/2)/\Gamma(L)$  and  $\mu^{-1}(\zeta) = 0$ , else.

In the construction of the statistical penalty we replace the unknown expected value  $\zeta = \mathbb{E}S_b(m)/\sigma$  by its estimate  $\tilde{\zeta} := \tilde{S}_b^{(k)}(m)/\hat{\sigma}$ . We observe that Equation (14) yields for the Kullback-Leibler divergence of the multivariate densities  $f_{\tilde{S}(m)/\hat{\sigma}}$  and  $f_{\tilde{S}(n)/\hat{\sigma}}$  that

$$\mathcal{KL}\left(\frac{\tilde{S}(m)}{\hat{\sigma}}, \frac{\tilde{S}(n)}{\hat{\sigma}}\right) = \sum_{b \in B_0} \mathcal{KL}\left(\frac{\tilde{S}_b(m)}{\hat{\sigma}}, \frac{\tilde{S}_b(n)}{\hat{\sigma}}\right),$$

where  $S := \mathcal{S}$  is as in Equation (10). Therefore, we may redefine the statistical penalty as

$$(17) \quad s_{mn}^{(k)} := \sum_{b \in B_0} \tilde{N}_{m,b}^{(k-1)} \mathcal{KL}\left(\frac{\tilde{S}_b^{(k-1)}(m)}{\hat{\sigma}}, \frac{\tilde{S}_b^{(k-1)}(n)}{\hat{\sigma}}\right)$$

Here,  $\tilde{N}_{m,b}^{(k-1)}$  relates for each  $b$ -value to the achieved variance reduction using the adaptive weights

$$(18) \quad \tilde{w}_{mn}^{(k)} := \bar{w}_{mn}^{(k)} \cdot K_{\text{ad}}\left(s_{mn}^{(k)}/\lambda\right),$$

where the adaptation parameter  $\lambda > 0$  determines the impact of the adaptivity. Its specific choice can be found in Section 2.5. Taking the impact of the spherical interpolation into account, we consider the same classification of cases as for the interpolation formula (11). For  $b = 0$ , we down-weight the influence of the  $S_0$ -images to compensate the already achieved variance reduction. Otherwise, the  $S_0$ -images would

dominate the adaptation in an undesirable manner. Hence, for  $m = (\vec{v}_m, \vec{g}_m)$  we set

$$(19) \quad \tilde{N}_{m,b}^{(k)} = \begin{cases} \max_{k' \leq k} \left( \sum_{n \in V \times G_b} \tilde{w}_{mn}^{(k')} \right) & \text{if } b > 0 \wedge \vec{g}_m \in G_b, \\ \max_{k' \leq k} \left( \sum_{l=1}^3 c_{b, \vec{g}_m}^{(l)} / \tilde{N}_{(\vec{v}_m, \mathfrak{g}_{b, \vec{g}_m}^{(l)}, b)}^{(k')} \right)^{-1} & \text{if } b > 0, \vec{g}_m \in G_{b'}, 0 < b' \neq b, \\ \max_{k' \leq k} \left( \sum_{\vec{g} \in G_b} 1 / \tilde{N}_{(\vec{v}_m, \vec{g}, b)}^{(k')} \right)^{-1} \cdot |G_b| & \text{if } b > 0 \wedge \vec{g}_m \in G_0, \\ \max_{k' \leq k} \left( \sum_{n \in V \times G_0} \tilde{w}_{mn}^{(k')} \right) \cdot |S_0|^{-1} & \text{if } b = 0 \wedge \vec{g}_m \in G_b \end{cases}$$

where  $|G_b|$  denotes the number of measured gradients on the shell with  $b$ -value  $b$  and  $|S_0|$  is the number of acquired  $S_0$ -images that form the mean image  $S_0$ . The constants  $c_{b, \vec{g}_m}^{(l)}$  and the gradients  $\mathfrak{g}_{b, \vec{g}_m}^{(l)}$  are given in Section 2.2. Note, that we consider the maximal variance reduction  $\tilde{N}_{m,b}^{(k)} := \max_{k' \leq k} (\cdot)$  until

step  $k$  in order to ensure that the sequence  $\{\tilde{N}_{m,b}^{(k)}\}_{k=0}^{k^*}$  is non-decreasing. This allows to preserve an achieved adaptation quality during iteration.

Finally, we may introduce the algorithm for multi-shell position-orientation adaptive smoothing (msPOAS). For simplicity we denote  $h^{(k)} := h^{(k)}(b, \vec{g})$  and  $\varkappa := \varkappa(h^{(k)})$ , see the discussion above.

- Input parameters: Sequence of location bandwidths  $\{h\}_{k=0}^{k^*}$ , balancing parameter  $\kappa$ , adaptation parameter  $\lambda > 0$ .
- Initialization:  $\tilde{S}_b^{(0)}(m) := \bar{S}_b^{(0)}(m)$  and  $\tilde{N}_{m,b}^{(0)} := \bar{N}_{m,b}^{(0)}$  for all  $m \in V \times G_b, b \in B_0$ .
- Iteration: For each  $b \in B_0$  and  $m := (\vec{v}_m, \vec{g}_m) \in V \times G_b$  do the following. Interpolate the missing values of  $\tilde{S}_{b'}^{(k-1)}(m)$  and  $\tilde{N}_{m,b'}^{(k)}, b' \in B \setminus \{b\}$ , according to Equations (11) and (19). Then, calculate the statistical penalty

$$s_{mn}^{(k)} = \sum_{b \in B_0} \tilde{N}_{m,b}^{(k-1)} \mathcal{KL} \left( \frac{\tilde{S}_b^{(k-1)}(m)}{\hat{\sigma}}, \frac{\tilde{S}_b^{(k-1)}(n)}{\hat{\sigma}} \right), \quad n \in V \times G_b,$$

the adaptive weights

$$\tilde{w}_{mn}^{(k)} = K_{\text{loc}} \left( \delta_{\kappa}(m, n) / h^{(k)} \right) \cdot K_{\text{ad}} \left( s_{mn}^{(k)} / \lambda \right), \quad n \in V \times G_b,$$

the corresponding sum over the adaptive weights

$$\tilde{N}_{m,b}^{(k)} = \max_{k' \leq k} \left( \sum_{n \in V \times G_b} \tilde{w}_{mn}^{(k')} \right),$$

and the adaptive estimator

$$\tilde{S}_b^{(k)}(m) = \sum_{n \in V \times G_b} \tilde{w}_{mn}^{(k)} S_b(n) / \tilde{N}_{m,b}^{(k)}.$$

- Stopping: Stop if  $k = k^*$  and return  $\tilde{S}_b^{(k^*)}(m)$  for each  $b \in B_0$  and all  $m \in V \times G_b$ , else set  $k := k + 1$ .

In Becker and Mathé [2013] theoretical properties of the Propagation-Separation approach have been studied. These results are restricted to noise distributions that belong to an exponential family and do not generalize to non-central chi-distributed observations. However, the examples in Section 4.1 give sufficient evidence, that the general behavior of the algorithm remains unchanged. We included them in this article in order to provide a better understanding for the procedure, its performance during iteration and the impact of the maximal location bandwidth  $h^{(k^*)}$ .

**2.5. Advanced choice of the adaptation bandwidth.** We present an advanced parameter choice strategy for the adaptation bandwidth  $\lambda$  which is the crucial parameter of the procedure. It determines the amount of adaptation. For  $\lambda \rightarrow \infty$  msPOAS simply is a non-adaptive filter with the kernel function  $K_{10c}$  and the bandwidth  $h^{(k^*)}$ . For  $\lambda \rightarrow 0$  the adaptive estimator only includes a non-vanishing weight  $\tilde{w}_{mn}^{(k)}$  at  $m = n$  and the msPOAS result coincides with the original data. We call this effect for small values of  $\lambda$  “adaptation to noise” as msPOAS then interprets variability of the signal caused by noise as discontinuities. Hence,  $\lambda$  should be sensible to the structure of the data allowing as much adaptation as possible, but avoiding adaptation to noise. The propagation condition [Polzehl and Spokoiny, 2006] enables a choice of  $\lambda$  which is independent of the data at hand. It has been reformulated and analyzed in a recent article by Becker and Mathé [2013]. The basic idea is to ensure, under homogeneity, for the adaptive estimator a similar behavior as for the non-adaptive estimator, which is optimal in this situation. That is, the impact of the adaptation should be negligible supposing a globally constant parameter function  $\theta(\cdot) \equiv \theta$ .

The parameter choice is formulated with respect to the parameter  $\theta \in \Theta \subseteq \mathbb{R}$  that determines the probability distribution  $\mathbb{P}_\theta$  of the observations. For dMRI data the rescaled observed signals  $S_b(\cdot)/\sigma \sim \chi_{2L}(\theta/\sigma)$ ,  $b \in B_0$ , relate to the non-centrality parameter  $\theta$  via

$$(20) \quad \theta_{m,b} := \mu^{-1}(\mathbb{E}[S_b(m)]/\sigma) \in \Theta, \quad m \in V \times G_b.$$

For the parameter choice we consider an artificial data set with parameters  $\{\theta_{m,b}\}_{m \in V \times G_b}$ ,  $b \in B_0$ , such that the (artificial) observations satisfy  $S_b(m) \sim \mathbb{P}_{\theta_{m,b}}$  with  $\theta_{m,b} \equiv \theta_b$  for some  $\theta_b \in \Theta$ . For  $m \in V \times G_b$ , we denote the resulting non-adaptive estimate by  $\bar{\theta}_{m,b}^{(k)}$  and the adaptive one by  $\tilde{\theta}_{m,b}^{(k)}$  replacing in Equation (20) the expectation  $\mathbb{E}S_b(m)$  by the corresponding estimate  $\bar{S}_b^{(k)}(m)$  or  $\tilde{S}_b^{(k)}(m)$ , respectively.

Now, we analyze the behavior of the non-adaptive estimator. This has been studied in Polzehl and Spokoiny [2006] supposing an exponential family model with unbiased observations  $\mathbb{E}Y_j = \theta_j$  instead of non-central chi-distributions. This model includes Gaussian, Gamma, Poisson, Bernoulli and many other probability distributions. Then, the non-adaptive estimator satisfies under homogeneity for each  $b$ -value  $b \in B_0$  and all  $m \in V \times G_b$  the exponential bound

$$(21) \quad \mathbb{P}\left(\bar{N}_{m,b}^{(k)} \mathcal{KL}(\bar{\theta}_{m,b}^{(k)}, \theta_b) > z\right) \leq 2e^{-z}, \quad S_b(n) \stackrel{\text{iid}}{\sim} \mathbb{P}_{\theta_b}, \theta_b \in \Theta, \text{ for all } n \in V \times G_b.$$

This means that the Kullback-Leibler divergence between the exact parameter  $\theta_b$  and the corresponding estimator  $\bar{\theta}_{m,b}^{(k)}$  decreases (in probability) at least with rate  $\bar{N}_{m,b}^{(k)}$ . Thus, we aim to ensure the same rate for the adaptive estimator supposing a non-central chi-distribution.

Let us consider the procedure for single-shell data, first. This can be done by simulating data

$$\{m, S_b(m)\}_{m \in V \times G_b} \subseteq \mathbb{R}, \quad S_b(m) \stackrel{\text{iid}}{\sim} \mathbb{P}_{\theta_b},$$

for some least favorable  $\theta_b \in \Theta$ , where  $b \in B_0$  is fixed. Then, we choose for  $b \in B_0$  the minimal value  $\lambda_b$  such that the following condition is satisfied for this fixed parameter  $\theta_b$ .

Let the function  $\mathfrak{Z}_\lambda : \{0, \dots, k^*\} \times (0, 1) \times \Theta \rightarrow \mathbb{R}^+$  with  $\lambda > 0$  and  $m \in V \times G_b$  fixed be defined by

$$\mathfrak{Z}_\lambda(k, p; \theta_b) := \inf \left\{ z > 0 : \mathbb{P}\left(\bar{N}_{m,b}^{(k)} \mathcal{KL}(\tilde{\theta}_b^{(k)}(m, \lambda), \theta_b) > z\right) \leq p \right\},$$

where  $\tilde{\theta}_b^{(k)}(m, \lambda)$  denotes the adaptive estimator resulting from the algorithm with adaptation bandwidth  $\lambda > 0$  and observations  $S_b(n) \stackrel{\text{iid}}{\sim} \mathbb{P}_{\theta_b}$  for all  $n \in V \times G_b$ . We say that  $\lambda$  is chosen in accordance with the propagation condition of level  $\epsilon > 0$  for  $\theta_b \in \Theta$  if the function  $\mathfrak{Z}_\lambda(\cdot, p; \theta_b)$  is non-increasing for all  $p \in (\epsilon, 1)$  [Becker and Mathé, 2013].

