

# Diffusion tensor smoothing through weighted Karcher means

Owen Carmichael

*Department of Neurology and Computer Science  
University of California, Davis  
e-mail: [ocarmichel@ucdavis.edu](mailto:ocarmichel@ucdavis.edu)*

and

Jun Chen, Debashis Paul, Jie Peng

*Department of Statistics  
University of California, Davis  
e-mail: [jchen@wald.ucdavis.edu](mailto:jchen@wald.ucdavis.edu); [debpaull@ucdavis.edu](mailto:debpaull@ucdavis.edu); [jiepeng@ucdavis.edu](mailto:jiepeng@ucdavis.edu)*

**Abstract:** Diffusion tensor magnetic resonance imaging (MRI) quantifies the spatial distribution of water diffusion at each voxel on a regular grid of locations in a biological specimen by diffusion tensors— $3 \times 3$  positive definite matrices. Removal of noise from DTI is an important problem due to the high scientific relevance of DTI and relatively low signal to noise ratio it provides. Leading approaches to this problem amount to estimation of weighted Karcher means of diffusion tensors within spatial neighborhoods, under various metrics imposed on the space of tensors. However, it is unclear how the behavior of these estimators varies with the magnitude of DTI sensor noise (the noise resulting from the thermal effects of MRI scanning) as well as the geometric structure of the underlying diffusion tensor neighborhoods. In this paper, we combine theoretical analysis, empirical analysis of simulated DTI data, and empirical analysis of real DTI scans to compare the noise removal performance of three kernel-based DTI smoothers that are based on Euclidean, log-Euclidean, and affine-invariant metrics. The results suggest, contrary to conventional wisdom, that imposing a simplistic Euclidean metric may in fact provide comparable or superior noise removal, especially in relatively unstructured regions and/or in the presence of moderate to high levels of sensor noise. On the contrary, log-Euclidean and affine-invariant metrics may lead to better noise removal in highly structured anatomical regions, especially when the sensor noise is of low magnitude. These findings emphasize the importance of considering the interplay of sensor noise magnitude and tensor field geometric structure when assessing diffusion tensor smoothing options. They also point to the necessity for continued development of smoothing methods that perform well across a large range of scenarios.

**AMS 2000 subject classifications:** Primary 62G08; secondary 62P10.

---

\*The authors thank the anonymous reviewers for their helpful comments.

†Carmichael’s research is partially supported by the NIH grants AG-10220, AG-10129, AG-030514, AG-031252 and AG-021028 and the NSF grant DMS-1208917.

‡Chen and Peng’s research is partially supported by the NSF grants DMS-0806128 and DMS-1007583.

§Paul’s research is partially supported by the NSF grants DMS-0806128, DMR-1035468 and DMS-1106690.

**Keywords and phrases:** Tensor space, diffusion MRI, Karcher mean, kernel smoothing, perturbation analysis.

Received October 2012.

## 1. Introduction

Diffusion magnetic resonance imaging (MRI) has emerged as a prominent technique for using magnetic field gradients to measure the directional distribution of water diffusion in a biological specimen (Bammer et al., 2009; Beaulieu, 2002; Chanraud et al., 2010; Mukherjee et al., 2008a,b) because the diffusion measurements are useful as non-invasive proxy measures for the structural organization of the underlying tissue. In diffusion MRI of the brain, the head of a person or animal is placed inside a stationary magnetic field, and the tissue of the brain is excited by applying direction-specific magnetic field gradients and pulses of radio frequency energy. Measuring the frequency characteristics of energy emitted from the excited tissue allows us to estimate, at each location in the brain, the bulk amount of water diffusion occurring along each of the magnetic field gradient directions. Directional distributions of water diffusion across all possible 3D directions are then estimated by extrapolating from these direction-specific diffusion measurements. Diffusion tensor imaging (DTI) is the most widespread version of diffusion MRI, resulting in a representation of the directional distribution through a  $3 \times 3$  positive definite matrix (diffusion tensor) at each spatial location. For a more detailed description of diffusion tensor imaging measurements and models, see Section 2. However, DTI data include a substantial amount of noise, which results in noisy estimation of the diffusion tensors (Gudbjartsson and Patz, 1995; Hahn et al., 2006, 2009; Zhu et al., 2009). Consequently, noise adversely affects subsequent tracing of anatomical structures and mapping of inter-regional brain connectivity (Basser and Pajevic, 2000). For this reason, a variety of methods have been developed to filter noise from DTI data by enforcing spatial smoothness. One prominent approach replaces the tensor at each voxel by a weighted Karcher mean of tensors in a local spatial neighborhood under a Riemannian metric imposed on the space of tensors (Arsigny et al., 2005, 2006; Castaño Moraga et al., 2007). This amounts to kernel smoothing of the tensor field. In a notable development, Yuan et al. (2012) extended the weighted Karcher-mean approach to local polynomial smoothing in the tensor space with respect to different geometries, when the observations are taken at random locations. They also derived asymptotic mean squared errors of the estimated tensors. While other alternatives exist, including spatially regularizing the estimation of diffusion tensors from the raw diffusion weighted imaging (DWI) data (Tabelow et al., 2008), in this paper, we focus on the more common Karcher mean approach, with an emphasis on theoretical and numerical evaluation of smoothing results under various conditions.

In the existing literature on Karcher mean approaches, methods that impose a Euclidean metric on the space of tensors (hereafter “Euclidean smoothers”) are generally considered inferior to those that impose a log-Euclidean metric or

an affine-invariant metric (hereafter, “geometric smoothers”). The former gives rise to a *swelling effect* (Arsigny et al., 2005, 2006), which means smoothed tensors have larger determinants than those of the original tensors. This results in artificially inflated estimates of local water diffusion. However, in other important respects the performance characteristics of these smoothers are not well understood. For example, it is not clear how the geometric structure of the tensor neighborhoods impacts smoothing performance. Because this geometric structure is highly variable across the human brain, and because it provides the primary cue for the structural organization of the underlying brain tissue, understanding its effects on tensor smoothing is highly relevant. Secondly, it is not known how smoother performance varies by noise magnitude, which can vary substantially across DTI scanners. Clarifying the performance tradeoffs of competing smoothing algorithms could help to guide end users toward advantageous smoothers based on the biological sample and the operating characteristics of the MRI scanner. Given that the choice of DTI smoothing algorithm has a significant impact on scientific studies of DTI properties in various brain diseases, a deeper understanding of DTI smoothing performance has high practical importance (Viswanath et al., 2012).

We first study tensor estimation at each voxel based on the raw DWI data. The estimated tensors will then be used as input for the smoothing process. We derive asymptotic expansions for a nonlinear regression estimator and a linear regression estimator under the small noise asymptotic regime. We show that, compared with the more widely used linear estimator, the nonlinear estimator is more efficient. We also show that the additive noise resulting from thermal effects of MRI scanning (sensor noise) on the raw DWI data leads to approximately additive noise on the estimated tensors. We then study the properties of the Euclidean smoothers and geometric smoothers applied to these estimated tensors for the purpose of further denoising. The main goal of this study is to demonstrate the effect of different metrics on the performance of local constant smoothing under various local structures of the tensor field and various sensor noise levels. The major finding of this paper is that on contrary to conventional wisdom – Euclidean smoothers may in fact have superior performance than geometric smoothers depending on the interplay of the aforementioned two factors. More specifically, we use perturbation analysis to show that when sensor noise levels are relatively low, either Euclidean or geometric smoothers may have smaller bias depending on whether the tensor field is spatially homogeneous or inhomogeneous. Here we focus on asymptotic bias rather than variance since regardless of the choice of the metric, the variance is essentially inversely proportional to the neighborhood size and is proportional to noise level. However, even when the tensor field is constant within a neighborhood, the smoothers corresponding to different metrics can exhibit different bias characteristics (see Section 4.2 for a detailed discussion). We then use simulated tensor fields to show empirically that when sensor noise levels are relatively high, Euclidean smoothers tend to perform better regardless of the geometric structure of the tensor neighborhoods. Finally, we perform validation experiments on real DTI scans that confirm the theoretical and simulation findings. Together, these find-

ings suggest that we may need to revisit the conventional wisdom that Euclidean smoothers generally provide inferior smoothing performance.

The rest of the paper is organized as follows. In Section 2, we study tensor estimation from raw DWI data. In Section 3 we describe the three tensor smoothers, followed by Section 4 which presents a perturbation analysis to compare these smoothers under the small noise regime. Section 5 is for simulation studies and Section 6 is an application to human brain DTI data. The paper is concluded by a discussion in Section 7.

## 2. Tensor estimation

In this section, we derive asymptotic expansions of two regression based tensor estimators under the additive noise model for the raw DWI data. We consider the small noise asymptotic regime with a fixed number of gradient directions, which is different from the conventional statistical paradigm where the sample size increases to infinity. We choose to work under this setting because in DTI experiments, the number of gradient directions is usually small and fixed whereas the signal to noise ratio may be increased by increasing the magnetic field strength and scan time. In contrast, asymptotic analysis under the framework where the number of gradient directions grows to infinity has been conducted by Zhu et al. (2007) and Zhu et al. (2009). Zhu et al. (2007) proposed a weighted least squares estimate of the diffusion tensor and quantified the effects of noise on this estimator and their eigenvalues and eigenvectors, as well as on the subsequent morphological classification of the tensors. Zhu et al. (2009) studied a number of different regression models for characterizing stochastic noise in DWI and functional MRI (fMRI) and developed a diagnostic procedure for systematically exploring MR images to identify noise components other than simple stochastic noise.

### *DWI measurements and additive noise model*

Proton Nuclei Magnetic Resonance measures signals from the  $H^1$  nuclei, the majority of which in biological tissues is from water molecules. In *diffusion magnetic resonance (DT-MR)*, the signals are sensitized to water diffusion by applying a set of magnetic gradient fields to the tissue. The raw data obtained by MRI scanning are complex numbers representing the Fourier transformation of a magnetization distribution of a tissue at a certain point in time. The observed DWI data are the amplitude of the DT-MR signals corresponding to a set of magnetic gradients denoted by  $\mathcal{Q}$  – a set of unit norm vectors in  $\mathbb{R}^3$  referred to as *gradient directions*. Assuming Gaussian diffusion of water molecules at a given voxel, the noiseless diffusion weighted signal intensity in direction  $\mathbf{q} = (q_1, q_2, q_3)^T \in \mathcal{Q}$  is given by (Mori, 2007)

$$S_{\mathbf{q}} = S_0 \exp(-b\mathbf{q}^T D \mathbf{q}), \quad (1)$$

where  $S_0$  is the baseline intensity determined by the background constant field,  $b$  is a fixed experimental constant, and  $D$  is a  $3 \times 3$  positive definite matrix which is referred to as the *diffusion tensor* at that voxel. For simplicity of exposition, throughout this section,  $S_0$  is assumed to be known and fixed. We also absorb  $b$  into the tensor  $D$  and ignore it hereafter. The so called sensor noise in the observed (complex) signal at each voxel is mainly attributed to thermal noise in the MRI scanner and is modeled as independent and additive white noise on the real and imaginary parts of the signal (Gudbjartsson and Patz, 1995). Consequently, the actually observed, noise corrupted DWI data are

$$\hat{S}_{\mathbf{q}} = \| S_{\mathbf{q}} u_{\mathbf{q}} + \sigma \varepsilon_{\mathbf{q}} \|^2, \quad \mathbf{q} \in \mathcal{Q}, \tag{2}$$

where  $u_{\mathbf{q}}$ , a unit vector in  $\mathbb{R}^2$ , is the phase of the signal, the random vectors  $\varepsilon_{\mathbf{q}}$  have two independent coordinates with zero mean and unit variance and are also assumed to be independent across gradient directions  $\mathbf{q}$ 's. The parameter  $\sigma > 0$  controls the noise level. This model is referred to as the *additive noise model* for DWI data. If in model (2) the noise is further assumed to be Gaussian, i.e.,  $\varepsilon_{\mathbf{q}}$ 's are independent  $N(0, I_2)$ , then  $\hat{S}_{\mathbf{q}}$ 's follow the *Rician distribution* with probability density function  $p_{S_{\mathbf{q}}, \sigma}$ , where for  $\zeta > 0$ ,

$$p_{\zeta, \sigma}(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + \zeta^2}{2\sigma^2}\right) I_0\left(\frac{x\zeta}{\sigma^2}\right) \mathbf{1}(x > 0), \tag{3}$$

where  $I_0$  denotes the zero-th order modified Bessel function of the first kind.

*Tensor estimators*

Here, we consider estimating the tensor  $D$  based on the observed DWI signal intensities  $\{\hat{S}_{\mathbf{q}} : \mathbf{q} \in \mathcal{Q}\}$ . We use  $\text{vec}(D)$  to denote the  $6 \times 1$  vector  $(D_{11}, D_{22}, D_{33}, D_{12}, D_{13}, D_{23})^T$ , and define  $x_{\mathbf{q}} = (q_1^2, q_2^2, q_3^2, 2q_1q_2, 2q_1q_3, 2q_2q_3)^T$ . With a slight abuse of notation, in this section we use  $D$  to mean  $\text{vec}(D)$ . Then the quadratic form  $\mathbf{q}^T D \mathbf{q}$  can be written as  $x_{\mathbf{q}}^T D$ . In the following, we also assume that the matrix  $\sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T$  is well-conditioned, which is guaranteed by an appropriate choice of the gradient directions (in practice, an approximate uniform design on a sphere is often used). The first method is to use the log transformed DWI's to estimate  $D$  by a linear regression. The resulting estimator is referred to as the *linear regression estimator* and denoted by  $\hat{D}_{LS}$ .

$$\hat{D}_{LS} := \arg \min_{D \in \mathbb{R}^6} \sum_{\mathbf{q} \in \mathcal{Q}} (\log \hat{S}_{\mathbf{q}} - \log S_0 + x_{\mathbf{q}}^T D)^2, \tag{4}$$

The second method uses nonlinear regression based on the original DWI's. The resulting estimator is referred to as the *nonlinear regression estimator* and denoted by  $\hat{D}_{NL}$

$$\hat{D}_{NL} := \arg \min_{D \in \mathbb{R}^6} \sum_{\mathbf{q} \in \mathcal{Q}} (\hat{S}_{\mathbf{q}} - S_0 \exp(-x_{\mathbf{q}}^T D))^2. \tag{5}$$

Note that, in the above two estimates, we did not restrict  $D$  to be positive definite, but only that it is a symmetric matrix. Accordingly, Propositions 2.1 and 2.2 below, dealing with the asymptotic behavior of the estimated tensors, do not explicitly require positive-definiteness, even though they point to the positive-definiteness of the estimated tensors with a high probability. Imposing the positive definite constraint explicitly for tensor estimation at each voxel could be done, e.g., through a logarithm parametrization of the tensor.

Let  $X$  be the  $|\mathcal{Q}| \times 6$  matrix with rows  $x_{\mathbf{q}}^T$ , for  $\mathbf{q} \in \mathcal{Q}$ , where  $|\mathcal{Q}|$  denotes the number of gradient directions. Also let  $\ell_S = (\log \hat{S}_{\mathbf{q}} - \log S_0)_{\mathbf{q} \in \mathcal{Q}}$ . For future use, we also define  $\mathcal{D}_S = \text{diag}(S_{\mathbf{q}} : \mathbf{q} \in \mathcal{Q})$ . Then we have an explicit formula for  $\hat{D}_{LS}$ , namely,

$$\hat{D}_{LS} = - \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} (\log \hat{S}_{\mathbf{q}} - \log S_0) x_{\mathbf{q}} \right) = -(X^T X)^{-1} X^T \ell_S.$$

Though no explicit formula is available for  $\hat{D}_{NL}$ , it can be numerically solved by a nonlinear least square solver. We used the Levenberg-Marquardt method in the numerical study of this paper. With the linear regression estimator as the initial estimate, the algorithm usually converges in just a few iteration steps.

In the literature, it has been shown numerically that the nonlinear regression estimator often outperforms the linear regression estimator (Basser and Pajevic, 2000; Hahn et al., 2006). Also, when the noise associated with DWI measurements is Rician, it is shown in Polzehl and Tabelow (2008) that in the low signal-to-noise-ratio setting, the linear regression estimator is biased. In the following, we analyze the behavior of the two regression estimators under the additive noise model (2) as the noise parameter  $\sigma \rightarrow 0$  and show that the nonlinear regression estimator is asymptotically more efficient. The proofs of the stated asymptotic results can be found in the Appendix. Throughout, we use  $D^0$  to denote the true diffusion tensor.

#### Asymptotic expansions

**Proposition 2.1.** *Under the additive noise model, as  $\sigma \rightarrow 0$ ,  $\hat{D}_{LS} = D^0 + \sigma D_{1,LS} + \sigma^2 D_{2,LS} + O(\sigma^3 |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^3)$ , where the random vectors  $D_{1,LS}$  and  $D_{2,LS}$  satisfy*

$$\mathbb{E}(D_{1,LS}) = 0, \quad \mathbb{E}(D_{2,LS}) = 0,$$

and

$$\begin{aligned} \text{Var}(D_{1,LS}) &= \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^{-2} x_{\mathbf{q}} x_{\mathbf{q}}^T \right) \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \\ &= (X^T X)^{-1} (X^T \mathcal{D}_S^{-2} X) (X^T X)^{-1}. \end{aligned} \quad (6)$$

Thus, the bias in  $\hat{D}_{LS}$  is of the order  $O(\sigma^3)$ . Moreover, under the Rician noise model,  $D_{1,LS}$  is a normal random vector, whereas the coordinates of  $D_{2,LS}$  are weighted sums of differences of independent  $\chi_1^2$  random variables.

