

The Version of Record of this manuscript has been published and is available in *Journal of Applied Statistics* 23 September 2015 <http://www.tandfonline.com/doi/10.1080/02664763.2015.1080671>

Regularisation, interpolation and visualisation of diffusion tensor images using non-Euclidean statistics

Diwei Zhou^a, Ian L. Dryden^b, Alexey A. Koloydenko^{c*}, Koenraad M.R. Audenaert^c, and Li Bai^d

^a *Department of Mathematical Sciences, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK;* ^b *School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK;* ^c *Mathematics Department, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK;* ^d *School of Computer Science, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB, UK*

(Received 7 February 2013; revised 16 December 2013, 31 December 2014, 11 April 2015; accepted 4 August 2015)

Practical statistical analysis of diffusion tensor images is considered, and we focus primarily on methods that use metrics based on Euclidean distances between powers of diffusion tensors. First we describe a family of anisotropy measures based on a scale invariant power-Euclidean metric, which are useful for visualisation. Some properties of the measures are derived and practical considerations are discussed, with some examples. Second we discuss weighted Procrustes methods for diffusion tensor interpolation and smoothing, and we compare methods based on different metrics on a set of examples as well as analytically. We establish a key relationship between the principal-square-root-Euclidean metric and the size-and-shape Procrustes metric on the space of symmetric positive semi-definite tensors. We explain, both analytically and by experiments, why the size-and-shape Procrustes metric may be preferred in practical tasks of interpolation, extrapolation, and smoothing, especially when observed tensors are degenerate or when a moderate degree of tensor swelling is desirable. Third we introduce regularisation methodology, which is demonstrated to be useful for highlighting features of prior interest and potentially for segmentation. Finally, we compare several metrics in a dataset of human brain diffusion-weighted MRI, and point out similarities between several of the non-Euclidean metrics but important differences with the commonly used Euclidean metric.

Keywords: anisotropy; metric; positive definite; power; Procrustes; Riemannian; smoothing; weighted Fréchet mean;

Classification codes: 62G05; 62H11; 62H35; 62P10

1. Introduction

Diffusion tensor imaging (DTI) is an advanced magnetic resonance imaging (MRI) modality which provides a unique insight into tissue structure and organisation *in vivo*. DTI has been applied to the study of brain diseases such as multiple sclerosis, schizophrenia, and stroke [25], and white matter tractography [4] is a useful application of DTI for investigating brain connectivity.

There has been substantial interest in the development of approaches for diffusion tensor processing. For example, a regularisation scheme was proposed to process a ten-

*Corresponding author. Email: alexey.koloydenko@rhul.ac.uk

tor field using diffusion direction maps and diffusion anisotropy maps [10]. A k-means algorithm with the Mahalanobis distance has been proposed for clustering tensors in the thalamus [40]. The Euclidean metric was used in level set segmentation methods [38, 45] for grouping tensor data of particular interest. However, the usual Euclidean method may be unsatisfactory for diffusion tensors due to the non-Euclidean nature of the diffusion tensor space. A potential drawback of using the Euclidean distance may be a violation of the positive semi-definiteness, e.g. in the course of extrapolation [2]. Also, Euclidean averaging is prone to swelling, i.e. inflation of the tensor determinant. To overcome this problem, several non-Euclidean approaches were developed. The affine-invariant Riemannian [5, 14, 30] and log-Euclidean Riemannian [2] metrics have been proposed for diffusion tensor smoothing and interpolation.

Procrustes analysis is another promising non-Euclidean approach to diffusion tensor processing [11, 43]. The full Procrustes shape metric is also invariant under scaling of the individual tensors, and the Procrustes metrics can deal with rank-deficient tensors unlike the affine-invariant and log-Euclidean Riemannian metrics [11]. This is further elaborated in the paper.

Thus, this paper focuses on weighted mean tensor processing using power-Euclidean and Procrustes based distances. DTI and weighted tensor averaging are briefly reviewed, and we also develop a family of anisotropy measures, which is useful for visualising different aspects of structure in DT images. As already mentioned, it is of both theoretical and practical interest to compare tensor processing methods based on different metrics. For example, it is valuable to know that the Euclidean average preserves the trace and hence the mean diffusivity (MD) [29], whereas the affine-invariant and log-Euclidean Riemannian metrics instead preserve the determinant, and hence the geometric mean diffusivity (GMD) [2]. We also compare in this paper various properties of our methods, focusing on the (principal-)square-root-Euclidean member of the power-Euclidean family of metrics and the Procrustes size-and-shape metric. While tensor processing methods based on these metrics can sometimes give very similar or identical results, we attempt to explain how the two metrics differ and what consequences their differences have for tensor processing. In particular, we both prove analytically and show by simulation that the two approaches can be radically different when used for processing of degenerate tensors. Thereby, it is established that unlike its Euclidean sibling, the Procrustes averaging preserves matrix ranks, i.e. dimension of diffusion.

More recently, it has been argued that a certain degree of swelling may be acceptable to compensate for Johnson (Rician) noise, and that the above Riemannian methods may therefore be biased in such scenarios [29]. We show here that the Procrustes metric and the square-root-Euclidean metric reduce the Euclidean swelling effect less aggressively than do the aforementioned Riemannian methods. This makes the Procrustes and square-root-Euclidean methods interesting alternatives when some swelling may indeed need to be allowed. A full comparison of the interpolated mean diffusivity under the two Riemannian, Procrustes, and all power-Euclidean metrics is also presented analytically, and illustrated empirically.

We also establish a key relationship between the Procrustes and square-root-Euclidean metrics in the form of an analytic bound, which is also illustrated on synthetic examples.

We then also discuss a weighted regularisation model which incorporates smoothing and a generalised regularisation that forces the tensor field to favour a prescribed diffusion profile. The power-Euclidean and Procrustes metrics are again our main interest, and we discuss properties of the different methods using synthetic examples and human brain DT images.

We conclude by discussing several possible extensions and further applications of this work, such as tensor valued kernel smoothing and non-parametric regression.

Although the paper focuses on 3×3 matrices, all our theoretical results also hold true for the general $n \times n$ case, even when this is not explicitly stated below.

2. DTI and anisotropy

2.1 Diffusion tensor imaging

Diffusion MRI yields quantitative measures reflecting the directions of water diffusion in white matter fibre tracts [35]. DTI assumes that a water molecule displacement $\mathbf{x} \in \mathbb{R}^3$ over a fixed time t in a voxel follows a zero-mean multivariate Gaussian distribution [1] with covariance matrix $2\mathbf{D}t$, where \mathbf{D} is the diffusion tensor, which is a symmetric positive semi-definite matrix. We will write $\Omega_{\geq 0}(n)$ for the set of $n \times n$ such matrices. Some of our discussion will be restricted to symmetric positive definite matrices, and we will write $\Omega_{> 0}(n)$ for the set of $n \times n$ such matrices; $n = 3$ is our main case here. The diffusion tensor at each voxel can be estimated from diffusion MR images [6, 23, 24, 42]. The eigensystem of the diffusion tensor plays an important role in DTI studies. Thus, the MD of the voxel is defined as

$$\text{MD}(\mathbf{D}) = \frac{1}{3} \text{tr } \mathbf{D} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3},$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ are the eigenvalues of \mathbf{D} . Whenever $\lambda_1 > \lambda_2$, the principal eigenvector is defined as the one corresponding to λ_1 and is thought to be aligned with the dominant fibre orientation at the voxel. In biological tissues there are barriers such as cell walls and nerve fibres, and so it is easier for water molecules to diffuse along certain preferred directions. The dominance of the preferred direction of water molecule diffusion can be captured quantitatively using diffusion anisotropy measures [25, 39]. Fractional anisotropy (FA) is the most popular such measure [22], defined below as the proportion of the ‘magnitude’ of \mathbf{D} that can be ascribed to anisotropic diffusion:

$$\text{FA}(\mathbf{D}) = \sqrt{\frac{3 \left[(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2 \right]}{2 (\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} = \frac{s_\lambda}{\sqrt{\bar{\lambda}^2}},$$

where $\bar{\lambda} = \text{MD}(\mathbf{D})$ and $\bar{\lambda}^2 = \sum_{i=1}^3 \lambda_i^2/3$, and s_λ is the sample standard deviation of the spectrum of $\mathbf{D} \neq \mathbf{0}$, the zero tensor. FA ranges from 0 for complete isotropy to 1 for linear anisotropy, and planar diffusion ($\lambda_1 = \lambda_2 > \lambda_3 = 0$, i.e. oblate diffusion ellipsoid) has $\text{FA}(\mathbf{D}) = 1/\sqrt{2}$.

Procrustes anisotropy (PA) is an alternative anisotropy measure that has been proposed based on the full Procrustes metric [11], and is defined as follows:

$$\text{PA}(\mathbf{D}) = \sqrt{\frac{3 \left[(\sqrt{\lambda_1} - \sqrt{\bar{\lambda}})^2 + (\sqrt{\lambda_2} - \sqrt{\bar{\lambda}})^2 + (\sqrt{\lambda_3} - \sqrt{\bar{\lambda}})^2 \right]}{2 (\lambda_1 + \lambda_2 + \lambda_3)}} = \frac{s_{\sqrt{\lambda}}}{\sqrt{\bar{\lambda}}}, \quad (1)$$

where $\sqrt{\bar{\lambda}} = \sum_{i=1}^3 \sqrt{\lambda_i}/3$ and $s_{\sqrt{\lambda}}$ is the sample standard deviation of square roots of the eigen values. Thus, $\text{PA}(\mathbf{D}) = \text{FA}(\sqrt{\mathbf{D}})$, where $\sqrt{\mathbf{D}}$ is the principal square root of \mathbf{D} (see more below). As in the case of FA, the range of PA is $[0, 1]$ with 0, $1/\sqrt{2}$, and 1 corresponding to extreme isotropy, planar diffusion, and full anisotropy, respectively.

2.2 A family of anisotropy indices

The power-Euclidean metric was briefly introduced in [11], and for $a \neq 0$ ($a < 0$ is meaningful for full rank tensors only) is given by

$$d_A(\mathbf{D}_1, \mathbf{D}_2|a) = \frac{1}{|a|} \|\mathbf{D}_1^a - \mathbf{D}_2^a\|, \quad (2)$$

where $\mathbf{D}^a = \mathbf{E}\mathbf{\Lambda}^a\mathbf{E}^T$ and $\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$ is the spectral decomposition of \mathbf{D} , and $\|\mathbf{A}\|^2 = \text{tr}(\mathbf{A}^T\mathbf{A})$. As $a \rightarrow 0$ the metric becomes the log-Euclidean metric $d_L(\mathbf{D}_1, \mathbf{D}_2) = \|\log(\mathbf{D}_1) - \log(\mathbf{D}_2)\|$, which requires \mathbf{D}_1 and \mathbf{D}_2 to be positive definite, i.e. $\in \Omega_{>0}(3)$.

We introduce another distance called the scale invariant power metric for $a \neq 0$ given by

$$d_{AS}(\mathbf{D}_1, \mathbf{D}_2|a) = \frac{1}{|a|} \begin{cases} 0 & \text{if } \mathbf{D}_1 = \mathbf{D}_2 = \mathbf{0} \\ 1 & \text{if either } \mathbf{D}_1 = \mathbf{0} \text{ or } \mathbf{D}_2 = \mathbf{0} \\ \inf_{\beta \in \mathbb{R}} \|\beta \mathbf{D}_1^a - \frac{\mathbf{D}_2^a}{\|\mathbf{D}_2^a\|}\| & \text{otherwise,} \end{cases} \quad (3)$$

which involves scaling of a powered tensor to best match a unit size powered tensor. Writing $\langle \mathbf{D}_1, \mathbf{D}_2 \rangle$ for the inner product $\text{tr}(\mathbf{D}_1^T \mathbf{D}_2)$ and assuming neither \mathbf{D}_i equals $\mathbf{0}$, $d_{AS}(\mathbf{D}_1, \mathbf{D}_2|a) = \frac{1}{|a|} \sqrt{|\sin(\angle(\mathbf{D}_1^a, \mathbf{D}_2^a))|}$. The angle $\angle(\mathbf{D}_1^a, \mathbf{D}_2^a)$ is defined via its cosine $\cos(\angle(\mathbf{D}_1^a, \mathbf{D}_2^a)) = \langle \mathbf{D}_1^a, \mathbf{D}_2^a \rangle / (\|\mathbf{D}_1^a\| \|\mathbf{D}_2^a\|)$, which in the context of kernels [32] is known as the *alignment* between \mathbf{D}_1^a and \mathbf{D}_2^a .

The anisotropy measure based on the scale invariant power-Euclidean metric with power a is a generalisation of FA given by

$$\text{FA}(\mathbf{D}^a) = \sqrt{\frac{3 \left[(\lambda_1^a - \bar{\lambda}^a)^2 + (\lambda_2^a - \bar{\lambda}^a)^2 + (\lambda_3^a - \bar{\lambda}^a)^2 \right]}{2 (\lambda_1^{2a} + \lambda_2^{2a} + \lambda_3^{2a})}} = \frac{s_{\lambda^a}}{\sqrt{\lambda^{2a}}}, \quad (4)$$

where $\bar{\lambda}^a = \sum_{i=1}^3 \lambda_i^a / 3$ [11] and s_{λ^a} is the sample standard deviation of the a -th power of the eigenvalues of \mathbf{D} . Note that

$$\text{FA}(\mathbf{D}^a) = |a| \sqrt{\frac{3}{2}} d_{AS}(\mathbf{I}_{3 \times 3}, \mathbf{D}|a).$$

We see that $\text{FA}(\mathbf{D})$ and $\text{PA}(\mathbf{D})$ are both members of $\text{FA}(\mathbf{D}^a)$ with $a = 1$ and $a = 1/2$ respectively. Note that for any non-zero a , $\text{FA}((c\mathbf{D})^a) = \text{FA}(\mathbf{D}^a)$ for any constant $c > 0$, i.e. all the members of this family of anisotropy maps are scale invariant. Perhaps less obvious is the following result that for any fixed tensor \mathbf{D} , the family is monotonically increasing in a .

THEOREM 2.1 *For any non-zero symmetric semi-positive definite \mathbf{D} , $\text{FA}(\mathbf{D}^a)$ is an increasing function of a , where $a \in [0, \infty)$.*

For the proof, see Appendix A. Hence, we have

Corollary 2.1 $\text{PA}(\mathbf{D}) \leq \text{FA}(\mathbf{D})$ for any non-zero tensor \mathbf{D} .

Note that as $a \rightarrow 0^+$, $\text{FA}(\mathbf{D}^a)$ approximates a rank indicator function:

$$\lim_{a \rightarrow 0^+} \text{FA}(\mathbf{D}^a) = \begin{cases} 0 & \text{if rank}(\mathbf{D}) = 3 \\ 1/\sqrt{2} & \text{if rank}(\mathbf{D}) = 2 \\ 1 & \text{if rank}(\mathbf{D}) = 1. \end{cases}$$

Also

$$\lim_{a \rightarrow +\infty} \text{FA}(\mathbf{D}^a) = \begin{cases} 0 & \text{if } \lambda_1 = \lambda_2 = \lambda_3 > 0 \\ 1/\sqrt{2} & \text{if } \lambda_1 = \lambda_2 > \lambda_3 \\ 1 & \text{if } \lambda_1 > \lambda_2 \geq \lambda_3, \end{cases}$$

that is, any oblate tensor \mathbf{D} in the limit flattens to a disk, whereas a general tensor becomes linear.

We can also define log-Euclidean Anisotropy (LA) for full rank tensors (apart from ones with $\lambda_1 = \lambda_2 = \lambda_3 = 1$) as follows

$$\text{LA}(\mathbf{D}) = \text{FA}(\log(\mathbf{D})) = \sqrt{\frac{3 \sum_{i=1}^3 \left(\log(\lambda_i) - \overline{\log(\lambda)} \right)^2}{2 \sum_{i=1}^3 (\log(\lambda_i))^2}}.$$

Note that $\text{LA}(\mathbf{D})$ is not a member of the family (4), and $\text{FA}(\mathbf{D}^a)$ does not converge to $\text{LA}(\mathbf{D})$ as $a \rightarrow 0^+$. Hence the case of $\text{LA}(\mathbf{D})$ is not covered by Theorem 2.1, and indeed, for some tensors $\text{LA}(\mathbf{D}) \leq \text{FA}(\mathbf{D})$ whereas for others $\text{LA}(\mathbf{D}) \geq \text{FA}(\mathbf{D})$. Also, for any non-zero a , $\text{LA}(\mathbf{D}^a) = \text{LA}(\mathbf{D})$, and by varying a , it is possible to make $\text{FA}(\mathbf{D}^a)$ smaller or larger than $\text{LA}(\mathbf{D})$. Note also that as $\lambda_3 \rightarrow 0$ (while λ_1 and λ_2 are bounded away from 0 or vanish at slower rates), $\text{LA} \rightarrow 1$. At the same time, as $\lambda_1 \rightarrow \infty$ (while λ_2 and λ_3 are bounded or grow at slower rates), $\text{LA} \rightarrow 1$ as well. Thus, in the limit LA does not distinguish between planar diffusion and linear diffusion. Likewise, as $\lambda_1 = \lambda_2 \rightarrow \infty$ (while λ_3 is bounded or grows at a slower rate), $\text{LA} \rightarrow 1/\sqrt{2}$. Also, as $\lambda_2 = \lambda_3 \rightarrow 0$ (while λ_1 is bounded away from 0 or vanishes at a slower rate), $\text{LA} \rightarrow 1/\sqrt{2}$ as well. Thus LA in the limit does not distinguish between isotropic planar diffusion and linear diffusion either. These observations might also explain why FA and other members of its family (4) may be preferred in practice as elaborated in Section 2.3 below.

At the same time, for full rank tensors,

$$\lim_{a \rightarrow 0^+} \frac{\text{FA}(\mathbf{D}^a)}{a} = \sqrt{\frac{1}{2} \sum_{i=1}^3 \left(\log(\lambda_i) - \overline{\log(\lambda)} \right)^2} \quad (5)$$

which is just the sample standard deviation of the spectrum of $\log(\mathbf{D})$, and is also $\text{GA}(\mathbf{D})/\sqrt{2}$, where GA is the geodesic anisotropy [9].

2.3 Practical comparisons of anisotropy measures

One motivation for considering anisotropy other than FA is to enhance visual experience of the end user. In order to visualize the difference between FA and PA in a practical example we can see in Figure 1 much better contrast in the bright areas (high anisotropy) of the PA map compared to the FA map, and structure within the corpus callosum in the zoomed insets is easier to visualize using PA.

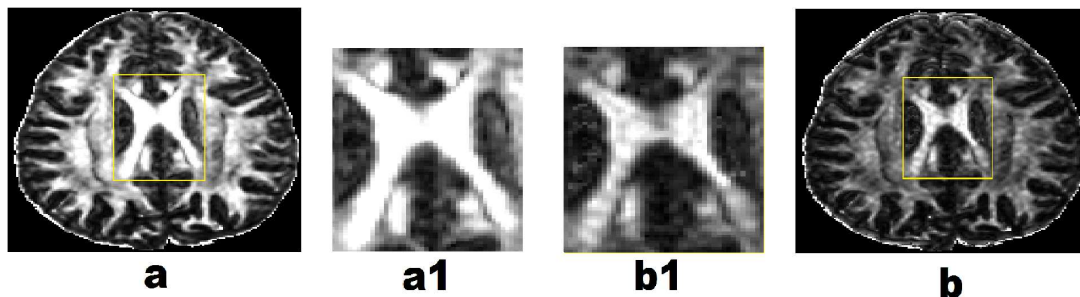


Figure 1. FA (a, a1) and PA (b, b1) maps; a and b are axial FA and PA maps; a1 and b1 are zoomed inset regions in the yellow box.

More generally, using values of a larger than 1 can increase the contrast in low anisotropy regions, while using values of a lower than 1, e.g. $a = 0.5$ as in PA, can increase the contrast in high anisotropy regions. It may then appear as if some brightness transformation of the $[0, 1]$ image intensity (anisotropy) range could increase the contrast within the bright structures of a1 to reveal the same amount of detail as in b1. However, for a general pair of anisotropy maps ($a, b > 0$, $a \neq b$ as in Theorem 2.1) such as FA ($a = 1$) and PA ($b = 0.5$), there need not exist a brightness transformation $\phi : [0, 1] \rightarrow [0, 1]$ such that $\phi(\text{FA}(\mathbf{D}^a)) = \text{FA}(\mathbf{D}^b)$ for all tensors \mathbf{D} . In fact, consider \mathbf{D}_1 with eigenvalues $[1, 0.1, 0.01]$ and \mathbf{D}_2 with eigenvalues $[1, \approx 0.1011, 0]$. It is easy to see that both tensors have the same FA value of ≈ 0.9486 , but $\text{FA}(\mathbf{D}_1^{0.025}) \approx 0.0864$ whereas $\text{FA}(\mathbf{D}_2^{0.025}) \approx 0.7077$, which would be impossible if ϕ as above existed. Therefore, computing $\text{FA}(\mathbf{D}^a)$ for a range of a values is important for providing the end user with the flexibility of highlighting specific anisotropy ranges in real time and subsequently optimising the end user's visual perception of specific features of the DT image. (Note that re-computing $\text{FA}(\mathbf{D}^a)$ in real time for a range of a values is not an issue.)

To illustrate the possibility of contrast adjustment with the help of the $\text{FA}(\mathbf{D}^a)$ family of maps, Figure 2 presents $\text{FA}(\mathbf{D}^a)$ maps of the following parametric family of tensors $\mathbf{D}(t)$:

$$\mathbf{D}(t) = \begin{cases} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 - 2t \end{pmatrix} & \text{if } t \in [0, 1/2], \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 - 2t & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{if } t \in (1/2, 1] \end{cases}$$

for $a \in \{\frac{1}{40}, \frac{1}{10}, \frac{1}{2}, 1, 2, 10, 40\}$. Note that $\text{FA}(\mathbf{D}(0)) = 0$, $\text{FA}(\mathbf{D}(\frac{1}{2})) = \frac{1}{\sqrt{2}}$, and $\text{FA}(\mathbf{D}(1)) = 1$, and for $t \in [1/2, 1)$ $\mathbf{D}(t)$ corresponds to planar diffusion, i.e. $\text{rank}(\mathbf{D}(t)) = 2$. The fact that FA does not distinguish between anisotropy of \mathbf{D}_1 (full rank) and \mathbf{D}_2 (rank-deficient) whereas PA and other members (with $a < 1$) do, is not surprising and adds to our motivation to use other members along with FA. Indeed, as already discussed in the context of Equation 1, PA is based on the full Procrustes metric, which, unlike some other metrics, can deal with some rank-deficient situations. For more details of comparing FA with PA, see [41].