Here, the behavior of  $\mathfrak{J}_\lambda$  is invariant w.r.t. the choice of  $m \in V \times G_b$  since Equation (21) and the error distribution hold on the whole measurement space  $V \times G_b$ . The parameter  $\theta_b$  is appropriate if it yields a sufficiently large value  $\lambda_b$  such that the propagation condition remains valid for a certain range of parameters  $\Theta_b^* \subseteq \Theta$  including the true but unknown parameters  $\theta_{b,m}$  with  $m \in V \times G_b$  with high probability. Simulations show that  $\theta_b \approx 3$  satisfies the desired.

For the multi-shell data the same strategy has been implemented with

$$\sum_{b \in B_0} \overline{N}_{m,b}^{(k)} \mathcal{KL}(\tilde{\theta}_b^{(k)}(m, \lambda), \theta_b) \quad \text{in place of} \quad N_{m,b}^{(k)} \mathcal{KL}(\tilde{\theta}_b^{(k)}(m, \lambda), \theta_b),$$

where the missing values of  $\overline{N}_{m,b}^{(k)}$  and  $\tilde{\theta}_b^{(k)}(m)$  follow again by interpolation, see Equations (10) and (19). This is motivated by the equivalent modification of the statistical penalty, see Equation (17). Selection of  $\lambda$  is then done using the gradient schemes and  $b$ -values at hand and specifying appropriate values of  $\theta_b$  for the different shells. In practice, we approximate the probability

$$\mathbb{P} \left( \sum_{b \in B_0} \overline{N}_{m,b}^{(k)} \mathcal{KL}(\tilde{\theta}_b^{(k)}(m, \lambda), \theta_b) > z \right)$$

by the corresponding relative frequency

$$|V \times G|^{-1} \sum_{m \in V \times G} \mathbf{1}_{M_\lambda^{(k)}(z)}(m),$$

where  $|V \times G|$  denotes the number of observations  $m$  and

$$M_\lambda^{(k)}(z) := \left\{ m \in V \times G : \sum_{b \in B_0} \overline{N}_{m,b}^{(k)} \mathcal{KL}(\tilde{\theta}_b^{(k)}(m, \lambda), \theta_b) > z \right\}.$$

The approximation and the propagation condition are restricted to the interior of the design space in order to avoid boundary effects. Simulations are carried out using an artificial data set that ensures a sufficiently large number of effectively independent regions for estimating the propagation level on the basis of a single realization.

For our real datasets described in Section 3.2, the adaptation bandwidth  $\lambda = 20$  yield the propagation level  $\epsilon = 5 \cdot 10^{-5}$ . This means, that (in probability) 5 of  $10^5$  estimates would adapt to noise in a homogeneous setting. One of the considered datasets contains single-shell measurements and the other one multi-shell data with three  $q$ -shells (including the  $S_0$ -image with  $b = 0$ ). Hence, the choice of  $\lambda$  seems to be invariant w.r.t. the number of shells. Furthermore, it shows a certain stability w.r.t. the number of measured gradients as the first dataset was measured with 60 gradient directions and the second one with 100 per  $q$ -shell. Probably, the choice of  $\varkappa_0$ , which is given in Section 2.4, has compensated the impact of this difference. Additionally, we observed for a fixed value of  $\lambda$  with varying degrees of freedom  $2L'$  always the same propagation level. In Section 5, we will discuss the mutual effects of the adaptation bandwidth  $\lambda$  and the noise variance  $\sigma$ .

The introduced choice of the adaptation bandwidth provides a better interpretability than our previous strategy. In particular, this enabled a detailed evaluation which provided the mentioned invariances of  $\lambda$  w.r.t. the numbers of shells, gradients and receiver coils. Additional theoretical benefits are discussed in [Becker and Mathé, 2013].

**2.6. Simplified discrepancy in the measurement space.** We introduce a new and simplified discrepancy on  $\mathbb{R}^3 \times \mathbb{S}^2$  in order to accelerate the algorithm and avoid possible artifacts which may be introduced by the local approximation of our recent discrepancy  $\Delta_{\kappa}$ .

In the original work [Becker et al., 2012], we embedded the measurement space  $\mathbb{R}^3 \times \mathbb{S}^2$  into the special Euclidean motion group  $\text{SE}(3)$  following an approach of Duits and Franken [2011]. For processing of dMRI data, the authors recommended left-invariant  $\text{SE}(3)$ -operations because of their close relation to Euclidean-invariant operations on the orientation-marginal. The embedding provides a group operation, a definition of a  $(\mathbb{R}^3 \times \mathbb{S}^2)$ -convolution and several discrepancy functions. The canonical discrepancy on  $\text{SE}(3)$  is the Riemannian 2-norm since it refers to the  $\text{SE}(3)$ -geodesics. We used it for the definition of the discrepancy  $\Delta_\kappa(\cdot, \cdot)$  on  $\mathbb{R}^3 \times \mathbb{S}^2$ , see Equation (2). Then, the squared non-adaptive estimator in Equation (3) corresponds to a (discrete) convolution on  $\mathbb{R}^3 \times \mathbb{S}^2$  [Becker et al., 2012, Appendix A.3]. Since the typical kernel estimation on  $\mathbb{R}$  can be described as (discrete) convolution on  $\mathbb{R}$  this property seems natural. Additionally, the method POAS yield quite convincing results on simulated and real dMRI data [Becker et al., 2012, Section 4].

However, the Riemannian 2-norm is defined via integral curves [Dungey et al., 2003]. In order to get an explicit formula, we switched to a local approximation. The special Euclidean motion group  $\text{SE}(3)$  is defined as the semi-direct product of the Euclidean space  $\mathbb{R}^3$  and the special orthogonal group  $\text{SO}(3)$ , such that

$$\text{SE}(3) := \mathbb{R}^3 \rtimes \text{SO}(3).$$

The former describes translations and the latter describes rotations in  $\mathbb{R}^3$ . The approximation of the Riemannian 2-norm bases upon the left-invariant coordinates  $k_i$  of  $\text{SE}(3)$  setting

$$(22) \quad \|\hat{m}\|_{R,\kappa} \approx \inf \left\{ \left( \sum_{i=1}^3 k_i^2 + \kappa^{-2}(k_4^2 + k_5^2 + |k_6|) \right)^{1/2} : \prod_{i=1}^6 \exp(k_i \mathbf{A}_i) = \mathbf{M}_{\hat{m}} \right\},$$

where  $\mathbf{M}_{\hat{m}}$  denotes the matrix representation of  $\hat{m} \in \text{SE}(3)$  and  $\mathbf{A}_i$  are the left-invariant vector fields of  $\text{SE}(3)$  in matrix representation. The parameter  $\kappa$  balances between distances on the sphere and in voxel space, see Section 2.5. For  $\hat{m} = (\hat{v}, \hat{\mathbf{R}}) \in \mathbb{R}^3 \rtimes \text{SO}(3)$ , the coordinates  $k_1, k_2$  and  $k_3$  relate to the spatial coordinates of  $\hat{\mathbf{R}}^{-1} \hat{v} \in \mathbb{R}^3$  and the others to the angular coordinates depending on the rotation matrix  $\hat{\mathbf{R}} \in \text{SO}(3)$ , only. We refer to Becker et al. [2012, Appendix A] for more details. While the exact discrepancy  $\Delta_\kappa(\cdot, \cdot)$  on  $\mathbb{R}^3 \times \mathbb{S}^2$  is well-defined w.r.t. the embedding of  $\mathbb{R}^3 \times \mathbb{S}^2$  into  $\text{SE}(3)$ , its local approximation violates a certain invariance. More precisely, the approximation depends on an additional rotation angle which appears in the parametrization of  $\text{SO}(3)$  but not in the parametrization of  $\mathbb{S}^2$ , see Figure 3. To avoid possible artifacts, we introduce a new discrepancy function that is also easier to compute.

By definition of the group product on  $\text{SE}(3)$

$$m_1 \cdot_{\text{SE}(3)} m_2 = (\vec{v}_1, \mathbf{R}_1) \cdot_{\text{SE}(3)} (\vec{v}_2, \mathbf{R}_2) = (\vec{v}_1 + \mathbf{R}_1 \vec{v}_2, \mathbf{R}_1 \mathbf{R}_2), \quad m_1, m_2 \in \text{SE}(3),$$

the rotation group  $\text{SO}(3)$  acts on the translation group  $\mathbb{R}^3$ , not vice versa. Hence,  $\mathbb{R}^3$  and  $\text{SO}(3)$  play different roles. Nevertheless, in the approximated discrepancy  $\Delta_\kappa(\cdot, \cdot)$  on  $\mathbb{R}^3 \times \mathbb{S}^2$ , see Equations (2), and (22) we get two well separated parts, the squared Euclidean metric

$$\sum_{i=1}^3 k_i^2 = \|\vec{v}_1 - \vec{v}_2\|^2$$

and the term  $k_4^2 + k_5^2 + |k_6|$  depending on the two gradients  $\vec{g}_1, \vec{g}_2 \in \mathbb{S}^2$ . Against the comprehensible conjecture that both parts would interact this implies that the Euclidean and the spherical distances of voxels and gradients respectively can be considered on its own. The Euclidean metric as distance in voxel space complies with our intuition, but we want to overcome the violation of the mentioned invariance. Hence, we propose the great-circle distance, which describes the shortest distance between two points on the sphere following the spherical geodesics. It equals the radian, that is, the arc length between the

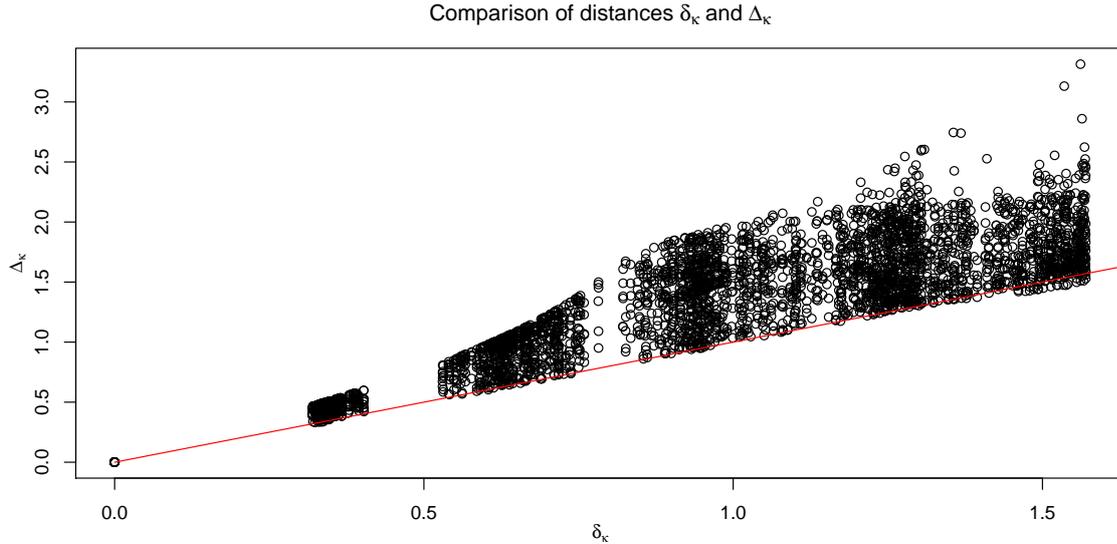


FIGURE 3. Pairwise scatterplot of discrepancies  $\delta_\kappa$  and  $\Delta_\kappa$  for  $\kappa = .5$ ,  $\|\vec{v}_1 - \vec{v}_2\| = 0$ . Each point corresponds to a gradient pair out of 60 gradients. The line where both values would coincide is shown in red.

vectors  $\vec{g}_1, \vec{g}_2 \in \mathbb{S}^2$  on the corresponding unit circle. It can be calculated by

$$d_{\mathbb{S}^2}(\vec{g}_1, \vec{g}_2) := \arccos \langle \vec{g}_1, \vec{g}_2 \rangle \in [0, \pi],$$

where  $\langle \cdot, \cdot \rangle$  denotes the Euclidean scalar product. Replacing in  $\Delta_\kappa(\cdot, \cdot)$  the  $\text{SO}(3)$ -term by the spherical geodesic we get

$$\left( \|\vec{v}_1 - \vec{v}_2\|^2 + \kappa^{-2} \arccos^2 \langle \vec{g}_1, \vec{g}_2 \rangle \right)^{1/2}$$

In order to identify each gradient  $\vec{g} \in \mathbb{S}^2$  with the opposite direction  $-\vec{g}$  we consider the absolute value of the scalar product in  $d_{\mathbb{S}^2}(\cdot, \cdot)$ . Further, we recommend a slight modification to ensure the triangle inequality, see Proposition 1. This leads to the definition

$$(23) \quad \delta_\kappa(m_1, m_2) := \|\vec{v}_1 - \vec{v}_2\| + \kappa^{-1} \arccos |\langle \vec{g}_1, \vec{g}_2 \rangle|, \quad m_1, m_2 \in \mathbb{R}^3 \times \mathbb{S}^2.$$

Now, we analyze the behavior of our new discrepancy  $\delta_\kappa(\cdot, \cdot)$ . It provides a number of theoretical properties which justify its appropriateness. The associated proof is given in Appendix A.