**Proposition 2.2.** *Under the additive noise model, as  $\sigma \rightarrow 0$ ,  $\widehat{D}_{NL} = D^0 + \sigma D_{1,NL} + \sigma^2 D_{2,NL} + O(\sigma^3 |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^3)$ , where,*

$$\mathbb{E}(D_{1,NL}) = 0, \quad \text{Var}(D_{1,NL}) = \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} = (X^T \mathcal{D}_S^2 X)^{-1}, \quad (7)$$

and

$$\mathbb{E}(D_{2,NL}) = -\frac{1}{2} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left[ \sum_{\mathbf{q} \in \mathcal{Q}} \left( 1 - S_{\mathbf{q}}^2 x_{\mathbf{q}}^T \left( \sum_{\mathbf{q}' \in \mathcal{Q}} S_{\mathbf{q}'}^2 x_{\mathbf{q}'} x_{\mathbf{q}'}^T \right)^{-1} x_{\mathbf{q}} \right) x_{\mathbf{q}} \right]. \quad (8)$$

In (8), at least a few the coefficients of  $x_{\mathbf{q}}$  are nonzero if  $|\mathcal{Q}| > 6$  and thus the mean of  $D_{2,NL}$  is non-vanishing. Therefore, the bias in  $\widehat{D}_{NL}$  is of order  $O(\sigma^2)$ . Moreover, under the Rician noise model,  $D_{1,NL}$  is a normal random vector, whereas the coordinates of  $D_{2,NL}$  are weighted sums of (dependent)  $\chi_1^2$  random variables.

In terms of the first order terms, both regression estimators are unbiased and  $\widehat{D}_{NL}$  has a smaller variance than  $\widehat{D}_{LS}$ , in the sense

$$a^T (X^T \mathcal{D}_S^2 X)^{-1} a \leq a^T (X^T X)^{-1} (X^T \mathcal{D}_S^{-2} X) (X^T X)^{-1} a, \quad \text{for all } a \in \mathbb{R}^6. \quad (9)$$

Indeed, the difference between the two asymptotic covariance matrices can be quite substantial if  $S_{\mathbf{q}}$ 's vary a lot. This situation may arise when the true diffusion tensor  $D^0$  is highly anisotropic. Under such a situation, only a few gradient directions are likely to be aligned to the leading eigen-direction, which results in small  $S_{\mathbf{q}}$ , whereas other gradient directions lead to large  $S_{\mathbf{q}}$ . The above analysis suggests that at least when the noise level is low, the nonlinear regression estimator is more preferable due to its smaller variance. In the Supplementary Material (Carmichael et al.), we present a numerical study which shows that for both small and large noise levels, the nonlinear estimator performs better than the linear estimator.

In terms of the second order terms, the linear regression estimator is also unbiased. If the number of gradient directions is not too small, we have the approximation for the second order bias term (8) of  $\widehat{D}_{NL}$ :

$$\mathbb{E}(D_{2,NL}) \approx -\frac{1}{2} \left( \frac{1}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \frac{1}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} \right) = -\frac{1}{2} (X^T \mathcal{D}_S^2 X)^{-1} X^T \mathbf{1}_{\mathcal{Q}}. \quad (10)$$

Notice that in this approximation, only the matrix inverse term involves the true tensor (through  $S_{\mathbf{q}}$ 's). So, as long as the design is uniform, no single gradient direction has a dominating influence on this bias.

### 3. Kernel smoothing in tensor space

In this section, we consider kernel smoothing on the space of  $N \times N$  positive definite matrices (hereafter referred to as the *tensor space*  $\mathcal{P}_N$ ). We first briefly

review the kernel smoothing idea in a Euclidean space (Fan and Gijbels, 1996). Let  $f$  be a function defined on  $\mathbb{D} \subset \mathbb{R}^d$  and taking values in  $\mathbb{R}^p$ . The data consist of pairs  $\{(s_i, X_i)\}_{i=1}^n$  with  $s_i$ 's being spatial positions in  $\mathbb{D}$ , and  $X_i \in \mathbb{R}^p$ 's being noise-corrupted versions of  $f(s_i)$ 's. One way to reconstruct  $f$  is to use weighted averages:

$$\widehat{f}(s) = \sum_{i=1}^n w_i(s) X_i / \sum_{i=1}^n w_i(s), \quad s \in \mathbb{D}.$$

Let  $\|\cdot\|$  denote a Euclidean norm. A common scheme for the weights is

$$\omega_i(s) := K_h(\|s_i - s\|), \quad i = 1, \dots, n, \quad (11)$$

where  $K_h(\cdot) := (1/h)K(\cdot/h)$ ,  $K(\cdot)$  is a nonnegative, integrable kernel and  $h > 0$  is the bandwidth which determines the size of the smoothing neighborhood. Note that,  $\widehat{f}(s)$  minimizes  $\sum_{i=1}^n \omega_i(s) d^2(X_i, \mathbf{c})$  with respect to  $\mathbf{c}$ , where  $d(X_i, \mathbf{c}) = \|X_i - \mathbf{c}\|$  is a Euclidean distance on  $\mathbb{R}^p$ . Thus kernel smoothing can be immediately generalized to a Riemannian manifold  $\mathcal{M}$  by replacing the Euclidean distance with the geodesic distance  $d_{\mathcal{M}}(\cdot, \cdot)$ . Specifically, to reconstruct a function  $f : \mathbb{D} \subset \mathbb{R}^d \rightarrow \mathcal{M}$ , we define

$$\widehat{f}(s) := \arg \min_{\mathbf{c} \in \mathcal{M}} \sum_{i=1}^n \omega_i(s) d_{\mathcal{M}}^2(X_i, \mathbf{c}), \quad s \in \mathbb{D}. \quad (12)$$

Kernel smoothing thus takes the form of a *weighted Karcher mean* of  $X_i$ 's (Karcher, 1977). From (12), it is obvious that different distance metrics on the manifold  $\mathcal{M}$  may lead to different kernel smoothers. Since  $\mathcal{P}_N$  is the interior of a cone in the Euclidean space  $\mathbb{R}^{N \times N}$ , one may simply impose a Euclidean metric on  $\mathcal{P}_N$ , for example,  $d_E(X, Y) := \{\text{trace}(X - Y)^2\}^{1/2}$ . Under Euclidean distances, (12) can be easily solved by a weighted average

$$\widehat{f}_E(s) = \sum_{i=1}^n \omega_i(s) X_i / \sum_{i=1}^n \omega_i(s). \quad (13)$$

Since the tangent space of  $\mathcal{P}_N$  is the space of  $N \times N$  symmetric matrices, as an alternative to the Euclidean metrics, logarithmic Euclidean (henceforth *log-Euclidean*) metrics have been proposed (Arsigny et al., 2005, 2006):

$$d_{LE}(X, Y) := \|\log X - \log Y\|. \quad (14)$$

Under  $d_{LE}$ , (12) can also be explicitly solved as

$$\widehat{f}_{LE}(s) = \exp \left( \sum_{i=1}^n \omega_i(s) \log(X_i) / \sum_{i=1}^n \omega_i(s) \right). \quad (15)$$

Here  $\exp(\cdot)$  and  $\log(\cdot)$  denote the matrix exponential and matrix logarithm functions. Log-Euclidean metric has also been used by Schwartzman (2006) to develop nonparametric test procedures in the context of DTI.

Moreover, since  $\mathcal{P}_N$  can be identified with a naturally reductive homogenous space (Absil et al., 2008), one may use a bi-invariant metric:

$$d_{Aff}(X, Y) := \left[ \text{tr} \left( \log(X^{-1/2} Y X^{-1/2}) \right)^2 \right]^{1/2}, \quad (16)$$

where  $\text{tr}(\cdot)$  is the trace operator. This metric is affine-invariant, i.e., for any  $g \in GL^+(N, \mathbb{R})$ , where  $GL^+(N, \mathbb{R})$  is the group of matrices (defined on  $\mathbb{R}$ ) with positive determinant, we have  $d_{Aff}(gXg^T, gYg^T) = d_{Aff}(X, Y)$ . Therefore, we refer to  $d_{Aff}$  as the *affine-invariant metric*. The affine-invariant geometry on  $\mathcal{P}_N$  has been extensively studied (Fletcher and Joshi, 2004, 2007; Förstner and Moonen, 1999) and has been applied to DTI data (Arsigny et al., 2005, 2006; Pennec et al., 2006). While a closed form solution for  $\hat{f}(s)$  is not available under  $d_{Aff}$ ,  $\hat{f}(s)$  may be computed using gradient descent methods (Pennec et al., 2006), Newton-Raphson methods (Ferreira et al., 2006; Fletcher and Joshi, 2007), or an approximate recursive procedure which is computationally much faster (outlined in Section S-1 of the Supplementary Material). Note that all three Karcher means considered here are scale-equivariant.

Euclidean smoothing is often being criticized due to its swelling effect where the determinant of the “average” tensor is larger than those of the tensors being averaged (Arsigny et al., 2005). Since the determinant of the tensor quantifies the overall diffusivity of water within the voxel, this property contradicts with principles of physics. On the other hand, under both geometric smoothing, averaging of tensors results in an averaging of their determinants, thus precluding swelling artifacts. However, as we shall see in the following sections, the relative merits of these smoothers in terms of estimation accuracy are rather complicated in that there is no one smoother performs the best under all situations. Indeed, the performance of these smoothers depends heavily on the nature of the noise and local geometric structures in the data.

#### 4. Comparison of smoothers under different geometries

In this section, we first use perturbation analysis to quantify the differences among the Euclidean mean and the two geometric means under an arbitrary target tensor. Arsigny et al. (2005) compared the log-Euclidean and affine-invariant means. However, their analysis was restricted to the setting where the target tensor is the identity matrix. We use the perturbation analysis results to compare the three tensor smoothers defined in Section 3. In particular, we study the effects of the noise in raw DWI data and the spatial heterogeneity of the tensor field on the bias associated with the smoothers. Yuan et al. (2012) considered local polynomial smoothing in tensor space using log-Euclidean and affine-invariant geometries when the observations are taken at random locations. In their setting, conditional mean and variance of the tensors are specified and asymptotic MSE is derived under the asymptotic regime where the number of sampling points in the neighborhood goes to infinity. In contrast, in this paper,

the tensor field is observed on a grid determined by voxelization. The noise in the observed tensors (which are used for smoothing) originates from the sensor noise in the diffusion weighted MRI data based on which the observed tensors are estimated. Local constant smoothing is applied to these noisy tensors and asymptotic analysis is carried out under a regime where the noise level converges to zero. Moreover, the focus of our asymptotic analysis is primarily on studying the bias characteristics of the local constant smoother under different geometries.

#### 4.1. Asymptotic expansions of Karcher means

Let  $\{D(\omega) \in \mathcal{P}_N : \omega \in \Omega\}$  be a set of random tensors, where  $\Omega$  is an arbitrary index set with a Borel  $\sigma$ -algebra. Let  $\mathbb{P}_\Omega$  be a probability measure on  $\Omega$ . Let  $\bar{D}_E$ ,  $\bar{D}_{LE}$  and  $\bar{D}_{Aff}$  denote the weighted Karcher mean of  $D$ :

$$\arg \min_{\mathbf{c} \in \mathcal{P}_N} \int d^2(\mathbf{c}, D(\omega)) d\mathbb{P}_\Omega(\omega) \quad (17)$$

with respect to the distance metrics  $d_E$ ,  $d_{LE}$  and  $d_{Aff}$ , respectively. We also use  $\mathbb{E}_\Omega$  to denote expectation with respect to the measure  $\mathbb{P}_\Omega$  under the Euclidean metric  $d_E$ , i.e., for any random symmetric matrix  $K$ ,  $\mathbb{E}_\Omega(K) := \int K(\omega) d\mathbb{P}_\Omega(\omega)$ . Thus  $\bar{D}_E = \mathbb{E}_\Omega(D)$ .

We take a matrix perturbation analysis approach to study the differences among these means under a small noise limit regime. Let  $D^0$  denote the underlying ‘‘true’’ or target tensor. Let  $\lambda_j$  denote the  $j$ -th largest distinct eigenvalue of  $D^0$ , and let  $P_j$  denote the corresponding eigen-projection. For simplicity of exposition, we consider only two scenarios: (a) when the eigenvalues of  $D^0$  are all distinct (anisotropic tensor); and (b) when all the eigenvalues of  $D^0$  are equal to  $\lambda_1$  (isotropic tensor). Let

$$C := \begin{cases} \max\{\max_j \lambda_j^{-1}, \max_{k \neq j} |\lambda_k - \lambda_j|^{-1}\}, & \text{if } D^0 \neq \lambda_1 I \\ \lambda_1^{-1} & \text{if } D^0 = \lambda_1 I, \end{cases}$$

where  $I$  denotes the  $N \times N$  identity matrix. We assume that the probability measure  $\mathbb{P}_\Omega$  satisfies

$$\mathbb{P}_\Omega(\sup_{\omega \in \Omega} \|\Delta(\omega)\| < C^{-1}t) = 1, \quad (18)$$

for some  $t > 0$ , where  $\Delta(\omega) := D(\omega) - D^0$ , and  $\|\cdot\|$  denotes the operator norm of matrices. Roughly speaking,  $C^{-1}$  indicates the scale of the signal and  $t$  is a parameter that controls the degree of deviation of the tensors from the target tensor. Also, we denote the difference between the Euclidean mean and the target tensor by

$$\bar{\Delta}_E := \mathbb{E}_\Omega(\Delta(\omega)) = \bar{D}_E - D^0.$$

The following theorems give expansions of the three means around the target  $D^0$  when  $t \rightarrow 0$ . For simplicity of exposition, in the anisotropic case, expansions are in terms of the logarithm of the determinant of the mean. The expansions of the logarithm of the mean are given in Propositions 8.1-8.3 in the Appendix. Proofs are also given in the Appendix.

**Theorem 4.1.** *Suppose that the tensors  $\{D(\omega) : \omega \in \Omega\}$  and the probability distribution  $\mathbb{P}_\Omega$  satisfy (18) and  $t \rightarrow 0$ .*

(a) *If  $D^0 = \lambda_1 I$ , then*

$$\log \bar{D}_E - \log D^0 = \frac{1}{\lambda_1} \bar{\Delta}_E - \frac{1}{2\lambda_1^2} \bar{\Delta}_E^2 + O(t^3). \tag{19}$$

(b) *If the eigenvalues of  $D^0$  are all distinct,*

$$\begin{aligned} & \log \det(\bar{D}_E) - \log \det(D^0) \\ &= \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E) \\ & \quad - \sum_{j=1}^N \sum_{k:k \neq j}^N \frac{1}{\lambda_j(\lambda_k - \lambda_j)} [\text{tr}(P_j \bar{\Delta}_E P_k \bar{\Delta}_E)] - \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j^2} [\text{tr}(P_j \bar{\Delta}_E)]^2 + O(t^3). \end{aligned} \tag{20}$$

**Theorem 4.2.** *Assume that the conditions of Theorem 4.1 hold.*

(a) *If  $D^0 = \lambda_1 I$ , then*

$$\log \bar{D}_{LE} - \log D^0 = \frac{1}{\lambda_1} \bar{\Delta}_E - \frac{1}{2\lambda_1^2} \mathbb{E}_\Omega(\Delta^2) + O(t^3). \tag{21}$$

(b) *If the eigenvalues of  $D^0$  are all distinct,*

$$\begin{aligned} & \log \det(\bar{D}_{LE}) - \log \det(D^0) \\ &= \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E) \\ & \quad - \sum_{j=1}^N \sum_{k:k \neq j}^N \frac{1}{\lambda_j(\lambda_k - \lambda_j)} \mathbb{E}_\Omega[\text{tr}(P_j \Delta P_k \Delta)] - \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j^2} \mathbb{E}_\Omega[\text{tr}(P_j \Delta)]^2 + O(t^3). \end{aligned} \tag{22}$$

**Theorem 4.3.** *Assume that the conditions of Theorem 4.1 hold.*

(a) *If  $D^0 = \lambda_1 I$ , then*

$$\log \bar{D}_{Aff} - \log D^0 = \frac{1}{\lambda_1} \bar{\Delta}_E - \frac{1}{2\lambda_1^2} \mathbb{E}_\Omega(\Delta^2) + O(t^3). \tag{23}$$

(b) If the eigenvalues of  $D^0$  are all distinct,

$$\begin{aligned}
& \log \det(\bar{D}_{Aff}) - \log \det(D^0) \\
= & \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E) \\
& - \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \frac{1}{\lambda_j \lambda_k} \mathbb{E}_\Omega[\text{tr}(P_j \Delta P_k \Delta)] \\
& + \sum_{j=1}^N \sum_{k:k \neq j}^N \frac{1}{\lambda_j} \left[ \frac{1}{2\lambda_k} - \frac{1}{\lambda_k - \lambda_j} \right] \text{tr}(P_j \bar{\Delta}_E P_k \bar{\Delta}_E) \\
& + \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j^2} [\text{tr}((P_j \bar{\Delta}_E)^2) - [\text{tr}(P_j \bar{\Delta}_E)]^2] + O(t^3). \quad (24)
\end{aligned}$$

These theorems show that, the three means are the same up to the first order terms. If  $D^0$  is isotropic, then the first order difference between these means and the truth (in logarithm scale) is  $(1/\lambda_1)\bar{\Delta}_E$ . Moreover,  $\bar{D}_{LE}$  and  $\bar{D}_{Aff}$  are the same even up to the second order terms. If  $D^0$  is anisotropic, then the first order difference in log-determinant between these means and the truth is  $\sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E)$ . Also, in this case, all three means differ in second order terms.

#### 4.2. Comparison of smoothers

In this subsection, we utilize the expansions in Section 4.1 to compare the three tensor smoothers described in Section 3. We assume that, the tensors being smoothed are derived from the raw DWI data by using one of the regression methods discussed in Section 2. For comparison purposes we focus on the asymptotic bias of the smoothed tensors, as the variance is inversely proportional to the size of the smoothing neighborhood and is not very sensitive to the choice of the smoother. The main objective here is to demonstrate the effect of different metrics on the performance of local constant smoothing under different local structures of the tensor field. Since we are dealing with DTI, throughout this subsection, we have  $N = 3$  where  $N$  is the dimension of the tensor space.

We first relate the notations in Section 4.1 to the context of tensor smoothing. Now,  $\omega$  is the voxel index and  $D(\omega)$  is the estimated tensor based on the raw DWI data at that voxel, while  $\Omega$  denotes the smoothing neighborhood of the target voxel, and  $\mathbb{P}_\Omega$  is the measure determined by the kernel weights. Thus, the mean tensor defined in (17) is the local constant estimate given in (12). In the following, we assume that a compact kernel is used. Note that, as the bandwidth of the kernel goes to zero, the parameter  $t$  in (18) goes to zero if the tensor field is sufficiently smooth and the noise in the raw DWI data goes to zero as well. We adopt an infill asymptotic framework in which as the bandwidth goes to zero, the number of data points in the neighborhood increases. Specifically, we assume that  $|\Omega|$  goes to infinity and  $r_\Omega := \sum_{\omega \in \Omega} (p(\omega))^2 = o(1)$  as  $t \rightarrow 0$

where  $\{p(\omega) : \omega \in \Omega\}$  denotes the probability mass function of  $\mathbb{P}_\Omega$ . This is true for example if  $\mathbb{P}_\Omega$  is the uniform measure on  $\Omega$  since then  $r_\Omega = 1/|\Omega|$ . More generally, this holds if the kernel is a continuous density that is compactly supported. We further assume that  $|\Omega| \leq t^{-K}$ , for some  $K > 0$ , which puts an upper bound on the growth rate of  $|\Omega|$  as  $t \rightarrow 0$ .