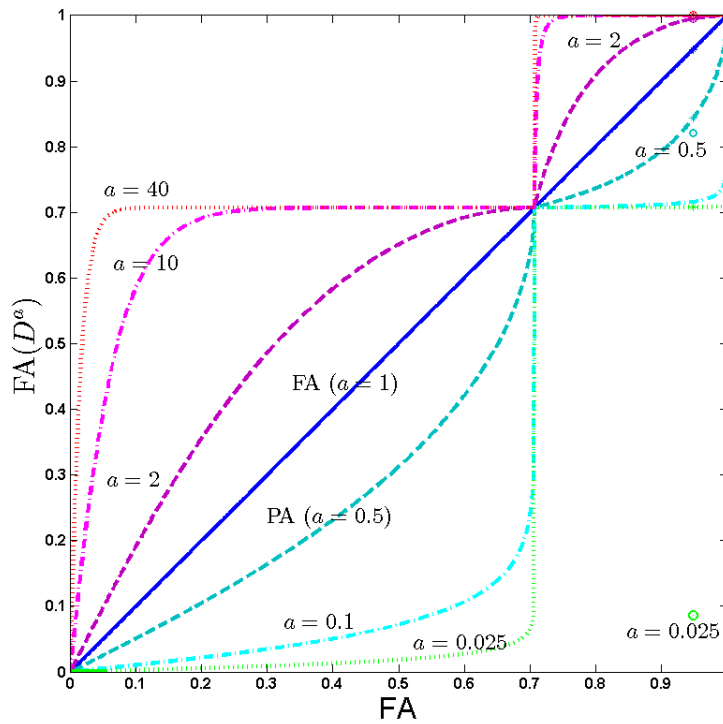


Figure 2. Members of the $FA(\mathbf{D}^a)$ family of anisotropy maps are ordered as explained by Theorem 2.1 but still meet at the points of full isotropy ($FA(\mathbf{D}^a) = 0$), planar diffusion ($FA(\mathbf{D}^a) = 1/\sqrt{2}$) and linear diffusion ($FA(\mathbf{D}^a) = 1$). The isolated circles correspond to the tensor \mathbf{D}_1 with eigenvalues $[1, 0.1, 0.01]$, which has the same FA value of ≈ 0.9486 as $\mathbf{D}_2 = \mathbf{D}(t \approx 0.1011)$ (intersecting the curves at $FA = 0.9486$), whereas the other anisotropy measures (e.g. $a = 1/2, 1/40$) disagree notably for these tensors.

3. Comparing distances

3.1 Weighted generalised Procrustes mean

We are often interested in weighted mean tensor estimation, for example when interpolating or smoothing tensors. A brief description of the method was given in [44] but here we supply more details and consider more applications and results. Given a suitable distance function d , the weighted Fréchet mean [2, 16] of a sample of N diffusion tensors $\mathbf{D}_1, \dots, \mathbf{D}_N$ is defined by

$$\hat{\Sigma} = \arg \inf_{\Sigma} \sum_{i=1}^N w_i d^2(\mathbf{D}_i, \Sigma), \quad (6)$$

where the weights w_i satisfy $0 \leq w_i \leq 1$ and $\sum_{i=1}^N w_i = 1$, and in applications can be, for example, a function of the Euclidean distance from the location of interest to the sampling locations (e.g., voxels). Naturally, different choices of $d(\cdot)$ generally result in different weighted mean diffusion tensors.

Procrustes analysis is a powerful shape analysis tool for matching configurations of points as closely as possible using the similarity transformations (rotation, translation and scaling) [12, 17]. In this study, a weighted Procrustes framework is proposed to compute the weighted Fréchet mean tensor.

To ensure the symmetry and positive semi-definiteness of \mathbf{D} , a reparameterisation $\mathbf{D} = \mathbf{Q}\mathbf{Q}^T$ is used [42], where \mathbf{Q} is a general 3×3 matrix. A Procrustes size-and-shape metric [11] was introduced to match \mathbf{Q}_1 (from \mathbf{D}_1) and \mathbf{Q}_2 (from \mathbf{D}_2) under rotation and

reflection while preserving scale. Thus, the Procrustes size-and-shape metric is defined as

$$d_S(\mathbf{D}_1, \mathbf{D}_2) = \inf_{\mathbf{R} \in O(3)} \|\mathbf{Q}_1 - \mathbf{Q}_2 \mathbf{R}\|, \quad (7)$$

where a 3×3 rotation and reflection matrix \mathbf{R} ranges over $O(3)$, the space of 3×3 orthogonal matrices. Note that d_S in Equation (7) is well-defined as a metric on the set of symmetric positive semi-definite matrices since $\mathbf{Q}\mathbf{Q}^T = \mathbf{P}\mathbf{P}^T$ implies $\mathbf{P} = \mathbf{Q}\mathbf{R}$ for some $\mathbf{R} \in O(3)$ and the Frobenius (Euclidean) norm $\|\cdot\|$ is rotation and reflection invariant.

Any orthogonal matrix $\hat{\mathbf{R}}$ that minimises the norm of the difference in Equation (7) is given by $\hat{\mathbf{R}} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}, \mathbf{V} \in O(3)$ are obtained from a singular value decomposition $\mathbf{Q}_1^T \mathbf{Q}_2 = \mathbf{V}\mathbf{\Delta}\mathbf{U}^T$, with $\mathbf{\Delta}$ a diagonal 3×3 matrix of singular values.

When a sample of N tensors is available, weighted generalised Procrustes analysis (WGPA) can be used to find the weighted mean $\hat{\mathbf{\Sigma}}$ when $d(\cdot) = d_S(\cdot)$ is the size-and-shape distance [11]. Specifically, WGPA computes $\hat{\mathbf{\Sigma}}_{WGPA} = \hat{\mathbf{Q}}_{WGPA} \hat{\mathbf{Q}}_{WGPA}^T$, with $\hat{\mathbf{Q}}_{WGPA} = \sum_{i=1}^N w_i \mathbf{Q}_i \hat{\mathbf{R}}_i$, where the orthogonal matrices $\hat{\mathbf{R}}_i, i = 1, \dots, N$ minimise the sum of weighted squared Euclidean norms given by

$$\begin{aligned} f_{WGPA}(\mathbf{R}_1, \dots, \mathbf{R}_N) &= \sum_{i=1}^N w_i \|\mathbf{Q}_i \mathbf{R}_i - \sum_{j=1}^N w_j \mathbf{Q}_j \mathbf{R}_j\|^2 \\ &= \sum_{i=1}^N w_i \|(1 - w_i) \mathbf{Q}_i \mathbf{R}_i - \sum_{j \neq i} w_j \mathbf{Q}_j \mathbf{R}_j\|^2 \\ &= \sum_{i=1}^N w_i (1 - w_i)^2 \|\mathbf{Q}_i \mathbf{R}_i - \frac{1}{(1 - w_i)} \sum_{j \neq i} w_j \mathbf{Q}_j \mathbf{R}_j\|^2. \end{aligned} \quad (8)$$

Note that \mathbf{Q}_i could be the Cholesky decomposition $\text{chol}(\mathbf{D}_i)$ or the symmetric, i.e. principal square root of \mathbf{D}_i or another choice such that $\mathbf{D}_i = \mathbf{Q}_i \mathbf{Q}_i^T$, since any two decomposition matrices of the same tensor are related via an orthogonal transformation. We prefer the symmetric square root not least because its uniqueness extends to degenerate matrices. An iterative algorithm to minimise f_{WGPA} was introduced in [44] and is used here to produce experimental results in Section 6.2.

An extension of the method to a power Procrustes approach is where $\mathbf{D} = \mathbf{Q}^{1/a}$, and $\mathbf{Q} = \mathbf{D}^a$ is the symmetric a root. The algorithm proceeds exactly as above for an even power of $1/a$, where above we have used $1/a = 2$.

3.2 Other tensor distances and their properties

The weighted tensor averaging can be applied to interpolate between two diffusion tensors. Weighted averages of \mathbf{D}_1 and \mathbf{D}_2 corresponding to different metrics are listed in Table 1. The weights are typically constrained by $w_1 + w_2 = 1$, enabling us to write $\mathbf{D}(w)$ for $\mathbf{D}(1 - w, w)$. For interpolation, we additionally require $w_i \geq 0, i = 1, 2$, but Table 1 does not make such assumptions. In the table, $\hat{\mathbf{R}}$ is the Procrustes minimiser of Equation (7), and transposition in the case of d_H is not necessary given the symmetry of the principal square root tensors. The definitions of the metrics used in this table were summarised in [11]. Note that the Euclidean root distance $d_H(\mathbf{D}_1, \mathbf{D}_2) = \|\mathbf{D}_1^{1/2} - \mathbf{D}_2^{1/2}\| = 2d_A(\mathbf{D}_1, \mathbf{D}_2|0.5)$.

Table 1. Weighted averages of \mathbf{D}_1 and \mathbf{D}_2 using different metrics.

Metric (notation)	Weighted Average $\mathbf{D}(w_1, w_2)$
Euclidean (d_E)	$w_1\mathbf{D}_1 + w_2\mathbf{D}_2$
log-Euclidean (d_L)	$\exp\{w_1 \log(\mathbf{D}_1) + w_2 \log(\mathbf{D}_2)\}$
Affine-invariant Riemannian (d_R)	$\mathbf{D}_1^{\frac{w_1+w_2}{2}} \left(\mathbf{D}_1^{-\frac{1}{2}} \mathbf{D}_2 \mathbf{D}_1^{-\frac{1}{2}} \right)^{w_2} \mathbf{D}_1^{\frac{w_1+w_2}{2}}$
Cholesky (d_C)	$[w_1 \text{chol}(\mathbf{D}_1) + w_2 \text{chol}(\mathbf{D}_2)]$ $[w_1 \text{chol}(\mathbf{D}_1) + w_2 \text{chol}(\mathbf{D}_2)]^T$
Euclidean root (d_H)	$(w_1 \mathbf{D}_1^{1/2} + w_2 \mathbf{D}_2^{1/2})^2$
Procrustes size-and-shape (d_S)	$(w_1 \mathbf{Q}_1 + w_2 \mathbf{Q}_2 \hat{\mathbf{R}})(w_1 \mathbf{Q}_1 + w_2 \mathbf{Q}_2 \hat{\mathbf{R}})^T$

All of the above interpolants $\mathbf{D}(w)$ satisfy $d(\mathbf{D}(w_1), \mathbf{D}(w_2)) = |w_1 - w_2|d(\mathbf{D}(0), \mathbf{D}(1))$ for all $w_1, w_2 \in [0, 1]$, and in particular describe the shortest path between \mathbf{D}_1 and \mathbf{D}_2 with respect to their metrics. We will refer to curves $\{\mathbf{D}(w), w \in I \supset [0, 1]\}$ as geodesics through \mathbf{D}_1 and \mathbf{D}_2 ; extrapolation corresponds to w outside $[0, 1]$.

From their definitions, it is clear that $d_H(\mathbf{D}_1, \mathbf{D}_2) \geq d_S(\mathbf{D}_1, \mathbf{D}_2)$ for any tensors $\mathbf{D}_1, \mathbf{D}_2$. Since we find these two metrics particularly useful, especially when dealing with degenerate tensors, and often producing similar results in our studies, we want to understand their relationship better. To this effect, we establish the following theorem.

THEOREM 3.1 *Let $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{\geq 0}(3)$. Then, $\sqrt{0.5}d_H(\mathbf{D}_1, \mathbf{D}_2) \leq d_S(\mathbf{D}_1, \mathbf{D}_2) \leq d_H(\mathbf{D}_1, \mathbf{D}_2)$. Moreover, if $\mathbf{D}_1 \neq \mathbf{D}_2$, and \mathbf{D}_1 and \mathbf{D}_2 are of rank 1, then $\sqrt{0.5}d_H(\mathbf{D}_1, \mathbf{D}_2) < d_S(\mathbf{D}_1, \mathbf{D}_2)$ and $d_S(\mathbf{D}_1, \mathbf{D}_2)/d_H(\mathbf{D}_1, \mathbf{D}_2) \rightarrow \sqrt{0.5}$ as $d(\mathbf{D}_1, \mathbf{D}_2) \rightarrow 0$ in any metric d .*

See Appendix B for the proofs of this and the following auxiliary result.

Proposition 3.1 Let $\mathbf{Q}_i, i = 1, 2$ be symmetric positive semi-definite $n \times n$ real matrices. Then

$$\min_{\mathbf{R} \in O(n)} \|\mathbf{Q}_1 - \mathbf{Q}_2 \mathbf{R}\|^2 \geq 0.5 \|\mathbf{Q}_1 - \mathbf{Q}_2\|^2. \quad (9)$$

Thus, in the rank 1 case, the two metrics diverge the most as the two tensors become indistinguishable. We also conjecture that in general, in order to approach the $1/\sqrt{2}$ bound, the two tensors must become indistinguishable and simultaneously approach the subspace of rank 1 tensors; see Appendix B for more details. Section 4.3.2 below gives relevant supporting experiments.

Despite these seemingly rigid bounds, the two distances can lead to significantly different results when used to average degenerate tensors, as established in the following proposition.

Proposition 3.2 Let $\text{rank}(\mathbf{D}_1) = \text{rank}(\mathbf{D}_2) < 3$. Then $\text{rank}(\mathbf{D}(w_1, w_2))$ is constant for all (general) $w_1, w_2 \in \mathbb{R}$, where $\mathbf{D}(w_1, w_2)$ is as in the d_S row of Table 1.

For the proof see Appendix C. The result generalises to degenerate tensors of higher dimensions. It is straightforward to see that neither the root-Euclidean nor the Euclidean averaging has this property, i.e. the rank of their weighted average tensor is generally higher than the rank of the reference tensors \mathbf{D}_1 and \mathbf{D}_2 when the latter are degenerate. Such jumps in rank can be thought of as an infinite extension of the swelling effect, and hence may be undesirable in practice. We show consequences of these observations in experiments in Section 4.3.2 below. We think that this is an important point in favour

of using d_S in practice when dealing with positive semi-definite tensors.

It is worth noting that the Euclidean, log-Euclidean Riemannian, and the power-Euclidean metrics are all special cases of the following general type of metric

$$d_g(\mathbf{D}_1, \mathbf{D}_2) = \|g(\mathbf{D}_1) - g(\mathbf{D}_2)\|. \quad (10)$$

Namely, let g be an injection from the non-negative real numbers $\mathbb{R}_{\geq 0}$ to the extended real line $\bar{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$. Consider the extension of g to the following matrix function from $\Omega_{\geq 0}(3)$ to the 3×3 symmetric $\bar{\mathbb{R}}$ -valued matrices:

$$g(\mathbf{D}) = \mathbf{E} \begin{pmatrix} g(\lambda_1) & 0 & 0 \\ 0 & g(\lambda_2) & 0 \\ 0 & 0 & g(\lambda_3) \end{pmatrix} \mathbf{E}^T,$$

where $\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$ is the spectral decomposition of \mathbf{D} .

Thus, in the Euclidean case $g(x) = x$, in the log-Euclidean case $g(x) = \log(x)$, and in the root-Euclidean case $g(x) = \sqrt{x}$.

We will refer to such metrics as *Euclidean-based*. In this case, the general weighted Fréchet mean Equation (6) has a unique solution as shown below:

$$\begin{aligned} \mathbf{D}(w_1, w_2) &= \arg \min_{\mathbf{D}} \{w_1 d_g^2(\mathbf{D}_1, \mathbf{D}) + w_2 d_g^2(\mathbf{D}_2, \mathbf{D})\} \\ \mathbf{D}(w_1, w_2) &= g^{-1}(w_1 g(\mathbf{D}_1) + w_2 g(\mathbf{D}_2)), \end{aligned} \quad (11)$$

which unifies the respective entries in Table 1 above, and will be extended and exploited further in the ensuing discussion.

We would also like to understand how various tensor interpolations behave and compare in terms of the tensor size. For example, the swelling effect under the Euclidean averaging has been frequently mentioned in the literature. Although this effect is indeed straightforward, below we give its detailed mathematical explanation, which we have not seen stated explicitly in DTI literature. Namely, writing $\mathbf{D}_E(w_1, w_2)$ for the weighted Euclidean average (with $w_1 + w_2 = 1$, $w_1, w_2 \geq 0$), the Minkowski Determinant Theorem [26] with dimension $n = 3$ gives

$$|\mathbf{D}_E(w_1, w_2)|^{\frac{1}{3}} \geq w_1 |\mathbf{D}_1|^{\frac{1}{3}} + w_2 |\mathbf{D}_2|^{\frac{1}{3}} \quad (12)$$

with the equality if and only if $\mathbf{D}_2 = c\mathbf{D}_1$ for some $c \geq 0$. Thus, $|\mathbf{D}_E(w_1, w_2)| \geq \min\{|\mathbf{D}_1|, |\mathbf{D}_2|\}$, and if $|\mathbf{D}_1| = |\mathbf{D}_2|$ but the tensors are distinct, the swelling is inevitable as $|\mathbf{D}_E(w_1, w_2)| > |\mathbf{D}_1| = |\mathbf{D}_2|$ for positive weights w_1, w_2 . Since both the affine-invariant and log-Euclidean Riemannian interpolations give $|\mathbf{D}_{L,R}(w_1, w_2)| = |\mathbf{D}_1|^{w_1} |\mathbf{D}_2|^{w_2}$ [2], the arithmetic-geometric mean inequality implies $|\mathbf{D}_{L,R}(w_1, w_2)| \leq |\mathbf{D}_E(w_1, w_2)|$ [26] for any positive definite tensors \mathbf{D}_1 and \mathbf{D}_2 and for all (probability) weights w_1, w_2 .

A simple but useful observation is that the root-Euclidean determinants $|\mathbf{D}_H(w_1, w_2)|$ are sandwiched between the Euclidean $|\mathbf{D}_E(w_1, w_2)|$ and geometric, i.e. log-Euclidean and affine-invariant Riemannian, ones $|\mathbf{D}_{L,R}(w_1, w_2)|$, as stated next.

Proposition 3.3 Let w_1, w_2 be probability weights, and let \mathbf{D}_1 and \mathbf{D}_2 be $n \times n$ positive definite symmetric matrices, and let $\mathbf{D}_{L,R}(w_1, w_2)$, $\mathbf{D}_H(w_1, w_2)$, and $\mathbf{D}_E(w_1, w_2)$ be their Riemannian (log-Euclidean or affine-invariant), root-Euclidean, and Euclidean weighted averages, respectively, where $w_1, w_2 \geq 0$. Then we have

$$|\mathbf{D}_{L,R}(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)| \leq |\mathbf{D}_E(w_1, w_2)|.$$

See Appendix D for the proof of this proposition and the following corollary and proposition.

Corollary 3.1 Under the assumptions of Proposition 3.3 above, we also have

$$|\mathbf{D}_L(w_1, w_2)| \leq |\mathbf{D}_S(w_1, w_2)|.$$

We establish next that in general $|\mathbf{D}_S(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)|$, therefore

$$|\mathbf{D}_{L,R}(w_1, w_2)| \leq |\mathbf{D}_S(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)| \leq |\mathbf{D}_E(w_1, w_2)|. \quad (13)$$

Proposition 3.4 Let \mathbf{D}_1 and \mathbf{D}_2 be $n \times n$ symmetric positive semi-definite real matrices, and let $\mathbf{D}_S(w_1, w_2)$ and $\mathbf{D}_H(w_1, w_2)$ be their Procrustes and root-Euclidean weighted averages, respectively, where $w_1, w_2 \geq 0$. Then

$$|\mathbf{D}_S(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)|.$$

We will see below that when \mathbf{D}_1 and \mathbf{D}_2 commute, $\mathbf{D}_S(w_1, w_2) = \mathbf{D}_H(w_1, w_2)$. Our experiments in Sections 4.3.1 and 6.2 show further that $|\mathbf{D}_H(w_1, w_2)|$ dominates $|\mathbf{D}_S(w_1, w_2)|$ by a relatively small margin only. This is not surprising given that comparison of the traces, or MDs, reverses the inequality as established next.

Proposition 3.5 For all $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{>0}(3)$, and for all probability weights w_1, w_2 , we have

$$\text{tr } \mathbf{D}_R(w_1, w_2) \leq \text{tr } \mathbf{D}_L(w_1, w_2) \leq \text{tr } \mathbf{D}_H(w_1, w_2) \leq \text{tr } \mathbf{D}_S(w_1, w_2) \leq \text{tr } \mathbf{D}_E(w_1, w_2).$$

Moreover, the last two inequalities remain valid for all $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{\geq 0}(3)$, and additionally we have

$$\text{tr } \mathbf{D}_S(w_1, w_2) - \text{tr } \mathbf{D}_H(w_1, w_2) \leq 0.5d_H^2(w_1^2\mathbf{D}_1, w_2^2\mathbf{D}_2). \quad (14)$$

See Appendix E for the proof. Even more is proved in [7], from which the following Corollary follows immediately.

Corollary 3.2 For any $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{\geq 0}(3)$, $\text{tr } \mathbf{D}_A(w_1, w_2|a)$ is an increasing function of $a \in (0, \infty)$.

Also, for samples of $N \geq 2$ tensors, we still have

$$\text{tr } \mathbf{D}_L(w_1, w_2, \dots, w_N) \leq \text{tr } \mathbf{D}_A(w_1, w_2, \dots, w_N|a) \leq \text{tr } \mathbf{D}_E(w_1, w_2, \dots, w_N),$$

for all $a \in (0, 1)$, and in particular for $a = 0.5$, which corresponds to the square root-Euclidean averaging.

Note that $\text{tr } \mathbf{D}_L(w_1, w_2) \leq \text{tr } \mathbf{D}_E(w_1, w_2)$ has been, at least implicitly, commonly acknowledged in DTI literature (e.g. [2, 29]); $\text{tr } \mathbf{D}_R(w_1, w_2) \leq \text{tr } \mathbf{D}_L(w_1, w_2)$ has also been mentioned (e.g. [2]). However, we have not seen the other inequalities stated explicitly in the DTI literature.

We also point out that the power-Euclidean interpolants satisfy $\text{tr}(\mathbf{D}_A(w_1, w_2|a)^a) = w_1 \text{tr}(\mathbf{D}_1^a) + w_2 \text{tr}(\mathbf{D}_2^a)$, so, for our main example of $a = 0.5$, $\text{tr}(\mathbf{D}_H(w_1, w_2)^{\frac{1}{2}}) = w_1 \text{tr } \mathbf{Q}_1 +$

$w_2 \operatorname{tr} \mathbf{Q}_2$. Thus, if \mathbf{D}_1 and \mathbf{D}_2 have equal traces of their square roots, then this will be preserved by the square root of the interpolant for all weights with $w_1 + w_2 = 1$.

Returning to the dominance of $|\mathbf{D}_H(w_1, w_2)|$ over $|\mathbf{D}_S(w_1, w_2)|$ (Proposition 3.4), and in view of $\operatorname{tr} \mathbf{D}_H(w_1, w_2) \leq \operatorname{tr} \mathbf{D}_S(w_1, w_2)$, we note that the positivity of the difference $|\mathbf{D}_H(w_1, w_2)| - |\mathbf{D}_S(w_1, w_2)|$ cannot be explained by the first-order approximation.