**Proposition 1.** *The function  $\delta_\kappa : (\mathbb{R}^3 \times \mathbb{S}^2) \times (\mathbb{R}^3 \times \mathbb{S}^2) \rightarrow \mathbb{R}$  satisfies the following properties:*

1  $\delta_\kappa$  is a pseudometric on  $\mathbb{R}^3 \times \mathbb{S}^2$ . In particular, we have

$$(24) \quad \delta_\kappa[m_1, m_2] = 0 \quad \text{iff} \quad m_1 = m_2 \text{ or } m_1 = (\vec{v}, \vec{g}) \text{ and } m_2 = (\vec{v}, -\vec{g}).$$

2  $\delta_\kappa$  is invariant w.r.t. the left-action of  $\text{SE}(3)$  on  $\mathbb{R}^3 \times \mathbb{S}^2$ , that is, Euclidean invariant in the voxel space  $\mathbb{R}^3$  and rotation invariant in the gradient space  $\mathbb{S}^2$ .

3 The non-adaptive estimator in Equation (16) can be described by an  $(\mathbb{R}^3 \times \mathbb{S}^2)$ -convolution.

4 The method msPOAS is invariant w.r.t. the left-action of  $\text{SE}(3)$  on  $\mathbb{R}^3 \times \mathbb{S}^2$  if the non-adaptive weights base upon the discrepancy  $\delta_\kappa$  in Equation (23).

5 The non-adaptive weights  $\bar{w}_{mn}^{(k)}$  resulting from Equation (4) with  $\delta_\kappa$  in place of  $\Delta_\kappa$  are increasing w.r.t.  $k$  for all points  $m, n \in \mathbb{R}^3 \times \mathbb{S}^2$ .

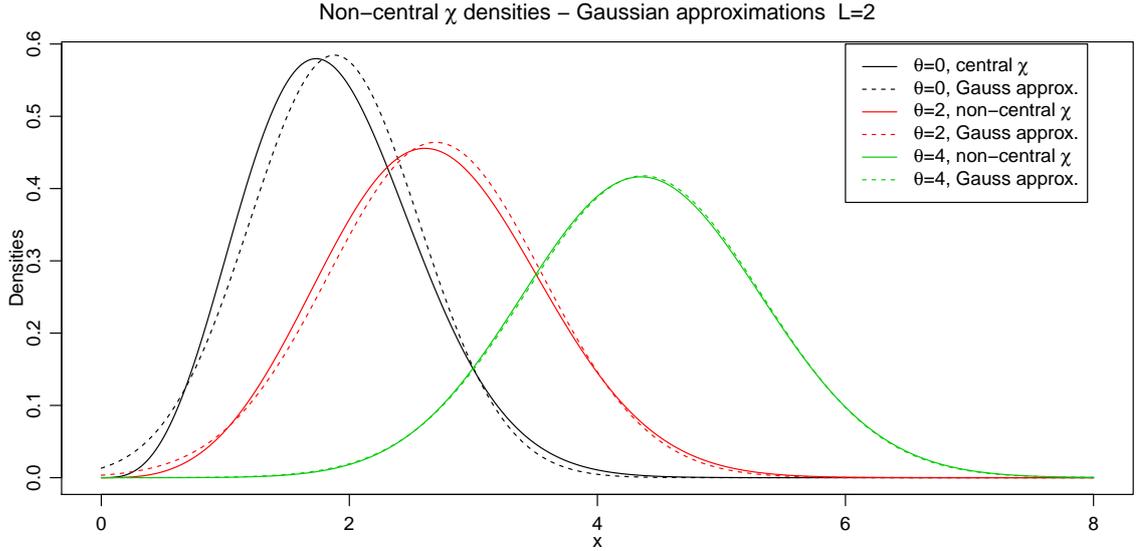


FIGURE 4. Densities of non-central  $\chi$ -distributions with  $2L = 4$  degrees of freedom and approximating Gaussian densities.

In contrast to a metric, a pseudometric allows the case  $\delta_\kappa(m_1, m_2) = 0$  with  $m_1 \neq m_2$ . Equation (24) is in agreement with the general interpretation of dMRI data, where we do not distinguish between one direction  $\vec{g}$  and its opposite  $-\vec{g}$  which follows from the symmetry of the diffusion process.

### 2.7. Simplified approximation of the statistical penalty.

The adaptation of the algorithm depends via the statistical penalty of the Kullback-Leibler divergence. This is based upon the probability distribution of the measurement observations, which we discussed in Section 2.3. As a result, for known noise variance  $\sigma^2$  the standardized data  $S_b(m)/\sigma$  may be considered as non-central  $\chi$ -distributed. Due to the lack of an explicit formula of the Kullback-Leibler divergence between two non-central  $\chi$ -distributions, we seek for an appropriate approximation. The following approach improves and accelerates the results in Becker et al. [2012, Appendix B].

We approximate the noncentral  $\chi$ -distribution with  $2L'$  degrees of freedom and noncentrality parameter  $\theta$  by a Gaussian distribution with mean

$$\mu(\theta, 2L') = \sqrt{\frac{\pi}{2}} \mathbb{I}_{1/2}^{(L'-1)} \left( -\frac{\theta^2}{2} \right)$$

and variance

$$v(\theta, 2L') = 2L' + \theta^2 - \mu^2(\theta, 2L')$$

such that the first and second moments of the original and approximating distributions coincide. This is motivated by Figure 4, where we compare various non-central  $\chi$ -distributions with their corresponding Gaussian approximation.

Then, we approximate the Kullback-Leibler divergence of two non-central  $\chi$ -distributions  $\chi_{2L',\theta_1}$  and  $\chi_{2L',\theta_2}$  by

$$\widetilde{\mathcal{KL}}(\chi_{2L',\theta_1}, \chi_{2L',\theta_2}) = \frac{[\mu(\theta_1, 2L') - \mu(\theta_2, 2L')]^2}{[v(\theta_1, 2L') + v(\theta_2, 2L')]} ,$$

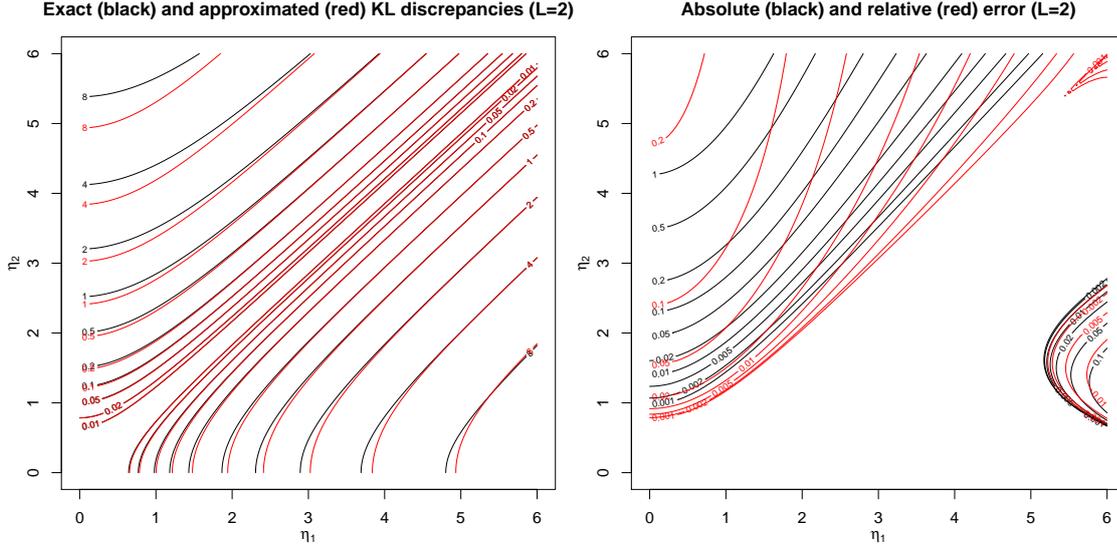


FIGURE 5. Left: Contour plots of Kullback-Leibler divergence between two non-central  $\chi$ -distributions with  $2L = 4$  degrees of freedom (black) and its symmetrized Gaussian approximations (red), Right: absolute (black) and relative (red) error of the approximation.

which can be interpreted as a symmetrized Kullback-Leibler divergence of the approximating Gaussian distributions. In the definition of the statistical penalty this leads to

$$(25) \quad \mathcal{KL} \left( \frac{\tilde{S}_b^{(k-1)}(m)}{\hat{\sigma}}, \frac{\tilde{S}_b^{(k-1)}(n)}{\hat{\sigma}} \right) \approx \frac{(\tilde{S}_b^{(k-1)}(m) - \tilde{S}_b^{(k-1)}(n))^2}{\hat{\sigma}^2 [\tilde{v}(\tilde{S}_b^{(k-1)}(m)/\hat{\sigma}) + \tilde{v}(\tilde{S}_b^{(k-1)}(n)/\hat{\sigma})]},$$

where

$$\tilde{v}(\tilde{S}_b^{(k-1)}(m)/\hat{\sigma}) = v(\tilde{\theta}_b(m), 2L') = 2L' + \left[ \mu^{-1} \left[ (\tilde{S}_b^{(k-1)}(m)/\hat{\sigma}) \right] \right]^2 - \left[ \tilde{S}_b^{(k-1)}(m)/\hat{\sigma} \right]^2,$$

with  $\tilde{\theta}_b(m) := \mu^{-1} \left( \tilde{S}_b^{(k-1)}(m)/\hat{\sigma} \right)$  and  $\mu$  as in Equation (12).

The relative and absolute error of this approximation are given in Figure 5, where the exact Kullback-Leibler divergence is calculated via a numerical integration. This would be prohibitive in msPOAS due to the computational workload. We show the results for 4 degrees of freedom. For other values of  $2L$  we observed a very similar behavior.

### 3. METHODS

**3.1. Simulations.** In order to provide some more intuition for the behavior of the algorithm, we show two simple examples on a one-dimensional design  $\mathcal{X} := \{1, \dots, 4000\}$ . Here, we use the R-package **aws** [Polzehl, 2013], where the Propagation-Separation approach has been implemented for one-, two- and three-dimensional designs. It differs from (ms)POAS via the applicable designs only. We consider a piecewise constant test function  $\theta_1(\cdot)$  and a piecewise polynomial test function  $\theta_2(\cdot)$ . Then, we simulate observations that follow a non-central chi-distribution with non-centrality parameter  $\theta_l$ ,  $l = 1, 2$  and  $2L' = 4$  degrees of freedom, i.e.,  $Y_i \sim \chi_4(\theta_l(X_i))$  for all  $X_i \in \mathcal{X}$ . The plots were provided by the function `aws()` setting `hmax := h^{(k*)} := 4000` and `lkern = "Triangle"` using the same kernel functions as in Eq. (15). The sizes of the considered test functions compare well with the parameters in our real data, see Section 3.2. Similar situations have been considered in Becker and Mathé [2013]

for Gaussian distributed noise. However, it was not sure that these results remain valid for  $\chi$ -distributed observations as the theoretical results in Becker and Mathé [2013] were restricted to another class of probability distributions.