We denote the underlying true tensor at each voxel  $\omega$  by  $D^0(\omega)$  and denote the true tensor at the target voxel for smoothing by  $D^0$ . The observed tensors  $D(\omega)$ 's may be viewed as noisy versions of  $D^0(\omega)$ 's. The analysis in Section 2 shows that if a regression estimator is used, then the noise in  $D(\omega)$  is approximately additive. Moreover, the spatial heterogeneity is another factor influencing estimator bias of the kernel smoothers. In the analysis of Case 2 described below, we treat  $D^0(\omega)$ 's as random perturbations of  $D^0$ . This perspective is in agreement with approaches in spatial statistics where the spatial field is treated as a random process. We look into the effects of these two factors separately by considering the following two cases.

- Case 1* There is no spatial inhomogeneity, i.e.,  $D^0(\omega) = D^0$  for all  $\omega \in \Omega$ . In this case, the sensor noise from the DTI scanner is the sole source of variation in  $D(\omega)$ 's leading to the bias in the kernel estimators.
- Case 2* The noise in the DWI data is small, such that the spatial inhomogeneity is the dominant source of variation in  $D(\omega)$ 's. Moreover, we study a particular case where the geometric mean of the tensors  $\{D^0(\omega) : \omega \in \Omega\}$  is approximately equal to the target tensor  $D^0$ .

For the subsequent analysis, we assume that, in the additive noise model (2) for DWI data,  $\varepsilon_{\mathbf{q}}(\omega)$ 's are i.i.d. across voxel  $\omega \in \Omega$  and gradient  $\mathbf{q} \in \mathcal{Q}$ , and the moment generating function of  $\|\varepsilon_{\mathbf{q}}(\omega)\|^2$  is finite in a neighborhood of zero, for example, when the coordinates of  $\varepsilon_{\mathbf{q}}(\omega)$  have sub-Gaussian tails. By Chebyshev's inequality, this implies that for every  $c > 0$ , there is a  $c' > 0$  such that, for any  $\delta \in (0, 1)$ ,

$$\mathbb{P}\left(\max_{\omega \in \Omega} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}(\omega)\|^3 \geq c' |\mathcal{Q}| (\log(1/\delta) + \log |\mathcal{Q}|)^{3/2}\right) \leq 2|\Omega| \delta^c. \quad (25)$$

This allows us to impose a uniform tail bound on the residual terms in the expansions of the regression estimators in Propositions 2.1 and 2.2. Since  $\mathcal{Q}$  is fixed, we ignore the terms involving  $\log |\mathcal{Q}|$  when using the bound (25) with a sequence  $\delta \rightarrow 0$ .

**Case 1.** Let  $\sigma$  be the standard deviation of the noise in raw DWI data (see equation (2)). By Propositions 2.1 and 2.2, and the fact that  $D^0(\omega) = D^0$ , for all  $\omega \in \Omega$ , we have

$$\Delta(\omega) = D(\omega) - D^0 = [\sigma D_1(\omega) + \sigma^2 D_2(\omega)] + R_\Delta(\omega) := \Delta^*(\omega) + R_\Delta(\omega), \quad (26)$$

where  $R_\Delta(\omega) = O(\sigma^3 (\log(1/\sigma))^{3/2})$  with high probability<sup>1</sup>, and  $D_1(\omega)$  and

<sup>1</sup>We say that an event holds with *high probability* if the complement of that event has probability  $O(\delta^c)$  for a given  $c > 0$  where  $\delta$  is a positive sequence converging to zero.

$D_2(\omega)$  denote the first order and second order terms in the expansion of  $D(\omega)$  around  $D^0$  (in matrix form). Thus, the parameter  $t$  may be taken as  $\sigma\sqrt{\log(1/\sigma)}$ , since this ensures that equation (18) is satisfied with high probability. Then

$$\bar{\Delta}_E = \mathbb{E}_\Omega(\Delta) = \bar{\Delta}_E^* + \mathbb{E}_\Omega(R_\Delta), \quad (27)$$

where  $\mathbb{E}_\Omega(R_\Delta) = O(\sigma^3(\log(1/\sigma))^{3/2})$  with high probability and  $\bar{\Delta}_E^* := \mathbb{E}_\Omega(\Delta^*) = \sigma\mathbb{E}_\Omega(D_1) + \sigma^2\mathbb{E}_\Omega(D_2)$ .

We first consider the case when  $D(\omega)$ 's are the nonlinear regression estimates. By Proposition 2.2,  $\mathbb{E}(D_1) = \mathbb{E}(\mathbb{E}_\Omega(D_1)) = 0$  and hence

$$\mathbb{E}(\bar{\Delta}_E^*) = \mathbb{E}(\mathbb{E}_\Omega(\Delta^*)) = \sigma^2\mathbb{E}(D_2). \quad (28)$$

Moreover, since  $\text{Var}(\text{vec}(\bar{\Delta}_E^*)) = r_\Omega \text{Var}(\text{vec}(\Delta^*))$ , we have

$$\begin{aligned} & \mathbb{E}[\text{vec}(\bar{\Delta}_E^*)\text{vec}(\bar{\Delta}_E^*)^T] \\ &= r_\Omega \text{Var}(\text{vec}(\Delta^*)) + \mathbb{E}(\text{vec}(\bar{\Delta}_E^*))\mathbb{E}(\text{vec}(\bar{\Delta}_E^*))^T \\ &= r_\Omega \mathbb{E}(\text{vec}(\Delta^*)\text{vec}(\Delta^*)^T) + (1 - r_\Omega)\mathbb{E}(\text{vec}(\Delta^*))\mathbb{E}(\text{vec}(\Delta^*))^T \\ &= r_\Omega \sigma^2 \text{Var}(\text{vec}(D_1)) + r_\Omega O(\sigma^3) + O(\sigma^4). \end{aligned} \quad (29)$$

where, in the last step we have used the fact that

$$\begin{aligned} & \mathbb{E}(\text{vec}(\Delta^*)\text{vec}(\Delta^*)^T) \\ &= \sigma^2 \text{Var}(\text{vec}(D_1)) + \sigma^3 [\mathbb{E}(\text{vec}(D_1)(\text{vec}(D_2))^T) + \mathbb{E}(\text{vec}(D_2)(\text{vec}(D_1))^T)] \\ & \quad + \sigma^4 \mathbb{E}(\text{vec}(D_2)\text{vec}(D_2)^T) \\ &= \sigma^2 \mathbb{E}(\text{vec}(D_1)\text{vec}(D_1)^T) + O(\sigma^3). \end{aligned} \quad (30)$$

Note that,  $\mathbb{E}(\text{vec}(D_2))$  and  $\text{Var}(\text{vec}(D_1))$  are given in Proposition 2.2.

*Asymptotic biases* of the logarithm of the smoothers with respect to  $\log D^0$ , denoted generically by  $ABias(\log \bar{D}; \log D^0)$ , are defined as expectations of the leading order terms (i.e., the terms that are linear or quadratic in  $\Delta^*$  or  $\bar{\Delta}_E^*$ ) in the asymptotic expansions of  $\log \bar{D}$  (see Theorems 4.1, 4.2 and 4.3). Note that, by the discussion above, the remainder terms in these expansions are of the order  $O(\sigma^3(\log(1/\sigma))^{3/2})$  with high probability.

If  $D^0 = \lambda_1 I_3$ , by part (a) of Theorems 4.1, 4.2 and 4.3, and equations (26) – (29), we conclude that:

**Corollary 4.1.** *If  $D^0$  is isotropic, then*

$$\begin{aligned} ABias(\log \bar{D}_E; \log D^0) &= \frac{1}{\lambda_1} \mathbb{E}(\bar{\Delta}_E^*) - \frac{1}{2\lambda_1^2} \mathbb{E}((\bar{\Delta}_E^*)^2) \\ &= \frac{1}{\lambda_1} \mathbb{E}(\bar{\Delta}_E^*) - r_\Omega \cdot \frac{1}{2\lambda_1^2} \mathbb{E}((\Delta^*)^2) + O(\sigma^4), \\ ABias(\log \bar{D}_{LE}; \log D^0) &= \frac{1}{\lambda_1} \mathbb{E}(\bar{\Delta}_E^*) - \frac{1}{2\lambda_1^2} \mathbb{E}((\Delta^*)^2), \\ ABias(\log \bar{D}_{Aff}; \log D^0) &= \frac{1}{\lambda_1} \mathbb{E}(\bar{\Delta}_E^*) - \frac{1}{2\lambda_1^2} \mathbb{E}((\Delta^*)^2). \end{aligned}$$

Also, Lemma 8.1 in the Appendix shows that if the gradient directions are uniform on the sphere, then  $\mathbb{E}(\bar{\Delta}_E^*) = \sigma^2 \mathbb{E}(D_2)$  is a negative definite matrix. Since  $\mathbb{E}((\Delta^*)^2)$  is positive definite and is of order  $O(\sigma^2)$ , under the assumption that  $r_\Omega = o(1)$ , the two geometric smoothers are more biased than the Euclidean smoother when  $D^0$  is isotropic and the gradient design is uniform.

If  $D^0$  is anisotropic, asymptotic expansions for the smoothers can be obtained from Propositions 8.1, 8.2 and 8.3 in the Appendix. Using these and equations (26)–(29), we have the following:

**Corollary 4.2.** *If  $D^0$  is anisotropic, then*

$$\begin{aligned} \text{ABias}(\log \bar{D}_E; \log D^0) &= \sigma^2(T_1 - r_\Omega T_2 + r_\Omega T_3) + O(r_\Omega \sigma^3) + O(\sigma^4), \\ \text{ABias}(\log \bar{D}_{LE}; \log D^0) &= \sigma^2(T_1 - T_2 + T_3) + O(\sigma^3), \\ \text{ABias}(\log \bar{D}_{Aff}; \log D^0) &= \sigma^2(T_1 - r_\Omega T_2 + r_\Omega T_3 + (1 - r_\Omega)T_4) + O(\sigma^3), \end{aligned}$$

where

$$\begin{aligned} T_1 &= \sum_{j=1}^3 \frac{1}{\lambda_j} \text{tr}(P_j \mathbb{E}(D_2)) P_j - \sum_{j=1}^3 \sum_{k \neq j}^3 \frac{\log \lambda_j}{\lambda_k - \lambda_j} [P_j \mathbb{E}(D_2) P_k + P_k \mathbb{E}(D_2) P_j] \\ T_2 &= \sum_{j=1}^3 \sum_{k: k \neq j}^3 \frac{1}{\lambda_j (\lambda_k - \lambda_j)} \mathbb{E}[\text{tr}(P_j D_1 P_k D_1)] P_j + \frac{1}{2} \sum_{j=1}^3 \frac{1}{\lambda_j^2} \mathbb{E}[\text{tr}(P_j D_1)]^2 P_j \\ T_3 &= \sum_{j=1}^3 \sum_{k: k \neq j}^3 \sum_{l: l \neq j}^3 \frac{\log \lambda_j}{(\lambda_k - \lambda_j)(\lambda_l - \lambda_j)} \\ &\quad \times \mathbb{E}[P_j D_1 P_k D_1 P_l + P_k D_1 P_j D_1 P_l + P_k D_1 P_l D_1 P_j] \\ &\quad - \sum_{j=1}^3 \sum_{k: k \neq j}^3 \frac{\log \lambda_j}{(\lambda_k - \lambda_j)^2} \mathbb{E}[P_j D_1 P_j D_1 P_k + P_j D_1 P_k D_1 P_j + P_k D_1 P_j D_1 P_j] \\ &\quad - \sum_{j=1}^3 \sum_{k: k \neq j}^3 \frac{1}{\lambda_j (\lambda_k - \lambda_j)} \mathbb{E}[\text{tr}(P_j D_1)(P_j D_1 P_k + P_k D_1 P_j)] \\ T_4 &= -\frac{1}{2} \sum_{j=1}^3 \sum_{k=1}^3 \frac{1}{\lambda_j \lambda_k} \mathbb{E}[\text{tr}(P_j D_1 P_k D_1)] P_j \\ &\quad + \frac{1}{2} \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l: l \neq j}^3 \frac{\log \lambda_j}{\lambda_k (\lambda_l - \lambda_j)} \mathbb{E}[P_j D_1 P_k D_1 P_l + P_l D_1 P_k D_1 P_j]. \end{aligned}$$

Corollary 4.2 shows that the asymptotic biases of all three estimators are of the order  $O(\sigma^2)$ . Note that, each summand in the  $T_i$ 's can be expressed as a product of two terms, one is a polynomial of inverses of  $\lambda_j$ 's and their differences (and multiplied by  $\log \lambda_j$ 's in some instances), and the other involves linear functions of  $\text{Var}(\text{vec}(D_1))$  and  $\mathbb{E}(\text{vec}(D_2))$ . By (7) and (10), the latter quantities are dependent on  $D^0$  essentially only through the matrix  $(\sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T)^{-1}$ ,

as long as the number of gradient directions is not too small. Now consider the case where  $D^0$  is highly anisotropic, such that its smallest eigenvalue approaches zero while assuming constant diffusivity, i.e., constant  $\det(D^0)$ . Under this setting, the terms  $T_2$ ,  $T_3$  and  $T_4$  diverge faster than  $T_1$  due to the presence of additional multipliers of the form  $1/\lambda_j$  or  $1/(\lambda_k - \lambda_j)$  in  $T_2$ ,  $T_3$  and  $T_4$ . This, and the fact that  $r_\Omega = o(1)$ , imply that  $\text{ABias}(\log \bar{D}_{LE}; \log D^0)$  grows faster than  $\text{ABias}(\log \bar{D}_E; \log D^0)$ . For comparing the asymptotic biases of  $\log \bar{D}_{Aff}$  and  $\log \bar{D}_E$ , first it can be checked that  $T_4$  diverges faster than  $r_\Omega T_2$ . We further suppose that  $\lambda_{j-1} - \lambda_j \geq c_0 \lambda_{j-1}$  for some fixed  $c_0 > 0$  for  $j = 2, 3$ . This condition ensures that the eigenvalues of  $D^0$  do not coalesce as  $D^0$  becomes more anisotropic and is needed to show that  $T_4$  diverges faster than  $r_\Omega T_3$ . Specifically, as long as  $r_\Omega \max\{|\log \lambda_1|, |\log \lambda_3|\} = o(1)$ , then under the above setting,  $\text{ABias}(\log \bar{D}_{Aff}; \log D^0)$  grows faster than  $\text{ABias}(\log \bar{D}_E; \log D^0)$ . These show that when the true tensor is highly anisotropic, the Euclidean smoother tends to have smaller second order bias than the geometric smoothers.

If linear regression estimates are used as input for smoothing, then it is easy to see that the Euclidean smoother has a bias of order  $O(r_\Omega \sigma^2) = o(\sigma^2)$ , while the biases in the two geometric smoothers are of the order  $O(\sigma^2)$ . This is because, from Proposition 2.1 we can deduce that  $\mathbb{E}(\bar{\Delta}_E^*) = 0$ , and hence the terms involving  $\mathbb{E}(\bar{\Delta}_E^*)$  do not contribute to the expansions of the asymptotic biases.

In summary, when the tensor field is locally homogeneous, the Euclidean smoother tends to have smaller second order bias than the geometric smoothers.

**Case 2.** We assume that the underlying true tensors  $D^0(\omega)$ 's are perturbed versions of the target tensor  $D^0$ , where the perturbation is additive on the logarithm of the eigenvalues of  $D^0$ , but does not alter the eigenvectors. Specifically, let the spectral decomposition of  $D^0$  be  $D^0 = G\Lambda G^T$ , where  $\Lambda$  is a diagonal matrix with elements being the ordered eigenvalues of  $D^0$ :  $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ . Then, we assume

$$D^0(\omega) = G \text{diag}(\lambda_1 e^{\tau Z_1(\omega)}, \lambda_2 e^{\tau Z_2(\omega)}, \lambda_3 e^{\tau Z_3(\omega)}) G^T, \quad (31)$$

where  $\tau > 0$  is a scale parameter and the random variables  $\{(Z_j(\omega))_{j=1}^3 : \omega \in \Omega\}$  satisfy the following:

**Condition M:**  $\mathbb{E}_\Omega(Z_j) = 0$  for  $j = 1, \dots, 3$ ;  $Z_j(\omega)$ 's are uniformly bounded in  $\omega$ .

Note that, under this model, the perturbations are larger along the dominant eigen-directions of  $D^0$ . This can be seen as an additive noise structure under the log-Euclidean geometry. Since by (31),  $D^0(\omega)$ 's commute with each other, the log-Euclidean and affine invariant means of  $\{D^0(\omega) : \omega \in \Omega\}$  are the same. Moreover, under condition **M**, the log-Euclidean mean can be easily seen to be equal to  $D^0$ . However, by Jensen's inequality,  $\mathbb{E}_\Omega(e^{Z_j}) > 1$  unless  $Z_j$  is degenerate at zero, which implies that Euclidean mean is not equal to  $D^0$ . Indeed, the difference between the two means is of order  $O(\tau^2)$ .

Assume that  $\sigma = o(\tau)$ , where  $\sigma$  is the standard deviation of the additive noise associated with the DWI data. This means that the noise in DWI data is

small compared with the degree of spatial variation. As in *Case 1*, we compare the asymptotic biases of the three smoothers under the assumption that  $r_\Omega := \sum_{\omega \in \Omega} (p(\omega))^2 = o(1)$  as  $\tau \rightarrow 0$ . For simplicity of exposition, we only state the result when  $D(\omega)$  is the nonlinear regression estimator and  $D^0$  is anisotropic (i.e., its eigenvalues are all distinct).