All in all, the Procrustes and root-Euclidean methods may be a good compromise as they do give some swelling (which is clear from Equation (12) with \mathbf{D}_i replaced by \mathbf{Q}_i), but their swelling is less aggressive than that of the Euclidean method. This may be helpful in view of [29] arguing that swelling may be desirable in certain scenarios.

Now, we consider a special situation of averaging an isotropic tensor with a general tensor, which is also illustrated by experiments in Section 4.3.1. Since this means that the reference tensors \mathbf{D}_1 and \mathbf{D}_2 commute, the log-Euclidean Riemannian and the affine-invariant Riemannian averages will be identical [2], and we now show other implications of this assumption.

Proposition 3.6 Assume that one of the tensors $\mathbf{D}_1, \mathbf{D}_2$ (as in Table 1) is isotropic; without loss of generality, let it be \mathbf{D}_1 , i.e. $\mathbf{D}_1 = \lambda \mathbf{I}_{3 \times 3}$ for some $\lambda \geq 0$. Let d_g be a Euclidean-based metric. Let $\lambda_1, \lambda_2, \lambda_3$ be eigenvalues of \mathbf{D}_2 ,

- (1) The weighted average tensor $\mathbf{D}(w_1, w_2)$ is given by Equation (15) below. Subsequently, if g is increasing, then the orientation of $\mathbf{D}(w_1, w_2)$ is the same as the orientation of \mathbf{D}_2 for all weights with $w_2 \neq 0$. In particular, if $\lambda_1 > \lambda_2$, i.e. \mathbf{D}_2 is prolate (has a well-defined direction), then $\mathbf{D}(w_1, w_2)$ is also prolate and has the same direction (provided $w_2 \neq 0$).
- (2) The Procrustes averaging is identical to the root-Euclidean averaging, and $\mathbf{D}(w_1, w_2)$ in this case is given by Equation (16) below.
- (3) The (log-Euclidean or affine-invariant) Riemannian average tensor $\mathbf{D}(w_1, w_2)$ is given by Equation (17) below.
- (4) $\mathbf{D}(w_1, w_2)$ for the Euclidean averaging is given by Equation (18) below.
- (5) The Euclidean, log-Euclidean, the affine-invariant Riemannian, root-Euclidean, and the Procrustes interpolations all preserve the orientation of \mathbf{D}_2 , provided $w_2 \neq 0$.

$$\mathbf{E} \begin{pmatrix} g^{-1}((w_1 g(\lambda) + w_2 g(\lambda_1))) & 0 & 0 \\ 0 & g^{-1}((w_1 g(\lambda) + w_2 g(\lambda_2))) & 0 \\ 0 & 0 & g^{-1}((w_1 g(\lambda) + w_2 g(\lambda_3))) \end{pmatrix} \mathbf{E}^T, \quad (15)$$

where $\mathbf{E} \mathbf{\Lambda} \mathbf{E}^T$ is the spectral decomposition of \mathbf{D}_2 with eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$.

$$\mathbf{E} \left(\lambda w_1^2 \mathbf{I}_{3 \times 3} + w_2^2 \mathbf{\Lambda} + 2w_1 w_2 \sqrt{\lambda} \sqrt{\mathbf{\Lambda}} \right) \mathbf{E}^T, \quad (16)$$

$$\mathbf{E} (\lambda^{w_1} \mathbf{\Lambda}^{w_2}) \mathbf{E}^T, \quad (17)$$

$$\mathbf{E} (\lambda w_1 \mathbf{I}_{3 \times 3} + w_2 \mathbf{\Lambda}) \mathbf{E}^T. \quad (18)$$

The proof of this proposition is straightforward and is given in Appendix F.

Clearly, the above results, including the equality of the root-Euclidean and Procrustes averages, immediately generalise to any commuting tensors \mathbf{D}_1 and \mathbf{D}_2 . However, it is understood that in general, eigenvalues of only one of the two tensors can be ordered at a time when generalising Equation (15) to Equation (19) below:

$$\mathbf{E} \begin{pmatrix} g^{-1}((w_1 g(\gamma_1) + w_2 g(\lambda_1))) & 0 & 0 \\ 0 & g^{-1}((w_1 g(\gamma_2) + w_2 g(\lambda_2))) & 0 \\ 0 & 0 & g^{-1}((w_1 g(\gamma_3) + w_2 g(\lambda_3))) \end{pmatrix} \mathbf{E}^T, \quad (19)$$

where $\gamma_1, \gamma_2, \gamma_3$ are eigenvalues of \mathbf{D}_1 . Thus, the direction of the average tensor may turn out to be that of a non-principal eigenvector of \mathbf{D}_1 or \mathbf{D}_2 .

Although we focus on interpolation, extrapolation outside the shortest path between \mathbf{D}_1 and \mathbf{D}_2 is easily included by relaxing the non-negativity constraint on the weights w_i (see Section 4.3.2).

Finally, we note that other relevant metrics and more general measures of divergence continue to be proposed for symmetric positive definite matrices. One such example is the root Stein Divergence [34]

$$d_{\text{Stein}}(\mathbf{D}_1, \mathbf{D}_2) = \sqrt{\log \det \left[\frac{1}{2} (\mathbf{D}_1 + \mathbf{D}_2) \right] - \frac{1}{2} \log \det (\mathbf{D}_1 \mathbf{D}_2)}, \quad (20)$$

but we do not attempt to overview all such proposals in this paper.

4. Interpolation and smoothing

We now apply the above ideas and results on weighted averages to the tasks of interpolation and smoothing.

4.1 Choice of weights

The choice of weights is application dependent. For example, w_i could be a decreasing function of the Euclidean distance from the location of interest to the sampling location i . One simple choice for the weights is the inverse distance function given by

$$w_i = \frac{d_i^{-1}}{\sum_{j=1}^N d_j^{-1}}, i = 1, \dots, N, \quad (21)$$

where d_i is the Euclidean distance from the location (voxel) at which the weighted mean is to be estimated, to the location (voxel) of i th tensor \mathbf{D}_i .

For more flexibility, the following exponential weight function is proposed and used in this paper:

$$w_i = \frac{\exp(-Ad_i^2) + B}{\sum_{j=1}^N [\exp(-Ad_j^2) + B]}, i = 1, \dots, N, \quad (22)$$

where $A, B \geq 0$. The two parameter exponential weight family is flexible since it allows us to adjust the w_i depending on the application, as seen in Section 4.2 where we discuss cross validation.

Given distances $d_i, i = 1, 2, \dots, N$, some properties of exponential weight function are listed below:

As $A \rightarrow +\infty$ and $B = 0$,

$$w_i \rightarrow \begin{cases} 1/k & \text{if } i \in \arg \min\{d_1, d_2, \dots, d_N\}, \\ 0 & \text{otherwise,} \end{cases}$$

where k is the size of the set $\arg \min\{d_1, d_2, \dots, d_N\}$. Under any of the following conditions, the weight distribution becomes uniform, i.e. $w_i \rightarrow 1/N$: (a) $A \rightarrow +\infty$ and $B \neq 0$ is fixed; (b) $A \rightarrow 0$ and B is fixed; (c) $B \rightarrow +\infty$ and A is fixed.

4.2 Cross-validation

To illustrate how parameters A and B in the weight function (22) can be chosen in practice, and to also assess the weighted mean estimators, a cross-validation procedure is carried out over a region of interest shown in Figure 3(a) and taken from the real human brain data described in Section 6 below. Voxels V_1, \dots, V_K ($K = 630$ for this example) shown in Figure 3(b) in black contain the validation tensors and remaining voxels contain training tensors. All tensors $\mathbf{D}_1, \dots, \mathbf{D}_K$ are initially estimated using a Bayesian framework [42] and the resulting estimates are considered the ground truth for the purposes of cross-validation. We re-estimate the tensor at voxel V_i , $i = 1, \dots, K$ by computing the weighted mean of its four first-order neighbours (i.e. tensors from two horizontally and two vertically adjacent voxels). The spatial resolution in the horizontal direction is twice that in the vertical direction, hence d_i in Equation (22) is taken to be 1 for horizontal neighbours and 2 for vertical neighbours, hence the weights are generally unequal.

To compare the effects of using the Euclidean d_E , log-Euclidean d_L , root-Euclidean d_H , and Procrustes d_S distances in Equation (6), we compute the root mean square distance (RMSD) between the re-estimated tensor \mathbf{D}_i^{cv} and the original estimate \mathbf{D}_i , as shown below:

$$RMSD(d) = \sqrt{\frac{1}{K} \sum_{i=1}^K d(\mathbf{D}_i, \mathbf{D}_i^{cv})^2},$$

where d can be any suitable metric, such as d_E , d_L , d_H , and d_S .

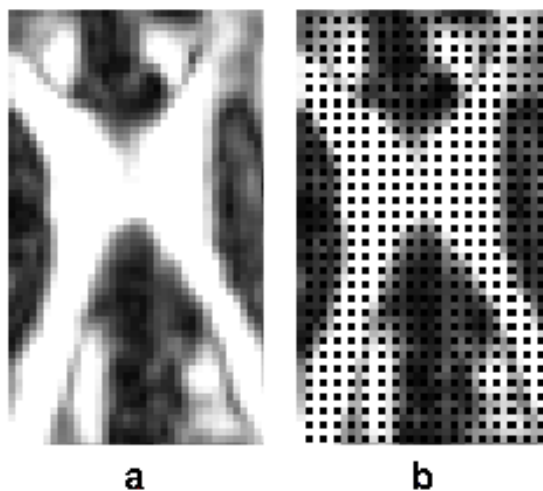


Figure 3. A two-dimensional region of interest; locations of validation tensors are shown in black (b) against the background of FA map (a) of the same region (the remaining locations provide training tensors).

To choose parameters A and B ($A > 0$, $B > 0$) for the weight function (22), we consider $RMSD(d_E)$, $RMSD(d_L)$, $RMSD(d_H)$, and $RMSD(d_S)$ simultaneously. A greedy algorithm is used, starting from small positive values for weight parameters ($A = 0.01$

and $B = 0.01$). We then first increase A in steps of 0.01 and calculate the RMSD values. If at least one RMSD value starts increasing, the previous value of A is retained as optimal. Next, we increase B in steps of 0.01 and re-calculate all the RMSD values. If at least one RMSD value starts increasing, the previous B value is declared optimal. The optimal parameter setting for the example in Figure 3 is found to be $A = 2$ and $B = 0.01$. We choose to optimise over A first since this parameter has a bigger impact on the weights.

Table 2 shows $RMSD(d_E)$, $RMSD(d_L)$, $RMSD(d_H)$, and $RMSD(d_S)$. The weighted Procrustes approach provides the smallest $RMSD(d_E)$, $RMSD(d_H)$, and $RMSD(d_S)$ measures in this example, whereas the weighted root-Euclidean method provides the smallest $RMSD(d_E)$ and $RMSD(d_L)$ and $RMSD(d_S)$ measures. Thus, in this example, the two methods perform best in general, although their performances are very similar.

Table 2. Measures of the cross-validation results with different methods.

	Euclidean	log-Euclidean	Euclidean root	Procrustes
$RMSD(d_E)$	0.00005	0.00006	0.00005	0.00005
$RMSD(d_L)$	0.32601	0.31696	0.29664	0.29881
$RMSD(d_H)$	0.00101	0.00094	0.00093	0.00092
$RMSD(d_S)$	0.00082	0.00081	0.00077	0.00077

4.3 Interpolation and extrapolation

In Sections 4.3.1 and 4.3.2 below we consider the simplest case of interpolation and extrapolation of two tensors, respectively, using the main metrics of interest. Then, in Section 4.3.3 we briefly illustrate the lack of invariance of the Cholesky interpolation to orthogonal changes of coordinates, and therefore its lack of practical application in DTI. This is followed by Section 4.3.4 with experiments on interpolation of four tensors.

4.3.1 Interpolation of two diffusion tensors

We illustrate the observations of Proposition 3.6 and also investigate more general paths obtained with the metrics listed in Table 1. Specifically, we choose two reference tensors $\mathbf{D}(0)$ and $\mathbf{D}(1)$ and then interpolate between them by sampling $N - 1$ additional tensors along the shortest path as follows:

$$\mathbf{D}(w_i) = \arg \min_{\mathbf{D}} [(1 - w_i)d^2(\mathbf{D}(0), \mathbf{D}) + w_i d^2(\mathbf{D}(1), \mathbf{D})], \quad (23)$$

resulting in the total of $N + 1$ tensors with weights $w_i = i/N$, $i = 0, 1, \dots, N$. The settings of $\mathbf{D}(0)$ and $\mathbf{D}(1)$ for each experiment are as follows:

Experiment I:

$$\mathbf{D}(0) = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \quad \mathbf{D}(1) = \begin{pmatrix} 8.50 & 7.50 & 0 \\ 7.50 & 8.50 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 16 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}^T.$$

Experiment II:

$$\mathbf{D}(0) = \begin{pmatrix} 5.50 & 4.50 & 0 \\ 4.50 & 5.50 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}^T,$$

$$\mathbf{D}(1) = \begin{pmatrix} 4.72 & -11.46 & 0 \\ -11.46 & 36.28 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} \sin(-0.1\pi) & 0 & \cos(1.1\pi) \\ \cos(-0.1\pi) & 0 & \sin(1.1\pi) \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 40 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sin(-0.1\pi) & 0 & \cos(1.1\pi) \\ \cos(-0.1\pi) & 0 & \sin(1.1\pi) \\ 0 & 1 & 0 \end{pmatrix}^T.$$

To help us compare the effects of using the different metrics on the size, orientation, and shape of the interpolated tensor $\mathbf{D}(w_i)$, we measure its volume (determinant $|\mathbf{D}(w_i)|$) and perimeter ($\text{tr } \mathbf{D}(w_i)$), angle $\phi(w_i)$ of relative orientation, and fractional (FA) and Procrustes (PA) anisotropies. The angle $\phi(w_i)$ is the (non-obtuse) angle between the direction (principal eigenvector $\mathbf{p}\mathbf{v}(w_i)$) of $\mathbf{D}(w_i)$ and the direction (principal eigenvector $\mathbf{p}\mathbf{v}(1)$) of $\mathbf{D}(1)$:

$$\phi(w_i) = \arcsin(\|\mathbf{p}\mathbf{v}(1) \times \mathbf{p}\mathbf{v}(w_i)\|) \times 180^\circ/\pi, i = 0, \dots, N. \quad (24)$$

Experiment I, where $\mathbf{D}(0)$ is isotropic, illustrates the observations of Proposition 3.6. Since $\mathbf{D}(0)$ has no orientation, $\phi(0)$ is undefined in this case and will be replaced by $\lim \phi(w)$ as $w \rightarrow 0$.

The left pane of Figure 4 shows six geodesic paths in Experiment I with $N = 8$. Unlike $\mathbf{D}(0)$, $\mathbf{D}(1)$ is strongly anisotropic in this experiment, and the two tensors have the same volume. The right pane displays the above measures of size, orientation, and shape, using $N = 20$ for higher accuracy. The well-known swelling effect of the Euclidean interpolation is clearly seen from the plot of the GMD given by the cubic root of the determinant of $\mathbf{D}(w_i)$. Further, in the case of Procrustes interpolation (which is identical here to the root-Euclidean interpolation), Equation (16) gives $\mathbf{D}(w_i) = \mathbf{E} \left(4(1-w_i)^2 \mathbf{I}_{3 \times 3} + w_i^2 \mathbf{\Lambda} + 4w_i(1-w_i)\sqrt{\mathbf{\Lambda}} \right) \mathbf{E}^T$.

Since $\mathbf{D}(1)$ is prolate (in the direction $\mathbf{p}\mathbf{v}(1)$), so is each interpolant $\mathbf{D}(w_i)$, i.e. $\mathbf{p}\mathbf{v}(w_i)$ aligns with $\mathbf{p}\mathbf{v}(1)$ and $\phi_i = 0^\circ$ for all $i = 1, 2, \dots, 20$, for all but the Cholesky interpolation (the Cholesky method cannot be expressed in terms of matrix functions defined via the spectral decomposition) as promised by item (5) of Proposition 3.6.

In the case of the log-Euclidean interpolation (which is identical here to the affine-invariant Riemannian interpolation), Equation (17) gives $\mathbf{D}(w_i) = \mathbf{E} \left(4^{1-w_i} \mathbf{\Lambda}^{w_i} \right) \mathbf{E}^T$.

With the Cholesky metric (d_C) the angle $\phi(w_i)$ decreases gradually from about 15.5° ($\phi(w_1 = 0.05)$) to 0° ($\phi(w_{20} = 1)$) along the geodesic path. It is worth recalling that the Cholesky metric is problematic in practice due to its lack of invariance to orthogonal transformations.

Since the log-Euclidean and affine-invariant Riemannian methods provide the same monotonic geometric interpolation of determinants (even when their geodesics are otherwise distinct) $|\mathbf{D}(w)| = |\mathbf{D}(0)|^{1-w} |\mathbf{D}(1)|^w$, $|\mathbf{D}(0)| = |\mathbf{D}(1)|$ implies that these methods preserve the volume and subsequently GMD along their geodesic [2], whereas the other methods result in some swelling. The Euclidean averaging instead gives monotonic linear interpolation of traces $\text{tr } \mathbf{D}(w) = (1-w) \text{tr } \mathbf{D}(0) + w \text{tr } \mathbf{D}(1)$ (which, provided $\text{tr } \mathbf{D}(0) = \text{tr } \mathbf{D}(1)$, preserves the trace and subsequently the MD) [29]. The GMD and MD plots also illustrate the dominance of the arithmetic mean over the geometric mean, as well as the general inequality (13) and Proposition 3.5, respectively.

FA and PA values are all very similar for all the four distinct geodesics, and they are nearly identical for the Cholesky interpolation and the root-Euclidean/Procrustes interpolation. Although all of the four distinct FA (PA) graphs are very similar, the Euclidean interpolants are more anisotropic than the root-Euclidean/Procrustes inter-

polants, which in turn dominate the non-Euclidean Riemannian interpolants. The same ordering holds for the PA graphs, whose differences are even less noticeable in this example. Note that this ordering is not general, and, for example, increasing λ (the eigenvalue of $\mathbf{D}(0)$) to 6 reverses this order, and setting $\lambda = 5$ makes the distinct FA (PA) curves cross (not shown). Note also that for the non-Euclidean Riemannian interpolants we have $\text{FA}(\mathbf{D}(w)) = \text{FA}(\mathbf{\Lambda}^w)$; in particular, $\text{PA}(\mathbf{D}(1)) = \text{FA}(\mathbf{D}(0.5))$ for this geodesic.

Thus, the root-Euclidean/Procrustes geodesic disturbs the shape slightly more than does the log-Euclidean and affine-invariant Riemannian geodesic, but still notably less compared with the Euclidean geodesic.

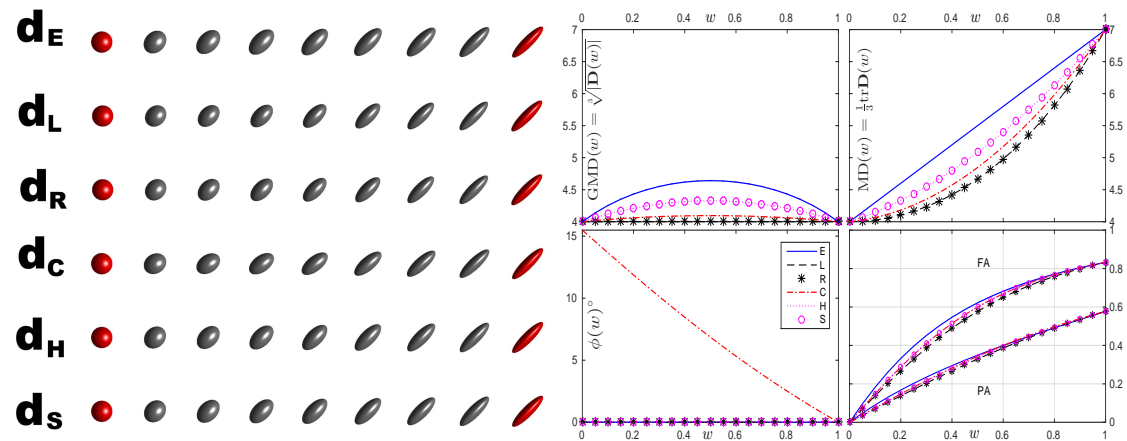


Figure 4. Experiment I: geodesic paths between two tensors of the same volume (red). Tensor $\mathbf{D}(0)$ (left) is fully isotropic and tensor $\mathbf{D}(1)$ (right) is strongly anisotropic. The geodesic paths are obtained with $d_E(\cdot)$, $d_L(\cdot)$, $d_R(\cdot)$, $d_C(\cdot)$, $d_H(\cdot)$ and $d_S(\cdot)$. As explained by Proposition 3.6, d_H and d_S yield the same results here; likewise, the results from using d_L and d_R are the same in this case [2]. ($\phi(0)$ is undefined and replaced by $\lim_{w \rightarrow 0} \phi(w)$ as $w \rightarrow 0$). The GMD and MD of the different interpolants compare according to Inequality (13) and Proposition 3.5, respectively. The non-Euclidean Riemannian interpolants satisfy $\text{FA}(\mathbf{D}(w)) = \text{FA}(\mathbf{\Lambda}^w)$ in this example, implying $\text{PA}(\mathbf{D}(1)) = \text{FA}(\mathbf{D}(0.5))$ for this geodesic.

In Experiment II, tensors $\mathbf{D}(0)$ and $\mathbf{D}(1)$ are neither orthogonal nor collinear, and are of different shapes and sizes. Figure 5 shows six distinct geodesic paths between $\mathbf{D}(0)$ and $\mathbf{D}(1)$. The Euclidean metric again suffers significant swelling. The geometric (log-Euclidean and affine-invariant Riemannian) interpolation of the determinants ($|\mathbf{D}(w)| = |\mathbf{D}_1|^{1-w} |\mathbf{D}_0|^w$) yields the lowest volume evolution among the considered methods. The Cholesky interpolation gives slightly larger volumes. The Procrustes determinant (GMD) curve is uniformly and notably above these three, and it is dominated, albeit less notably, by the root-Euclidean determinants.

The Cholesky path has a significant cusp (nearly 90°) in the orientation in this example. The Euclidean orientational angle curve is everywhere dominated by the others except for the affine-invariant Riemannian one, which it crosses on approach to $\mathbf{D}(1)$. Orientational angle curves of the non-Euclidean methods intersect non-trivially; the log-Euclidean and affine-invariant Riemannian curves tend to dominate the root-Euclidean and to a lesser extent the Procrustes orientation curves. The log-Euclidean and affine-invariant Riemannian orientation curves clearly reveal a sigmoidal shape, which is more pronounced in the affine-invariant case.