**3.2. Experimental data.** We first re-analyzed the 7T dataset from a whole body MR scanner (MAGNETOM 7T, Siemens Healthcare, Erlangen, Germany) used and described already in Becker et al. [2012]. The scanner was equipped with gradients with a peak amplitude of  $70\text{mT/m}$  and a maximum slew rate of  $200\text{T/m/s}$  (SC72, Siemens Healthcare, Erlangen, Germany). Diffusion weighting gradients were applied for 60 directions at a  $b$ -value of  $1000\text{s/mm}^2$ . 7 interspersed  $S_0$  images were acquired. The scan was repeated 4 times. An optimized monopolar Stejskal-Tanner sequence according to Morelli et al. [2010] together with the ZOOPPA approach described in Heidemann et al. [2012] has been used for the scan. The experiment was performed using a single channel transmit, 24-channel receive phased array head coil (Nova Medical, Wilmington, MA, USA). 91 slices with 10% overlap were acquired at a field-of-view (FoV) of  $143 \times 147\text{mm}^2$  resulting in an isotropic high resolution of  $800\mu\text{m}$ . Further imaging protocol parameters were: TR 14.1s, TE 65ms, BW  $1132\text{Hz/pixel}$ , ZOOPPA acceleration factor of 4.6. A healthy young volunteer was scanned four times using this protocol in one session after obtaining written informed consent in accordance with the ethical approval from the University of Leipzig. Total acquisition time was 65min.

The second example data has been acquired on a 3T MAGNETOM Trio scanner (Siemens AG, Healthcare Sector, Erlangen, Germany) using a reduced FoV-technique as described in Heidemann et al. [2010]. The FoV was  $161 \times 58\text{mm}$  around the motor cortex resulting in an isotropic in-plane resolution of  $1.2\text{mm}$ . 34 slices of  $1.3\text{mm}$  slice thickness. Diffusion weighted data were acquired at 2 different  $b$ -values:  $b = 800\text{s/mm}^2$  and  $b = 2000\text{s/mm}^2$  each with 100 gradient directions as suggested by Caruyer et al. [2011]. 21 interspersed  $S_0$ -images at  $b = 20\text{s/mm}^2$  were acquired. One healthy adult volunteer (male, age: 36) participated in the study approved by the local ethics committee after giving written informed consent. The total scan time was 22min. Data was corrected for eddy currents and motion artifacts [Mohammadi et al., 2010] using the ACID toolbox (<http://www.diffusiontools.com/>) for SPM.

Both datasets have been smoothed using the msPOAS method described in this article. The adaptation bandwidth  $\lambda$  of the procedure has been fixed at a value of 20 as detailed in Section 2.5. The number of effective coils, that determines the degrees of freedom of the non-central  $\chi$ -distribution, see Section 2.3, is very difficult to estimate from the data. Fortunately, msPOAS is relatively robust against misspecification of  $L$ , as will be shown for the second dataset in Figure 8. We therefore use  $L' = 2$  for the first dataset to mimic an average influence of two coils to the resulting distribution, and  $L' = 1, 4, 16$  for the second dataset to analyze the dependence of the results from the choice of  $L'$  which is difficult to estimate from the data. We do not explicitly analyze the dependence of the results on the choice of the noise variance  $\sigma$  or on the adaptation bandwidth  $\lambda$  as the qualitative behavior can be extracted from the definition of the adaptive weights in Eq. (18), see also the discussion section.  $\kappa_0$  has been chosen in accordance with our suggestions in Section 2.4 based on the number of gradient directions for both datasets, where 0.5 for the first and 0.3 for the second dataset have shown best performance of msPOAS.

We estimated the noise standard deviation  $\sigma$  using the method described in Becker et al. [2012] and the methods “Bk-M1- $\chi$ ” and “Bk-M2- $\chi$ ” described in Aja-Fernández et al. [2009]. The estimates for  $\sigma$  from the former method are rather independent of  $L'$ . For the first dataset we get consistently values around 75, for the second dataset values around 30 have been estimated. The results from the latter two methods depend on  $L$ , providing

$$\hat{\sigma}_{Bk-M1-\chi} = 69.2 \frac{\Gamma(L')}{\sqrt{2L'}\Gamma(L'+1/2)}, \quad \hat{\sigma}_{Bk-M2-\chi} = 58.3/(2L')$$

for the first and

$$\hat{\sigma}_{Bk-M1-\chi} = 41.7 \frac{\Gamma(L')}{\sqrt{2L'}\Gamma(L' + 1/2)}, \quad \hat{\sigma}_{Bk-M2-\chi} = 33.7/(2L')$$

for the second data set. However, estimation of  $L'$  is very difficult. Based on our experience with these variance estimates we decided to use for our calculations, i.e.,  $\sigma = 75$  for the single-shell data and  $\sigma = 30$  for the double-shell data. Finally we note, that based on these estimates for  $\sigma$  the noncentrality parameter of the considered  $\chi$ -distributions is in the range of  $[0, 8]$ , therefore we considered these values in the simulated univariate example and for the propagation level.

For the first (single-shell) dataset we explicitly compare the msPOAS results with the results of POAS obtained in Becker et al. [2012]. There, we used  $\lambda = 12$ ,  $\kappa_0 = 0.6$ ,  $\sigma = 66$ , and  $L' = 2$ . The different choices of the adaptation bandwidth  $\lambda$  in POAS and msPOAS may result from the explicit coupling of the  $S_0$ -images in the adaptation. The simplified discrepancy and the new approximation of the Kullback-Leibler divergence should have only minor impact. For comparison we estimate the diffusion tensor model, as it is the most widely used in practice and we also did in our previous publication [Becker et al., 2012]. The diffusion tensor has been estimated using a non-linear method described in Polzehl and Tabelow [2009]. Then, FA maps have been calculated based on the tensor estimates.

For the second dataset we compared msPOAS with the results of a simple approach that applies POAS separately on each shell of the double-shell dataset. We denote this latter approach for smoothing multi-shell data the *naive POAS method*.

For the naive POAS method we used the same variance estimate  $\sigma = 30$  as for msPOAS and a medium  $L' = 4$ . We chose an adaptation parameter  $\lambda$  for POAS such that a fair comparison can be made. For the naive POAS approach on the multi-shell data we searched for a value of  $\lambda$  that provided optimal results, i.e.,  $\lambda = 6$ . The recombined smoothed diffusion weighted data again forms a double-shell dataset.

For the multi-shell dataset we estimated a one-stick-one-ball model [Behrens et al., 2003] to evaluate the variability of directional estimates in a model that goes beyond the diffusion tensor model. To do this we used the FSL standard parameters and sampled 50 directions for the sticks. The model has been evaluated for the original data, for the msPOAS result, and for the result from the naive POAS approach.

We also calculated fiber tracks using a streamline FACT algorithm [Mori et al., 1999] implemented in the package **dti** based on diffusion tensor modeling of the multi-shell data.

**3.3. Software.** For data preparation, motion and eddy-current correction we used SPM [Friston et al., 2006] and the ACID-toolbox [Mohammadi et al., 2010].

The computations for msPOAS and POAS in this study were performed with our **R**-package **dti** [Tabelow and Polzehl, 2013] (version 1.1-5). For the simulations we used our **R**-package **aws** [Polzehl, 2013]. Both packages are freely available on CRAN (<http://cran.r-project.org>). A detailed description of the usage of the package **dti** can be found in Polzehl and Tabelow [2011]. The implementation uses FORTRAN and native **R**-code. Diffusion tensor estimates and FA maps were produced using the package **dti**.

The one-stick-one-ball model has been computed using FSL [Smith et al., 2004].

## 4. RESULTS

**4.1. Simulations.** In Figure 6 we present the results for the one-dimensional test-functions  $\theta_1(\cdot)$  and  $\theta_2(\cdot)$  with  $\lambda = 20$ . For both examples, we show two plots with increasing location bandwidths  $h_1 = 485, 3610$  and  $h_2 = 81, 3610$  corresponding to the iteration steps  $k_1 = 30, 39$  and  $k_2 = 22, 39$ . This

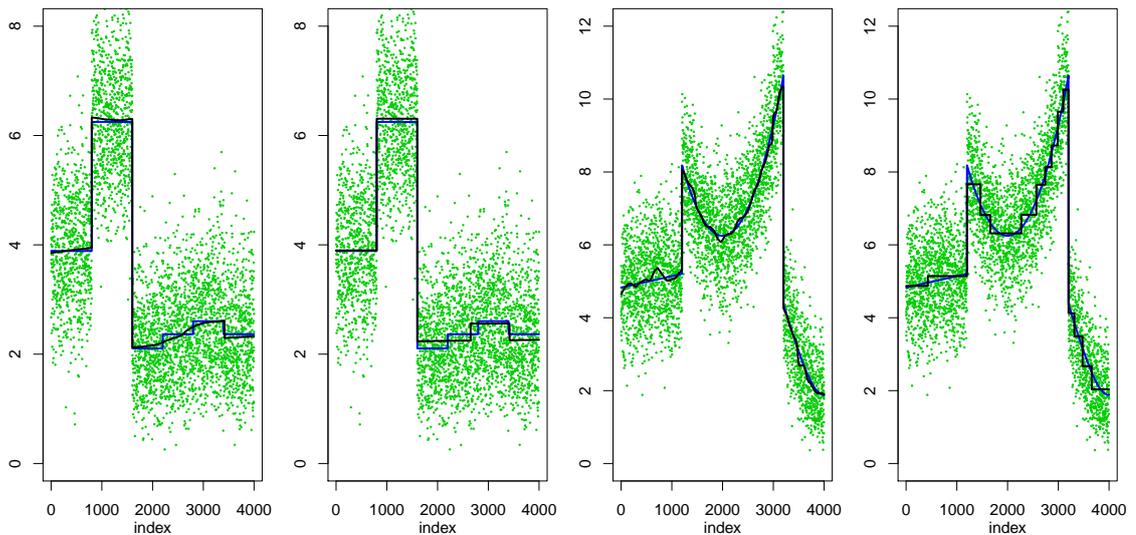


FIGURE 6. Univariate examples for adaptive smoothing of noncentral  $\chi$ -distributed observations for an optimal and for an extremely large bandwidth. From left to right: Local constant noncentrality parameter (NCP),  $\theta_1(\cdot)$ , with location bandwidths  $h_1 = 485, 3610$ , and locally smooth NCP,  $\theta_2(\cdot)$ , with location bandwidths  $h_2 = 81, 3610$ . Observations are presented as green dots, the expected values as blue lines, smoothing results are shown as black lines.

illustrates the progress of the estimation function during iteration. The maximal number of iterations is  $k^* = 39$ . In the steps  $k_1 = 30$  and  $k_2 = 22$  the mean-squared error is minimal.

We may conclude that msPOAS provides the following (heuristic) properties.

- The algorithm separates homogeneous compartments with sufficiently large discontinuities. This allows (almost) consistent estimation of the unknown expected value, which refers to the unknown parameter function  $\theta_1(\cdot)$ . This is illustrated in Figure 6, see  $x \in \{1, \dots, 1600\}$  in the local constant example.
- The separation property fails if the discontinuities are small. In this case, the algorithm leads to a bounded estimation bias since different homogeneous compartments are treated as one, see  $x \in \{1601, \dots, 4000\}$  in the same plots as before.
- For model misspecification, i.e., for piecewise smooth parameter functions the estimator is forced into a step function, see the piecewise smooth example  $\theta_2(\cdot)$  in Figure 6. This can be interpreted as an intrinsic stopping criterion of the algorithm as it ensures a bounded estimation bias. Therefore, the maximal location bandwidth  $h^{(k^*)}$  and as a consequence the number of iteration steps  $k^*$  are restricted by the resulting computational workload, only. The choice of  $h^{(k^*)}$  should ensure a reasonable computation time still providing a certain smoothness within homogeneous regions.