**Corollary 4.3.** *Suppose that  $D^0(\omega)$  is given by (31) and  $D(\omega)$  denotes the nonlinear regression estimate. If  $\sigma = o(\tau)$ , and  $r_\Omega = o(1)$ , as  $\tau \rightarrow 0$ , and condition **M** holds, then if the eigenvalues of  $D^0$  are all distinct,*

$$\text{ABias}(\log \bar{D}_E, \log D^0) = \frac{\tau^2}{2} \left( \sum_{j=1}^3 \mathbb{E}_\Omega(Z_j^2) P_j \right) + \sigma^2(T_1 - r_\Omega T_2 + r_\Omega T_3) + O(\tau^3)$$

$$\text{ABias}(\log \bar{D}_{LE}, \log D^0) = \sigma^2(T_1 - T_2 + T_3) + O(\tau^3)$$

$$\text{ABias}(\log \bar{D}_{Aff}, \log D^0) = \sigma^2(T_1 - r_\Omega T_2 + r_\Omega T_3 + (1 - r_\Omega)T_4) + O(\tau^3)$$

where the terms  $T_i$ ,  $i = 1, \dots, 4$ , are as in Corollary 4.2.

Corollary 4.3 shows that the asymptotic biases of the geometric smoothers are smaller than that of the Euclidean smoother in terms of the logarithm of the mean. Corollary 4.3 can be proved by using arguments similar to those used in proving Corollary 4.2 but with a different expansion for  $\Delta(\omega)$ . The details are given in the Appendix.

**Remark 4.1.** Qualitatively similar statements hold under a more general perturbation model, with the spectral decomposition for  $D^0(\omega)$  given by

$$D^0(\omega) = \exp(\tau X(\omega)) G \cdot \text{diag}(\lambda_1 \exp(\tau Z_1(\omega)), \lambda_2 \exp(\tau Z_2(\omega)), \lambda_3 \exp(\tau Z_3(\omega))) \cdot G^T \exp(-\tau X(\omega)) \tag{32}$$

where  $X(\omega)$ 's are skew-symmetric random matrices which are uniformly bounded with respect to  $\omega$  and are independent of  $\{Z_j\}_{j=1}^3$ . Thus, under this model the perturbation is in terms of both eigenvalues and eigenvectors and the log-Euclidean mean of  $\{D^0(\omega) : \omega \in \Omega\}$  is only approximately the same as  $D^0$ . As long as the magnitude of eigenvector perturbations is of a smaller order than that of the eigenvalue perturbations (specifically,  $\mathbb{E}_\Omega(X) = o(\tau)$  and  $\mathbb{E}_\Omega(X^2) = o(1)$ , as  $\tau \rightarrow 0$ ), the geometric smoothers have smaller bias than the Euclidean smoother.

**Remark 4.2.** Propositions 8.2 and 8.3 (in the Appendix) provide clues as for when the affine-invariant smoother is likely to be more efficient than the log-Euclidean smoother. Suppose that the noise structure is such that in the expansions of  $\log \bar{D}_{LE}$  and  $\log \bar{D}_{Aff}$ , the terms that are quadratic in  $\bar{\Delta}_E$  can be ignored. Then, the dominant terms in the asymptotic biases are  $\hat{T}_1 - \hat{T}_2 + \hat{T}_3$  for  $\log \bar{D}_{LE}$  and  $\hat{T}_1 + \hat{T}_4$  for  $\log \bar{D}_{Aff}$ , where  $\hat{T}_i$ 's has the same functional form as  $\tilde{T}_i$ 's in (59) with  $\bar{\Delta}_E^*$  replaced by  $\bar{\Delta}_E$  and  $\Delta^*$  replaced by  $\Delta$ . Since the terms in the expression for  $\hat{T}_3$  involve inverses of quadratic terms in the spacings between the eigenvalues, while all the other terms involve inverses of at most linear terms

in the spacings, this suggests that the affine-invariant smoother tends to have smaller bias compared to the log-Euclidean smoother provided that  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are of similar magnitude while the spacing  $(\lambda_{j-1} - \lambda_j)$  is of smaller order than  $\lambda_{j-1}$  for at least one  $j \in \{2, 3\}$ .

The analysis in this subsection suggests that under the small noise setting, the Euclidean smoother may or may not outperform the geometric smoothers in terms of bias, depending on whether raw DWI sensor noise or spatial heterogeneity of the tensor field dominate.

## 5. Simulation studies

In this section, we conduct simulation studies that explore how the differences among the three kernel smoothers extend to scenarios with relatively larger levels of sensor noise.

Besides the choice of metrics, one also needs to choose a scheme to assign weights  $w_i(s)$ 's in (12). The conventional approach is to simply set weights as in (11) where  $K$  is a fixed kernel. This is referred to as *isotropic smoothing*. However, the tensor field often shows various degrees of anisotropy in different regions and the tensors tend to be more homogeneous along the leading anisotropic directions (e.g., along the fiber tracts). Thus it makes sense to set the weights larger if the vector  $s_i - s$  is aligned to the leading diffusion direction at  $s$ . Therefore, we propose the following *anisotropic weighting scheme*:

$$\omega_i(s) := K_h \left( \sqrt{\text{tr}(\widehat{D})(s_i - s)^T \widehat{D}^{-1}(s_i - s)} \right), \quad (33)$$

where  $\widehat{D}$  is the current estimate of the tensor at voxel location  $s$ , and  $K_h(\cdot) := K(\cdot/h)$  is a nonnegative, integrable kernel with  $h > 0$  being the bandwidth. The use of  $\text{tr}(\widehat{D})$  in (33) is to set the weights scale-free with respect to  $\widehat{D}$ . There are other schemes for anisotropic weights. For example, in Tabelow et al. (2008), the term  $\text{tr}(\widehat{D})$  is replaced by  $\det(\widehat{D})$  in (33), which is supposed to capture not only the directionality of the local tensor field, but also the degree of anisotropy. Chung et al. (2003, 2005) also propose kernel smoothing under Euclidean geometry with a different scheme of anisotropic kernel weights.

In the following, we conduct two simulation studies. Simulation design I corresponds to *Case 1* studied in Section 4.2, where the true tensor field is locally homogeneous and the dominant source of variability in the tensor field comes from the raw DWI data. In the second simulation study, we consider a case with substantial spatial variations in the underlying tensor field which is generated based on a real human brain diffusion MRI data set. Thus this setting may be seen as a realistic generalization of *Case 2*. For both simulations, we consider different levels of observational noise in the raw DWI data. For performance measure, we use the median (across a region of the tensor field) of the squared affine-invariant distances between true and smoothed tensors at a variety of combinations of preliminary isotropic smoothing bandwidth and

secondary anisotropic smoothing bandwidth. We choose median distance over mean distance because of its robustness and the fact that the log-Euclidean and affine-invariant metrics occasionally produce extreme estimates in scenarios where the true tensor is near singular and/or the noise level is high. Moreover, using Euclidean distance as error measure leads to qualitatively similar results.

### 5.1. Simulation I

Here we construct a simulated tensor field on a  $128 \times 128 \times 4$  three-dimension grid consisting of the *background regions* with identical isotropic tensors and the *banded regions* with three parallel vertical bands and three horizontal bands (for each of the four slices), where within each band tensors are identical and aligned in either the  $x$ - or the  $y$ - direction. The bands are of various widths and degrees of anisotropy (see Figure S-1 and Table S-1 of the Supplementary Materials for details). For a clearer comparison of different smoothers, we examine their performance on four sets of tensors: (i) the “whole set” – the entire set of tensors; (ii) the “crossings,” where pairs of bands intersect or bands intersect with the background; (iii) the “background interior” regions that are within the background and are at least four voxels away from any crossing; and (iv) the “band interior” regions that are within a band and are at least four voxels away from the background. The purpose is to compare smoother performance on homogeneous regions where diffusion is isotropic (background interior) and anisotropic (band interior), as well as on heterogeneous regions where the diffusion direction and the degree of anisotropy vary within individual neighborhoods (crossings).

At each voxel, we simulate the raw DWI data  $\hat{S}_{\mathbf{q}}$ 's using the true tensor at that voxel and the Rician noise model (equations (1) to (3)). Specifically, we set  $S_0 = 1,000$  and use 9 gradient directions each repeated twice, which are normalized versions of the following vectors

$$(1, 0, 1), (1, 1, 0), (0, 1, 1), (3, 2, 1), (0.9, 0.45, 0.2), (1, 0, 0), (0, 1, 0), \\ (0, 0, 1), (2, 1, 1.3).$$

This gradient design is from a real DT-MRI study performed in UC Davis. We consider three different values of the noise parameter  $\sigma = 10, 50, 100$ , which correspond to  $\text{SNR} := S_0/\sigma = 100, 20$  and  $10$ . These are referred to as “low”, “moderate” and “high” noise levels, respectively. Such signal to noise ratios are typical for real DTI studies with  $\text{SNR} = 100$  at the higher end and  $\text{SNR} = 10$  at the lower end (Farrell et al., 2007; Parker et al., 2000; Tuch et al., 2002). Finally, at each voxel, a nonlinear regression procedure is applied to the DWI data to derive the observed tensors as inputs for the smoothing procedure.

Errors in terms of affine-invariant distance between kernel-smoothed and ground-truth tensors are summarized in Figures 1, 2 and 3. At the low noise level ( $\sigma = 10$ ), the Euclidean smoother works marginally better than the geometric smoothers on the isotropic homogeneous region – “background interior”, and its advantage is more pronounced on the anisotropic homogeneous region – “bands interior”. This is consistent with the analysis in Section 4.2. Moreover, smooth-

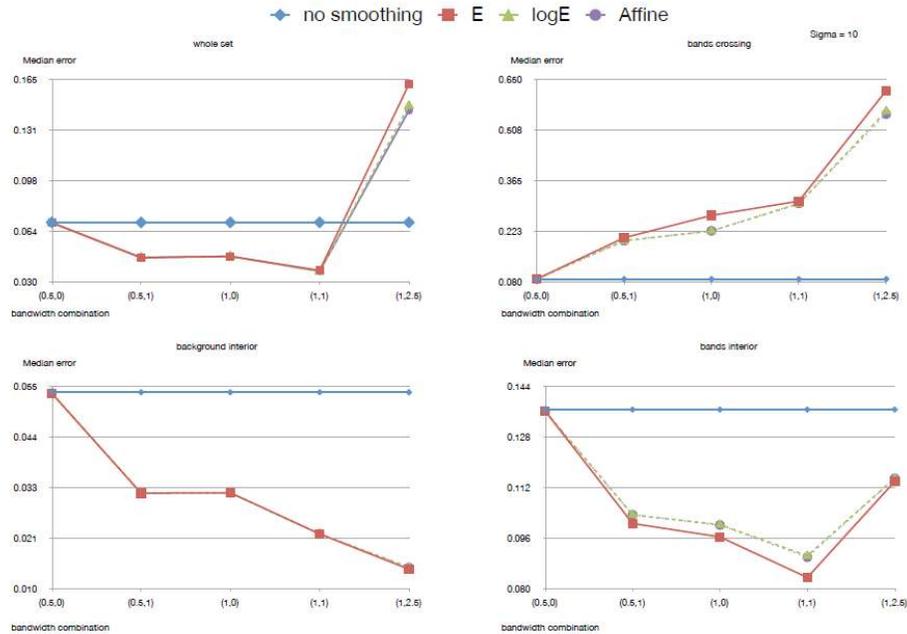


FIG 1. **Simulation I:**  $\sigma = 10$ . Comparison of median errors measured by affine invariant distance over different regions for “observed” tensors (obtained by **nonlinear regression**) and Euclidean, log-Euclidean and Affine smoothers. In “bandwidth combination”: the first number denotes the isotropic bandwidth and the second number denotes the anisotropic bandwidth.

ing is not beneficial on the heterogenous “bands crossing” regions at low noise level. At higher noise levels ( $\sigma = 50, 100$ ), Euclidean smoothing substantially outperforms geometric smoothing in anisotropic regions (“bands interior” and “bands crossing”), and is slightly better for the isotropic region (“background interior”). The two geometric smoothers perform comparably regardless of noise levels and regional heterogeneity, although affine-invariant smoothing is slightly better at higher noise levels. Also, anisotropic smoothing is seen to be beneficial in the anisotropic “bands interior” regions when noise level is low.

## 5.2. Simulation II

Here, we first smoothed four axial slices of a real DTI scan for one human subject from the data set described in Section 6 using affine-invariant smoothing. We then used the resulting tensors as the underlying true tensors and simulated the raw DWI data using the 18 gradient directions as described in Section 5.1 and adding Rician noise (Figure 4). We set the baseline signal strength  $S_0 = 1000$  and consider two noise levels, namely “low noise” –  $\sigma = 10$  and “moderate noise” –  $\sigma = 50$ .

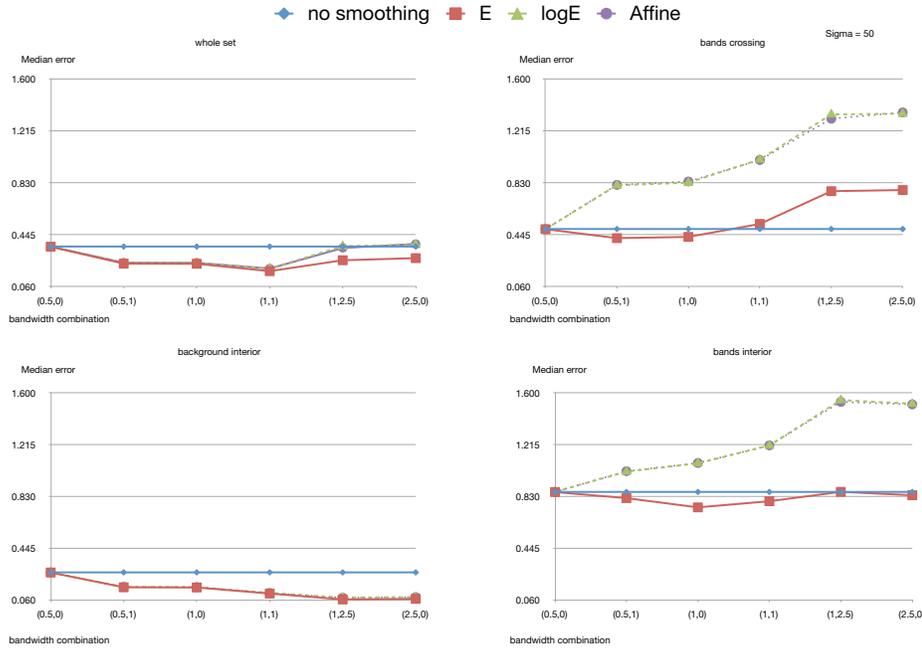


FIG 2. **Simulation I:**  $\sigma = 50$ . Comparison of median errors measured by affine invariant distance over different regions for “observed” tensors (obtained by **nonlinear regression**) and Euclidean, log-Euclidean and Affine smoothers. In “bandwidth combination”: the first number denotes the isotropic bandwidth and the second number denotes the anisotropic bandwidth.

Errors in terms of affine-invariant distance between kernel-smoothed and ground-truth tensors are shown in Figure 5 (for  $\sigma = 10$ ) and Figure 6 (for  $\sigma = 50$ ). As can be seen from these figures, for each isotropic bandwidth smaller than 0.9 for  $\sigma = 10$  and smaller than 1.3 for  $\sigma = 50$ , there is an optimal anisotropic smoothing bandwidth, and bandwidths larger and smaller than the optimal one suffer from over- and under-smoothing, respectively. For  $\sigma = 10$ , the geometric smoothers outperform the Euclidean smoother, especially at larger bandwidths. In contrast, when  $\sigma = 50$ , the Euclidean smoother has smaller error than the geometric smoothers. This is consistent with the findings in Section 4, that is, when spatial heterogeneity is dominant over the sensor noise in DWI data, the geometric smoothers may be more advantageous, while the Euclidean smoothers are more advantageous when sensor noise is dominant. We also observe that at the low noise level, the affine-invariant smoother generally performs better than the log-Euclidean smoother. It is conjectured that a presence of the features described in Remark 4.2 in Section 4.2 in a significant portion of the tensor field may be partly responsible for the observed phenomenon.

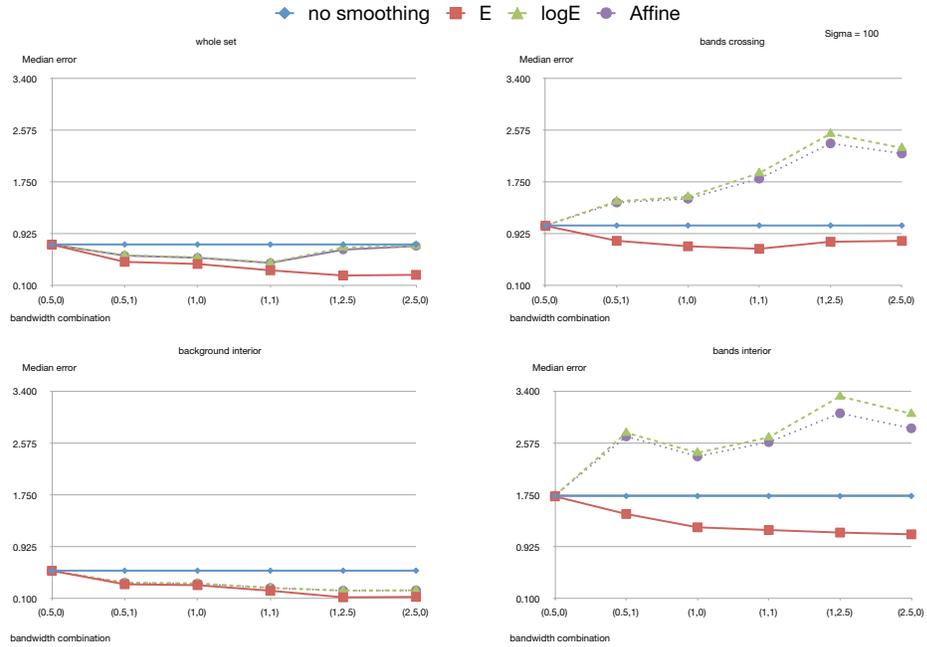


FIG 3. **Simulation I:**  $\sigma = 100$ . Comparison of median errors measured by affine invariant distance over different regions for “observed” tensors (obtained by nonlinear regression) and Euclidean, log-Euclidean and Affine smoothers. In “bandwidth combination”: the first number denotes the isotropic bandwidth and the second number denotes the anisotropic bandwidth.