In terms of anisotropy (FA and PA maps), the affine-invariant Riemannian interpolation is most variable after the Cholesky interpolation. The log-Euclidean, root-Euclidean, and the Euclidean interpolations are all structurally similar, and the Procrustes interpolation is the least variable of the six.

In summary, the Procrustes size-and-shape metric and to a lesser extent the Euclidean root metric offer overall least variable interpolations of the size, orientation and shape.

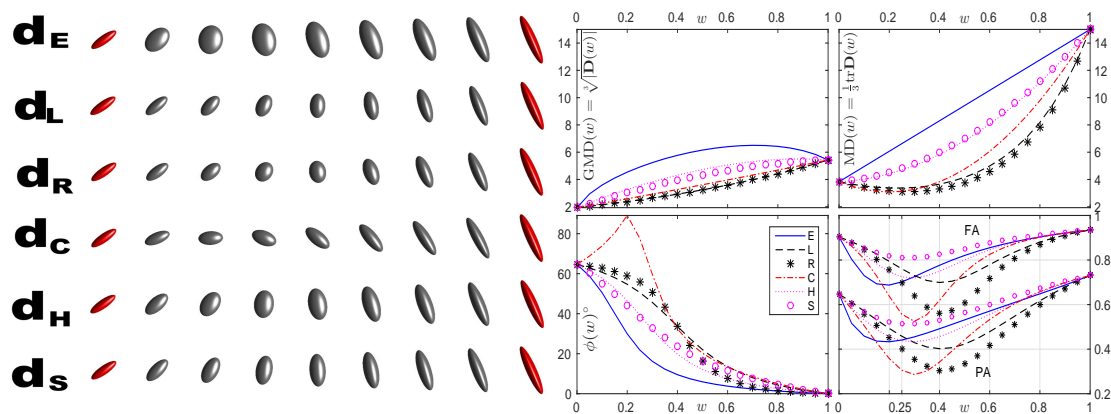


Figure 5. Geodesic paths in Experiment II between two general reference tensors $\mathbf{D}(0)$ (left, red) and $\mathbf{D}(1)$ (right, red) with general (i.e. non-collinear non-orthogonal) orientation, different shape and size. The geodesic paths are obtained with $d_E(\cdot)$, $d_L(\cdot)$, $d_R(\cdot)$, $d_C(\cdot)$, $d_H(\cdot)$ and $d_S(\cdot)$. The Geometric and Arithmetic Mean Diffusivities of the different interpolants compare according to Inequality (13) and Proposition 3.5, respectively. PA helps resolving ambiguity of FA at $w = 0.25$ where FA of the Euclidean and affine-invariant interpolants coincide, but their PA values do not. There is no general ordering of anisotropies apart from the log-Euclidean interpolants dominating the affine-invariant ones [2].

From the examples discussed above, and several more in [41] and other related studies (e.g. [2]), it does seem that the Euclidean metric may be very problematic, especially due to the swelling of the determinant (apart from situations where a certain degree of swelling may help compensate for a previously suffered shrinkage [29]).

That the Cholesky metric should never be used in the standard DTI setting due to its lack of orthogonal invariance is illustrated in Section 4.3.3 below, and the above experiments also highlighted other potential issues with this metric. The log-Euclidean and affine-invariant Riemannian metrics generally give similar geodesic paths [2], and have a small advantage in Experiment I. The Procrustes size-and-shape and Euclidean root metrics offer similar interpolated tensors, and they are preferable in Experiment II.

4.3.2 Rank deficiency and extrapolation

The log-Euclidean and affine-invariant Riemannian metrics, as well as the root Stein Divergence (20) all give infinite distances between any positive definite tensor and any degenerate tensor. Furthermore, such metrics are undefined if both the tensors are rank-deficient. Therefore, when working with rank-deficient tensors, the root-Euclidean and Procrustes metrics may be advantageous. Generally, $d_H(\mathbf{D}_1, \mathbf{D}_2)$ and $d_S(\mathbf{D}_1, \mathbf{D}_2)$ give similar results, however, their (relative) disagreement tends to its supremum ($d_H/d_S = \sqrt{2}$) when the tensors become indistinguishable and simultaneously approach the rank 1 subspace; see also Theorem 3.1 above.

The potential advantage of the Procrustes size-and-shape metric when extrapolating towards rank-deficient tensors has been noted in [31] in the infinite dimensional case of covariance operators. Similar to the geodesic through two covariance operators defined in [31], geodesics through two diffusion tensors with the Euclidean root metric and the Procrustes metric are given by $\mathbf{D}_H(w) = \mathbf{D}_H(1 - w, w)$ and $\mathbf{D}_S(w) = \mathbf{D}_S(1 - w, w)$ from the respective rows of Table 1, where the weights are no longer constrained to be non-negative, i.e. $w \in \mathbb{R}$.

The following example shows geodesic paths through two planar (i.e. rank 2) diffusion

tensors, using d_H and d_S , and illustrating Proposition 3.2. The reference tensors are defined as follows

$$\mathbf{D}(0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{D}(1) = \mathbf{V} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}^T, \text{ with } \mathbf{V} = \begin{pmatrix} -0.5441 & 0.7040 & 0.4565 \\ 0.8391 & 0.4565 & 0.2960 \\ 0 & -0.5440 & 0.8391 \end{pmatrix}.$$

The angle between the principal planes of these tensors is about 33° , and the (2D) shapes of the tensors are also different: $\text{FA}(\mathbf{D}(0)) = 1/\sqrt{2} \approx 0.7071$ (2D isotropy), and $\text{FA}(\mathbf{D}(1)) = 0.7746$. Figure 6 shows $d_H(\mathbf{D}(0), \mathbf{D}_H(w))$ and $d_S(\mathbf{D}(0), \mathbf{D}_H(w))$ against $w \in [0, 5]$; the $d_H/\sqrt{2} \leq d_S \leq d_H$ bounds are also illustrated. The difference between d_H and d_S along the Euclidean root geodesic is hardly noticeable for $w > 3$, whereas in the case of the Procrustes size-and-shape geodesic the distances disagree more noticeably over the entire $(0, 5]$ range. Figure 7 shows the minimum eigenvalue λ_1 , determinant,

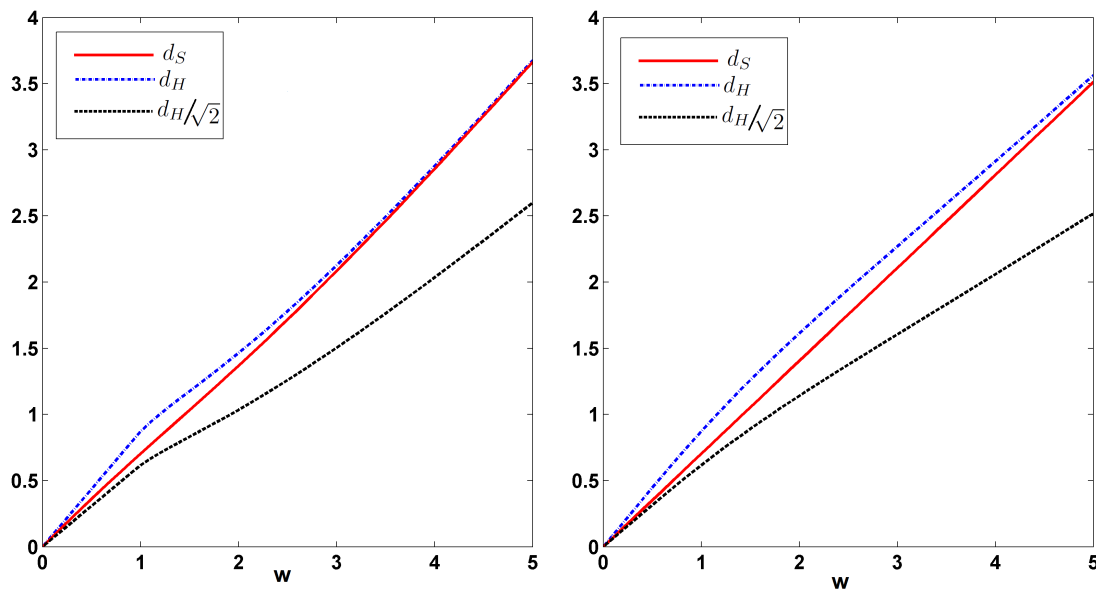


Figure 6. The Procrustes size-and-shape d_S and the Euclidean root d_H distances from $\mathbf{D}(0)$ to the averaged (interpolated or extrapolated) tensor $\mathbf{D}(w)$ along the Euclidean root geodesic (left) and the Procrustes size-and-shape geodesic (right).

and FA of $\mathbf{D}_H(w)$ and $\mathbf{D}_S(w)$ versus $w \in (0, 5)$. The Euclidean root geodesic yields positive, albeit very small, λ_1 's already in the course of interpolation ($w \in (0, 1)$). This would imply emergence of 3D diffusion from 2D diffusion, which may be hard to justify from the viewpoint of physics. The minimal eigenvalue grows very rapidly in the course of extrapolation ($w > 1$). A similar behaviour is observed with the determinant $|\mathbf{D}_H(w)|$, even though it appears to be nearly zero for $w \in (1, 2]$, before a steep take-off over $w > 3$. As explained in Proposition 3.2, the Procrustes geodesic does not suffer from this problem, i.e., *it does not require full rank (3D) tensors to either interpolate or extrapolate through rank 2 tensors, which makes its action more realistic*. Figure 8 shows the product $\lambda_1(w) \times \lambda_2(w)$ of the two largest eigenvalues, which exhibits the growth of the effective volume (area) of $\mathbf{D}_S(w)$ in its principal non-degenerate plane; the Euclidean extrapolation in this Figure appears misleadingly regular while in reality it goes out of space $\Omega_{\geq 0}(3)$ ($\lambda_3(w) < 0$ for $w > 1$).

FA along the root-Euclidean path changes significantly, especially in the course of extrapolation ($w > 1$) where it falls rapidly (the full rank extrapolation tensor loses

its shape). FA of the Procrustes interpolations and extrapolations varies little, always remaining within the $[1/\sqrt{2}, 1)$ range of the planar diffusion. For larger w values, i.e. $w > 5$, $\text{FA}(\mathbf{D}_H(w))$ may also be observed (not shown in the figure) to behave non-monotonically, yet never rising back to the $1/\sqrt{2}$ level of the planar diffusion. In summary,

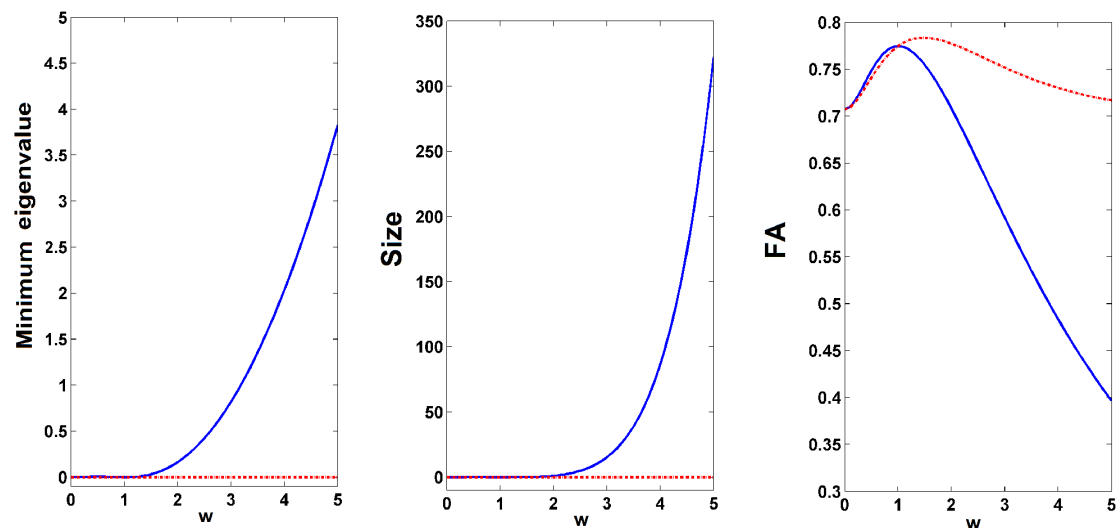


Figure 7. Minimum eigenvalue λ_3 (left), determinant (middle), and FA (right) of $\mathbf{D}_S(w)$ (dotted line) and $\mathbf{D}_H(w)$ (solid line) versus the (interpolation $[0, 1]$ and extrapolation $(1, 5]$) weight w .

both $\mathbf{D}_H(w)$ and $\mathbf{D}_S(w)$ expand in the course of extrapolation beyond \mathbf{D}_2 ($w > 1$), eventually becoming increasingly isotropic. However, $\mathbf{D}_S(w)$ confines its expansion to its principal plane ($|\mathbf{D}_S(w)| = 0$ while $\lambda_{S1}(w) \times \lambda_{S2}(w)$ grows). $\mathbf{D}_H(w)$, on the other hand, does not have this ($\lambda_3(w) = 0$) constraint and expands in the entire 3D space.

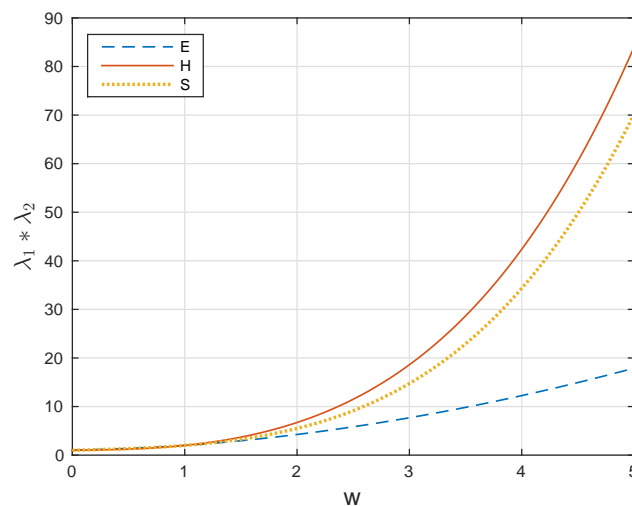


Figure 8. Product of the two largest eigenvalues of $\mathbf{D}_S(w)$ (dotted line), $\mathbf{D}_H(w)$ (solid line), and $\mathbf{D}_E(w)$ (dash line) versus the (interpolation $[0, 1]$ and extrapolation $(1, 5]$) weight w .

Hence, although the choice between which metrics to use is application dependent and generally there may be not much difference between the log-Euclidean and affine-invariant Riemannian, Procrustes and root-Euclidean metrics, it appears that the Procrustes size-and-shape metric is a good choice and is widely applicable, even in rank-deficient or near rank-deficient cases.

4.3.3 Interpolation under simultaneous rotation

It is known [11] that all of the metrics considered in this paper, except d_C , are invariant to orthogonal transformations of the underlying 3D space, i.e. $d(\mathbf{U}\mathbf{D}_1\mathbf{U}^T, \mathbf{U}\mathbf{D}_2\mathbf{U}^T) = d(\mathbf{D}_1, \mathbf{D}_2)$ for all $\mathbf{U} \in O(3)$. This is very important in practice as any method for diffusion tensor processing must be independent of the choice of the reference frame [2]. Below we include a set of simple orthogonal change of frame experiments to demonstrate this issue.

Consider two mutually orthogonal tensors $\mathbf{D}(0)$ and $\mathbf{D}(1)$ with the same eigenvalues (40, 2, 1). Since the largest eigenvalue is much greater than the other two, $\mathbf{D}(0)$ and $\mathbf{D}(1)$ are nearly linear (and their principal eigenvectors are orthogonal); the corresponding ellipsoids are placed at the left and right top corners, respectively, of each of the six plots of Figure 9. Next, we simulate five orthogonal changes of the reference frame by rotating it by 15° , 30° , 45° , 60° , and 75° in the plane of view, so that the corresponding representations of $\mathbf{U}\mathbf{D}(w)\mathbf{U}^T$ with $w = 0$ and $w = 1$ are placed (top down) in the left and right most columns (red), respectively, of each of the six plots of Figure 9. Each plot corresponds to one of the aforementioned metrics, and samples along the geodesic paths (rows) for each reference frame and each metric are defined according to Equation (23) with $N = 8$.

It can be seen, e.g. by looking at the middle ellipsoids of the top and bottom rows of the boxed plot, that the Cholesky geodesic paths are not invariant under the orthogonal transformations of the reference frame. Namely, the shape and size (determinant) of the interpolated tensor differ across the transformed geodesic paths. Although if viewed individually outside the change of frame context each of these Cholesky paths appears to be very reasonable, the lack of invariance renders the method unreliable in practice. The other metrics, as expected, give invariant geodesic paths, which are also included in this figure for the sake of completeness of the demonstration.

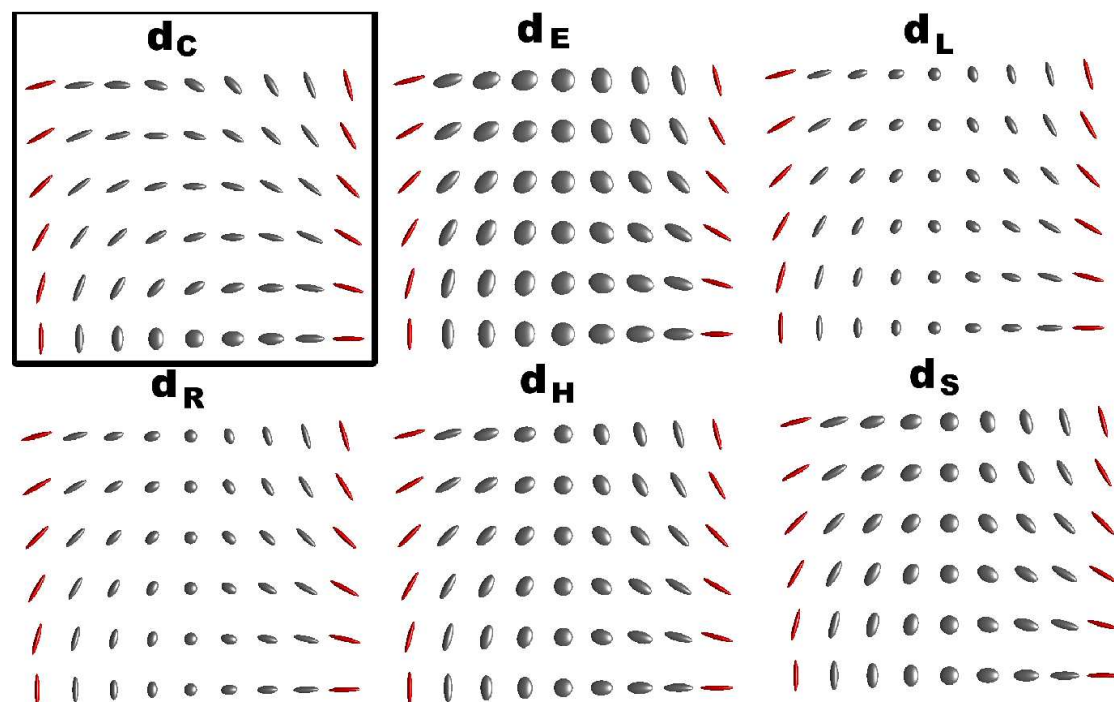


Figure 9. Geodesic paths under simultaneous rotation of reference tensors (red) with different metrics. The boxed example demonstrates the violation of the invariance to orthogonal changes of coordinates in the case of the Cholesky method.

4.3.4 Interpolation of four tensors

The next experiment is to interpolate in a 2-dimensional region. Specifically, we mesh the region with a regular 5×5 grid, and place four synthetic tensors $\mathbf{D}_{1,1}$, $\mathbf{D}_{1,5}$, $\mathbf{D}_{5,1}$, $\mathbf{D}_{5,5}$ at the corners of the grid. At each of the 21 remaining nodes (i, j) , a weighted average of these four tensors is computed according to Equation (6) and using the exponential weight function (22) with $A = 2$ and $B = 0.01$ (Section 4.2) and the Euclidean distances between the nodes of the grid in the definition of the weight function. Figure 10 shows results of the interpolations using the Euclidean, Procrustes and log-Euclidean metrics. At the top left corner, the diffusion tensor is isotropic (spherical) and of a small volume. Tensors at the other three corners are anisotropic with different volumes and orientations. The interpolated tensors from the Euclidean results have large volumes. Interpreting the Euclidean interpolation is also problematic in terms of the orientation and shape. The log-Euclidean interpolation tends to produce significantly smaller tensors. The Procrustes metric appears to provide a reasonable interpolation of the tensor volume, orientation and anisotropy.

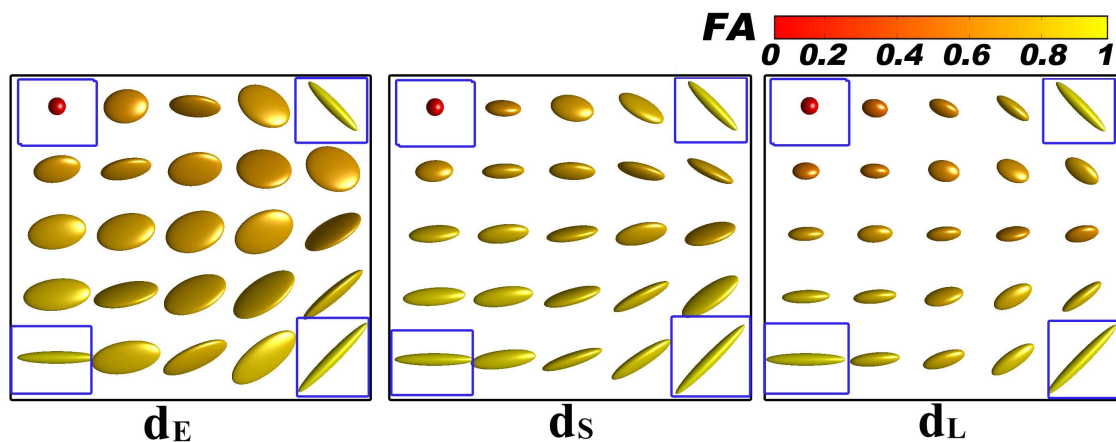


Figure 10. Interpolation of four tensors at the corners of a grid. Left: Euclidean interpolation (d_E). Middle: Procrustes interpolation (d_S). Right: log-Euclidean interpolation (d_L). FA is used for the colouring of the tensors.

5. Tensor regularisation

In the following, we develop a weighted regularisation model which incorporates the smoothness of local diffusion and regularisation by imitation of a prescribed diffusion behaviour. Specifically, the Procrustes size-and-shape metric is adapted in the regularisation model.