**4.2. Experimental data.** The multi-shell data has been processed by our implementation of msPOAS within 15 minutes on a single core of a HP SL390s compute server with an Intel Xeon, Six-Core 3467 MHz. The same computation using POAS for each shell separately required more than one hour. This significant acceleration effect is due to the simplifications and approximations for the discrepancy and the Kullback-Leibler divergence. For the much larger single-shell dataset msPOAS used 3 hours and 18 minutes, while POAS used 4 hours and 37 minutes computation time on a single core of the same

machine. Our implementation has been parallelized using OpenMP which significantly speeds up the computation compared with these single core results.

We start with the analysis of the first (single-shell) dataset, that has been used in Becker et al. [2012] to introduce POAS. In Figure 7 we show a color-coded FA map for an axial slice of the original data (a).

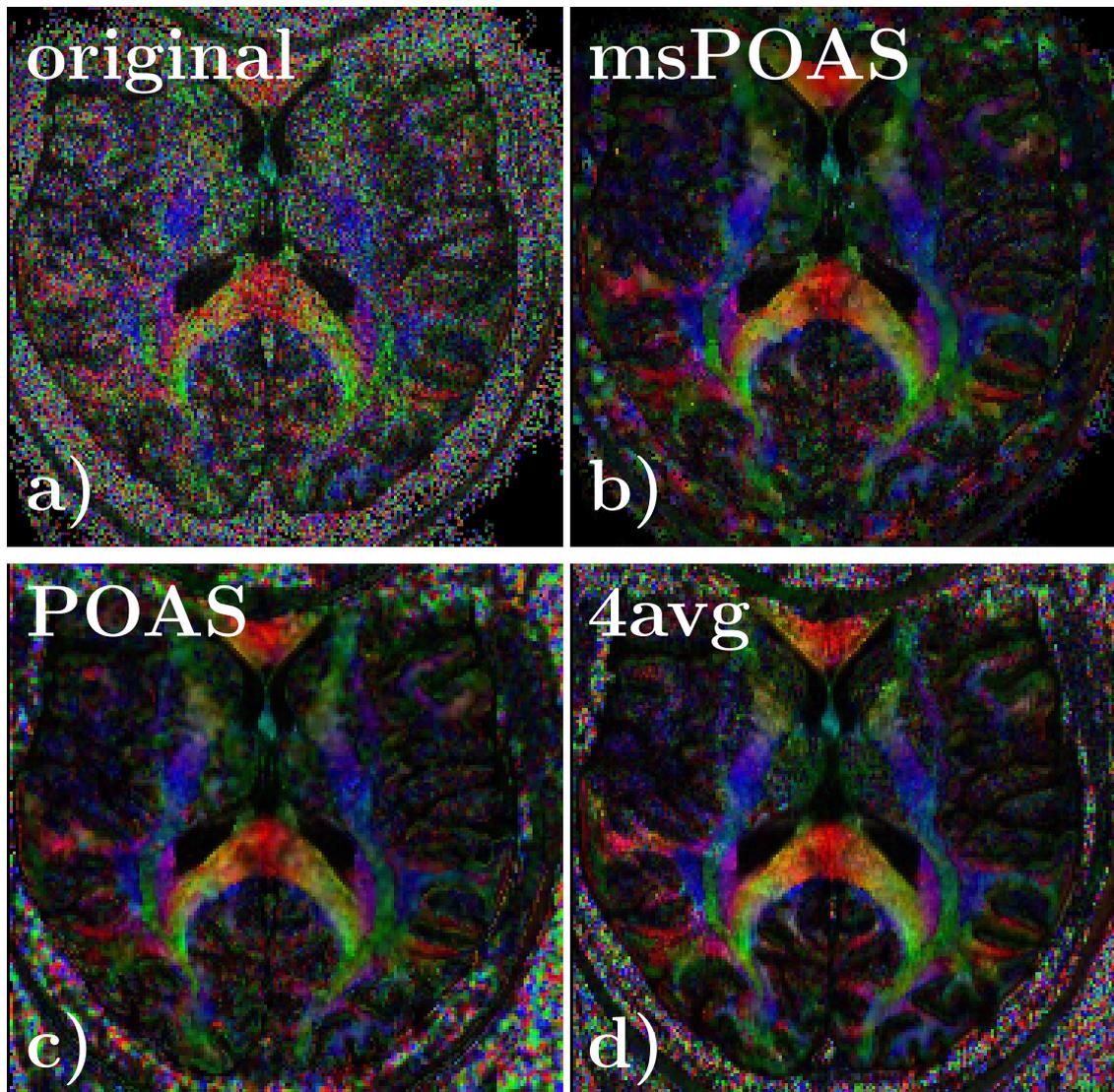


FIGURE 7. Comparison of color-coded FA maps of the single shell dataset also used in Becker et al. [2012]. a) Original noisy data, b) msPOAS reconstruction, c) POAS reconstruction as in Becker et al. [2012], d) mean image of four repeated measurements from the same session.

The smoothing effect by msPOAS is obvious in Figure 7(b), while there is no blurring effect on borders. For comparison we give again the result of the single-shell POAS from Becker et al. [2012] for the same slice. There are of course small differences in detail: The first reason is, that msPOAS for a single shell is not identical with POAS, as information from diffusion weighted and  $S_0$  images is now combined to improve adaptivity, see Section 2.2. Second, the parameter choice for  $\lambda$  is different in accordance with the advanced parameter choice strategy detailed in Section 2.5. Third, we choose a smaller value for  $\kappa_0$  as the new discrepancy leads to slightly smaller distances on the sphere, see Figure 3. In Figure 7(d) we

show the color-coded FA obtained from all four (unsmoothed) scans. This image can to a certain degree serve as a ground truth for the evaluation of the msPOAS (or POAS) result.

We now provide the results for the multi-shell dataset considering the  $b$ -values  $b = 800\text{s}/\text{mm}^2$  and  $b = 2000\text{s}/\text{mm}^2$ . In Figure 8 we show the diffusion weighted data for a slice before and after reconstruction using msPOAS. The adaptive smoothing effect of msPOAS is apparent for both shells. Simultaneously, we demonstrate that the result of msPOAS is rather robust against misspecification of the effective number of coils  $L'$  by showing corresponding results for  $L' = 1, 4, 16$ . This is a very helpful property of msPOAS as the estimation of  $L'$  from the data is a highly non-trivial task and the value of  $L'$  might not be homogeneous in voxel space. For the question, how the msPOAS results depend on the estimate of  $\sigma$  and on the choice of  $\lambda$  see the discussion section.

In the last row e) we show the result of a naive POAS analysis. For a fair comparison of msPOAS with the naive POAS approach, we choose the adaptation parameter  $\lambda = 6$  for POAS, such that the shell at  $b = 800\text{s}/\text{mm}^2$  shows similar results as the msPOAS approach. The outer shell at  $b = 2000\text{s}/\text{mm}^2$  has a significantly lower signal-to-noise ratio and the size of discontinuities approximately scales with the  $b$ -value. As a result, the diffusion-weighted image is somewhat blurred as the image contrast is not sufficient for the adaptation to work. Our new method msPOAS is therefore able to achieve much better results than the naive approach, as the larger contrast at the inner shells is used to improve adaptation also at the shells with higher  $b$ -value.

In Figure 9 we show the fiber track reconstruction using a streamline FACT algorithm [Mori et al., 1999] after DTI modeling for the multi-shell data. Comparison is given for smoothed multi-shell data after the naive POAS approach (left) and after msPOAS (right). We include only fiber tracks with a minimal length of 25 line segments for a better visibility. The Figure shows, that after msPOAS the reconstruction of the fibers even with this very simple algorithm is much richer than the one obtained from the naive POAS approach, see for example the occurrence of the U-Fibers. This confirms the observation from Figure 8 that msPOAS indeed leads to improved results compared to the naive single-shell POAS approach.

Finally, we illustrate quantitatively that directional information is much less variable allowing, e.g., more precise fiber tracking, see also Figure 9. We therefore analyze in Figure 10 the variability of the estimated directions in a one-stick-one-ball model for a small region. The central slice of the region consisting of 10 slices is shown in Figure 10a). The white square illustrates the location of the  $20 \times 20$  voxel region-of-interest. The one-stick-one-ball model estimated a sample of 50 directions for the stick. In Figure 10b) we illustrate for one voxel these directions by plotting them using their representation by spherical angles  $\phi$  and  $\theta$ . In light red we show the directions for the original non-smoothed data, in black we show the corresponding points after smoothing with msPOAS. We now want to quantify the observation in Figure 10b) that these points are much more concentrated. We therefore compare the mean angular deviation (MAD) for the sample directions from its mean direction in the original data and after msPOAS reconstruction. For the voxel considered in Figure 10b) the circles illustrate the location of the mean direction and this MAD which corresponds to the radius. The mean directions for the original data and the data after msPOAS of course differ. In total, after msPOAS the MAD is much smaller implying that the estimate is much less variable. In Figure 10c) we plot the ratio of the MAD for the msPOAS reconstructed data with the MAD of the original non-smoothed data as a function of the fractional anisotropy in a DTI model for the data. If there is some improvement in the sense, that the directional estimates are less variable after msPOAS, this ratio is expected to be less than 1. The horizontal line in Figure 10c) represents this value. We observe that for most voxels especially for those with higher anisotropy the ratio is indeed smaller than 1 (mean value 0.57 for  $FA > 0.3$ ). In Figure 10d) we show the same plot for the data analyzed with the naive POAS approach (mean value 0.64 for  $FA > 0.3$ ). Obviously for more isotropic voxels the variability of the directional estimate is much larger due to the lower anisotropy. In Figure 10e) we show the improvement for the variability of the directional estimates after msPOAS for the ‘‘white matter’’ voxels only, which we loosely characterize by  $FA > 0.3$ ) here. These voxel may of course also contain gray

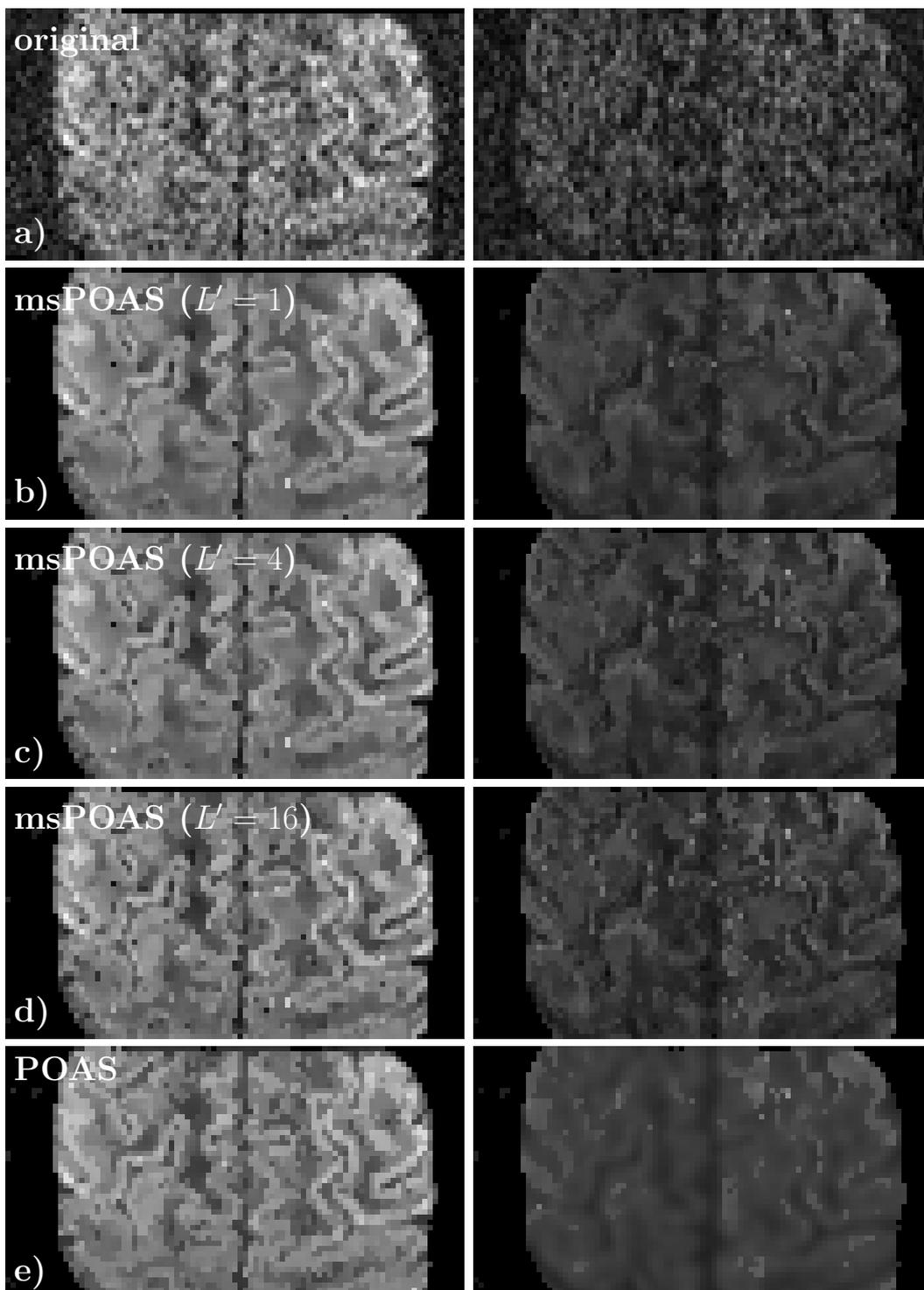


FIGURE 8. Comparison of the diffusion weighted data for a slice of the double-shell experimental dataset. The left row shows the data at the  $b = 800\text{s/mm}^2$ -shell, the right row at the  $b = 2000\text{s/mm}^2$ -shell: a) Original noisy data, b) msPOAS reconstruction ( $L=1$ ), c) msPOAS reconstruction ( $L=4$ ), d) msPOAS reconstruction ( $L=16$ ), e) reconstruction using POAS on each shell separately.