In summary, the results in Sections 4 and 5 show that the choice of the best kernel smoother depends on the sensor noise level in the raw DWI data and the degree and nature of spatial variation of the underlying tensor field. When the noise from the raw DWI data dominates, the Euclidean smoother is less biased and tends to perform better. On the other hand, if spatial variation of the tensor field is dominant, geometric smoothers may perform better. Moreover, the simulation results also show that anisotropic smoothing is often beneficial in anisotropic and heterogeneous regions.

## 6. Application to DT-MRI scans of human brain

A third evaluation of the kernel smoothers was performed on a set of 33 real DTI scans of elderly individuals who volunteered for research at the UC Davis Alzheimer’s Disease Center. The purpose of the experiments was to evaluate the degree to which the smoothers enhanced or diminished the biological plau-

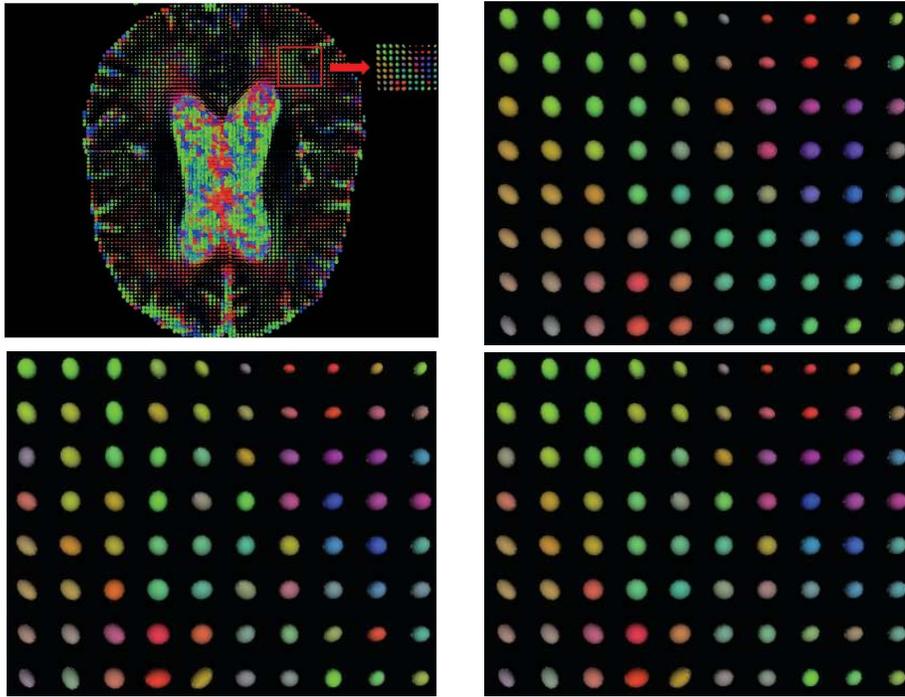


FIG 4. **Simulation II:** Upper left panel: raw tensors in slice 10; Upper right panel: smoothed tensors used as ground truth in the simulation in the rectangle region shown in the upper left panel; Lower left panel: noisy tensors after adding Rician noise ( $\sigma = 10$ ); Lower right panel: smoothed tensors by Affine smoother

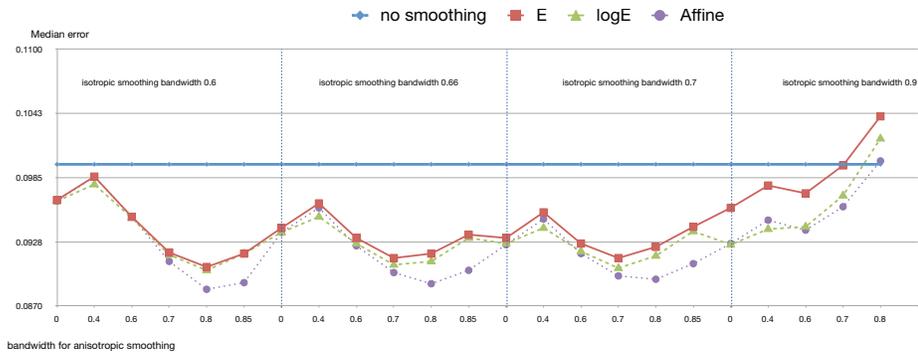


FIG 5. **Simulation II:**  $\sigma = 10$ . Comparison of median errors measured by affine-invariant distance across the whole tensor field for “observed” tensors (obtained by nonlinear regression) and Euclidean, log-Euclidean and Affine smoothers.

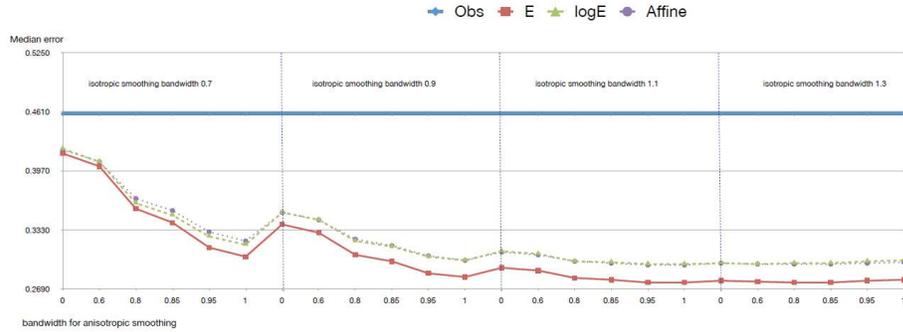


FIG 6. **Simulation II:**  $\sigma = 50$ . Comparison of median errors measured by affine-invariant distance across the whole tensor field for “observed” tensors (obtained by **nonlinear regression**) and Euclidean, log-Euclidean and Affine smoothers.

sibility of white matter integrity and inter-regional connectivity measures calculated from the diffusion tensors. We assessed the plausibility of the DTI-based measures in terms of a few bedrock neuroanatomical principles: (a) in the cerebrospinal fluid (CSF), water is free to diffuse in any direction, and therefore fractional anisotropy (FA, defined in equation (34) below), which measures the degree to which diffusion tensor suggests an anisotropic water diffusion distribution, should be near zero in the CSF; (b) the corpus callosum is a highly-organized white matter tract, and therefore its FA should be relatively high; (c) white matter tracts do not travel through CSF compartments, and therefore the number of fibers traced by DTI tractography that intersect CSF spaces should be low; and (d) the parieto-temporo-occipital subregion of the corpus callosum connects the left and right parietal, temporal, and occipital lobes to each other, and therefore the number of tractography fibers that connect those lobar regions via the parieto-temporo-occipital subregion of the corpus callosum should be relatively high. Noise in DTI acquisition may cause the collected DTI data to violate these basic principles. Our goal is to evaluate the degree to which the removal of noise via kernel smoothing reduces the instances of such violations.

### 6.1. Data

Imaging was performed at the UC Davis Imaging Research Center on a 1.5T GE Signa Horizon LX Echospeed system. Subjects were scanned in a supine position with an 8-channel head coil placed surrounding the head. After placement of the head coil, subjects were inserted into the MRI scanner magnetic field and two MRI sequences were acquired: a three-dimensional T1-weighted coronal spoiled gradient-recalled echo acquisition (T1) used for parcellating the brain into tissue classes and regions of interest; and a single-shot spin-echo echo planar imaging DTI sequence used for estimating diffusion tensors.

The T1-weighted sequence was an axial-oblique 3D Fast Spoiled Gradient Recalled Echo sequence with the following parameters: TE: 2.9 ms (min), TR: 9 ms (min), Flip angle: 15 deg, Slice thickness: 1.5 mm, Number of Slices: 128, FOV: 25 cm x 25 cm, Matrix: 256 x 256. Data from the T1 sequence gave an indication of the tissue type at each location in the brain: white matter appeared brightest, cerebrospinal fluid (CSF) appeared darkest, and gray matter appeared as an intermediate gray. This contrast between tissue types enabled an expert rater to manually trace the corpus callosum (a white matter structure), along with a subdivision of the corpus callosum into four sub-regions, using established protocols on a population-averaged brain called the T1-weighted Minimum Deformation Template (MDT) (Kochunov et al., 2001). The subdivision partitioned the corpus callosum into four zones that carry interhemispheric axonal connections within prefrontal cortex, premotor and supplementary motor cortex, sensory-motor cortex, and parieto-temporo-occipital cortex without post-central gyrus respectively. The MDT represents the brain of a prototypical elderly individual whose brain anatomy has been warped to represent the approximate average anatomy of a large group of cognitively-healthy elderly individuals. In addition, an established method was used to segment the MDT into gray, white, and CSF tissue compartments (Rajapakse et al., 1996). The MDT was nonlinearly warped to the skull-stripped T1-weighted scan of each individual (Rueckert et al., 1999), thus allowing the delineation of the corpus callosum and its sub-regions to be transferred to the brain of each individual in the study.

Relevant DTI acquisition parameters include: TE: 94 ms, TR: 8000 ms, Flip angle: 90 degrees, Slice thickness: 5 mm, slice spacing: 0.0 mm, FOV: 22 cm x 22 cm, Matrix: 128 x 128, B-value: 1000 s/mm<sup>2</sup>. Each acquisition included collection of 2 B0 images and 4 diffusion-weighted images acquired along each of 6 gradient directions. The directions vectors were:

$$(1, 0, 1), (1, 1, 0), (0, 1, 1), (-1, 0, 1), (-1, 1, 0), (0, -1, 1).$$

Geometric distortions were removed from the diffusion-weighted images (De Crespigny and Moseley, 1998), and diffusion tensors were estimated using a linear least squares estimator (Basser and Pierpaoli, 2005). The eigenvalues of the diffusion tensors were then estimated, and FA was calculated at each voxel as follows:

$$FA = \frac{1}{\sqrt{2}} \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \quad (34)$$

where  $\lambda_j$ 's are the eigenvalues of the tensor. The T1-weighted scan was affinely aligned to the image of FA values derived from the corresponding DTI data. This allowed the CSF, corpus callosum, and corpus callosum subdivision labels to be transferred from the space of the MDT image, to the space of the subject T1-weighted scan, and then to the space of the corresponding DTI data. Label maps in the DTI space were eroded by one voxel to remove spurious labels generated by partial volume effects.

## 6.2. Experimental design

For each of the 33 DTI scans, we calculated FA at all voxels labeled as CSF, prior to smoothing. FA is a commonly used measure of the degree of anisotropy of diffusion: if diffusion is isotropic, then  $FA = 0$ ; if diffusion is highly anisotropic, then FA is near one. Then, for each smoother, we calculated FA at CSF voxels after smoothing. We used the same two-stage isotropic-anisotropic smoothing procedure described in the simulation study (Section 5), with isotropic bandwidth of 0.8 and anisotropic bandwidth of 1.8. Analogous results with a smaller anisotropic bandwidth of 1.2 are very similar. For each scan, we summarized smoothing-induced shifts in the distribution of CSF FA values by calculating the difference in the median CSF FA values before and after smoothing (Results are similar for smoothing-induced differences in the 75th and 90th percentiles of CSF FA values). Greater reductions in median CSF FA caused by smoothing would indicate greater smoothing performance in terms of encouraging biological plausibility. Similarly, for each scan and smoother, we calculated the difference in median corpus callosum FA before and after smoothing; greater reductions in callosal FA caused by smoothing are considered detrimental, since relatively higher FA there is more plausible.

We then used the MedINRIA software package (Toussaint et al., 2007), and its implementation of the TensorLines algorithm (Weinstein et al., 1999), to trace white matter tract fibers throughout the brain based on the raw diffusion tensors and tensors that were smoothed using each of the three smoothers. All tract fibers were stored as trajectories of 3D points in the DTI space. For each individual, in-house software was used to select only those fibers that intersect selected regions of interest defined by our anatomical labels. We first isolated the tract fibers that intersected voxels labeled as CSF, and counted the number of such CSF-intersecting tract fibers before and after smoothing. Greater reduction of such spurious tract fibers is an indicator of higher performance of the smoother. We then isolated tract fibers that intersected both the parieto-temporo-occipital portion of the corpus callosum, and the white matter of the occipital lobe; greater addition of such plausible fibers is an indicator of higher performance. However, we guarded against the possibility that smoothers increased the number of such plausible fibers simply by increasing the total number of fibers, both plausible and implausible, that connect the parieto-temporo-occipital corpus callosum to any and all parts of the brain— to do this, we calculated the number of implausible fibers connecting the parieto-temporo-occipital corpus callosum to the prefrontal cortex before and after smoothing, to be sure that this number did not increase with smoothing.

## 6.3. Results

A boxplot of smoothing-induced reductions in median CSF FA, across all 33 individuals for the three smoothers is shown in the upper left panel of Figure 7. As in the simulation study, all three smoothers performed similarly in this highly

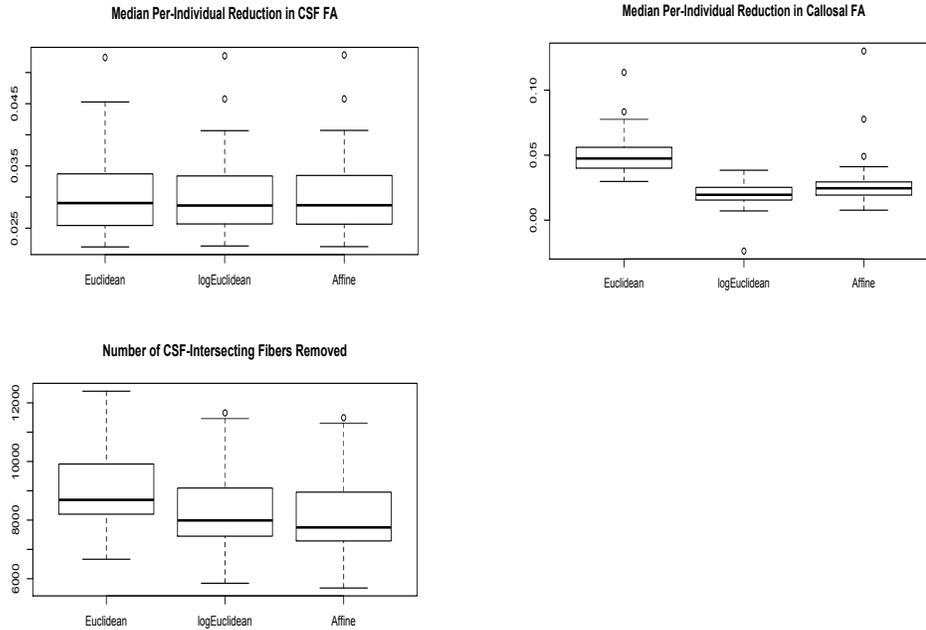


FIG 7. **Application.** **Upper left panel:** Boxplots of the reduction in median CSF FA before and after smoothing for the three smoothers across 33 scans. **Upper right panel:** Boxplots of the reduction in median callosal FA before and after smoothing for the three smoothers. **Lower left panel:** Boxplots of reduction in the number of spurious tract fibers intersecting the CSF before and after smoothing for the three smoothers.

isotropic region. However, both geometric smoothers succeeded in maintaining diffusion anisotropy in the corpus callosum (CC) much better than the Euclidean smoother (Figure 7, upper right), with Euclidean smoothing erroneously reducing the high FA inherent in this structure by approximately 0.05 in terms of median FA. Analogous histograms for differences in other per-individual FA summary measures are similar. This result reinforces the findings of the simulation study: the geometric smoothers may be more effective at maintaining the structure of real-world, highly-organized tensor fields while removing low levels of noise.

A boxplot of smoothing-induced reductions in number of CSF-intersecting fibers is shown in the lower left panel of Figure 7. The three methods perform similarly, although the Euclidean smoother may be superior for removing such spurious fibers. The left panel of Figure 8 shows how many additional fibers between occipital cortex and occipital corpus callosum were traced by tractography after smoothing compared to before. While the performance of the three smoothers is similar, the number of fibers added by the geometric smoothers is slightly higher, suggesting greater performance in encouraging such biologi-

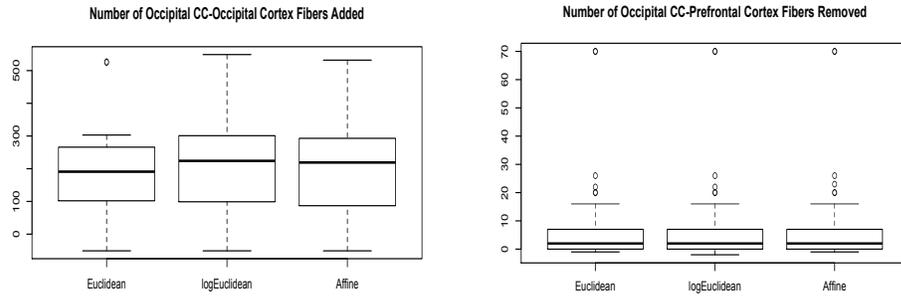


FIG 8. **Application.** **Left panel:** Boxplots of the addition in the number of plausible tract fibers intersecting the occipital corpus callosum and occipital lobe before and after smoothing for the three smoothers across 33 scans. **Right panel:** Boxplots of the differences in the number of implausible tract fibers intersecting the occipital corpus callosum and prefrontal lobe before and after smoothing for the three smoothers.

cally plausible fibers. Meanwhile, the three methods are nearly identical in their ability to prevent the number of spurious fibers intersecting occipital corpus callosum and prefrontal cortex from increasing (right panel of Figure 8).

## 7. Discussion

Based on both theoretical and numerical analysis, the key finding of this study is that the performance of diffusion tensor smoothers depends heavily on the characteristics of both the noise from the MRI sensor and the geometric structure of the underlying tensor field. The asymptotic expansion of the linear and nonlinear regression estimators in Section 2 quantifies the advantage of the latter, which suggests that the nonlinear estimator should be incorporated in standard software packages for DTI analysis. The perturbation analysis in Sections 2 and 4 shows that under the small noise regime, if the tensors in the smoothing neighborhood is homogeneous then the Euclidean smoother tends to have a smaller bias than the geometric smoothers. On the other hand, if the local heterogeneity of the tensor field is the dominant component of variation of the tensors, and the noiseless tensors follow a certain geometric regularity, then the geometric smoothers may lead to smaller bias. In addition, the simulation studies show that if the sensor noise from DWI data is the dominant source of variation of the tensors, the Euclidean smoother outperforms the geometric smoothers as the latter tend to fit the spurious structure induced by the noise under such a case. Together, these findings suggest that DTI users may need to revisit the conventional wisdom that geometric smoothers are generally superior to Euclidean smoothers. In fact, optimal DTI smoothing may need to be spatially adaptive, applying varying DTI metrics depending on local geometric structures and levels of sensor noise. The simulation results also point to the benefits of anisotropic smoothing in the highly anisotropic regions.