5.1 Weighted regularisation model

Consider a sample of diffusion tensors $\mathbf{D}_1, \dots, \mathbf{D}_N$ from a noisy tensor field containing N voxels with coordinates $x_i \in \mathbb{Z}^3$, $i = 1, \dots, N$. Now suppose we wish to regularise this tensor field. We propose a weighted regularisation model, which is defined by minimising the following function with respect to $\Sigma_j \in \Omega_{>0}(3)$, $j = 1, \dots, N$,

$$f(\Sigma_1, \Sigma_2, \dots, \Sigma_N) = \sum_{j=1}^N \sum_{i=1}^N w_{ij} d_1^p(\mathbf{D}_i, \Sigma_j) + \lambda \sum_{j=1}^N d_2^q(\mathbf{\Pi}, \Sigma_j)$$

$$= \sum_{j=1}^N \left[\sum_{i=1}^N w_{ij} d_1^p(\mathbf{D}_i, \boldsymbol{\Sigma}_j) + \lambda d_2^q(\boldsymbol{\Pi}, \boldsymbol{\Sigma}_j) \right]. \quad (25)$$

Here the weights w_{ij} can be again obtained via some decreasing function of the Euclidean distance between x_i and x_j (see Section 4.1), $p, q \geq 0$, and λ is a regularisation parameter ($\lambda > 0$), and $\boldsymbol{\Pi}$ is a reference tensor, representing the prescribed diffusion behaviour. Users can either define their own reference tensor with the expected diffusion profile or choose a representative tensor from the tensor field as a reference tensor. For a simple example, by employing a multiple of the identity tensor as the reference tensor we can regularise the tensor field to be both more isotropic and less variable in intensity. Note that d_1 and d_2 are general, i.e. can be non-Euclidean and need not be the same. When appropriate, $\Omega_{>0}(\mathfrak{3})$ can be replaced by $\Omega_{\geq 0}(\mathfrak{3})$.

The proposed regularisation model (25) simultaneously smooths each tensor \mathbf{D}_i , $i = 1, \dots, N$, by averaging it with its neighbours (endogenous regularisation) and also makes \mathbf{D}_i where $i = 1, \dots, N$ imitate the reference tensor (exogenous regularisation). The flexibility of having a reference tensor can be exploited to introduce additional information about the expected diffusion profile in the given region, or to highlight, and eventually segment, local structures. Allowing the end user to modify $\boldsymbol{\Pi}$, e.g. by varying its spectrum and the Euler angles of the eigenvectors, in real time in a visualisation session might help pick up otherwise unnoticed details.

It is clear that each $\hat{\boldsymbol{\Sigma}}_j$, $j = 1, 2, \dots, N$, of the solution to the weighted regularisation model (25) must minimise

$$\sum_{i=1}^N w_{ij} d_1^p(\mathbf{D}_i, \boldsymbol{\Sigma}) + \lambda d_2^q(\boldsymbol{\Pi}, \boldsymbol{\Sigma}).$$

Therefore,

$$\hat{\boldsymbol{\Sigma}}_j = \arg \inf_{\boldsymbol{\Sigma} \in \Omega_{>0}(\mathfrak{3})} \sum_{i=1}^N w_{ij} d_1^p(\mathbf{D}_i, \boldsymbol{\Sigma}) + \lambda' d_2^q(\boldsymbol{\Pi}, \boldsymbol{\Sigma}), \quad (26)$$

where $w_i = w_{ij}/(w_{1j} + w_{2j} + \dots + w_{Nj} + \lambda)$ and $\lambda' = \lambda/(w_{1j} + w_{2j} + \dots + w_{Nj} + \lambda)$.

Consider first the case where $d_1 = d_2$ is a Euclidean-based metric d_g (Equation (10)).

In this case, Equation (11) for $N = 2$ generalises and gives a unique solution $\hat{\boldsymbol{\Sigma}}_j = g^{-1}(\hat{\mathbf{Q}}_j)$ to this model, where

$$\hat{\mathbf{Q}}_j = \frac{\sum_{i=1}^N w_{ij} g(\mathbf{D}_i) + \lambda g(\boldsymbol{\Pi})}{\sum_{i=1}^N w_{ij} + \lambda}. \quad (27)$$

For example, if $g(\mathbf{D}) = \frac{1}{a} \mathbf{D}^a$, then $\hat{\boldsymbol{\Sigma}}_j = a \hat{\mathbf{Q}}_j^{1/a}$. If $g(\mathbf{D}) = \log(\mathbf{D})$, then $\hat{\boldsymbol{\Sigma}}_j = \exp\{\hat{\mathbf{Q}}_j\}$. Thus, using identical Euclidean-based metrics for the weighted regularisation is straightforward. Note that the Cholesky method also admits the same closed form solution (Equation (27)) even though g in that case is not defined in terms of the special decomposition.

Note also that $p = 2$ and $q = 0$ gives the weighted Fréchet mean, if $q = 0$ we have a type of M-estimator [12, 21]. If $p = 1$ and $q = 0$ we have an L_1 -type regularisation where we have the geometric median [15] instead of the mean. If $p = 2$ and $q = 2$ this gives a type of ridge-regression [19, (3.43)]. Indeed, while the penalty term in ridge regression represents the square of the Euclidean distance to the origin, our generalized penalty term uses the

square of the general distance d_2 to the general reference tensor $\mathbf{\Pi}$. Note also that our formulation is primarily concerned with the multiple output case [19, Section 3.7], i.e. N would usually be more than one and a single regularisation parameter λ would be used with each of the N penalty terms. If $p = 2$ and $q = 1$ then we have a type of least absolute shrinkage and selection operator (LASSO) [19, (3.52)], i.e. replacing the L_1 distance to the origin by the general d_2 distance to the general reference tensor $\mathbf{\Pi}$. Similarly to how ridge regression and LASSO can be viewed as Bayes estimates [19, p.72], we can also embed our generalized tensor regularisation in the Bayesian framework. In the cases of the Euclidean-based metrics the connection with ridge regression and LASSO is particularly appropriate, and so, for example, the least angle regression (LARS) algorithm can be used for fast implementation [13] of the LASSO. At the same time, the flexibility of the choice of d_2 can also be seen as a generalization of the grouped LASSO method [19, Section 3.8.4].

Note that non-Euclidean distances generally require more computations. In particular, calculation of the log-Euclidean, affine-invariant Riemannian and Euclidean root and other Euclidean based distances d_g requires the spectral decomposition. The Cholesky distance requires the Cholesky decomposition. When computing weighted averages (Equation (6)), any Euclidean-based averaging as well as the Cholesky method have the closed form solution (Equation (27)) and therefore all such metrics require the same amount of computations in terms of g and g^{-1} . The affine-invariant Riemannian averaging and Procrustes-based averaging can be significantly more costly as they require iterative numerical optimisations, see [30] and [44] respectively.

5.2 Weighted Procrustes regularisation

Now consider the special case that $(p, q) = (2, 2)$ and $d_1 = d_2 = d_S$ where d_S is the Procrustes size-and-shape metric (7). Then, the weighted Procrustes regularisation objective function is given by

$$f_{WPS}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \dots, \mathbf{\Sigma}_N) = \sum_{j=1}^N \sum_{i=1}^N w_{ij} d_S^2(\mathbf{D}_i, \mathbf{\Sigma}_j) + \lambda \sum_{j=1}^N d_S^2(\mathbf{\Pi}, \mathbf{\Sigma}_j)$$

Using Equation (26), the solution is given by

$$\widehat{\mathbf{\Sigma}}_j = \arg \min_{\mathbf{\Sigma} \in \Omega_{\geq 0}(3)} \sum_{i=1}^N w_i d_S^2(\mathbf{D}_i, \mathbf{\Sigma}_j) + \lambda' d_S^2(\mathbf{\Pi}, \mathbf{\Sigma}_j) = \arg \min_{\mathbf{\Sigma} \in \Omega_{\geq 0}(3)} \sum_{i=1}^{N+1} w_i d_S^2(\mathbf{D}_i, \mathbf{\Sigma}),$$

where $\mathbf{D}_{N+1} = \mathbf{\Pi}$ and $w_{N+1} = \lambda'$, $j = 1, 2, \dots, N$, and w_i 's and λ' also depend on j . As pointed out in Section 3.1, the solution $(\widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_2, \dots, \widehat{\mathbf{\Sigma}}_N)$ can be computed using the Weighted Generalised Procrustes Algorithm [44], which is also used in the following experiments.

6. Experiments using real human brain data

The diffusion-weighted MR data are from a healthy human brain. The data were acquired using a 3T Phillips Achieva clinical imaging system, using a spin-echo echo-planar imaging sequence with the Uniform 32 DTI diffusion weighting gradient scheme [33] applied with a weighting factor $b = 1000$ s/mm². In total, 52 interleaved contiguous transaxial slices were acquired throughout the subject's head in a 112x112 matrix (interpolated to

224x224) with an acquisition voxel size of $1 \times 1 \times 2 \text{mm}^3$. The diffusion tensor field has been computed using a Bayesian estimation framework [42].

6.1 Application to tensor regularisation and segmentation

The weighted Procrustes regularisation method proposed in Section 5.2 above is applied to help with the segmentation of two prominent white matter structures, the corpus callosum and cingulum, using real DTI data. The role of the regularisation term here may be reminiscent of template matching in computer vision. Figure 11 illustrates the process and the results. The original FA map of the region of interest is shown in pane a. Panes b and d highlight the corpus callosum with $\lambda = 0.6$ and 1.5 respectively. The reference tensor to be matched is taken to be $\mathbf{\Pi}_1 = [0.0022, 0, 0; 0, 0.0004, 0; 0, 0, 0.0004]^1$, which is highly anisotropic and oriented along the main diffusion direction in the corpus callosum. The weights are set again with $A = 2$ and $B = 0.01$ in the weighting function (22). As λ increases, the influence of the template, or probing, tensor becomes stronger and so the FA map becomes brighter since the probing tensor is strongly anisotropic here. While this also results in the overall loss of contrast, the contrast between the brightest corpus callosum and the now relatively homogeneous and darker background should still be sufficient for a simple thresholding to finish the task. Panes c and e highlight the cingulum with the reference tensor $\mathbf{\Pi}_2 = [0.0003, 0, 0; 0, 0.0015, 0; 0, 0, 0.0002]$, which is also anisotropic and oriented in the direction of diffusion in the cingulum. When $\lambda = 1.5$, the cingulum appears to be easily segmentable from the background.

6.2 Comparing metrics for smoothing and interpolation

Next, we apply the weighted tensor averaging with the Euclidean, log-Euclidean, Procrustes and Euclidean root metrics to interpolate and smooth a tensor field within a region of interest (ROI). Figure 12 shows FA maps of three adjacent axial slices from the original diffusion tensor data and zoomed 36×71 insets. This $36 \times 71 \times 3$ ROI contains a part of the cingulum and a part of the corpus callosum.

As before, we compare the Euclidean, Procrustes, log-Euclidean and Euclidean root approaches. First we interpolate the $36 \times 71 \times 3$ region of the original tensor field at two equally spaced locations between each pair of original tensors. Each interpolation tensor is then the weighted average of its six nearest (i.e. from the first-order 3D neighbourhood) original tensors. As discussed in Section 4.1, we use $A = 2$ and $B = 0.01$ in the weight function (22).

The resulting $106 \times 211 \times 7$ field is then smoothed by applying the $3 \times 3 \times 3$ (second order 3D neighbourhood) moving average with equal weights ($w = 1/27$). This simple combination of interpolation and smoothing was chosen as a compromise between simplicity and computational efficiency of the procedure on the one hand, and perceptual appeal of the resulting images, on the other.

We then also calculate at each voxel of the resulting field arithmetic (MD) and geometric (GMD, i.e. $\sqrt[3]{|\mathbf{D}|}$) mean diffusivity, anisotropy measures FA and PA, and principal directions.

Table 3 summarises the mean GMD, MD, FA, and PA over the tensor field processed with the four methods. The Euclidean processing results in notably larger tensors, whereas the log-Euclidean approach gives the smallest tensors, with the Procrustes and root-Euclidean results in between. Although Inequality (13) and Proposition 3.5 concern

¹The reference tensor is already on the real scale that is determined by the b-value of 1000 s/mm^2 .

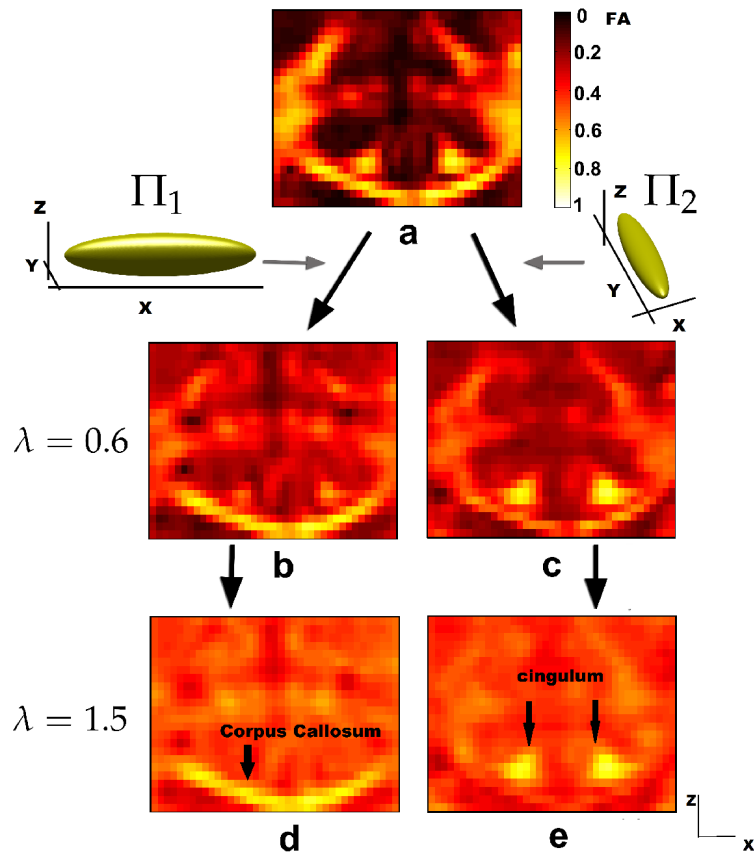


Figure 11. Tensor field preprocessing with a view towards segmentation of the corpus callosum (b,d) and cingulum (c,e).

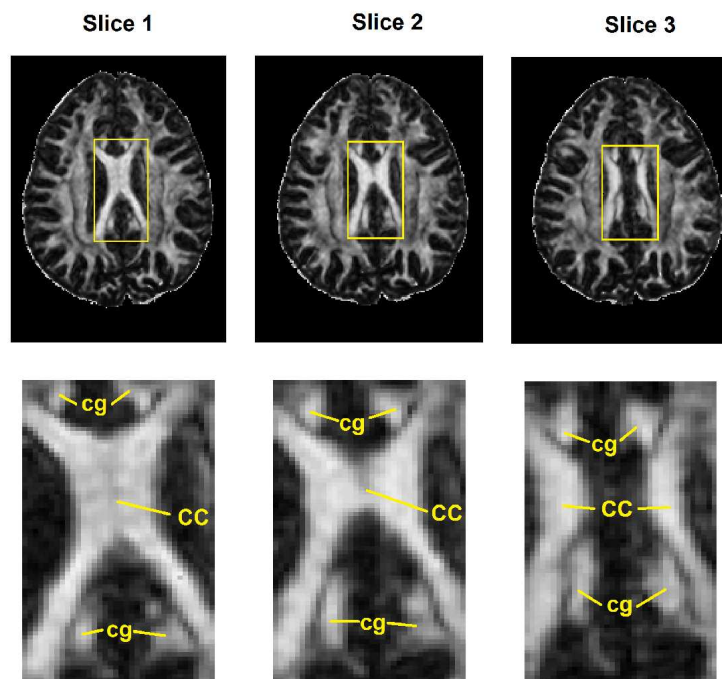


Figure 12. FA maps of 224×224 original axial slices (top) and corresponding 36×71 zoomed insets (bottom) with white matter tracts from the cingulum (cg) and corpus callosum (cc).

the interpolation of two tensors only and the present composition of interpolation and smoothing involves two distinct sets of locations with more than two tensors each, the same orderings are also observed at most locations and for the ROI averages. It is also expected [2] that the mean FA of the log-Euclidean method is higher than that of the Euclidean method. PA, on the other hand, is seen to be able to reverse this comparison in our processing, although by a small margin. The Procrustes and Euclidean root methods have similar averages, as expected.

Table 3. Averaged GMD, MD, FA, and PA under the different processing methods.

	Euclidean	log-Euclidean	Euclidean Root	Procrustes
GMD $\times 10^{-4}$	9.4887	<i>8.7383</i>	9.0704	9.0687
MD $\times 10^{-4}$	10.2081	<i>9.4262</i>	9.7217	9.7223
FA	0.3834	0.3868	<i>0.3749</i>	0.3755
PA	0.2145	0.2143	<i>0.2065</i>	0.2069

We also compare local variation of each of the four procedures. Specifically, for each of the above tensor characteristics, and also for the tensor orientation, we calculate within each (overlapping) $3 \times 3 \times 3$ neighbourhood with centre at (x, y, z) the root mean square deviation δ from the value at the centre. Thus, for example, at each interior voxel (x, y, z) this variation measure for MD is given by

$$\delta_{\text{MD}}(x, y, z) = \sqrt{\frac{1}{26} \sum_{i=-1}^1 \sum_{j=-1}^1 \sum_{k=-1}^1 (\text{MD}(x, y, z) - \text{MD}(x+i, y+j, z+k))^2},$$

and for the tensor orientation, the differences are replaced by the angles ϕ (in degrees) between each tensor in the neighbourhood and the tensor at the centre. The (density) histograms of the logarithms of the resulting angular variations, or dispersions, are shown in Figure 13. Note the distinct separations of the major modes in each of the three non-Euclidean cases. We believe that the larger mode, separated from the smaller one at about $1.25 \approx \log(3.5^\circ)$, is due to the smoothing across heterogeneous fibre structures, which is inevitable given the simplicity of our non-adaptive approach. The Euclidean method does not allow us to see this separation.

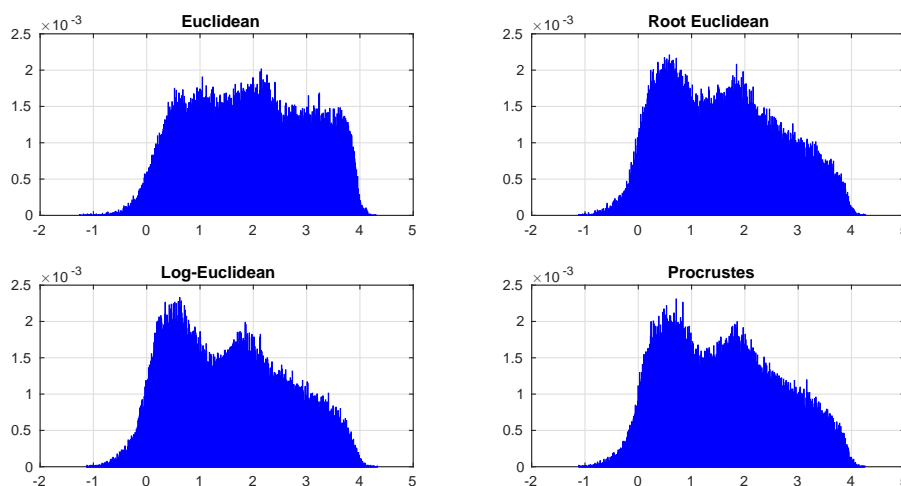


Figure 13. Histograms of the logarithms of the local angular dispersions.

Table 4 averages local variations of each characteristic and for each method over ROI. Thus, the Euclidean method consistently gives a higher mean. The least mean variations in orientation and size are observed with the log-Euclidean processing, although the respective results from the Procrustes and square-root-Euclidean methods results are very similar. At the same time, the Procrustes and square-root-Euclidean methods result in lower mean variations of anisotropy. Overall, the three non-Euclidean methods appear to produce more spatially homogeneous results as compared to the Euclidean method.

Table 4. Local variations $\delta(x, y, z)$ averaged over ROI.

	Euclidean	log-Euclidean	Euclidean Root	Procrustes
$\bar{\delta}_\phi$	12.0113°	<i>8.3890°</i>	8.5486°	8.5360°
$\bar{\delta}_{\text{GMD}} \times 10^{-4}$	1.2817	<i>0.8051</i>	0.8301	0.8303
$\bar{\delta}_{\text{MD}} \times 10^{-4}$	1.1464	<i>0.7257</i>	0.7458	0.7457
$\bar{\delta}_{\text{FA}}$	0.0767	0.0539	0.0536	<i>0.0535</i>
$\bar{\delta}_{\text{PA}}$	0.0483	0.0341	<i>0.0334</i>	<i>0.0334</i>

Figure 14 shows colour-coded orientation maps (with brightness scaled by FA) of three adjacent original diffusion tensor images and interpolations between those. Recall that the orientation of the diffusion tensor models the main fibre orientation at a given location. Some differences in fibre orientations are pointed out. For example, in column 1 the red region (pointed out by grey arrows) is the left lower part of the corpus callosum. The Euclidean approach has more horizontally oriented tensors in this region compared with the other methods. In column 6, we have the feature that the corpus callosum appears disconnected (clearly separated from the middle, as pointed out by yellow arrows) in the Euclidean map. The non-Euclidean methods, on the other hand, deliver a seemingly smoother transition from the original Slice 2 with the corpus callosum appearing connected to Slice 3 with the corpus callosum appearing disconnected.

Figure 15 shows angular deviations between tensor orientations obtained using different approaches, which are from the slice of smoothed interpolants in column 2 of Figure 14. For example, ϕ_{EL} is the smaller angle between the orientations of tensors processed with the Euclidean and log-Euclidean methods. It is clear from the ϕ_{EL} , ϕ_{EH} and ϕ_{ES} maps that the Euclidean tensor tends to be oriented notably differently from the non-Euclidean ones. Mainly these large deviations occur at the gap between the corpus callosum and the cingulum. The Procrustes and the Euclidean root approaches are again similar.

Figure 16 shows greyscale diffusion tensor volume ($|\mathbf{D}(x, y, z)|$) maps. The Euclidean approach gives notably larger tensors. For example, in columns 1 and 2 the bright region marked by yellow arrows from the Euclidean approach is significantly larger than in the respective maps from the non-Euclidean methods. Also in column 1, the bright region marked by a green arrow is also larger in the the Euclidean map.

Figure 17 contrasts the approaches by the tensor volume using the more convenient scale of the decadic logarithm. For example, $\log_{10}(|\mathbf{D}_E|/|\mathbf{D}_L|)$ contrasts the Euclidean and the log-Euclidean results. The Euclidean approach gives significantly larger tensors at the boundary of the corpus callosum. In particular, the contrast between the Euclidean and the log-Euclidean results peaks at at the arches of the corpus callosum. The root-Euclidean volumes are only negligibly larger than the Procrustes volumes, which is expected.