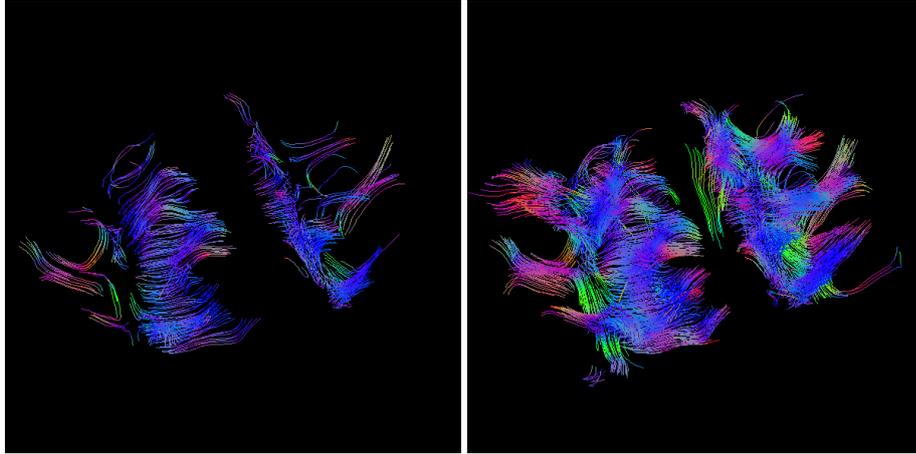


FIGURE 9. Fiber tracks from a diffusion tensor model of the double-shell data using a FACT algorithm with a minimal fiber length of 25 segments for better visibility. Left: Tracks from POAS reconstruction. Right: Tracks after msPOAS reconstruction.

matter. The main point, however, is that it possesses a larger enough directional anisotropy. The MAD is less after msPOAS for most of the voxels. Finally, we demonstrate, that a slight improvement of msPOAS compared to the naive POAS approach can be found: In Figure 10f) we plot the improvement of the MAD compared to the non-smoothed data (ratio) for the naive POAS method versus the msPOAS approach. Most points lie below the identity line.

## 5. DISCUSSION

We developed a novel approach for noise reduction in multi-shell diffusion-weighted data (msPOAS). For a maximum usage of information it exploits the geometry of the measurement space formed by (voxel) position and (gradient) orientation as well as the vector structure of the measurements on the different shells, including the  $S_0$ -images. The method avoids blurring of the structures observed in dMRI and considerably enhances the signal-to-noise ratio. We have demonstrated that msPOAS is relatively robust against the choices of method parameters and against misspecification of the number of effective coils at a position, which is the parameter that is most difficult to get from the data.

The various examples in Section 4 nicely confirm the effectiveness of our new noise reduction procedure msPOAS. Through a series of examples with simulated data and real dMRI data we illustrated its behavior. The method avoids blurring and consequently preserves observed structure by restricting smoothing to (almost) homogeneous compartments. To our knowledge, msPOAS is the first procedure which applies directly to the diffusion weighted images of multi-shell measurements by smoothing simultaneously all  $q$ -shells. This includes an explicit coupling of the smoothing of the  $S_0$ -image with the other shells. The advantages of this approach become obvious when comparing smoothing results of msPOAS with our former procedure POAS for single-shell data, applied to each  $q$ -shell separately.

Re-using data from our previous publication [Becker et al., 2012] we demonstrated that msPOAS achieved at least similar results on single-shell data as POAS, which was explicitly designed for this kind of data. We then presented a multi-shell dataset with high spatial and angular resolution. This not only showed that msPOAS indeed reduces the noise in the data, but also provides superior results than the naive POAS approach for each shell. The reason is, that the diffusion-weighted signal at the shells with higher  $b$ -value is further attenuated. Discontinuities are scaled down, such that with a comparable noise level, the contrast diminishes. This provides a more complicated situation for adaptation using (single-shell) POAS on higher shells. In contrast, msPOAS uses a vector representation of the data for all shells, and

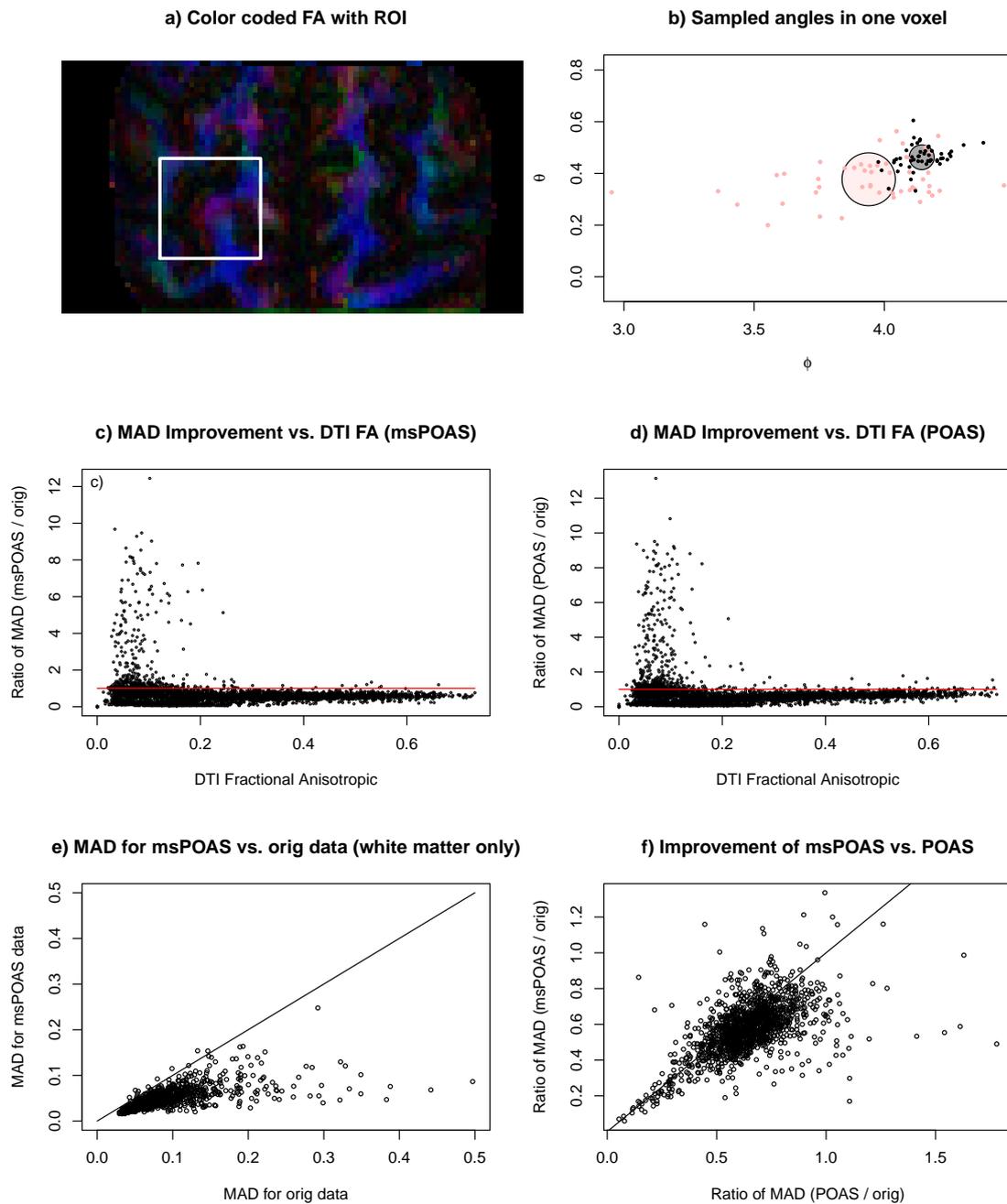


FIGURE 10. Mean angular deviation (MAD) in a 1-stick-1-ball model from the estimated mean direction of the stick: a) Color coded FA from diffusion tensor model with the region-of interest that has been analyzed, b) Sampled “stick directions” expressed in spherical angles  $\phi$  and  $\theta$  for the original data (light red) and the msPOAS reconstruction (black) for one white matter voxel, c) Ratio of the MAD for the msPOAS reconstruction and for the original data, d) Ratio of the MAD for the POAS reconstruction for each shell and for the original data, e) Comparison of the MAD for the msPOAS reconstruction and for the original data using voxels with a minimal fractional anisotropy of 0.3, f) Comparison of the ratio of the MAD for the msPOAS and the original data with its counterpart from the POAS reconstruction on each shell using voxels with a minimal fractional anisotropy of 0.3.

is thus able to achieve the same quality of adaptation on all shells. We demonstrated this superiority by comparing the smoothed diffusion weighted images directly, by showing resulting fiber tracks, and by quantitatively analyzing the variability of directional estimates in the sticks-and ball model.

As msPOAS and POAS are both based on the Propagation-Separation approach the general behavior is quite similar for example in case of partial volume effects or in their relation to smoothing methods based on anisotropic diffusion. We therefore refer the reader also to the discussion of POAS in Becker et al. [2012, Section 5] for comparison.

In order to simplify and accelerate msPOAS compared to the original POAS approach, we introduced several modifications. These modification can be directly used also in the original POAS approach. Due to computational complexity, the Kullback-Leibler divergence had to be approximated already in POAS [Becker et al., 2012]. Our new approximation, introduced in this article, considerably accelerates the computations while also behaving better (not shown here). The new discrepancy provides several useful mathematical properties. It is a pseudometric and ensures Euclidean invariance in voxel space and rotation invariance in gradient space. In contrast to the original POAS approach the embedding of  $\mathbb{R}^3 \times \mathbb{S}^2$  into the special Euclidean motion group  $SE(3)$  is only needed for the mathematical theory, but no longer for definition of the algorithm. Therefore, it is much simpler to understand than the previous approach.

MsPOAS is a method that can be applied in case of low signal-to-noise ratio. In such situations modeling the original data by any method leads to a high variability of estimated parameters and characteristics. This variability is essentially reduced by the proposed adaptive smoothing procedures without seriously compromising the structural information in the data. Note, that in case of high signal-to-noise ratio the algorithm essentially leaves the data unchanged.

The presented procedure benefits from a high number of measured gradient directions, and also from sampling additional shells. If all  $q$ -shells have an identical gradient scheme then the algorithm could be accelerated as no spherical interpolation is needed. In contrast, varying gradient schemes benefit from a higher angular resolution, but possibly suffer from a slightly biased statistical penalty due to the interpolation. In any case, increasing the number of gradients limits a possible bias introduced by assuming local constant parameters on the sphere. If the number of gradient directions is too low, say less than 20, msPOAS reduces to a separate adaptive smoothing of the diffusion weighted images with fixed gradients. This requires a sufficient image contrast.