These results are mostly in agreement with the qualitative features of our findings for the real DT-MRI data (Section 6), even though there the metrics for performance and comparison are different and are based on the biological plausibility of the subsequent FA maps and tractography results. As in the simulations, the geometric smoothers appear to perform better in highly structured, anisotropic regions such as the corpus callosum and the fiber tract pathways in the occipital lobe; meanwhile, all smoothers perform similarly in fairly isotropic regions, such as the CSF. In addition, the two geometric smoothers perform similarly in all aspects. The Euclidean smoother does an exceptional job at removing spurious tract fibers that travel through both unstructured regions (CSF) and structured regions. An explanation for this, suggested by the Euclidean smoother’s reduction of callosal FA and occipital fiber counts, is that it discourages highly structured tensor regions globally, including those that accidentally occur in the CSF. The geometric smoothers, meanwhile, may tend to preserve such structure.

In this paper, we have not addressed the issue of choosing smoothing parameters as the primary goal here is to demonstrate the comparative performance of different smoothers. In practice, one may use an AIC type criterion or a generalized cross validation criterion, as proposed in (Yuan et al., 2012). However, when the tensor field is inhomogeneous, any “global” choice of the bandwidth may not be very effective. How to choose the bandwidth adaptively, while accounting for spatial inhomogeneity is a future direction of study. The simulation results also demonstrate some degrees of boundary effect. For example, at the boundary of a band of anisotropic tensors and the background of isotropic tensors, the performance of tensor smoothers is different from that in the interior of the bands. Such boundary effect may be mitigated by a careful choice of the anisotropic neighborhoods (for example, using a multi-stage scheme as in Tabor et al. (2008)), or by employing higher order polynomial approximation of the tensor field as in Yuan et al. (2012).

In this paper, we did not explicitly specify a probabilistic model for the tensor distribution. Rather, it is implicitly determined by the noise model of DWI data and the tensor estimation procedure at each voxel. In Section 4.2, under “Case 2”, we considered a specific local probabilistic structure while comparing the bias characteristics of geometric smoothers with that of the Euclidean smoother. This structure is based on a spectral decomposition of the tensor and with independent random variations in the eigenvalues and eigenvectors. Possible extensions of such a probabilistic framework is a future direction of research.

## 8. Appendix

### 8.1. Proofs of Proposition 2.1 and Proposition 2.2

Let the phase vector for the DWI signal corresponding to gradient direction  $\mathbf{q}$  be denoted by  $u_{\mathbf{q}}$ . Then  $u_{\mathbf{q}}$  is a unit vector in  $\mathbb{R}^2$ . Let  $v_{\mathbf{q}}$  be a unit vector in  $\mathbb{R}^2$  such that  $v_{\mathbf{q}}^T u_{\mathbf{q}} = 0$  so that  $u_{\mathbf{q}} u_{\mathbf{q}}^T + v_{\mathbf{q}} v_{\mathbf{q}}^T = I_2$ .

Next, for an arbitrary tensor  $D$  (written in the vectorized form), an arbitrary gradient direction  $\mathbf{q}$  and an arbitrary  $w \in \mathbb{R}^2$ ,

$$f_{\mathbf{q}}(w, D) := \| S_0 e^{-x_{\mathbf{q}}^T D} u_{\mathbf{q}} + w \| . \quad (35)$$

Then,  $f_{\mathbf{q}}(0, D^0) = S_{\mathbf{q}}$  and  $f_{\mathbf{q}}(\sigma \varepsilon_{\mathbf{q}}, D^0) = \hat{S}_{\mathbf{q}}$ . We denote the partial derivatives with respect to the first and second arguments by  $\nabla_{ij} f_b$  etc, where  $1 \leq i, j \leq 2$ . Then,

$$\nabla_1 f_{\mathbf{q}}(w, D^0) = \frac{1}{f_{\mathbf{q}}(w, D^0)} (S_{\mathbf{q}} u_{\mathbf{q}} + w) \implies \nabla_1 f_{\mathbf{q}}(0, D^0) = \frac{1}{S_{\mathbf{q}}} S_{\mathbf{q}} u_{\mathbf{q}} = u_{\mathbf{q}}. \quad (36)$$

Thus, using (36),

$$\begin{aligned} \nabla_{11} f_{\mathbf{q}}(w, D^0) &= \frac{1}{f_{\mathbf{q}}(w, D^0)} I_2 - \frac{1}{(f_{\mathbf{q}}(w, D^0))^3} (S_{\mathbf{q}} u_{\mathbf{q}} + w)(S_{\mathbf{q}} u_{\mathbf{q}} + w)^T \\ \implies \nabla_{11} f_{\mathbf{q}}(0, D^0) &= \frac{1}{S_{\mathbf{q}}} (I_2 - u_{\mathbf{q}} u_{\mathbf{q}}^T) = \frac{1}{S_{\mathbf{q}}} v_{\mathbf{q}} v_{\mathbf{q}}^T. \end{aligned} \quad (37)$$

We also have,

$$\nabla_2 f_{\mathbf{q}}(0, D^0) = -S_0 \exp(-x_{\mathbf{q}}^T D^0) x_{\mathbf{q}} = -S_{\mathbf{q}} x_{\mathbf{q}} \quad (38)$$

and

$$\nabla_{22} f_{\mathbf{q}}(0, D^0) = S_0 \exp(-x_{\mathbf{q}}^T D^0) x_{\mathbf{q}} x_{\mathbf{q}}^T = S_{\mathbf{q}} x_{\mathbf{q}} x_{\mathbf{q}}^T. \quad (39)$$

*Proof of Proposition 2.1*

We show that, as  $\sigma \rightarrow 0$ ,  $\hat{D}_{LS} = D^0 + \sigma D_{1,LS} + \sigma^2 D_{2,LS} + O(\sigma^3 |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^3)$  where the random vectors  $D_{1,LS}$  and  $D_{2,LS}$  are given by

$$D_{1,LS} = -\left(\sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T\right)^{-1} \left(\sum_{\mathbf{q} \in \mathcal{Q}} \frac{1}{S_{\mathbf{q}}} (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}) x_{\mathbf{q}}\right), \quad (40)$$

and

$$D_{2,LS} = -\frac{1}{2} \left(\sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T\right)^{-1} \left(\sum_{\mathbf{q} \in \mathcal{Q}} \frac{1}{S_{\mathbf{q}}^2} ((v_{\mathbf{q}}^T \varepsilon_{\mathbf{q}})^2 - (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}})^2) x_{\mathbf{q}}\right). \quad (41)$$

Proposition 2.1 then follows from (40) and (41) by taking appropriate expectations and using the independence of  $\varepsilon_{\mathbf{q}}$ 's.

We use the representation

$$\begin{aligned} \hat{D}_{LS} &= -\left(\sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T\right)^{-1} \left(\sum_{\mathbf{q} \in \mathcal{Q}} (\log \hat{S}_{\mathbf{q}} - \log S_0) x_{\mathbf{q}}\right) \\ &= -\left(\sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T\right)^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} (\log f_{\mathbf{q}}(\sigma \varepsilon_{\mathbf{q}}, D^0) - \log f_{\mathbf{q}}(0, D^0) - x_{\mathbf{q}}^T D^0) x_{\mathbf{q}} \end{aligned}$$

$$\begin{aligned}
 &= D^0 - \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} \sigma \frac{(\nabla_1 f_{\mathbf{q}}(0, D^0))^T \varepsilon_{\mathbf{q}}}{f_{\mathbf{q}}(0, D^0)} x_{\mathbf{q}} \right) \\
 &\quad - \frac{1}{2} \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \sigma^2 \left[ \frac{\varepsilon_{\mathbf{q}}^T \nabla_{11} f_{\mathbf{q}}(0, D^0) \varepsilon_{\mathbf{q}}}{f_{\mathbf{q}}(0, D^0)} - \frac{(\varepsilon_{\mathbf{q}}^T \nabla_1 f_{\mathbf{q}}(0, D^0))^2}{(f_{\mathbf{q}}(0, D^0))^2} \right] x_{\mathbf{q}} \\
 &\quad + O(\sigma^3 |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^3).
 \end{aligned}$$

By invoking (36) and (37), we obtain (40) and (41).

*Proof of Proposition 2.2*

Unlike  $\widehat{D}_{LS}$ , there is no explicit expression for  $\widehat{D}_{NL}$ . Instead, it satisfies the normal equation:

$$\sum_{\mathbf{q} \in \mathcal{Q}} (\hat{S}_{\mathbf{q}} - S_0 \exp(-x_{\mathbf{q}}^T \widehat{D}_{NL})) S_0 \exp(-x_{\mathbf{q}}^T \widehat{D}_{NL}) x_{\mathbf{q}} = 0. \tag{42}$$

We prove that, as  $\sigma \rightarrow 0$ ,  $\widehat{D}_{NL} = D^0 + \sigma D_{1,NL} + \sigma^2 D_{2,NL} + O(\sigma^3 |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^3)$ , where

$$D_{1,NL} = - \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}} (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}) x_{\mathbf{q}} \right), \tag{43}$$

and

$$\begin{aligned}
 D_{2,NL} &= \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left[ \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 (x_{\mathbf{q}}^T D_{1,NL})^2 x_{\mathbf{q}} \right. \\
 &\quad \left. + \frac{1}{2} \sum_{\mathbf{q} \in \mathcal{Q}} (S_{\mathbf{q}} x_{\mathbf{q}}^T D_{1,NL} + u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}})^2 x_{\mathbf{q}} - \frac{1}{2} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^2 x_{\mathbf{q}} \right]. \tag{44}
 \end{aligned}$$

From (42) and the definition of  $\widehat{D}_{NL}$ , it can be shown using standard arguments that  $\|\widehat{D}_{NL} - D^0\| = O(\sigma |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|)$  as  $\sigma \rightarrow 0$ . Thus, expanding the LHS of (42) in Taylor series, we have

$$\begin{aligned}
 &\sum_{\mathbf{q} \in \mathcal{Q}} \left[ \sigma \varepsilon_{\mathbf{q}}^T \nabla_1 f_{\mathbf{q}}(0, D^0) + \frac{\sigma^2}{2} \varepsilon_{\mathbf{q}}^T \nabla_{11} f_{\mathbf{q}}(0, D^0) \varepsilon_{\mathbf{q}} - (\nabla_2 f_{\mathbf{q}}(0, D^0))^T (\widehat{D}_{NL} - D^0) \right. \\
 &\quad \left. - \frac{1}{2} (\widehat{D}_{NL} - D^0)^T \nabla_{22} f_{\mathbf{q}}(0, D^0) (\widehat{D}_{NL} - D^0) \right] \\
 &\quad \cdot \left( \nabla_2 f_{\mathbf{q}}(0, D^0) + \nabla_{22} f_{\mathbf{q}}(0, D^0) (\widehat{D}_{NL} - D^0) \right) \\
 &= O(\sigma^3 |\mathcal{Q}|^{-2} \left( \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^3 \right)).
 \end{aligned}$$

We can express  $\widehat{D}_{NL}$  as  $\widehat{D}_{NL} = D^0 + \sigma D_{1,NL} + \sigma^2 D_{2,NL} + O(\sigma^3 |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}\|^3)$ , where  $D_{1,NL}$  and  $D_{2,NL}$  involve terms that are only linear and only quadratic in  $\varepsilon_{\mathbf{q}}$ 's, respectively. Substituting this in the above expression and equating terms with multiplier  $\sigma$ , we have

$$D_{1,NL} = \left( \sum_{\mathbf{q} \in \mathcal{Q}} \nabla_2 f_{\mathbf{q}}(0, D^0) (\nabla_2 f_{\mathbf{q}}(0, D^0))^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} \varepsilon_{\mathbf{q}}^T \nabla_1 f_{\mathbf{q}}(0, D^0) \nabla_2 f_{\mathbf{q}}(0, D^0) \right)$$

which equals (43) by virtue of (38). Also, collecting terms with multiplier  $\sigma^2$ ,

$$\begin{aligned} D_{2,NL} = & \left( \sum_{\mathbf{q} \in \mathcal{Q}} \nabla_2 f_{\mathbf{q}}(0, D^0) (\nabla_2 f_{\mathbf{q}}(0, D^0))^T \right)^{-1} \cdot \\ & \left[ \frac{1}{2} \sum_{\mathbf{q} \in \mathcal{B}} (\varepsilon_{\mathbf{q}}^T \nabla_{11} f_{\mathbf{q}}(0, D^0) \varepsilon_{\mathbf{q}} - D_{1,NL}^T \nabla_{22} f_{\mathbf{q}}(0, D^0) D_{1,NL}) \nabla_2 f_{\mathbf{q}}(0, D^0) \right. \\ & + \sum_{\mathbf{q} \in \mathcal{Q}} (\varepsilon_{\mathbf{q}}^T \nabla_1 f_{\mathbf{q}}(0, D^0)) \nabla_{22} f_{\mathbf{q}}(0, D^0) D_{1,NL} \\ & \left. - \sum_{\mathbf{q} \in \mathcal{Q}} (D_{1,NL}^T \nabla_2 f_{\mathbf{q}}(0, D^0)) \nabla_{22} f_{\mathbf{q}}(0, D^0) D_{1,NL} \right]. \end{aligned}$$

Now, using (36)-(39) and (43), we can simplify the expression for  $D_{2,NL}$  as

$$\begin{aligned} D_{2,NL} = & \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left[ \frac{3}{2} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 (x_{\mathbf{q}}^T D_{1,NL})^2 x_{\mathbf{q}} - \frac{1}{2} \sum_{\mathbf{q} \in \mathcal{Q}} (v_{\mathbf{q}}^T \varepsilon_{\mathbf{q}})^2 x_{\mathbf{q}} \right. \\ & \left. + \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}} (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}) (x_{\mathbf{q}}^T D_{1,NL}) x_{\mathbf{q}} \right] \end{aligned}$$

which can be rearranged to get (44) by using the identity  $v_{\mathbf{q}} v_{\mathbf{q}}^T + u_{\mathbf{q}} u_{\mathbf{q}}^T = I_2$ .

### Bias in $\widehat{D}_{NL}$

The following shows that when the tensor  $D^0$  is isotropic, the second order bias in  $\widehat{D}_{NL}$  is negative definite and does not depend on  $D^0$ .

**Lemma 8.1.** *If  $D^0$  is isotropic, the gradient directions  $\{\mathbf{q} : \mathbf{q} \in \mathcal{Q}\}$  are uniformly distributed on  $\mathbb{S}^2$ , then under the framework of Proposition 2.2,  $\mathbb{E}(D_{2,NL}) = \sigma^2 \mathbf{N}$  where  $\mathbf{N}$  is a  $3 \times 3$  negative definite matrix and  $\tilde{D}_{2,NL}$  is the second order bias in  $\widehat{D}_{NL}$  written as a  $3 \times 3$  symmetric matrix.*

*Proof.* First note that  $D_{2,NL}$  is the vectorization of  $\tilde{D}_{2,NL}$  so that  $\tilde{D}_{2,NL}(i, i) = D_{2,NL}(i)$  for  $i = 1, \dots, 3$  and for  $\tilde{D}_{2,NL}(1, 2) = D_{2,NL}(4)$ ,  $\tilde{D}_{2,NL}(1, 3) = D_{2,NL}(5)$  and  $\tilde{D}_{2,NL}(2, 3) = D_{2,NL}(6)$ .

Suppose that  $D^0 = \lambda_1 I_3$  and the design is uniform. Without loss of generality, let  $S_0 = 1$ . Thus,  $S_{\mathbf{q}} = e^{-\lambda_1}$  for all  $\mathbf{q} \in \mathcal{Q}$ . Also, uniform design implies that the terms  $1 - x_{\mathbf{q}}^T (\sum_{\mathbf{q}' \in \mathcal{Q}} x_{\mathbf{q}'} x_{\mathbf{q}'}^T)^{-1} x_{\mathbf{q}}$  are the same for all  $\mathbf{q} \in \mathcal{Q}$ . In order to prove this, we use the following facts about the uniform design: (i)  $|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} q_j^4 = 1/5$  for  $j = 1, 2, 3$ ; (ii)  $|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} q_j q_k = 0$  for  $j \neq k$ ; (iii)  $|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} q_j^2 q_k^2 = 1/15$  for  $j \neq k$  (iv)  $|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} q_j^3 q_k = 0$  for  $j \neq k$ ; and (v)  $|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} q_j^2 q_k q_l = 0$  for  $j \neq k \neq l$ . (Note that this imposes a restriction on the minimum value of  $|\mathcal{Q}|$ .) Then it can be checked that  $x_{\mathbf{q}}^T (\sum_{\mathbf{q}' \in \mathcal{Q}} x_{\mathbf{q}'} x_{\mathbf{q}'}^T)^{-1} x_{\mathbf{q}} = 6/|\mathcal{Q}|$ . Thus, from Proposition 2.2, we obtain that  $\mathbb{E}(D_{2,NL})$  is a positive multiple of (with multiplicative factor  $(1 - 6/|\mathcal{Q}|)$ )

$$-\frac{1}{2} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} \right) = -\frac{1}{2} e^{2\lambda_1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} \right). \tag{45}$$

Let  $y_{\mathbf{q}} = (q_1^2, q_2^2, q_3^2)^T$  and  $z_{\mathbf{q}} = (2q_1 q_2, 2q_1 q_3, 2q_2 q_3)$ . Define,

$$L_{11} = |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} y_{\mathbf{q}} y_{\mathbf{q}}^T, \quad L_{12} = |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} y_{\mathbf{q}} z_{\mathbf{q}}^T, \quad \text{and} \quad L_{22} = |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} z_{\mathbf{q}} z_{\mathbf{q}}^T.$$

Then

$$\left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right) = |\mathcal{Q}| \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}$$

where  $L_{21} = L_{12}^T$ . If the design is uniform, then  $L_{12} = O$  (zero matrix) and

$$\sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} = \frac{|\mathcal{Q}|}{3} \begin{bmatrix} \mathbf{1}_3 \\ \mathbf{0}_3 \end{bmatrix}.$$

Using these, we have

$$\left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} x_{\mathbf{q}} \right) = \frac{1}{3} \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_3 \\ \mathbf{0}_3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} L_{11}^{-1} \mathbf{1}_3 \\ \mathbf{0}_3 \end{bmatrix}.$$

Note that, the diagonal entries of  $L_{11}$  are all equal and off-diagonal entries are also equal among themselves, due to the uniformity of the design. Thus,  $\mathbf{1}_3$  is an eigenvector of  $L_{11}$  and hence also of  $L_{11}^{-1}$ . Consequently,  $L_{11}^{-1} \mathbf{1}_3$  is a positive multiple of  $\mathbf{1}_3$  since  $L_{11}^{-1}$  is positive definite. Now we conclude the proof of Lemma 8.1 by invoking (45).  $\square$

**8.2. Proofs of Theorem 4.1, Theorem 4.2 and Theorem 4.3**

In the case that  $D^0$  is anisotropic, define  $H_j$  to be the matrix

$$H_j := \sum_{k \neq j} \frac{1}{\lambda_k - \lambda_j} P_k. \tag{46}$$

Note that  $H_j P_j = P_j H_j = \mathbf{0}$  for all  $j$ .