Figure 18 shows MD maps from the different approaches. The log-Euclidean approach generally tends to have lower MD values, and especially at regions pointed out by the green and yellow arrows. From Figure 19, it is clear that MD values from the Euclidean approach are generally higher and in particular at the boundary of the corpus callosum (red and yellow regions). MD values from the Euclidean method are slightly lower than

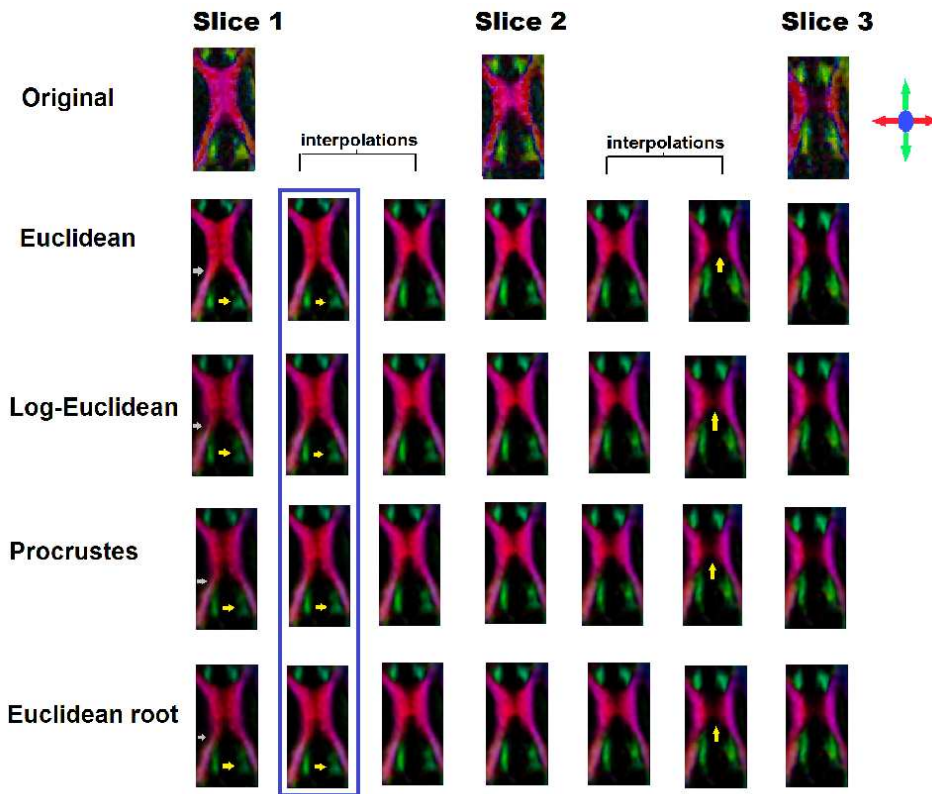


Figure 14. Orientation maps. First row: orientation maps from three adjacent original diffusion tensor images. Rows 2 to 5 are orientation maps from the processed diffusion tensor images using the Euclidean, log-Euclidean, Procrustes and Euclidean root methods. Columns 2, 3, 5 and 6 are from interpolation images. Differences are pointed out using arrows. Comparisons between orientation maps in the blue box are displayed in Figure 15.

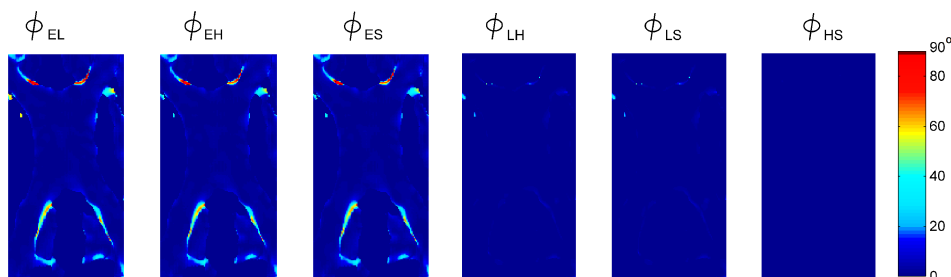


Figure 15. Deviations in orientation. ϕ_{EL} , ϕ_{EH} , ϕ_{ES} , ϕ_{LH} , ϕ_{LS} and ϕ_{HS} are the angular deviations between the Euclidean and log-Euclidean, the Euclidean and Euclidean root, the Euclidean and Procrustes, the log-Euclidean and Euclidean root, the log-Euclidean and the Procrustes, the Euclidean root and Procrustes results, respectively.

the others in the middle part of the corpus callosum (dark blue). The log-Euclidean tensors have notably lower MD values at the boundary of the corpus callosum than both the Euclidean root and Procrustes tensors. The Procrustes and Euclidean root approaches are very similar in MD, although the former do tend to be higher than the latter as suggested by Proposition 3.5.

Figure 20 shows PA maps from the four methods. At the boundary of the corpus callosum in the first and second column in the figure, the log-Euclidean, Procrustes and Euclidean root approaches have higher PA values than the Euclidean method, as the latter is known to lower anisotropy in interpolation. For the same reason, in the sixth and seventh columns the corpus callosum appears totally disconnected in the middle of the Euclidean maps. The other maps show some activity there. The PA contrast maps

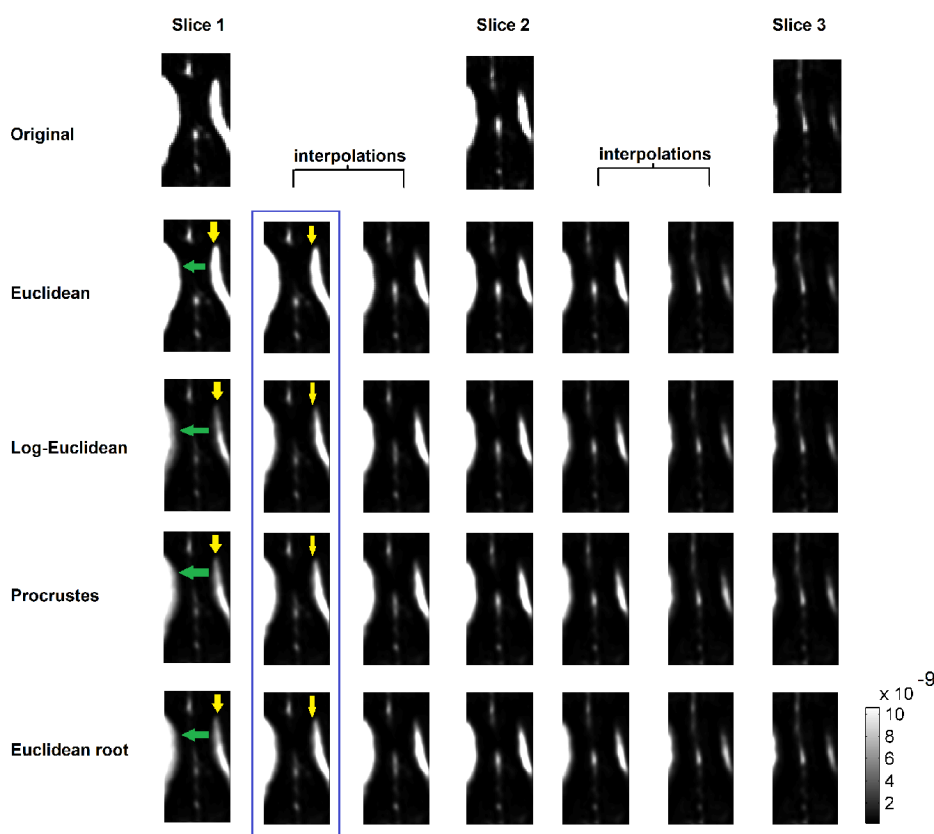


Figure 16. Volume maps. First row: The three original adjacent slices. Rows 2 to 5 are from the processed diffusion tensor images using the Euclidean, log-Euclidean, Procrustes and Euclidean root methods. Columns 2, 3, 5 and 6 are the interpolated slices. Differences are pointed out using arrows. Contrasts between the boxed (in blue) results processed by the different methods are further displayed in Figure 17.

of Figure 21 show how at the lateral boundaries of the corpus callosum (the blue arches) the Euclidean tensors tend to be significantly more isotropic than the non-Euclidean ones. This comparison reverses at the top and bottom boundaries (red horseshoes), in particular closer to the gap between the corpus callosum and the cingulum. The log-Euclidean approach has higher PA values at the left and right boundaries of the corpus callosum. The Procrustes and the Euclidean root results are similar as usual. It may be possible to gain more insight into these comparisons by using the alternative approach of the absolute value tensor of differences, as proposed in [2].

To summarise, the Euclidean method tends to give notably larger tensors and more variable (locally) processed fields. The non-Euclidean methods are similar regarding the size and orientation. There are some small differences between all of the non-Euclidean methods, and the Procrustes and Euclidean root methods are very similar overall in these examples. Comparisons of all the four methods are in line with the theoretical rankings of Inequality (13) and Proposition 3.5.

7. Conclusion and discussion

In this paper, we have considered non-Euclidean statistical methods for diffusion tensor regularisation, interpolation and visualisation. A family of anisotropy indices has been proposed using the scale invariant power-Euclidean metric, which are useful for visualisation. We have also considered the weighted tensor averaging using non-Euclidean metrics,

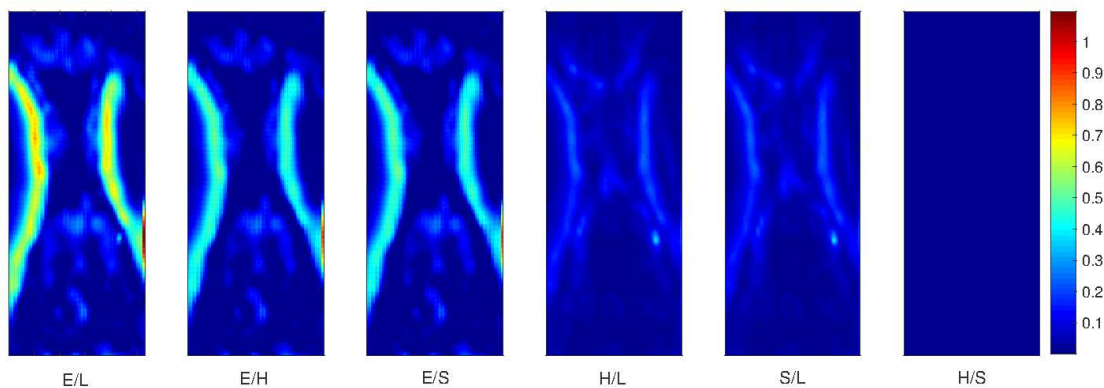


Figure 17. Comparisons of tensor volumes on the decadic logarithmic scale. As expected, the highest contrast is observed from $\log_{10}(|\mathbf{D}_E|) - \log_{10}(|\mathbf{D}_L|)$ of the Euclidean and log-Euclidean Riemannian methods. Contrasts $\log_{10}(|\mathbf{D}_H|) - \log_{10}(|\mathbf{D}_L|)$ and $\log_{10}(|\mathbf{D}_S|) - \log_{10}(|\mathbf{D}_L|)$ between the square root method and the log-Euclidean method, and between the Procrustes method and the log-Euclidean method, respectively, are notably lower. While the square root tensors are bigger than or equal to the Procrustes tensors, their differences $\log_{10}(|\mathbf{D}_H|) - \log_{10}(|\mathbf{D}_S|)$ appear negligible on the common scale.

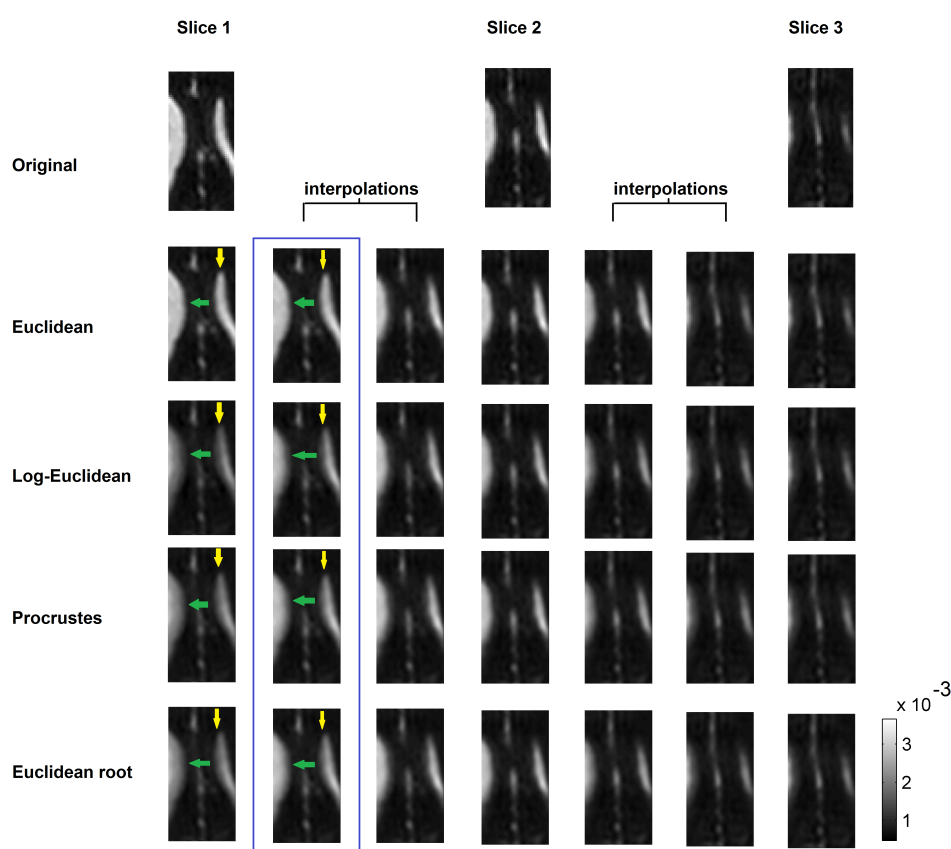


Figure 18. MD maps. First row: MD maps from three adjacent original diffusion tensor images. Rows 2 to 5 are MD maps from the processed diffusion tensor images using the Euclidean, log-Euclidean, Procrustes and Euclidean root methods. Columns 2, 3, 5 and 6 correspond to interpolated slices. Differences are pointed out using arrows.

including the Procrustes size-and-shape, log-Euclidean and affine-invariant Riemannian, Cholesky and Euclidean root metrics. We have also discussed how to use these non-Euclidean methods for diffusion tensor smoothing and interpolation. Then a weighted regularisation model has been developed, which combines smoothing and regularisation by matching a prescribed diffusion behaviour. Finally, we have compared several metrics

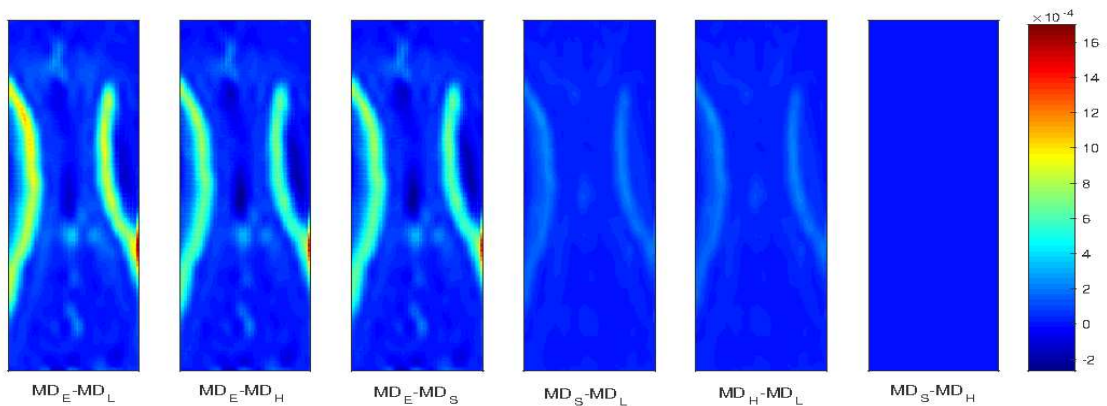


Figure 19. MD contrasts. Although some negative values do arise, the vast majority of the contrasts are consistent with Proposition 3.5. As in the case of the volume contrasts, the Procrustes and root-Euclidean results are nearly the same, hence their differences are negligible on the common scale (even though the Procrustes MD values do tend to be higher than the corresponding root-Euclidean MD values).

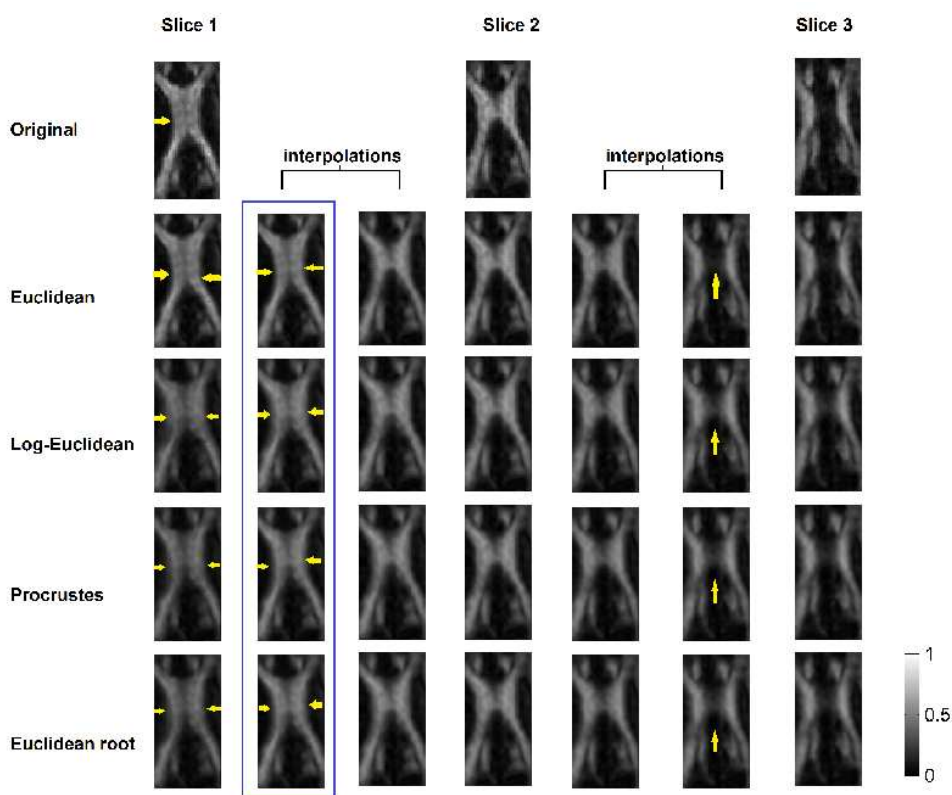


Figure 20. PA maps.

and the corresponding weighted averages, and have established several key analytic results concerning properties of these objects. Those have also been illustrated in simulation studies and on a real dataset from a human brain diffusion-weighted MRI scan.

We have presented several applications of non-Euclidean weighted averaging methods for diffusion tensor processing. How to process images with a level of optimality is a central question in many computer vision applications. For diffusion tensor smoothing, the main difficulty lies in the ability to predict how much smoothing needs to be applied at various locations. This is generally difficult, even when the brain atlas is available for guidance. It is therefore interesting to investigate algorithms which could modify the

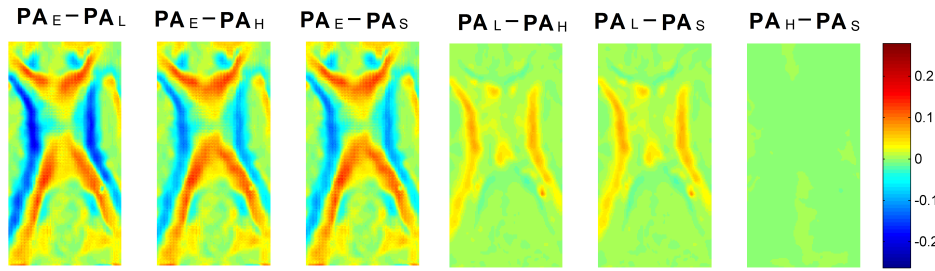


Figure 21. PA contrasts.

weights of individual voxels.

A concrete example could be $\Omega_{>0}(n)$ -valued kernel-based regression with, say, the Gaussian kernel defined via a distance d_1 on some set M of input, or predictor, variables ψ , where M may or may not be $\Omega_{>0}(n)$. Clearly, this is more general than DTI as symmetric positive semi-definite matrices occur in various other contexts, such as covariance descriptors in computer vision and even graph Laplacians. Hence, we write $\Omega_{>0}$ in short, understanding that degenerate matrices may also be included. (Indeed, all the key results we have given in this paper for 3×3 tensors, remain valid in general.) The idea is then to generalise the weighted tensor averaging (Equation (6)) in the spirit of the Nadaraya-Watson regression to predict a tensor valued response Σ corresponding to the input variable ψ :

$$\hat{\Sigma}(\psi) = \arg \inf_{\Sigma \in \Omega_{>0}} \sum_{i=1}^N \exp\{-\gamma d_1^2(\psi_i, \psi)\} d_2^2(\mathbf{D}_i, \Sigma),$$

So, in effect $\{\psi_i, \mathbf{D}_i\}$, $i = 1, 2, \dots, N$, is the training (observed) dataset, and ψ_i 's are observed feature values. A simple special case is when $M = \mathbb{R}^3$, so that ψ is simply the set of coordinates of locations where matrix-valued observations \mathbf{D}_i 's are collected, d_1 can then be the Euclidean distance (as used in this paper), unless M is a different manifold, such as a sphere (in geosciences). In that case, we get a generalised tensor-valued kernel Nadaraya-Watson-smoothed Kriging prediction at location ψ . An important point has been made recently [20] that in order for the above exponential kernel to be positive-definite for all $\gamma > 0$ (as required to guarantee the usual kernel properties used in statistical and machine learning), d_1^2 must be representable by an inner product on some inner product space. Certainly, the Euclidean distance satisfies this condition. Also, if we take our input variable ψ to be tensor valued, then all the Euclidean based metrics on $\Omega_{>0}$ also satisfy this condition. The Procrustes size-and-shape metric d_S appears to not satisfy this condition. At least, if we use the natural representation of \mathbf{D}_i by its principal square root \mathbf{Q}_i , the trace of the Δ matrix of the singular values of $\mathbf{Q}_1 \mathbf{Q}_2$ would have to be an inner product, which it is not (although it is a norm). Nonetheless, d_S is still a good choice for d_2 , as shown in this paper. All in all, the ready to use machinery of kernel regression then becomes available to address issues such as adapting γ to a specific task. The weighted averaging methods described in this work then gain a new perspective as the weights $w_i \propto \exp\{-\gamma d_1^2(\psi_i, \psi)\}$ are now defined in a principled manner, which may not have been clear from the function (22).