One important aspect of the presence of noise in dMRI data is the bias in the estimated quantities. The parameter characterizing the diffusion weighted signal, i.e.  $\theta_{m,b}$ , differs from the expected value  $\mathbb{E}S_b(m)$  of the signal distribution in a noisy situation, see Equations (12) and (13) in Section 2.3. It is, however, the expected value, that is actually measured. The relative difference between  $\mathbb{E}S_b(m)$  and  $\theta_{m,b}$  is especially large in case of a small non-centrality parameter, i.e., small signal-to-noise ratio. MsPOAS essentially leads to largely improved, i.e., less variable estimates for the expectation value of the distribution. Therefore, Equation (20) can be in principle used for assessing the true parameter  $\theta_{m,b}$  and correct for this bias. In practice, this requires a precise determination of the noise variance  $\sigma^2$  and the effectively applied number of receiver coils  $L'$ , which is a challenging problem in itself and goes beyond the scope of this article.

The msPOAS algorithm uses  $\sigma$  and  $L'$  as data-dependent input parameters. They should be estimated separately by any method that is available and suitable for the data. In this article we assumed homogeneous  $\sigma$  and  $L'$  over the voxel space. Fortunately, msPOAS has proven to be relatively robust with respect to misspecification of  $L'$ . On the other hand, msPOAS can be easily adapted to a heteroscedastic situation at the cost of more complicated presentation and larger need for memory and computation time. If  $\sigma$  is underestimated, the adaptation will be very restrictive such that msPOAS does not change

the data at all. In contrast, if  $\sigma$  is overestimated, oversmoothing occurs with blurring of discontinuities. An accurate estimate for  $\sigma$  can therefore improve the results of msPOAS.

Besides  $\sigma$  and  $L'$  there are several other parameters that influence the results of msPOAS. Most of them can be chosen independent of the specific realization of the data. Here, we summarize our recommendations for their selection. The kernel functions have only very minor impact on the results, see Section 2.4 and can, e.g., be chosen as in Eq. (15) for efficient computation. For the adaptation bandwidth  $\lambda$  there exist two extreme cases,  $\lambda = \infty$  and  $\lambda = 0$  as can be seen from Eq. (18) and already mentioned in Section 2.5: In the former case, the msPOAS result is a non-adaptive estimate as the adaptation term always equals 1. In the latter case, the data will not be changed by msPOAS as the statistical penalty will be infinity for any pair of distinct locations in measurement space.  $\lambda$  can be basically chosen using the propagation condition developed in Section 2.5. Fortunately, it does not depend on the realization of the noise and only weakly on the noise distribution which in our case also includes the values of  $\sigma$  and  $L'$  for the data at hand. However, from our experience the  $\lambda$ -values used in this paper are a good starting point for any dataset. Adjustment can be easily done having in mind the two limits of msPOAS, that is,  $\lambda = \infty$  and  $\lambda = 0$ . It is an important point that the choice of  $\lambda$  and the estimate of  $\sigma$  influence msPOAS in a similar manner such that a misspecification of  $\sigma$  can be to some degree balanced by the choice of  $\lambda$ . The choice of  $\kappa_0$  influences the amount of smoothing on the sphere in the first iteration steps. Again, the values used in this paper can serve as a good reference. For a larger number of gradients or relatively good signal-to-noise ratio it can be chosen slightly smaller. The number of iteration steps  $k^*$  relates to the last value in the sequence of increasing location bandwidths  $h^{(k)}$ . Its choice should balance the computation time and the desired smoothness within homogeneous regions, see Section 4.1. As discussed next, possible consequences of a violated structural assumption can be reduced by diminishing  $k^*$ .

We should critically discuss the assumption of msPOAS, that dMRI data is characterized by regions with homogeneous diffusion-weighted signal separated by discontinuities. This assumption is certainly only an approximation of a more realistic piecewise smooth model. In Figure 6 we showed, that in case of a violation of this assumption msPOAS forces the final estimator into a step function if the maximal location bandwidth  $h^{(k^*)}$  is sufficiently large. The same Figure also shows, that intermediate steps of the iteration of msPOAS show results with less error compared to the true situation. Thus, in practice a careful choice of  $k^*$  and visual inspection might improve the results. Even, if  $k^*$  is chosen large, the analysis shows, that the final stable msPOAS result is a step function approximation.

Additionally, we recall that msPOAS requires independent data in each point of the measurement space. As registration introduces spatial correlation into the data, applying msPOAS before registration may seem preferable. On the other hand, unregistered data might lead to spurious discontinuities, which msPOAS may identify. In our experience, msPOAS benefits from registered data without being harmed by the small spatial correlation caused by it. It would be a nice piece of future research to combine registration methods with msPOAS to further improve results.

Finally, we emphasize that msPOAS should not be combined with other smoothing methods. Previous application of, e.g., a Gaussian filter would only hamper the performance of msPOAS due to the induced spatial correlation and the resulting blurring. In fact, msPOAS performs very similar to a non-adaptive filter within homogeneous regions, while it does not blur the observed structure at discontinuities.

## 6. CONCLUSION

Multi-shell dMRI acquisition has become a wide-spread application. In this article we introduced a new noise reduction method msPOAS (multi-shell position orientation adaptive smoothing) for this data. The method does not mask the real structure by blurring borders. It outperforms naive approaches, which smooth each  $q$ -shell separately, by being the first algorithm that combines information from all shells for structural adaptive smoothing. Due to its computational efficiency it can be also applied in practical

applications. One of the strengths of msPOAS is that it is applied directly to the dMRI data and does not use any dMRI model, see Assemlal et al. [2011] for an overview. Thus, the method does not introduce a bias towards any of them. As a consequence any model can be used after smoothing for further data analysis.

## 7. ACKNOWLEDGMENT

We thank A. Anwander and R. Heidemann for the permission to re-use the first (single-shell) dataset for comparison.

## APPENDIX A. THEORETICAL PROPERTIES OF THE SIMPLIFIED METRIC

We want to verify Proposition 1 in Section 2.6. For that purpose, we remind some results from Duits and Franken [2011]. First, we define a  $(\mathbb{R}^3 \times \mathbb{S}^2)$ -convolution via the embedding of  $\text{SE}(3)$  into  $\mathbb{R}^3 \times \mathbb{S}^2$  by

$$(26) \quad (p *_{\mathbb{R}^3 \times \mathbb{S}^2} q)(\vec{v}_1, \vec{g}_1) := \int_{\mathbb{R}^3} \int_{\mathbb{S}^2} p(\mathbf{R}_{\vec{g}_2}^{-1}(\vec{v}_1 - \vec{v}_2), \mathbf{R}_{\vec{g}_2}^{-1} \vec{g}_1) q(\vec{v}_2, \vec{g}_2) d\mu(\vec{g}_2) d\vec{v}_2,$$

where  $\mathbf{R}_{\vec{g}} \in \text{SO}(3)$  is any rotation with  $\mathbf{R}_{\vec{g}} \mathbf{e}_z = \vec{g} \in \mathbb{S}^2$  and  $\mu$  denotes the surface measure on  $\mathbb{S}^2$ . Next, we consider the left-regular action  $L : \text{SE}(3) \rightarrow \mathcal{B}(\mathbb{R}^3 \times \mathbb{S}^2)$  of  $\text{SE}(3)$  on  $\mathbb{R}^3 \times \mathbb{S}^2$ , where  $\mathcal{B}(H)$  denotes the space of all linear, bounded and invertible functions  $H \rightarrow H$ . This action is defined by

$$L_n(m) := (\mathbf{R}^{-1}(\vec{v} - \vec{w}), \mathbf{R}^{-1} \vec{g}) \in \mathbb{R}^3 \times \mathbb{S}^2, \quad n := (\vec{w}, \mathbf{R}) \in \text{SE}(3), m := (\vec{v}, \vec{g}) \in \mathbb{R}^3 \times \mathbb{S}^2.$$

Using the same notation we describe the left-regular action  $\mathcal{L}$  of  $\text{SE}(3)$  on  $\mathbb{L}_2(\mathbb{R}^3 \times \mathbb{S}^2)$  by

$$\mathcal{L}_n[U](m) := U(L_n(m)) = U(\mathbf{R}^{-1}(\vec{v} - \vec{w}), \mathbf{R}^{-1} \vec{g}), \quad U \in \mathbb{L}_2(\mathbb{R}^3 \times \mathbb{S}^2).$$

An operator

$$\Phi : \mathbb{L}_2(\mathbb{R}^3 \times \mathbb{S}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^3 \times \mathbb{S}^2)$$

is called left-invariant if for all  $n \in \text{SE}(3)$  it holds

$$\mathcal{L}_n \circ \Phi = \Phi \circ \mathcal{L}_n.$$

In Becker et al. [2012], we limited ourselves to left-invariant  $\text{SE}(3)$ -methods. In the following, we show that msPOAS is indeed left-invariant.

*Proof sketch of Proposition 1.*

### 1 Pseudometric:

- $\delta_\kappa(m_1, m_2) = 0$  iff  $\vec{v}_1 = \vec{v}_2$  and  $\arccos |\langle \vec{g}_1, \vec{g}_2 \rangle| = 0$ . The latter holds iff  $|\langle \vec{g}_1, \vec{g}_2 \rangle| = 1$ , that is, iff  $\vec{g}_1 = \vec{g}_2$  or  $\vec{g}_1 = -\vec{g}_2$  by definition of  $\mathbb{S}^2$ .
- $\delta_\kappa$  is non-negative since the Euclidean metric  $\|\cdot - \cdot\|$  and  $\arccos : [0, 1] \rightarrow [0, \pi/2]$  are non-negative.
- The symmetry of  $\delta_\kappa$  follows from the symmetry of the Euclidean metric and the scalar product.
- Triangle inequality:  $\arccos |\langle \vec{g}_1, \vec{g}_2 \rangle|$  measures the arc length of the great circle passing through the gradients  $\vec{g}_1, \vec{g}_2 \in \mathbb{S}^2$ . The great circles are known to be the geodesics of  $\mathbb{S}^2$  relating to the shortest surface-path between the two points  $\vec{g}_1, \vec{g}_2 \in \mathbb{S}^2$ . Hence, it satisfies the triangle inequality such as the Euclidean metric and as a consequence  $\delta_\kappa$ :

$$\delta_\kappa(m_1, m_3) \leq \delta_\kappa(m_1, m_2) + \delta_\kappa(m_2, m_3), \quad m_1, m_2, m_3 \in \mathbb{R}^3 \times \mathbb{S}^2.$$

2 Left-invariance of  $\delta_\kappa$ : Let  $n := (\vec{w}, \mathbf{R}) \in \text{SE}(3)$  and  $m_i := (\vec{v}_i, \vec{g}_i) \in \mathbb{R}^3 \times \mathbb{S}^2$ ,  $i = 1, 2$ . Then, we get by the rotation invariance of the Euclidean metric and of the scalar product

$$\begin{aligned} \delta_\kappa(L_n[m_1], L_n[m_2]) &= \delta_\kappa[(\mathbf{R}^{-1}(\vec{v}_1 - \vec{w}), \mathbf{R}^{-1}\vec{g}_1), (\mathbf{R}^{-1}(\vec{v}_2 - \vec{w}), \mathbf{R}^{-1}\vec{g}_2)] \\ &= \|\mathbf{R}^{-1}(\vec{v}_1 - \vec{w} - \vec{v}_2 + \vec{w})\| + \varkappa^{-1} \arccos |\langle \mathbf{R}^{-1}\vec{g}_1, \mathbf{R}^{-1}\vec{g}_2 \rangle| \\ &= \|\vec{v}_1 - \vec{v}_2\| + \varkappa^{-1} \arccos |\langle \vec{g}_1, \vec{g}_2 \rangle| \\ &= \delta_\kappa(m_1, m_2). \end{aligned}$$

3 Convolution: We know from property (2) with  $\mathbf{R} := \mathbf{R}_{\vec{g}_2}$  that

$$\delta_\kappa(m_1, m_2) = \|\mathbf{R}_{\vec{g}_2}^{-1}(\vec{v}_1 - \vec{v}_2)\| + \varkappa^{-1} \arccos |\langle \mathbf{R}_{\vec{g}_2}^{-1}\vec{g}_1, (0, 0, 1)^T \rangle|.$$

This allows to describe the non-adaptive estimator, which equals a weighted mean, by a discrete convolution on  $\mathbb{R}^3 \times \mathbb{S}^2$ , see Equation (26).