**Proposition 8.1.** *Suppose that the tensors  $\{D(\omega) : \omega \in \Omega\}$  and the probability distribution  $\mathbb{P}_\Omega$  satisfy (18) and  $t \rightarrow 0$ . If the eigenvalues of  $D^0$  are all distinct, then*

$$\begin{aligned}
 & \log \bar{D}_E - \log D^0 \\
 = & \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E) P_j - \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E H_j \bar{\Delta}_E) P_j - \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j^2} [\text{tr}(P_j \bar{\Delta}_E)]^2 P_j \\
 & - \sum_{j=1}^N \log \lambda_j [P_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E P_j] \\
 & + \sum_{j=1}^N \log \lambda_j [P_j \bar{\Delta}_E H_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E P_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E H_j \bar{\Delta}_E P_j \\
 & \quad - P_j \bar{\Delta}_E P_j \bar{\Delta}_E H_j^2 - P_j \bar{\Delta}_E H_j^2 \bar{\Delta}_E P_j - H_j^2 \bar{\Delta}_E P_j \bar{\Delta}_E P_j] \\
 & - \sum_{j=1}^N \frac{1}{\lambda_j} [\text{tr}(P_j \bar{\Delta}_E)(P_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E P_j)] + O(t^3). \tag{47}
 \end{aligned}$$

Taking trace on both sides of (47), and using the fact that  $P_j H_j = \mathbf{0}$  and  $P_j^2 = P_j$ , we get (20) when the eigenvalues of  $D^0$  are all distinct.

**Proposition 8.2.** *Assume that the conditions of Theorem 4.1 hold. If the eigenvalues of  $D^0$  are all distinct, then*

$$\begin{aligned}
 & \log \bar{D}_{LE} - \log D^0 \\
 = & \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E) P_j - \sum_{j=1}^N \frac{1}{\lambda_j} \mathbb{E}_\Omega[\text{tr}(P_j \Delta H_j \Delta)] P_j - \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j^2} \mathbb{E}_\Omega[\text{tr}(P_j \Delta)]^2 P_j \\
 & - \sum_{j=1}^N \log \lambda_j [P_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E P_j] \\
 & + \sum_{j=1}^N \log \lambda_j \mathbb{E}_\Omega [P_j \Delta H_j \Delta H_j + H_j \Delta P_j \Delta H_j + H_j \Delta H_j \Delta P_j \\
 & \quad - P_j \Delta P_j \Delta H_j^2 - P_j \Delta H_j^2 \Delta P_j - H_j^2 \Delta P_j \Delta P_j] \\
 & - \sum_{j=1}^N \frac{1}{\lambda_j} \mathbb{E}_\Omega[\text{tr}(P_j \Delta)(P_j \Delta H_j + H_j \Delta P_j)] + O(t^3). \tag{48}
 \end{aligned}$$

Taking trace on both sides of (48), we get (22) when the eigenvalues of  $D^0$  are all distinct.

**Proposition 8.3.** *Assume that the conditions of Theorem 4.1 hold.*

(a)

$$\bar{D}_{Aff} - D^0 = \bar{\Delta}_E - \frac{1}{2} \mathbb{E}_\Omega [\Delta (D^0)^{-1} \Delta] + \frac{1}{2} \bar{\Delta}_E (D^0)^{-1} \bar{\Delta}_E + O(t^3). \tag{49}$$

(b) If the eigenvalues of  $D^0$  are all distinct, then

$$\begin{aligned}
 & \log \bar{D}_{Aff} - \log D^0 \\
 = & \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E) P_j - \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E H_j \bar{\Delta}_E) P_j - \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j^2} [\text{tr}(P_j \bar{\Delta}_E)]^2 P_j \\
 & - \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j} \mathbb{E}_\Omega [\text{tr}(P_j \Delta (D^0)^{-1} \Delta)] P_j + \frac{1}{2} \sum_{j=1}^N \frac{1}{\lambda_j} \text{tr}(P_j \bar{\Delta}_E (D^0)^{-1} \bar{\Delta}_E) P_j \\
 & - \sum_{j=1}^N \log \lambda_j [P_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E P_j] \\
 & + \frac{1}{2} \sum_{j=1}^N \log \lambda_j \mathbb{E}_\Omega [P_j \Delta (D^0)^{-1} \Delta H_j + H_j \Delta (D^0)^{-1} \Delta P_j] \\
 & - \frac{1}{2} \sum_{j=1}^N \log \lambda_j [P_j \bar{\Delta}_E (D^0)^{-1} \bar{\Delta}_E H_j + H_j \bar{\Delta}_E (D^0)^{-1} \bar{\Delta}_E P_j] \\
 & + \sum_{j=1}^N \log \lambda_j [P_j \bar{\Delta}_E H_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E P_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E H_j \bar{\Delta}_E P_j \\
 & \quad - P_j \bar{\Delta}_E P_j \bar{\Delta}_E H_j^2 - P_j \bar{\Delta}_E H_j^2 \bar{\Delta}_E P_j - H_j^2 \bar{\Delta}_E P_j \bar{\Delta}_E P_j] \\
 & - \sum_{j=1}^N \frac{1}{\lambda_j} [\text{tr}(P_j \bar{\Delta}_E) (P_j \bar{\Delta}_E H_j + H_j \bar{\Delta}_E P_j)] + O(t^3). \tag{50}
 \end{aligned}$$

Taking trace of both sides of (50), we get (24) when the eigenvalues of  $D^0$  are all distinct.

*Proof of Proposition 8.2 and Theorem 4.2*

First suppose that eigenvalues of  $D^0$  are all distinct. Let  $\mu_j(\omega)$  denote the  $j$ -th largest eigenvalue of  $D(\omega)$  and  $Q_j(\omega)$  denote the corresponding eigen-projection. Then under (18), for  $t$  sufficiently small, with probability 1,  $\mu_j(\omega)$  is of multiplicity 1, and so  $Q_j(\omega)$  is a rank 1 matrix. We can then use the following matrix perturbation analysis results (Kato, 1980)

$$\mu_j(\omega) = \lambda_j + \text{tr}(P_j \Delta(\omega)) - \text{tr}(P_j \Delta(\omega) H_j \Delta(\omega)) + O(\|\Delta(\omega)\|^3), \tag{51}$$

where

$$\begin{aligned}
 Q_j(\omega) = & P_j - (P_j \Delta(\omega) H_j + H_j \Delta(\omega) P_j) \\
 & + P_j \Delta(\omega) H_j \Delta(\omega) H_j + H_j \Delta(\omega) P_j \Delta(\omega) H_j + H_j \Delta(\omega) H_j \Delta(\omega) P_j \\
 & - P_j \Delta(\omega) P_j \Delta(\omega) H_j^2 - P_j \Delta(\omega) H_j^2 \Delta(\omega) P_j - H_j^2 \Delta(\omega) P_j \Delta(\omega) P_j \\
 & + O(\|\Delta(\omega)\|^3). \tag{52}
 \end{aligned}$$

We consider an asymptotic expansion of  $\log D(\omega)$  around  $\log D^0$  as  $t \rightarrow 0$ . By definition,

$$\begin{aligned} & \log D(\omega) - \log D^0 \\ &= \sum_{j=1}^N (\log \mu_j(\omega) Q_j(\omega) - \log \lambda_j P_j) \\ &= \sum_{j=1}^N (\log \mu_j(\omega) - \log \lambda_j) P_j + \sum_{j=1}^N \log \lambda_j (Q_j(\omega) - P_j) \\ & \quad + \sum_{j=1}^N (\log \mu_j(\omega) - \log \lambda_j) (Q_j(\omega) - P_j). \end{aligned} \tag{53}$$

Now, from (51) and (18), we have

$$\begin{aligned} & \log \mu_j(\omega) - \log \lambda_j \\ &= \log \left( 1 + \frac{\mu_j(\omega) - \lambda_j}{\lambda_j} \right) \\ &= \frac{1}{\lambda_j} [\text{tr}(P_j \Delta(\omega)) - \text{tr}(P_j \Delta(\omega) H_j \Delta(\omega))] - \frac{1}{2\lambda_j^2} [\text{tr}(P_j \Delta(\omega))]^2 + O(\|\Delta(\omega)\|^3), \end{aligned}$$

where we have used the series expansion  $\log(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} x^n/n$ . Substituting in (53), using (52), and finally taking expectation with respect to  $\mathbb{P}_\Omega$  and noticing that  $\log \bar{D}_{LE} = \mathbb{E}_\Omega(\log D)$ , we obtain (48). This concludes the proof of Proposition 8.2. Taking trace on both sides of (48) and using the fact that  $P_j H_j = H_j P_j = O$  (zero matrix), we have part (b) of Theorem 4.2.

Now, for part (a) of Theorem 4.2, we have  $D^0 = \lambda_1 I$ . Thus, from the expansion

$$\begin{aligned} \log D(\omega) &= (\log \lambda_1) I + \log \left( I + \frac{1}{\lambda_1} \Delta(\omega) \right) \\ &= \log D^0 + \frac{1}{\lambda_1} \Delta(\omega) - \frac{1}{2\lambda_1^2} (\Delta(\omega))^2 + O(\|\Delta(\omega)\|^3) \end{aligned}$$

we obtain (21) by taking expectation with respect to  $\mathbb{P}_\Omega$ .

*Proof of Proposition 8.1 and Theorem 4.1*

The proof is almost identical to the proof of Proposition 8.2 and Theorem 4.2, the only difference being that we replace  $\Delta(\omega)$  by  $\bar{\Delta}_E$  in every step.

*Proof of Proposition 8.3 and Theorem 4.3*

There is no closed form expression for  $\bar{D}_{Aff}$ . But it satisfies the barycentric equation (Arsigny et al., 2005)

$$\mathbb{E}_\Omega \left[ \log \left( \bar{D}_{Aff}^{-1/2} D \bar{D}_{Aff}^{-1/2} \right) \right] = O. \tag{54}$$

Define,

$$L(\omega) = \log(\bar{D}_{Aff}^{-1/2}D(\omega)\bar{D}_{Aff}^{-1/2}), \quad \omega \in \Omega.$$

Note that, (54) implies that  $\mathbb{E}_\Omega(L) = O$ . First, we show that

$$\|L\|_\infty := \sup_{\omega \in \Omega} \|L(\omega)\| = O(t). \tag{55}$$

This implies that

$$\begin{aligned} \bar{D}_E &= \mathbb{E}_\Omega(D) = \mathbb{E}_\Omega(\bar{D}_{Aff}^{1/2}e^L\bar{D}_{Aff}^{1/2}) \\ &= \bar{D}_{Aff} + \bar{D}_{Aff}^{1/2}\mathbb{E}_\Omega(L)\bar{D}_{Aff}^{1/2} + \frac{1}{2}\bar{D}_{Aff}^{1/2}\mathbb{E}_\Omega(L^2)\bar{D}_{Aff}^{1/2} + O(t^3), \end{aligned}$$

from which, after invoking the fact that  $\mathbb{E}_\Omega(L) = O$ , we get

$$\bar{D}_{Aff} = \bar{D}_E - \frac{1}{2}\bar{D}_E^{1/2}\mathbb{E}_\Omega(L^2)\bar{D}_E^{1/2} + O(t^3). \tag{56}$$

Now we express  $\mathbb{E}_\Omega(L^2)$  in terms of an expectation involving  $\tilde{\Delta}(\omega) = D(\omega) - \bar{D}_E$ . In order to do this, observe that

$$\begin{aligned} \tilde{\Delta}(\omega) &= D(\omega) - \bar{D}_E = \bar{D}_{Aff}^{1/2}e^{L(\omega)}\bar{D}_{Aff}^{1/2} - \bar{D}_{Aff} + \bar{D}_{Aff} - \bar{D}_E \\ &= \bar{D}_{Aff}^{1/2}L(\omega)\bar{D}_{Aff}^{1/2} + \frac{1}{2}\bar{D}_{Aff}^{1/2}((L(\omega))^2 - \mathbb{E}_\Omega(L^2))\bar{D}_{Aff}^{1/2} + O(t^3) \\ &= \bar{D}_E^{1/2}L(\omega)\bar{D}_E^{1/2} + \frac{1}{2}\bar{D}_E^{1/2}((L(\omega))^2 - \mathbb{E}_\Omega(L^2))\bar{D}_E^{1/2} + O(t^3), \end{aligned}$$

where, in the second and third steps we have used (55) and (56). Therefore, again using (55),

$$\mathbb{E}_\Omega(L^2) = \bar{D}_E^{-1/2}\mathbb{E}_\Omega(\tilde{\Delta}\bar{D}_E^{-1}\tilde{\Delta})\bar{D}_E^{-1/2} + O(t^3), \tag{57}$$

which, together with (56), leads to the representation

$$\begin{aligned} \bar{D}_{Aff} &= \bar{D}_E - \frac{1}{2}\mathbb{E}_\Omega(\tilde{\Delta}\bar{D}_E^{-1}\tilde{\Delta}) + O(t^3) \\ &= D^0 + \bar{\Delta}_E - \frac{1}{2}\mathbb{E}_\Omega(\Delta(D^0)^{-1}\Delta) + \frac{1}{2}\bar{\Delta}_E(D^0)^{-1}\bar{\Delta}_E + O(t^3), \end{aligned} \tag{58}$$

where, in the last step we use the fact that  $\tilde{\Delta}(\omega) = \Delta(\omega) - \bar{\Delta}_E$  and  $\mathbb{E}_\Omega(\Delta) = O$  and  $\|\bar{\Delta}_E\| = O(t)$ .

For part (a) of Theorem 4.3, since  $D^0 = \lambda_1 I$ , using the Taylor series expansion of  $\log(I + K)$ , where  $K$  is the sum of the terms after  $D^0$  on the RHS of (58) and is a symmetric matrix, (23) immediately follows.

For the proof of Proposition 8.3, using similar perturbation analysis as in the proof of Proposition 8.2, with  $\Delta(\omega)$  replaced by the expression for  $\bar{D}_{Aff} - D^0$  obtained from (58), we obtain (50). Part (b) of Theorem 4.3 now follows by taking trace.

*Proof of (55)*

From the fact that  $\|\tilde{\Delta}\|_\infty = O(t)$ , it easily follows that  $\mathbb{E}_\Omega(d_{Aff}^2(\bar{D}_E, D)) = O(t^2)$ . Then, by definition of  $\bar{D}_{Aff}$  it follows that  $\mathbb{E}_\Omega(d_{Aff}^2(\bar{D}_{Aff}, D)) = O(t^2)$  which implies  $\mathbb{E}_\Omega(d_{Aff}(\bar{D}_{Aff}, D)) = O(t)$ . Now, writing

$$\log(\bar{D}_{Aff}^{-1/2} D(\omega) \bar{D}_{Aff}^{-1/2}) = \log(\bar{D}_{Aff}^{-1/2} \bar{D}_E \bar{D}_{Aff}^{-1/2} + \bar{D}_{Aff}^{-1/2} \tilde{\Delta}(\omega) \bar{D}_{Aff}^{-1/2})$$

and using the Baker-Campbell-Hausdorff formula (Varadarajan, 1984) (which gives an expansion of  $\log A - \log B$  where  $A$  and  $B$  are positive definite matrices), together with the fact that  $\sup_{\omega \in \Omega} \|\tilde{\Delta}(\omega)\| = O(t)$ , we conclude that

$$d_{Aff}(\bar{D}_{Aff}, \bar{D}_E) = O(t) \quad \text{so that} \quad \|\bar{D}_{Aff} - \bar{D}_E\| = O(t).$$

From this it is easy to deduce (55).

### 8.3. Proofs of Corollary 4.2 and Corollary 4.3

*Proof of Corollary 4.2*

It can be seen from Propositions 8.1, 8.2 and 8.3 and equations (26) and (27) that

$$\begin{aligned} \text{ABias}(\log \bar{D}_E; \log D^0) &= T_1 - r_\Omega T_2 + r_\Omega T_3 + O(\sigma^4), \\ \text{ABias}(\log \bar{D}_{LE}; \log D^0) &= T_1 - T_2 + T_3, \\ \text{ABias}(\log \bar{D}_{Aff}; \log D^0) &= T_1 - r_\Omega T_2 + r_\Omega T_3 + (1 - r_\Omega) T_4 + O(\sigma^4), \end{aligned}$$

where

$$\begin{aligned} \tilde{T}_1 &= \sum_{j=1}^3 \frac{1}{\lambda_j} \text{tr}(P_j \mathbb{E}(\bar{\Delta}_E^*)) P_j - \sum_{j=1}^3 \sum_{k \neq j}^3 \frac{\log \lambda_j}{\lambda_k - \lambda_j} [P_j \mathbb{E}(\bar{\Delta}_E^*) P_k + P_k \mathbb{E}(\bar{\Delta}_E^*) P_j] \\ \tilde{T}_2 &= \sum_{j=1}^3 \sum_{k \neq j}^3 \frac{1}{\lambda_j (\lambda_k - \lambda_j)} \mathbb{E}[\text{tr}(P_j \Delta^* P_k \Delta^*)] P_j + \frac{1}{2} \sum_{j=1}^3 \frac{1}{\lambda_j^2} \mathbb{E}[\text{tr}(P_j \Delta^*)]^2 P_j \\ \tilde{T}_3 &= \sum_{j=1}^3 \sum_{k: k \neq j}^3 \sum_{l: l \neq j}^3 \frac{\log \lambda_j}{(\lambda_k - \lambda_j)(\lambda_l - \lambda_j)} \\ &\quad \times \mathbb{E}[P_j \Delta^* P_k \Delta^* P_l + P_k \Delta^* P_j \Delta^* P_l + P_k \Delta^* P_l \Delta^* P_j] \\ &\quad - \sum_{j=1}^3 \sum_{k: k \neq j}^3 \frac{\log \lambda_j}{(\lambda_k - \lambda_j)^2} \mathbb{E}[P_j \Delta^* P_j \Delta^* P_k + P_j \Delta^* P_k \Delta^* P_j + P_k \Delta^* P_j \Delta^* P_j] \\ &\quad - \sum_{j=1}^3 \sum_{k: k \neq j}^3 \frac{1}{\lambda_j (\lambda_k - \lambda_j)} \mathbb{E}[\text{tr}(P_j \Delta^*) (P_j \Delta^* P_k + P_k \Delta^* P_j)] \end{aligned}$$

$$\begin{aligned} \tilde{T}_4 &= -\frac{1}{2} \sum_{j=1}^3 \sum_{k=1}^3 \frac{1}{\lambda_j \lambda_k} \mathbb{E}[\text{tr}(P_j \Delta^* P_k \Delta^*)] P_j \\ &\quad + \frac{1}{2} \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l:l \neq j}^3 \frac{\log \lambda_j}{\lambda_k (\lambda_l - \lambda_j)} \mathbb{E}[P_j \Delta^* P_k \Delta^* P_l + P_l \Delta^* P_k \Delta^* P_j]. \end{aligned} \tag{59}$$

Now, the result is obtained by applying (28), (29) and (30).