If applied to DTI, this framework would still need to be fine tuned to respect various structures. We could start with some large region of interest and then alternate iteratively between the following phases. First, segment the region into more or less homogeneous subregions, which will, however, generally be irregular. Some standard classification tool can be used for this, especially since it has recently been shown that

essentially all statistical and machine learning methods, such as discriminant analysis and clustering, also work well with tensor valued data using positive-definite kernels, such as the exponential one above. Homogeneity within putative subregions will then be growing. These classification/segmentation can actually be “soft”, i.e. we would not be drawing any definitive boundaries, but rather be estimating class-conditional probabilities. Then we would do smoothing separately within each of the subregions, incorporating these probabilities into the weights. So, the voxels that have higher probability of being part of the currently chosen subregion will be contributing more. In the beginning we could allow our segmentation algorithm to have many distinct segments/subregions and very peaked distributions of the weights within each. That is, we would start with very little smoothing. Once the smoothed values are computed, we would rerun the segmentation algorithm, and so on, gradually causing the segmentation to be increasingly more conservative in terms of the number of distinct subregions. We might even allow for disconnected segments/subregions, which could be important to prevent oversmoothing of structure. Ideally we would like to be able to smooth distinct fibre bundles independently of one another, or, at least, to be able to smooth a segment of a major fibre tract preserving its boundaries.

The weighted regularisation framework proposed in Section 5 is very general. Presently, we have considered the special case with the power parameters $p = q = 2$ and the Procrustes size-and-shape metric being used for both d_1 and d_2 . It would be interesting to vary the parameters in the weighted regularisation model. For example, when $p = 2$ and $q = 2$, we have a generalised type of ridge-regression model. When $p = 2$ and $q = 1$, the method gives a generalised type of LASSO [36], and so on. To develop an efficient and more automated method for choosing the regularisation parameter λ is also of interest.

There is a special challenge to process tensor fields containing multiple diffusion profiles, especially when more than one distinct profile is observed at a single voxel (crossing fibres, etc). For example, to interpolate a pair of tensors at one voxel with a pair of tensors at another voxel is a basic and difficult problem. An intuitive idea is to find the optimal matches between two pairs, i.e. to match each tensor at the first voxel with a tensor at the second voxel. Then, the processing can be carried out between matched tensors. Other, more object-oriented, approaches may also be found not requiring explicit matching of individual tensors. Also, when more angular directions are available such as in HARDI data [37] more flexible models than diffusion tensors can be explored, which also present more challenges for non-Euclidean statistical analysis.

Acknowledgement

The diffusion MR image data used in this paper is provided by the Division of Academic Radiology, University of Nottingham and Queen’s Medical Centre, UK, and the authors would like to thank Dr Paul Morgan and Professor Dorothee Auer for their assistance. The authors are thankful to the editor and anonymous referees for their helpful comments and suggestions on improving the manuscript. The third author thanks God for making this work possible.

Funding

This work was supported by the European Commission FP6 Marie Curie Action Programme under the CMIAG (Collaborative Medical Image Analysis on Grid) project. The second author acknowledges support from a Royal Society Wolfson Research Merit Award and EPSRC grant EP/K022547/1. The third author acknowledges support from

NIHR i4i FPD2 grant.

Appendix A. Proof of Theorem 2.1 (See page 4)

THEOREM 2.1 *For any non-zero symmetric semi-positive definite \mathbf{D} , $\text{FA}(\mathbf{D}^a)$ is an increasing function of a , where $a \in [0, \infty)$.*

Proof. Let $a, b \geq 0$ be such that $a < b$. We need to show that

$$\frac{\sum_{i=1}^3 (\lambda_i^a - \bar{\lambda}^a)^2}{\sum_{i=1}^3 \lambda_i^{2a}} \leq \frac{\sum_{i=1}^3 (\lambda_i^b - \bar{\lambda}^b)^2}{\sum_{i=1}^3 \lambda_i^{2b}}, \text{ which is equivalent to } [\bar{\lambda}^b]^2 \bar{\lambda}^{2a} \leq [\bar{\lambda}^a]^2 \bar{\lambda}^{2b}. \quad (\text{A1})$$

Let $q = \frac{2b-a}{b}$, $p = \frac{2b-a}{b-a}$, so that $\frac{1}{q} + \frac{1}{p} = 1$, $p \geq 1$, $q \geq 1$, then write $\lambda_i^b = \lambda_i^{s+t}$, $i \in \{1, 2, 3\}$ where $s = \frac{a}{q} = \frac{ab}{2b-a}$, $t = \frac{2b}{p} = \frac{2b(b-a)}{2b-a}$.

Apply Hölder's inequality [28] with $g_i = \lambda_i^s$, $f_i = \lambda_i^t$, $i = 1, 2, 3$,

$$\sum_{i=1}^3 |g_i f_i| \leq \left[\sum_{i=1}^3 g_i^q \right]^{1/q} \left[\sum_{i=1}^3 f_i^p \right]^{1/p}$$

to obtain

$$\begin{aligned} \bar{\lambda}^b &\leq [\bar{\lambda}^a]^{\frac{b}{2b-a}} [\bar{\lambda}^{2b}]^{\frac{b-a}{2b-a}} \\ \implies [\bar{\lambda}^b]^2 &\leq [\bar{\lambda}^a]^{\frac{2b}{2b-a}} [\bar{\lambda}^{2b}]^{\frac{2b-2a}{2b-a}}. \end{aligned} \quad (\text{A2})$$

Next, take $q' = \frac{2b-a}{2b-2a}$, $p' = \frac{2b-a}{a}$, note that $p', q' \geq 1$, $\frac{1}{q'} + \frac{1}{p'} = 1$. Then write $\lambda_i^{2a} = \lambda_i^{s'+t'}$, and take $s' = \frac{a}{q'} = \frac{2a(b-a)}{2b-a}$, $t' = \frac{2b}{p'} = \frac{2ab}{2b-a}$. Again, let $g'_i = \lambda_i^{s'}$, $f'_i = \lambda_i^{t'}$, $i = 1, 2, 3$, and apply Hölder's inequality to get

$$\bar{\lambda}^{2a} = \bar{\lambda}^{s'+t'} \leq [\bar{\lambda}^a]^{\frac{2(b-a)}{2b-a}} [\bar{\lambda}^{2b}]^{\frac{a}{2b-a}}. \quad (\text{A3})$$

Finally, multiply both sides of Inequality (A2) by respective sides of Inequality (A3) to obtain Inequality (A1). \blacksquare

Appendix B. Proof of Theorem 3.1 (See page 9)

THEOREM 3.1 *Let $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{\geq 0}(3)$. Then, $\sqrt{0.5}d_H(\mathbf{D}_1, \mathbf{D}_2) \leq d_S(\mathbf{D}_1, \mathbf{D}_2) \leq d_H(\mathbf{D}_1, \mathbf{D}_2)$. Moreover, if $\mathbf{D}_1 \neq \mathbf{D}_2$, and \mathbf{D}_1 and \mathbf{D}_2 are of rank 1, then $\sqrt{0.5}d_H(\mathbf{D}_1, \mathbf{D}_2) < d_S(\mathbf{D}_1, \mathbf{D}_2)$ and $d_S(\mathbf{D}_1, \mathbf{D}_2)/d_H(\mathbf{D}_1, \mathbf{D}_2) \rightarrow \sqrt{0.5}$ as $d(\mathbf{D}_1, \mathbf{D}_2) \rightarrow 0$ in any metric d .*

Proof. The $\sqrt{0.5}d_H(\mathbf{D}_1, \mathbf{D}_2) \leq d_S(\mathbf{D}_1, \mathbf{D}_2)$ inequality actually holds for any dimension and follows immediately from Proposition 3.1.

Proposition 3.1 Let \mathbf{Q}_i , $i = 1, 2$ be symmetric positive semi-definite $n \times n$ real matrices. Then

$$\min_{\mathbf{R} \in O(n)} \|\mathbf{Q}_1 - \mathbf{Q}_2 \mathbf{R}\|^2 \geq 0.5 \|\mathbf{Q}_1 - \mathbf{Q}_2\|^2. \quad (9)$$

Proof. It is proved in [8] that for any unitarily invariant matrix norm $\|\cdot\|$ and for any positive semi-definite $n \times n$ matrices \mathbf{A} and \mathbf{B} , we have the following version of the arithmetic-geometric mean inequality:

$$\|\mathbf{AB}\| \leq \left\| \left(\frac{\mathbf{A} + \mathbf{B}}{2} \right)^2 \right\|. \quad (\text{B1})$$

Now, $\min_{\mathbf{R} \in O(3)} \|\mathbf{Q}_1 - \mathbf{Q}_2 \mathbf{R}\|^2$ can be seen [18, p.30] to be equal to

$$\|\mathbf{Q}_1\|^2 + \|\mathbf{Q}_2\|^2 - 2 \operatorname{tr}(\mathbf{\Delta}), \quad (\text{B2})$$

where $\mathbf{\Delta}$ is the matrix of the singular values of $\mathbf{Q}_1 \mathbf{Q}_2$, i.e. from the singular value decomposition $\mathbf{Q}_1 \mathbf{Q}_2 = \mathbf{V} \mathbf{\Delta} \mathbf{U}^T$ with $\mathbf{U}, \mathbf{V} \in O(n)$. Note that $\operatorname{tr}(\mathbf{\Delta})$ is the Schatten $p = 1$, or the trace norm (also known as the nuclear norm or Ky Fan n norm), of $\mathbf{Q}_1 \mathbf{Q}_2$. Let us denote this norm by $\|\mathbf{Q}_1 \mathbf{Q}_2\|_1$. Since this norm is also unitarily invariant, Inequality B1 applies to give

$$\|\mathbf{Q}_1 \mathbf{Q}_2\|_1 \leq \left\| \left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2} \right)^2 \right\|_1. \quad (\text{B3})$$

Now, since \mathbf{Q}_i 's are symmetric positive semi-definite, then so is $\left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2} \right)^2$. In this case, the trace norm $\left\| \left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2} \right)^2 \right\|_1$ coincides with $\operatorname{tr} \left[\left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2} \right)^2 \right]$. Therefore,

$$\begin{aligned} \min_{\mathbf{R} \in O(3)} \|\mathbf{Q}_1 - \mathbf{Q}_2 \mathbf{R}\|^2 &\geq \|\mathbf{Q}_1\|^2 + \|\mathbf{Q}_2\|^2 - 2 \operatorname{tr} \left[\left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2} \right)^2 \right] = \\ &\operatorname{tr} (0.5 \mathbf{Q}_1^2 + 0.5 \mathbf{Q}_2^2 - \mathbf{Q}_1 \mathbf{Q}_2) = \\ &0.5 \|\mathbf{Q}_1 - \mathbf{Q}_2\|^2. \end{aligned}$$

■

To prove the remaining result, it suffices to consider

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{D}_2 = \mathbf{V} \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}^T, \quad \text{for all } \mathbf{V} \in O(3) \text{ and for all } \lambda > 0.$$

Thus, the first column of \mathbf{V} , i.e. the principal eigenvector of \mathbf{D}_2 , parameterises an arbitrary direction, and the remaining two vectors (columns 2 and 3 of \mathbf{V}) are an arbitrary orthonormal basis of the zero rank subspace. (As both the distances scale the same under

$\mathbf{D} \mapsto c\mathbf{D}$ ($c > 0$), there is no need to consider non-unitary parameters in both \mathbf{D}_1 and \mathbf{D}_2 , hence general λ appears in \mathbf{D}_2 only.) Note that $\mathbf{Q}_1 = \mathbf{D}_1$, and $\mathbf{Q}_2 = \mathbf{V} \begin{pmatrix} \sqrt{\lambda} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}^T$.

To compute $d_S(\mathbf{D}_1, \mathbf{D}_2)$, note that in the singular value decomposition of $\mathbf{Q}_1\mathbf{Q}_2$, the matrix $\mathbf{\Delta}$ of singular values is given as $\begin{pmatrix} \sqrt{\lambda}|v_{11}| & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. With $|v_{11}|$ replaced by x , $x \in [0, 1]$,

Equation (B2) then yields

$$d_S^2(\mathbf{D}_1, \mathbf{D}_2) = 1 + \lambda - 2\sqrt{\lambda}x. \quad (\text{B4})$$

Similarly, to get $d_H^2(\mathbf{D}_1, \mathbf{D}_2)$, note that

$$\|\mathbf{Q}_1 - \mathbf{Q}_2\|^2 = \|\mathbf{Q}_1\|^2 + \|\mathbf{Q}_2\|^2 - 2\text{tr}(\mathbf{Q}_1\mathbf{Q}_2) = 1 + \lambda - 2\sqrt{\lambda}x^2. \quad (\text{B5})$$

Let b stand for $\frac{1+\lambda}{2\sqrt{\lambda}}$. Note that $b \geq 1$ for all $\lambda > 0$. Then we have

$$\frac{d_S^2(\mathbf{D}_1, \mathbf{D}_2)}{d_H^2(\mathbf{D}_1, \mathbf{D}_2)} = \frac{b-x}{b-x^2},$$

which, for any $x \in [0, 1]$ is a non-decreasing function of b , $b \geq 1$. Since the tensors are assumed to be distinct, the case of $b = x = 1$ is disregarded. We also disregard the case of $x = 1$ and $b > 1$ since it gives $d_S(\mathbf{D}_1, \mathbf{D}_2) = d_H(\mathbf{D}_1, \mathbf{D}_2)$. Therefore, for any $x \in [0, 1)$, we have

$$\frac{d_S^2(\mathbf{D}_1, \mathbf{D}_2)}{d_H^2(\mathbf{D}_1, \mathbf{D}_2)} \geq \frac{1-x}{1-x^2} = \frac{1}{1+x}.$$

Since $\inf_{x \in [0, 1)} \frac{1}{1+x} = \lim_{x \rightarrow 1} \frac{1}{1+x} = 1/2$, we get the claimed (strict) bound $d_H^2(\mathbf{D}_1, \mathbf{D}_2)/2 < d_S^2(\mathbf{D}_1, \mathbf{D}_2)$, and we also get that $\frac{d_S^2(\mathbf{D}_1, \mathbf{D}_2)}{d_H^2(\mathbf{D}_1, \mathbf{D}_2)} \rightarrow 0.5^+$ as $\lambda \rightarrow 1$ and $x \rightarrow 1^-$, which is equivalent to $d_H(\mathbf{D}_1, \mathbf{D}_2) \rightarrow 0$ (and $d_S(\mathbf{D}_1, \mathbf{D}_2) \rightarrow 0$ and $\mathbf{D}_2 \rightarrow \mathbf{D}_1$ component-wise). \blacksquare

Therefore, in the rank 1 case, the two metrics diverge the most as the two tensors become indistinguishable. We also conjecture that in general, in order to approach the 1/2 bound, the two tensors must become indistinguishable and simultaneously approach the subspace of rank 1 tensors. For example, if \mathbf{D}_2 is general (and \mathbf{D}_1 is still as in the proof of the Theorem), then we have

$$d_S^2(\mathbf{D}_1, \mathbf{D}_2) = 1 + \text{trace}(\mathbf{D}_2) - 2\sqrt{\lambda_1 v_{11}^2 + \lambda_2 v_{12}^2 + \lambda_3 v_{13}^2} \quad (\text{B6})$$

$$d_H^2(\mathbf{D}_1, \mathbf{D}_2) = 1 + \text{trace}(\mathbf{D}_2) - 2(\sqrt{\lambda_1}v_{11}^2 + \sqrt{\lambda_2}v_{12}^2 + \sqrt{\lambda_3}v_{13}^2), \quad (\text{B7})$$

which become Equations (B4) and (B5) respectively as $\lambda_2, \lambda_3 \rightarrow 0$ (and $\lambda_1 = \lambda$).

Appendix C. Proof of Proposition 3.2 (See page 9)

Proposition 3.2 Let $\text{rank}(\mathbf{D}_1) = \text{rank}(\mathbf{D}_2) < 3$. Then $\text{rank}(\mathbf{D}(w_1, w_2))$ is constant for all (general) $w_1, w_2 \in \mathbb{R}$, where $\mathbf{D}(w_1, w_2)$ is as in the d_S row of Table 1.

Proof. The result is immediately verified in the rank 1 case. Indeed, without loss of generality, assume the coordinate frame of \mathbf{D}_1 , i.e.

$$\mathbf{D}_1 = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{D}_2 = \mathbf{V} \begin{pmatrix} \gamma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}^T, \quad \text{for some } \mathbf{V} \in O(3) \text{ and for some } \lambda, \gamma > 0.$$

It is easy to see that $\mathbf{D}(w_1, w_2) = \mathbf{Q}\mathbf{Q}^T$, where

$$\mathbf{Q} = \left[w_1 \sqrt{\lambda} \mathbf{I}_{3 \times 3} + w_2 \sqrt{\gamma} \mathbf{V} \right] \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

hence $\text{rank}(\mathbf{D}(w_1, w_2)) = 1$ (except for the special case when the two tensors have the same orientation and $\lambda w_1 = -w_2 \gamma$, making the average the zero tensor).

Consider now the rank 2 case, namely:

$$\mathbf{D}_1 = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{D}_2 = \mathbf{V} \begin{pmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}^T, \quad \text{for some } \mathbf{V} \in O(3) \text{ and for some } \lambda_1 \geq \lambda_2 > 0,$$

and for some $\gamma_2 \geq \gamma_2 > 0$. Using the symmetric square roots $\mathbf{Q}_i = \mathbf{D}_i^{\frac{1}{2}}$, $i = 1, 2$, note that

$$\mathbf{Q}_1 \mathbf{Q}_2 = \begin{pmatrix} v_{11} \sqrt{\lambda_1 \gamma_1} & v_{12} \sqrt{\lambda_1 \gamma_2} & 0 \\ v_{21} \sqrt{\lambda_2 \gamma_1} & v_{22} \sqrt{\lambda_2 \gamma_2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}^T = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{V}^T,$$

where we have used the 2×2 singular value decomposition

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} v_{11} \sqrt{\lambda_1 \gamma_1} & v_{12} \sqrt{\lambda_1 \gamma_2} \\ v_{21} \sqrt{\lambda_2 \gamma_1} & v_{22} \sqrt{\lambda_2 \gamma_2} \end{pmatrix}.$$

Therefore, $\hat{\mathbf{R}} = \mathbf{V} \begin{pmatrix} b_{11} & b_{21} & 0 \\ b_{12} & b_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{V} \begin{pmatrix} c_{11} & c_{12} & 0 \\ c_{21} & c_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}$, where $c_{ij} = a_{j1} b_{1i} + a_{j2} b_{2i}$, $i, j = 1, 2$. Thus, $\mathbf{D}(w_1, w_2) = \mathbf{Q}\mathbf{Q}^T$, where \mathbf{Q} is given by

$$w_1 \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & 0 \end{pmatrix} + w_2 \mathbf{V} \begin{pmatrix} \sqrt{\gamma_1} & 0 & 0 \\ 0 & \sqrt{\gamma_2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} & 0 \\ c_{21} & c_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ \begin{pmatrix} w_1 \sqrt{\lambda_1} + w_2 (v_{11} c_{11} \sqrt{\gamma_1} + v_{12} c_{21} \sqrt{\gamma_2}) & w_2 (v_{11} c_{12} \sqrt{\gamma_1} + v_{12} c_{22} \sqrt{\gamma_2}) & 0 \\ w_2 (v_{21} c_{11} \sqrt{\gamma_1} + v_{22} c_{21} \sqrt{\gamma_2}) & w_1 \sqrt{\lambda_2} + w_2 (v_{21} c_{12} \sqrt{\gamma_1} + v_{22} c_{22} \sqrt{\gamma_2}) & 0 \\ w_2 (v_{31} c_{11} \sqrt{\gamma_1} + v_{32} c_{21} \sqrt{\gamma_2}) & w_2 (v_{31} c_{12} \sqrt{\gamma_1} + v_{32} c_{22} \sqrt{\gamma_2}) & 0 \end{pmatrix}.$$

Thus, $\text{rank}(\mathbf{D}(w_1, w_2)) \leq 2$ for any $w_1, w_2 \in \mathbb{R}$, and for a general setting of \mathbf{D}_1 , \mathbf{D}_2 , and w_1, w_2 this rank is exactly two. \blacksquare

Appendix D. Proofs of Proposition 3.3, Corollary 3.1, and Proposition 3.4 (See pages 10–11)

Proposition 3.3 Let w_1, w_2 be probability weights, and let \mathbf{D}_1 and \mathbf{D}_2 be $n \times n$ positive definite symmetric matrices, and let $\mathbf{D}_{L,R}(w_1, w_2)$, $\mathbf{D}_H(w_1, w_2)$, and $\mathbf{D}_E(w_1, w_2)$ be their Riemannian (log-Euclidean or affine-invariant), root-Euclidean, and Euclidean weighted averages, respectively, where $w_1, w_2 \geq 0$. Then we have

$$|\mathbf{D}_{L,R}(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)| \leq |\mathbf{D}_E(w_1, w_2)|.$$

Proof. Using the same Minkowski Determinant Theorem [26] (see also Equation (12)), followed by the arithmetic-geometric mean inequality, we obtain

$$|w_1 \mathbf{Q}_1 + w_2 \mathbf{Q}_2|^{\frac{1}{n}} \geq w_1 |\mathbf{Q}_1|^{\frac{1}{n}} + w_2 |\mathbf{Q}_2|^{\frac{1}{n}} \geq |\mathbf{Q}_1|^{\frac{w_1}{n}} |\mathbf{Q}_2|^{\frac{w_2}{n}}.$$

Therefore, $|w_1 \mathbf{Q}_1 + w_2 \mathbf{Q}_2| \geq |\mathbf{Q}_1|^{w_1} |\mathbf{Q}_2|^{w_2}$ and subsequently $|\mathbf{D}_L(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)|$.