- 4 The left-invariance of msPOAS follows with property (3) analogous to Becker et al. [2012, Appendix A.3] since the choice of the parameters  $\{h^{(k)}\}_{k=0}^{k^*}$  and  $\varkappa$  is defined as for POAS.
- 5 The monotonicity of the non-adaptive weights follows from the definition of  $\delta_\kappa$  and the monotonicity of the sequence of location bandwidths  $\{h^{(k)}\}_{k=0}^{k^*}$  since  $K_{\text{loc}}$  is non-decreasing. □

#### REFERENCES

- I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel. Reconstruction of the orientation distribution function in single and multiple shell q-ball imaging within constant solid angle. *Magnetic Resonance in Medicine*, 64:554–566, 2010.
- S. Aja-Fernández, A. Tristán-Vega, and C. Alberola-López. Noise estimation in single- and multiple-coil magnetic resonance data based on statistical models. *Magnetic Resonance Imaging*, 27(10):1397–1409, 2009.
- S. Aja-Fernández, A. Tristan-Vega, and W.S. Hoge. Statistical noise analysis in grappa using a parametrized noncentral chi approximation model. *Magnetic Resonance in Medicine*, 65(4):1195–1206, 2011.
- J.L.R. Andersson and S. Skare. A model-based method for retrospective correction of geometric distortions in diffusion-weighted epi. *Neuroimage*, 16(1):177–199, 2002.
- J.L.R. Andersson, S. Skare, and J. Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888, 2003.
- Y. Assaf and P. J. Basser. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *Neuroimage*, 27(1):48–58, 2005.
- H.-E. Assemlal, D. Tschumperlé, L. Brun, and K. Siddiqi. Recent advances in diffusion MRI modeling: Angular and radial reconstruction. *Med Image Anal*, 15(4):369–396, 2011.
- 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, 1994a.
- P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994b.
- S. Becker and P. Mathé. A new perspective on the propagation-separation approach: Taking advantage of the propagation condition. Technical Report 1766, WIAS, Berlin, 2013.
- S.M.A. Becker, K. Tabelow, H.U. Voss, A. Anwander, R.M. Heidemann, and J. Polzehl. Position-orientation adaptive smoothing of diffusion weighted magnetic resonance data (POAS). *Medical Image Analysis*, 16(6):1142–1155, 2012.
- T.E.J. Behrens, M.W. Woolrich, M. Jenkinson, H. Johansen-Berg, R.G. Nunes, S. Clare, P.M. Matthews, J.M. Brady, and S.M. Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50:1077–1088, 2003.

- P.T. Callaghan. *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford Science Publications, 1991.
- M.F. Carfora. Interpolation on spherical geodesic grids: a comparative study. *Journal of Computational and Applied Mathematics*, 210:99–105, 2007.
- E. Caruyer, J. Cheng, C. Lenglet, G. Sapiro, T. Jiang, and R. Deriche. Optimal design of multiple q-shells experiments for diffusion MRI. In *MICCAI Workshop on Computational Diffusion MRI- CDMRI 2011*, 2011.
- J. Cheng, A. Ghosh, T. Jiang, and R. Deriche. Model-free and analytical EAP reconstruction via spherical polar fourier diffusion MRI. In *13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'10), Beijing, September 20-24, 2010*, 2010.
- M. Descoteaux, R. Deriche, D. Le Bihan, J.-F. Mangin, and C. Poupon. Multiple q-shell diffusion propagator imaging. *Medical Image Analysis*, 15(4):603–621, 2011.
- O. Dietrich, J.G. Raya, S.B. Reeder, M. Ingrisch, M.F. Reiser, and S.O. Schoenberg. Influence of multi-channel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. *Magnetic Resonance Imaging*, 26(6):754 – 762, 2008.
- Z. Ding, J.C. Gore, and A.W. Anderson. Reduction of noise in diffusion tensor images using anisotropic smoothing. *Magnetic Resonance in Medicine*, 53(2):485–490, 2005.
- R. Duits and E. Franken. Left-invariant diffusions on the space of positions and orientations and their application to crossing-preserving smoothing of HARDI images. *International Journal of Computer Vision*, 92(3):231–264, 2011.
- N. Dungey, A.F.M. ter Elst, and D.W. Robinson. *Analysis on Lie groups with polynomial growth*, volume 214 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 2003. ISBN 0-8176-3225-5.
- A.M.A. El-Sayed. Laguerre polynomials of arbitrary (fractional) orders. *Applied Mathematics and Computation*, 109(1):1–9, 2000. ISSN 0096-3003.
- P.T. Fletcher. *Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI*. PhD thesis, University of North Carolina at Chapel Hill, 2004.
- KJ Friston, JT Ashburner, SJ Kiebel, TE Nichols, and WD Penny. *Statistical parametric mapping: the analysis of functional brain images*. Academic Press, 2006.
- M.A. Griswold, P.M. Jakob, R.M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002. ISSN 1522-2594.
- R.M. Heidemann, D.A. Porter, A. Anwander, T. Feiweier, K. Heberlein, T.R. Knösche, and R. Turner. Diffusion imaging in humans at 7T using readout-segmented EPI and GRAPPA. *Magnetic Resonance in Medicine*, 64(1):9–14, 2010.
- R.M. Heidemann, A. Anwander, T. Feiweier, T.R. Knösche, and R. Turner. k-space and q-space: Combining ultra-high spatial and angular resolution in diffusion imaging using ZOOPPA at 7T. *NeuroImage*, 60(2):967–978, 2012.
- J.H. Jensen, J.A. Helpert, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53(6):1432–1440, 2005.
- P. Jezzard, A. S. Barnett, and C. Pierpaoli. Characterization of and correction for eddy current artifacts in echo planar diffusion imaging. *Magn Reson Med*, 39(5):801–812, 1998.
- H. Johansen-Berg and T.E.J. Behrens, editors. *Diffusion MRI: From Quantitative Measurement to In-Vivo Neuroanatomy*. Academic Press, 2009.
- D.K. Jones, editor. *Diffusion MRI: Theory, Methods, and Applications*. Oxford University Press, 2010.
- D.K. Jones, T.R. Knösche, and R. Turner. White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion mri. *NeuroImage*, 73:239–254, 2012.
- A. Kamali, A.E. Flanders, J. Brody, J.V. Hunter, and K.M. Hasan. Tracing superior longitudinal fasciculus connectivity in the human brain using high resolution diffusion tensor tractography. *Brain Struct Funct*, 2013.

- M. Kleinnijenhuis, M. Barth, D.C. Alexander, A.-M. van Cappellen van Walsum, and D.G. Norris. Structure tensor informed fiber tractography (STIFT) by combining gradient echo MRI and diffusion weighted imaging. *Neuroimage*, 59(4):3941–3954, 2012.
- A. Kristoffersen. Estimating non-Gaussian diffusion model parameters in the presence of physiological noise and rician signal bias. *J Magn Reson Imaging*, 35(1):181–189, 2012.
- C. Liu, R. Bammer, and M.E. Moseley. Generalized diffusion tensor imaging (GDTI): a method for characterizing and imaging diffusion anisotropy caused by non-gaussian diffusion. *Israel Journal of Chemistry*, 43:145–154, 2003.
- P.P. Mitra and P.N. Sen. Effects of microgeometry and surface relaxation on NMR pulsed-field-gradient experiments: Simple pore geometries. *Physical Review B*, 45:143–156, 1992.
- S. Mohammadi, H.E. Möller, H. Kugel, D.K. Müller, and M. Deppe. Correcting eddy current and motion effects by affine whole-brain registrations: evaluation of three-dimensional distortions and comparison with slicewise correction. *Magn Reson Med*, 64(4):1047–1056, 2010.
- S. Mohammadi, C. Hutton, Z. Nagy, O. Josephs, and N. Weiskopf. Retrospective correction of physiological noise in DTI using an extended tensor model and peripheral measurements. *Magnetic Resonance in Medicine*, 2012a.
- S. Mohammadi, Z. Nagy, H.E. Möller, M.R. Symms, D.W. Carmichael, O. Josephs, and N. Weiskopf. The effect of local perturbation fields on human DTI: characterisation, measurement and correction. *Neuroimage*, 60(1):562–570, 2012b.
- S. Mohammadi, P. Freund, T. Feiweier, A. Curt, and N. Weiskopf. The impact of post-processing on spinal cord diffusion tensor imaging. *Neuroimage*, 70:377–385, 2013.
- J.N. Morelli, V.M. Runge, T. Feiweier, J.E. Kirsch, K.W. Williams, and U.I. Attenberger. Evaluation of a modified stejskal-tanner diffusion encoding scheme, permitting a marked reduction in  $t_e$ , in diffusion-weighted imaging of stroke patients at 3 t. *Investigative Radiology*, 45(1):29–35, 2010.
- S. Mori. *Introduction to Diffusion Tensor Imaging*. Elsevier, 2007.
- S. Mori, B.J. Crain, V.P. Chacko, and P.C.M. van Zijl. Three dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neurol*, 45:265–269, 1999.
- E. Özarlan and T.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.
- E. Özarlan, T.M. Shepherd, B.C. Vemuri, S.J. Blackband, and T.H. Mareci. Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). *NeuroImage*, 31:1086–1103, 2006.
- 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.
- J. Polzehl. *aws: Adaptive Weights Smoothing*, 2013. URL <http://cran.r-project.org/package=aws>. R-package version 1.9-1.
- J. Polzehl and V. Spokoiny. Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135:335–362, 2006.
- J. Polzehl and K. Tabelow. Structural adaptive smoothing in diffusion tensor imaging: The **R** package *dti*. *J Stat Software*, 31(9):1–23, 2009.
- J. Polzehl and K. Tabelow. Beyond the gaussian model of diffusion-weighted imaging: The package *dti*. *J Stat Software*, 44(12):1–26, 2011.
- K.P. Pruessmann, M. Weiger, M.B. Scheidegger, and P. Boesiger. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, Nov 1999.
- P.B. Roemer, W.A. Edelstein, C.E. Hayes, S.P. Souza, and O.M. Mueller. The NMR phased array. *Magnetic Resonance in Medicine*, 16(2):192–225, 1990. ISSN 1522-2594.
- L. Ruthotto, H. Kugel, J. Olesch, B. Fischer, J. Modersitzki, M. Burger, and C. H. Wolters. Diffeomorphic susceptibility artifact correction of diffusion-weighted magnetic resonance images. *Phys Med Biol*, 57(18):5715–5731, Sep 2012.

- D.W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, 1992.
- S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R.K. Niazy, J.Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23:S208–S219, 2004.
- P. Storey, F.J. Frigo, R.S. Hinks, B.J. Mock, B.D. Collick, N. Baker, J. Marmurek, and S.J. Graham. Partial k-space reconstruction in single-shot diffusion-weighted echo-planar imaging. *Magn Reson Med*, 57(3):614–619, Mar 2007.
- K. Tabelow and J. Polzehl. *dti: DTI/DWI Analysis*, 2013. URL <http://CRAN.R-project.org/package=dti>. R package version 1.1-5.
- K. Tabelow, H.U. Voss, and J. Polzehl. Modeling the orientation distribution function by mixtures of angular central gaussian distributions. *Journal of Neuroscience Methods*, 203(1):200–211, 2012.
- A. Tabesh, J.H. Jensen, B.A. Ardekani, and J.A. Helpert. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. *Magn Reson Med*, 65(3):823–836, Mar 2011.
- P. Thunberg and P. Zetterberg. Noise distribution in SENSE- and GRAPPA-reconstructed images: a computer simulation study. *Magnetic Resonance Imaging*, 25(7):1089–1094, 2007. ISSN 0730-725X.
- D.S. Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52:1358–1372, 2004.
- 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.
- L. Zhan, D. Franc, V. Patel, N. Jahanshad, Y. Jin, B.A. Mueller, M.A. Bernstein, B.J. Borowski, C.R. Jack Jr, A.W. Toga, K.O. Lim, and P.M. Thompson. How do spatial and angular resolution affect brain connectivity maps from diffusion MRI? *Proc IEEE Int Symp Biomed Imaging*, pages 1–6, 2012.