*Proof of Corollary 4.3*

We use the following analog of (26):

$$\begin{aligned} \Delta(\omega) &= \sigma D_1(\omega) + \sigma \tau D_Z(\omega) + \sigma^2 D_2(\omega) \\ &\quad + \sum_{j=1}^3 \lambda_j \left( \tau Z_j(\omega) + \frac{\tau^2}{2} (Z_j(\omega))^2 \right) P_j + \tilde{R}_\Delta(\omega) \end{aligned} \tag{60}$$

where  $D_1(\omega)$  and  $D_2(\omega)$  are exactly as in (26), and

$$\begin{aligned} &\text{vec}(D_Z(\omega)) \\ &= A^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}} \left( \sum_{j=1}^3 \lambda_j Z_j(\omega) \mathbf{q}^T P_j \mathbf{q} \right) (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}(\omega)) x_{\mathbf{q}} \right) \\ &\quad - A^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 \left( \sum_{j=1}^N \lambda_j Z_j(\omega) \mathbf{q}^T P_j \mathbf{q} \right) x_{\mathbf{q}} x_{\mathbf{q}}^T \right) A^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}} u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}(\omega) \right), \end{aligned} \tag{61}$$

with  $A = \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T$ , and  $\tilde{R}_\Delta(\omega) = O(\tau^3 (\log(1/\sigma))^{3/2})$  with high probability.

We only prove (60), from which Corollary 4.3 follows by calculations similar to those used in proving Corollary 4.2.

To see why (60) holds, first define  $\Delta_0(\omega) = D^0(\omega) - D^0$ . Then we can write

$$\Delta_0(\omega) = \sum_{j=1}^3 \left( \tau Z_j(\omega) + \frac{\tau^2}{2} Z_j^2(\omega) \right) P_j + O(\tau^3).$$

Let us denote the first order term in the expansion of the nonlinear regression estimator, when  $D^0(\omega)$  is the true tensor, by  $D_1(D^0(\omega); \omega)$ , and the corresponding term, when  $D^0$  is the true tensor, by  $D_1(D^0; \omega)$  ( $= D_1(\omega)$ ). Similarly, we define  $D_2(D^0(\omega); \omega)$  and  $D_2(D^0; \omega)$  ( $= D_2(\omega)$ ). Our strategy is to first find the expansion of the terms  $D_j(D^0(\omega); \omega)$  around  $D_j(D^0; \omega)$  and then use Proposition 2.2 to deal with the remainder terms.

Define  $S_{\mathbf{q}}(\omega) = \exp(-x_{\mathbf{q}}^T D^0(\omega))$  (where, by convention,  $D^0(\omega)$  means  $\text{vec}(D^0(\omega))$ ). Then, from (43) we have

$$\text{vec}(D_1(D^0(\omega); \omega)) = - \left( \sum_{\mathbf{q} \in \mathcal{Q}} (S_{\mathbf{q}}(\omega))^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \right)^{-1} \left( \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}(\omega) (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}(\omega)) x_{\mathbf{q}} \right). \quad (62)$$

We observe that, for each  $\omega \in \Omega$ ,

$$\begin{aligned} S_{\mathbf{q}}(\omega) &= S_{\mathbf{q}} - S_{\mathbf{q}} \left[ \mathbf{q}^T \Delta_0(\omega) \mathbf{q} - \frac{1}{2} (\mathbf{q}^T \Delta_0(\omega))^2 + O(\|\Delta_0(\omega)\|^3) \right] \\ (S_{\mathbf{q}}(\omega))^2 &= S_{\mathbf{q}}^2 - 2S_{\mathbf{q}}^2 \left[ \mathbf{q}^T \Delta_0(\omega) \mathbf{q} - (\mathbf{q}^T \Delta_0(\omega))^2 + O(\|\Delta_0(\omega)\|^3) \right] \end{aligned}$$

so that,

$$\frac{1}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} (S_{\mathbf{q}}(\omega))^2 x_{\mathbf{q}} x_{\mathbf{q}}^T = A - B(\omega),$$

where  $A = |\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 x_{\mathbf{q}} x_{\mathbf{q}}^T$  and

$$\begin{aligned} B(\omega) &= \frac{2}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 (\mathbf{q}^T \Delta_0(\omega) \mathbf{q}) x_{\mathbf{q}} x_{\mathbf{q}}^T - \frac{2}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}^2 (\mathbf{q}^T \Delta_0(\omega) \mathbf{q})^2 x_{\mathbf{q}} x_{\mathbf{q}}^T \\ &\quad + O(\|\Delta_0(\omega)\|^3); \end{aligned}$$

and

$$\frac{1}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}}(\omega) (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}(\omega)) x_{\mathbf{q}} = W(\omega) - C(\omega),$$

where

$$W(\omega) = \frac{1}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}} (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}(\omega)) x_{\mathbf{q}}$$

and

$$\begin{aligned} C(\omega) &= \frac{1}{|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}} (\mathbf{q}^T \Delta_0(\omega) \mathbf{q}) (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}(\omega)) x_{\mathbf{q}} \\ &\quad - \frac{1}{2|\mathcal{Q}|} \sum_{\mathbf{q} \in \mathcal{Q}} S_{\mathbf{q}} (\mathbf{q}^T \Delta_0(\omega) \mathbf{q})^2 (u_{\mathbf{q}}^T \varepsilon_{\mathbf{q}}(\omega)) x_{\mathbf{q}} + O(\|\Delta_0(\omega)\|^3). \end{aligned}$$

Substituting these in (62) and simplifying,

$$\begin{aligned} \text{vec}(D_1(D^0(\omega); \omega)) &= -(A - B(\omega))^{-1} (W(\omega) - C(\omega)) \\ &= -A^{-1} W(\omega) + A^{-1} C(\omega) - A^{-1} B(\omega) A^{-1} W(\omega) \\ &\quad + O(\|A^{-1}\|^2 \|B(\omega)\| \|C(\omega)\|) \\ &\quad + O(\|A^{-1}\|^3 \|B(\omega)\|^2 (\|W(\omega)\| + \|C(\omega)\|)). \end{aligned}$$

Now observe that  $A^{-1}W(\omega) = \text{vec}(D_1(D^0; \omega))$  and

$$A^{-1}C(\omega) - A^{-1}B(\omega)A^{-1}W(\omega) = \text{vec}(D_Z(\omega)) + O(\tau^2(|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}(\omega)\|^2)^{1/2})$$

where  $\text{vec}(D_Z(\omega))$  is defined in (61). Moreover,

$$\begin{aligned} \|B(\omega)\| &= O(\|\Delta_0(\omega)\| \|A\|) \\ \|C(\omega)\| &= O(\|\Delta_0(\omega)\| \|A\|^{1/2} (|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}(\omega)\|^2)^{1/2}) \\ \|W(\omega)\| &= O(\|A\|^{1/2} (|\mathcal{Q}|^{-1} \sum_{\mathbf{q} \in \mathcal{Q}} \|\varepsilon_{\mathbf{q}}(\omega)\|^2)^{1/2}). \end{aligned}$$

Combining these, together with a first order expansion of  $D_2(D^0(\omega); \omega)$  around  $D_2(D^0; \omega)$  and recalling the definition of  $\Delta(\omega)$ , we have

$$\begin{aligned} \Delta(\omega) &= \sigma D_1(D^0(\omega); \omega) + \sigma^2 D_2(D^0(\omega); \omega) + \Delta_0(\omega) + R_1(\omega) \\ &= (\sigma(D_1(D^0; \omega) + \tau D_Z(\omega) + \sigma^2 D_2(D^0; \omega) + R_2(\omega)) \\ &\quad + \left( \sum_{j=1}^3 (\tau Z_j(\omega) + \frac{\tau^2}{2} Z_j^2(\omega)) P_j + R_3(\omega) \right) + R_1(\omega), \end{aligned}$$

where  $R_1(\omega) = O(\sigma^3(\log(1/\sigma))^{3/2})$  and  $R_2(\omega) = O(\sigma\tau^2(\log(1/\sigma))^{3/2})$  with high probability, and  $R_3(\omega) = O(\tau^3)$ . Collecting terms, we obtain (60).

## Supplementary Material

### Supplement to “Diffusion tensor smoothing through weighted Karcher means”

(doi: [10.1214/00-EJS825SUPP](https://doi.org/10.1214/00-EJS825SUPP); .pdf).

## References

- ABSIL, P.-A., MAHONY, R., AND SEPULCHRE, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. [MR2364186](#)
- ARSIGNY, V., FILLARD, P., PENNEC, X., AND AYACHE, N. (2005). Fast and simple computations on tensors with log-euclidean metrics. Technical report, Institut National de Recherche en Informatique et en Automatique.
- ARSIGNY, V., FILLARD, P., PENNEC, X., AND AYACHE, N. (2006). Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56:411–421.
- BAMMER, R., HOLDSWORTH, S. J., VELDHUIS, W. B., AND SKARE, S. (2009). New methods in diffusion-weighted and diffusion tensor imaging. *Magnetic Resonance Imaging Clinics of North America*, 17:175–204.

- BASSER, P. AND PAJEVIC, S. (2000). Statistical artifacts in diffusion tensor mri (dt-mri) caused by background noise. *Magnetic Resonance in Medicine*, 44:41–50.
- BASSER, P. J. AND PIERPAOLI, C. (2005). A simplified method to measure the diffusion tensor from seven mr images. *Magnetic Resonance in Medicine*, 39(6):928–934.
- BEAULIEU, C. (2002). The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in Biomedicine*, 15:435–455.
- CARMICHAEL, O., CHEN, J., PAUL, D., AND PENG, J. Supplement to “Diffusion tensor smoothing through weighted Karcher means”. DOI: [10.1214/00-EJS825SUPP](https://doi.org/10.1214/00-EJS825SUPP).
- CASTAÑO MORAGA, C. A., LENGLET, C., DERICHE, R., AND RUIZ-ALZOLA, J. (2007). A Riemannian approach to anisotropic filtering of tensor fields. *Signal Processing*, 87:263–276.
- CHANRAUD, S., ZAHR, N., SULLIVAN, E. V., AND PFEFFERBAUM, A. (2010). MR diffusion tensor imaging: a window into white matter integrity of the working brain. *Neuropsychology Review*, 20:209–225.
- CHUNG, M. K., LAZAR, M., ALEXENDER, A. L., LU, Y., AND DAVIDSON, R. (2003). Probabilistic connectivity measure in diffusion tensor imaging via anisotropic kernel smoothing. Technical report, University of Wisconsin, Madison. Technical Report.
- CHUNG, M. K., LEE, J. E., AND ALEXANDER, A. L. (2005). Anisotropic kernel smoothing in diffusion tensor imaging: theoretical framework. Technical report, University of Wisconsin, Madison. Technical Report.
- DE CRESPIGNY, A. AND MOSELEY, M. (1998). Eddy current induced image warping in diffusion weighted epi. In *Proceedings of the 6th Meeting of the International Society for Magnetic Resonance in Medicine. Berkeley, Calif, ISMRM*.
- FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC. [MR1383587](https://doi.org/10.1002/9781118133587)
- FARRELL, J. A. D., LANDMAN, B. A., JONES, C. K., SMITH, S. A., PRINCE, J. L., VAN ZIJL, P. C. M., AND MORI, S. (2007). Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *Journal of Magnetic Resonance Imaging*, 256:756–767.
- FERREIRA, R., XAVIER, J., COSTEIRA, J. P., AND BARROSO, V. (2006). Newton method for riemannian centroid computation in naturally reductive homogeneous spaces. In *Proceedings of ICASSP 2006 - IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 77–85.
- FLETCHER, P. T. AND JOSHI, S. (2004). Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, pages 87–98. Springer.
- FLETCHER, P. T. AND JOSHI, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87:250–262.

- FÖRSTNER, W. AND MOONEN, B. (1999). Qua vadis geodesia...? In Krumm, F. and Schwarze, V. S., editors, *Festschrift for Erik W. Grafarend*, pages 113–128.
- GUDBJARTSSON, H. AND PATZ, S. (1995). The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34:910–914.
- HAHN, K. R., PRIGARIN, S., HEIM, S., AND HASAN, K. (2006). Random noise in diffusion tensor imaging, its destructive impact and some corrections. In Weickert, J. and Hagen, H., editors, *Visualization and Processing of Tensor Fields*, pages 107–117. Springer. [MR2210513](#)
- HAHN, K. R., PRIGARIN, S., RODENACKER, K., AND HASAN, K. (2009). Denoising for diffusion tensor imaging with low signal to noise ratios: method and monte carlo validation. *International Journal for Biomathematics and Biostatistics*, 1:83–81.
- KARCHER, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications in Pure and Applied Mathematics*, 30:509–541. [MR0442975](#)
- KATO, T. (1980). *Perturbation Theory of Linear Operators*. Springer-Verlag.
- KOCHUNOV, P., LANCASTER, J. L., THOMPSON, P., WOODS, R., MAZZIOTTA, J., HARDIES, J., AND FOX, P. (2001). Regional spatial normalization: toward an optimal target. *Journal of Computer Assisted Tomography*, 25:805–816.
- MORI, S. (2007). *Introduction to Diffusion Tensor Imaging*. Elsevier.
- MUKHERJEE, P., BERMAN, J. I., CHUNG, S. W., HESS, C., AND HENRY, R. (2008a). Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. *American Journal of Neuroradiology*, 29:632–641.
- MUKHERJEE, P., CHUNG, S. W., BERMAN, J. I., HESS, C. P., AND HENRY, R. G. (2008b). Diffusion tensor MR imaging and fiber tractography: technical considerations. *American Journal of Neuroradiology*, 29:843–852.
- PARKER, G. J. M., SCHNABEL, J. A., SYMMS, M. R., WERRING, D. J., AND BARKER, G. J. (2000). Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 11:702–710.
- PENNEC, X., FILLARD, P., AND AYACHE, N. (2006). A riemannian framework for tensor computing. *Journal of Computer Vision*, 66:41–66.
- POLZEHL, J. AND TABELOW, K. (2008). Structural adaptive smoothing in diffusion tensor imaging: the R package *dti*. Technical report, Weierstrass Institute, Berlin. Technical Report.
- RAJAPAKSE, J. C., GIEDD, J. N., DECARLI, C., SNELL, J. W., MCLAUGHLINAND, A., VAUSS, Y. C., KRAIN, A. L., HAMBURGER, S., AND RAPOPORT, J. L. (1996). A technique for single-channel MR brain tissue segmentation: application to a pediatric sample. *Magnetic Resonance in Medicine*, 14:1053–1065.
- RUECKERT, D., SONODA, L. I., HAYES, C., HILL, D. L., LEACH, M. O., AND HAWKES, D. J. (1999). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18:712–721.

- SCHWARTZMAN, A. (2006). *Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data*. PhD thesis, Stanford University. [MR2708811](#)
- TABELOW, K., POLZEHL, J., SPOKOINY, V., AND VOSS, H. U. (2008). Diffusion tensor imaging: structural adaptive smoothing. *Neuroimage*, 39:1763–1773.
- TOUSSAINT, N., SOUPLLET, J. C., AND FILLARD, P. (2007). MedINRIA: medical image navigation and research tool by INRIA. In *Proceedings of the MIC-CAI'07 Workshop on Interaction in medical image analysis and visualization, Brisbane, Australia*.
- TUCH, D. S., REESE, T. G., WIEGELL, M. R., MAKRIS, N., BELLIVEAU, J. W., AND WEDEEN, V. J. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48:577–582.
- VARADARAJAN, V. S. (1984). *Lie groups, Lie algebras, and their representations*. Springer. [MR0746308](#)
- VISWANATH, V., FLETCHER, E., SINGH, B., SMITH, N., PAUL, D., PENG, J., CHEN, J., AND CARMICHAEL, O. T. (2012). Impact of DTI smoothing on the study of brain aging. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2012)*.
- WEINSTEIN, D. M., KINDLMANN, G. L., AND LUNDBERG, E. C. (1999). Tensorlines: advection-diffusion based propagation through diffusion tensor fields. In *Proceedings of the IEEE Visualization*, pages 249–253.
- YUAN, Y., ZHU, H., LIN, W., AND MARRON, J. (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. [MR2965956](#)
- ZHU, H. T., LI, Y., IBRAHIM, I. G., SHI, X., AN, H., CHEN, Y., GAO, W., LIN, W., ROWE, D. B., AND PETERSON, B. S. (2009). Regression models for identifying noise sources in magnetic resonance images. *Journal of the American Statistical Association*, 104:623–637. [MR2751443](#)
- ZHU, H. T., ZHANG, H. P., IBRAHIM, J. G., AND PETERSON, B. (2007). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data. *Journal of the American Statistical Association*, 102:1081–1110. [MR2412533](#)