To prove the other side of the double inequality, note that according to the Löwner–Heinz Theorem [27], the matrix square function $g(\mathbf{Q}) = \mathbf{Q}^2$ is operator convex on the set of symmetric square matrices. Therefore, $\mathbf{D}_H(w_1, w_2) \leq \mathbf{D}_E(w_1, w_2)$ (in the sense that the difference $\mathbf{D}_E(w_1, w_2) - \mathbf{D}_H(w_1, w_2)$ is semi-positive definite). Hence, $|\mathbf{D}_H(w_1, w_2)| \leq |\mathbf{D}_E(w_1, w_2)|$, as claimed. ■

Corollary 3.1 Under the assumptions of Proposition 3.3 above, we also have

$$|\mathbf{D}_L(w_1, w_2)| \leq |\mathbf{D}_S(w_1, w_2)|.$$

Proof. First note that when using the (principal) square root decomposition, we have $\hat{\mathbf{R}} \in SO(3)$, therefore the proof of $|\mathbf{D}_L(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)|$ from the Proposition applies with \mathbf{Q}_2 replaced by $\mathbf{Q}_2 \hat{\mathbf{R}}$, which does not affect the determinant. ■

Proposition 3.4 Let \mathbf{D}_1 and \mathbf{D}_2 be $n \times n$ symmetric positive semi-definite real matrices, and let $\mathbf{D}_S(w_1, w_2)$ and $\mathbf{D}_H(w_1, w_2)$ be their Procrustes and root-Euclidean weighted averages, respectively, where $w_1, w_2 \geq 0$. Then

$$|\mathbf{D}_S(w_1, w_2)| \leq |\mathbf{D}_H(w_1, w_2)|.$$

Proof. Without loss of generality assume $w_1 = w_2 = 1$ and write \mathbf{Q}_S for $\mathbf{Q}_1 + \mathbf{Q}_2 \hat{\mathbf{R}}$ and \mathbf{Q}_H for $\mathbf{Q}_1 + \mathbf{Q}_2$. Indeed, the weight w_i can be absorbed in the matrix \mathbf{Q}_i , since this does not affect $\hat{\mathbf{R}}$. By the Minkowski Determinant Theorem, $|\mathbf{Q}_S| \geq 0$. Therefore, since $|\mathbf{Q}_S| = \sqrt{|\mathbf{D}_S|}$ and $|\mathbf{Q}_H| = \sqrt{|\mathbf{D}_H|}$, it suffices to show that $|\mathbf{Q}_S| \leq |\mathbf{Q}_H|$.

Below we give a complete proof for the 2×2 case, which is straightforward but does not easily extend to larger matrices. A general proof is given in [3] and uses the theory of majorization. Specifically, it establishes $|\mathbf{I}_{n \times n} + \mathbf{A} \# \mathbf{B}| \geq |\mathbf{I}_{n \times n} + \mathbf{A}^{1/2} \mathbf{B}^{1/2}|$, where $\mathbf{A} \# \mathbf{B}$ stands for the affine-invariant Riemannian average ($w_1 = w_2 = 0.5$) of positive definite symmetric matrices \mathbf{A} and \mathbf{B} . With $\mathbf{A} = \mathbf{D}_1^{-1}$ and $\mathbf{B} = \mathbf{D}_2$, the above inequality yields $|\mathbf{Q}_1 \mathbf{Q}_S| \leq |\mathbf{Q}_1 \mathbf{Q}_H|$ for (strictly) positive definite \mathbf{Q}_1 , which finally extends to the boundary of the cone $\Omega_{\geq 0}(n)$ by continuity of the determinant map.

Expressing the determinants via traces in the 2×2 case, and recalling that Δ stands

for the matrix of the singular values of $\mathbf{Q}_1\mathbf{Q}_2$, we obtain

$$\begin{aligned} |\mathbf{Q}_H| - |\mathbf{Q}_S| &= \frac{1}{2} \left\{ \text{tr}^2(\mathbf{Q}_H) - \text{tr}^2(\mathbf{Q}_S) + \text{tr}(\mathbf{Q}_S^2) - \text{tr}(\mathbf{Q}_H^2) \right\} \\ &= \frac{1}{2} \left\{ \left[\text{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) - \text{tr}(\mathbf{Q}_1 + \mathbf{Q}_2\hat{\mathbf{R}}) \right] \left[\text{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) + \text{tr}(\mathbf{Q}_1 + \mathbf{Q}_2\hat{\mathbf{R}}) \right] + \right. \\ &\quad \left. \text{tr} \left[\mathbf{Q}_1^2 + \mathbf{Q}_2\hat{\mathbf{R}}\mathbf{Q}_1 + \mathbf{Q}_1\mathbf{Q}_2\hat{\mathbf{R}} + \mathbf{Q}_2\hat{\mathbf{R}}\mathbf{Q}_2\hat{\mathbf{R}} - \mathbf{Q}_1^2 - \mathbf{Q}_1\mathbf{Q}_2 - \mathbf{Q}_2\mathbf{Q}_1 - \mathbf{Q}_2^2 \right] \right\} \\ &= \frac{1}{2} \left\{ \left[\text{tr} \mathbf{Q}_2 - \text{tr}(\mathbf{Q}_2\hat{\mathbf{R}}) \right] \left[2 \text{tr} \mathbf{Q}_1 + \text{tr} \mathbf{Q}_2 + \text{tr}(\mathbf{Q}_2\hat{\mathbf{R}}) \right] + \right. \\ &\quad \left. \text{tr} [2\Delta - 2\mathbf{Q}_1\mathbf{Q}_2] + \text{tr} \left[(\mathbf{Q}_2\hat{\mathbf{R}})^2 \right] \right\}. \end{aligned}$$

Noticing that

$$(\text{tr} \mathbf{Q}_2)^2 - \text{tr}(\mathbf{Q}_2^2) = 2|\mathbf{Q}_2| = 2|\mathbf{Q}_2\hat{\mathbf{R}}| = \left[\text{tr}(\mathbf{Q}_2\hat{\mathbf{R}}) \right]^2 - \text{tr} \left[(\mathbf{Q}_2\hat{\mathbf{R}})^2 \right],$$

we therefore have

$$|\mathbf{Q}_H| - |\mathbf{Q}_S| = \text{tr} \mathbf{Q}_1 \left[\text{tr} \mathbf{Q}_2 - \text{tr}(\mathbf{Q}_2\hat{\mathbf{R}}) \right] + \text{tr} [\Delta - \mathbf{Q}_1\mathbf{Q}_2] \geq 0,$$

since for any orthonormal matrix \mathbf{R} and $\text{tr} \mathbf{Q} \geq \text{tr}(\mathbf{Q}\mathbf{R})$ for any positive semi-definite matrix \mathbf{Q} , and since $\text{tr} \Delta \geq \text{tr}(\mathbf{Q}_1\mathbf{Q}_2)$. \blacksquare

Appendix E. Proof of Proposition 3.5 (See page 11)

Proposition 3.5 For all $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{>0}(3)$, and for all probability weights w_1, w_2 , we have

$$\text{tr} \mathbf{D}_R(w_1, w_2) \leq \text{tr} \mathbf{D}_L(w_1, w_2) \leq \text{tr} \mathbf{D}_H(w_1, w_2) \leq \text{tr} \mathbf{D}_S(w_1, w_2) \leq \text{tr} \mathbf{D}_E(w_1, w_2).$$

Moreover, the last two inequalities remain valid for all $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{\geq 0}(3)$, and additionally we have

$$\text{tr} \mathbf{D}_S(w_1, w_2) - \text{tr} \mathbf{D}_H(w_1, w_2) \leq 0.5d_H^2(w_1^2\mathbf{D}_1, w_2^2\mathbf{D}_2). \quad (14)$$

Proof. Assume $\mathbf{D}_1, \mathbf{D}_2 \in \Omega_{\geq 0}(3)$. Consider a type of Pythagoras' Theorem:

$$\begin{aligned} (w_1\mathbf{Q}_1 + w_2\mathbf{Q}_2\mathbf{R})(w_1\mathbf{Q}_1 + w_2\mathbf{Q}_2\mathbf{R})^T + (w_1\mathbf{Q}_1 - w_2\mathbf{Q}_2\mathbf{R})(w_1\mathbf{Q}_1 - w_2\mathbf{Q}_2\mathbf{R})^T = \\ 2(w_1^2\mathbf{D}_1 + w_2^2\mathbf{D}_2), \end{aligned} \quad (\text{E1})$$

which holds for all weights and all orthonormal matrices \mathbf{R} , in particular for probability weights and for $\mathbf{R} = \hat{\mathbf{R}}$ and $\mathbf{R} = \mathbf{I}_{3 \times 3}$. Note that it is the same $\hat{\mathbf{R}}$ that gives $d_S(\mathbf{D}_1, \mathbf{D}_2)$ and $d_S(w_1^2\mathbf{D}_1, w_2^2\mathbf{D}_2)$ according to Equation (7) of the main manuscript. Instantiating Equation (E1) with those two choices of \mathbf{R} and applying the trace operator, we arrive at

$$\text{tr} \mathbf{D}_S(w_1, w_2) + d_S^2(w_1^2\mathbf{D}_1, w_2^2\mathbf{D}_2) = \text{tr} \mathbf{D}_H(w_1, w_2) + d_H^2(w_1^2\mathbf{D}_1, w_2^2\mathbf{D}_2).$$

Since $d_H^2(w_1^2 \mathbf{D}_1, w_2^2 \mathbf{D}_2) \geq d_S^2(w_1^2 \mathbf{D}_1, w_2^2 \mathbf{D}_2)$, we thereby obtain the middle inequality $\text{tr } \mathbf{D}_S(w_1, w_2) \geq \text{tr } \mathbf{D}_H(w_1, w_2)$. Theorem 3.1 gives us more information regarding this comparison, in particular $0.5d_H^2(w_1^2 \mathbf{D}_1, w_2^2 \mathbf{D}_2) \leq d_S^2(w_1^2 \mathbf{D}_1, w_2^2 \mathbf{D}_2)$, which gives Inequality (14).

Next, we prove that $\text{tr } \mathbf{D}_S(w_1, w_2) \leq \text{tr } \mathbf{D}_E(w_1, w_2)$. To see that, note first that $\text{tr } \mathbf{D}_S(w_1, w_2) = w_1^2 \text{tr } \mathbf{D}_1 + 2w_1 w_2 \text{tr } \mathbf{\Delta} + w_2^2 \text{tr } \mathbf{D}_2$, where $\mathbf{\Delta}$ is the matrix of singular values of $\mathbf{Q}_1 \mathbf{Q}_2$. Using Inequality (B3) then gives

$$\text{tr } \mathbf{D}_S(w_1, w_2) \leq w_1^2 \text{tr } \mathbf{D}_1 + 2w_1 w_2 \text{tr} \left(\frac{\mathbf{Q}_1 + \mathbf{Q}_2}{2} \right)^2 + w_2^2 \text{tr } \mathbf{D}_2.$$

Using the fact that $\text{tr} (\mathbf{Q}_1 \mathbf{Q}_2 + \mathbf{Q}_2 \mathbf{Q}_1) \leq \text{tr} (\mathbf{Q}_1^2 + \mathbf{Q}_2^2)$, we get

$$\text{tr } \mathbf{D}_S(w_1, w_2) \leq (w_1^2 + w_1 w_2) \text{tr } \mathbf{D}_1 + (w_2^2 + w_1 w_2) \text{tr } \mathbf{D}_2 = \text{tr } \mathbf{D}_E(w_1, w_2).$$

Finally (restricting \mathbf{D}_1 and \mathbf{D}_2 to be in $\Omega_{>0}(3)$), we quote [7, Corollary 2], which in our notation writes as:

$$\|\mathbf{D}_R(w_1, w_2)\| \leq \|\mathbf{D}_L(w_1, w_2)\| \leq \|\mathbf{D}_A(w_1, w_2|a)\|, \quad (\text{E2})$$

where $\mathbf{D}_A(w_1, w_2|a)$ is the tensor averaged using the power-Euclidean metric $d_A(\cdot, \cdot|a)$ for any $a > 0$, and $\|\cdot\|$ is any unitarily invariant matrix norm. Thus, with $a = 0.5$, and the trace norm (see proof of Proposition 3.1), we obtain $\text{tr } \mathbf{D}_R(w_1, w_2) \leq \text{tr } \mathbf{D}_L(w_1, w_2) \leq \text{tr } \mathbf{D}_H(w_1, w_2)$. ■

Appendix F. Proof of Proposition 3.6 (See page 12)

Proposition 3.6 Assume that one of the tensors $\mathbf{D}_1, \mathbf{D}_2$ (as in Table 1) is isotropic; without loss of generality, let it be \mathbf{D}_1 , i.e. $\mathbf{D}_1 = \lambda \mathbf{I}_{3 \times 3}$ for some $\lambda \geq 0$. Let d_g be a Euclidean-based metric. Let $\lambda_1, \lambda_2, \lambda_3$ be eigenvalues of \mathbf{D}_2 ,

- (1) The weighted average tensor $\mathbf{D}(w_1, w_2)$ is given by Equation (15) below. Subsequently, if g is increasing, then the orientation of $\mathbf{D}(w_1, w_2)$ is the same as the orientation of \mathbf{D}_2 for all weights with $w_2 \neq 0$. In particular, if $\lambda_1 > \lambda_2$, i.e. \mathbf{D}_2 is prolate (has a well-defined direction), then $\mathbf{D}(w_1, w_2)$ is also prolate and has the same direction (provided $w_2 \neq 0$).
- (2) The Procrustes averaging is identical to the root-Euclidean averaging, and $\mathbf{D}(w_1, w_2)$ in this case is given by Equation (16) below.
- (3) The (log-Euclidean or affine-invariant) Riemannian average tensor $\mathbf{D}(w_1, w_2)$ is given by Equation (17) below.
- (4) $\mathbf{D}(w_1, w_2)$ for the Euclidean averaging is given by Equation (18) below.
- (5) The Euclidean, log-Euclidean, the affine-invariant Riemannian, root-Euclidean, and the Procrustes interpolations all preserve the orientation of \mathbf{D}_2 , provided $w_2 \neq 0$.

$$\mathbf{E} \begin{pmatrix} g^{-1}((w_1 g(\lambda) + w_2 g(\lambda_1))) & 0 & 0 \\ 0 & g^{-1}((w_1 g(\lambda) + w_2 g(\lambda_2))) & 0 \\ 0 & 0 & g^{-1}((w_1 g(\lambda) + w_2 g(\lambda_3))) \end{pmatrix} \mathbf{E}^T, \quad (15)$$

where $\mathbf{E} \mathbf{\Lambda} \mathbf{E}^T$ is the spectral decomposition of \mathbf{D}_2 with eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$.

$$\mathbf{E} \left(\lambda w_1^2 \mathbf{I}_{3 \times 3} + w_2^2 \mathbf{\Lambda} + 2w_1 w_2 \sqrt{\lambda} \sqrt{\mathbf{\Lambda}} \right) \mathbf{E}^T, \quad (16)$$

$$\mathbf{E} (\lambda^{w_1} \mathbf{\Lambda}^{w_2}) \mathbf{E}^T, \quad (17)$$

$$\mathbf{E} (\lambda w_1 \mathbf{I}_{3 \times 3} + w_2 \mathbf{\Lambda}) \mathbf{E}^T. \quad (18)$$

Proof. Equations (16), (17), and (18) are all straightforward instantiations of Equation (11) of the main manuscript. If g is increasing and $w_2 \neq 0$, then $\mathbf{D}(w_1, w_2)$ preserves the ordering of the eigenvalues of \mathbf{D}_2 , thus preserving the orientation of \mathbf{D}_2 .

To see that the Procrustes weighted averaging is identical to the root-Euclidean weighted averaging, note that the matrix $\hat{\mathbf{R}}$ in the definition of $d_S(\mathbf{D}_1, \mathbf{D}_2)$ becomes $\mathbf{I}_{3 \times 3}$ (and therefore $d_S(\mathbf{D}_1, \mathbf{D}_2) = d_H(\mathbf{D}_1, \mathbf{D}_2)$).

Since all of the functions g ($g(x) = x$, $g(x) = \log(x)$, $g(x) = \sqrt{x}$) involved in the definitions of the considered Euclidean-based weighted averages are increasing, these weighted averages preserve the orientation of \mathbf{D}_2 (given $w_2 \neq 0$). Since the Procrustes and the affine-invariant Riemannian averages in this case coincide with the respective Euclidean-based averages, they also preserve the orientation of \mathbf{D}_2 . ■

References

- [1] D. Alexander, *Multiple-fibre reconstruction algorithms for diffusion MRI*, Annals of the New York Academy of Sciences 1046 (2005), pp. 113–133, Available at <http://www.cs.ucl.ac.uk/staff/D.Alexander/Papers/NYAS05.pdf>.
- [2] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*, Magnetic Resonance in Medicine 56 (2006), pp. 411–421.
- [3] K.M.R. Audenaert, *A Determinantal Inequality for the Geometric Mean with an Application in Diffusion Tensor Imaging*, Tech. Rep. March 2015, Available at <http://arxiv.org/abs/1502.06902>.
- [4] P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, *In vivo fiber tractography using DT-MRI data*, Magnetic Resonance in Medicine 44 (2000), pp. 625–632.
- [5] P.G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly, *A rigorous framework for diffusion tensor calculus*, Magnetic Resonance in Medicine 53 (2005), pp. 221–225.
- [6] T.E.J. Behrens, M.W. Woolrich, M. Jenkinson, H. Johansen-Berg, R.G. Nunes, S. Clare, P.M. Matthews, J.M. Brady, and S.M. Smith, *Characterization and propagation of uncertainty in Diffusion-Weighted MR Imaging*, Magnetic Resonance in Medicine 50 (2003), pp. 1077–1088.
- [7] R. Bhatia and P. Grover, *Norm inequalities related to the matrix geometric mean*, Linear Algebra and its Applications 437 (2012), pp. 726 – 733, Available at <http://www.sciencedirect.com/science/article/pii/S0024379512001814>.
- [8] R. Bhatia and F. Kittaneh, *Notes on matrix arithmeticgeometric mean inequalities*, Linear Algebra and its Applications 308 (2000), pp. 203–211, Available at <http://linkinghub.elsevier.com/retrieve/pii/S0024379500000483>.
- [9] I. Corouge, P.T. Fletcher, S. Joshi, S. Gouttard, and G. Gerig, *Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis*, Medical Image Analysis 10 (2006), pp. 786 – 798, Available at <http://www.sciencedirect.com/science/article/pii/S1361841506000557>.
- [10] O. Coulon, D. Alexander, and S. Arridge, *Diffusion tensor magnetic resonance image regularization*, Medical Image Analysis 8 (2004), pp. 47–67, Available at <http://www.sciencedirect.com/science/article/pii/S1361841503000720>.
- [11] I.L. Dryden, A.A. Koloydenko, and D. Zhou, *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, Ann. Appl. Statist. 3 (2009), pp. 1102–1123.
- [12] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*, Wiley & Sons, Chichester, 1998, Available at <https://www.maths.nottingham.ac.uk/personal/ild/book/>.
- [13] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Annals of Statistics 32 (2004), pp. 407–499, Available at <http://arxiv.org/pdf/math/0406456.pdf>.
- [14] P.T. Fletcher and S. Joshi, *Riemannian geometry for the statistical analysis of diffusion tensor data*, Signal Processing 87 (2007), pp. 250–262, Available at <http://www.sciencedirect.com/science/article/pii/S0165168406001691>.

- [15] P.T. Fletcher, S. Venkatasubramanian, and S. Joshi, *The geometric median on Riemannian manifolds with application to robust atlas estimation.*, NeuroImage 45 (2009), pp. S143–S152, Available at <http://dx.doi.org/10.1016/j.neuroimage.2008.10.052>.
- [16] M. Fréchet, *Les éléments aléatoires de nature quelconque dans un espace distancié*, Annals of the Institute of Henri Poincaré 10 (1948), pp. 215–310.
- [17] J. Gower, *Generalized procrustes analysis*, Psychometrika 40 (1975), pp. 33–51, Available at <http://dx.doi.org/10.1007/BF02291478>.
- [18] J. Gower and G. Dijksterhuis, *Procrustes Problems*, Series 30, Oxford Statistical Science, 2004.
- [19] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, 10th ed., Springer Series in Statistics, Springer, New York, 2013, Available at <http://dx.doi.org/10.1007/978-0-387-84858-7>, data mining, inference, and prediction.
- [20] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, *Kernel methods on the Riemannian manifold of symmetric positive definite matrices* (2014), Available at <http://arxiv.org/abs/1412.4172>.
- [21] J.T. Kent, *New directions in shape analysis*, in *The art of statistical science*, K.V. Mardia, ed., Wiley Ser. Probab. Math. Statist., Wiley, Chichester, 1992, pp. 115–127.
- [22] P.B. Kingsley, *Introduction to diffusion tensor imaging mathematics: Part II. Anisotropy, diffusion-weighting factors, and gradient encoding schemes*, Concepts in Magnetic Resonance Part A 28A (2006), pp. 123–154, Available at <http://dx.doi.org/10.1002/cmr.a.20049>.
- [23] P.B. Kingsley, *Introduction to diffusion tensor imaging mathematics: Part III. Tensor calculation, noise, simulations, and optimization*, Concepts in Magnetic Resonance Part A 28A (2006), pp. 155–179, Available at <http://dx.doi.org/10.1002/cmr.a.20050>.
- [24] C.G. Koay, J.D. Carew, A.L. Alexander, P.J. Basser, and M.E. Meyerand, *Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging*, Magnetic Resonance in Medicine 55 (2006), pp. 930–936, Available at <http://dx.doi.org/10.1002/mrm.20832>.
- [25] D. Le Bihan, J.F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat, *Diffusion tensor imaging: Concepts and applications*, Journal of Magnetic Resonance Imaging 13 (2001), pp. 534–546, Available at <http://dx.doi.org/10.1002/jmri.1076>.
- [26] M. Marcus and H. Minc, *A survey of matrix theory and matrix inequalities*, Dover Publications, Inc., New York, 1992, reprint of the 1969 edition.
- [27] M. Marcus and H. Minc, *A survey of matrix theory and matrix inequalities*, Dover Publications, Inc., New York, 1992, (English summary) Reprint of the 1969 edition.
- [28] L. Mirsky, *An introduction to linear algebra*, Oxford, at the Clarendon Press, 1955.
- [29] O. Pasternak, N. Sochen, and P. Basser., *Metric selection and diffusion tensor swelling*, Mathematics and Visualization, Springer-Verlag Berlin Heidelberg, 2012, pp. 323–336.
- [30] X. Pennec, P. Fillard, and N. Ayache, *A Riemannian framework for tensor computing*, International Journal of Computer Vision 66 (2006), pp. 41–66, Available at <http://dx.doi.org/10.1007/s11263-005-3222-z>.
- [31] D. Pigoli, J.A.D. Aston, I.L. Dryden, and P. Secchi, *Distances and inference for covariance operators*, Biometrika 101 (2014), pp. 409–422, Available at <http://dx.doi.org/10.1093/biomet/asu008>.
- [32] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004, Available at www.kernel-methods.net.
- [33] S.N. Sotiropoulos, L. Bai, P.S. Morgan, D.P. Auer, C.S. Constantinescu, and C.R. Tench, *A regularized two-tensor model fit to low angular resolution diffusion images using basis directions*, Journal of Magnetic Resonance Imaging 28 (2008), pp. 199–209, Available at <http://dx.doi.org/10.1002/jmri.21380>.
- [34] S. Sra, *Positive definite matrices and the S-divergence* (2013), Available at <http://arxiv.org/abs/1110.1773>, preprint.
- [35] P. Sundgren, Q. Dong, D. Gómez-Hassan, S. Mukherji, P. Maly, and R. Welsh, *Diffusion tensor imaging of the brain: review of clinical applications*, Neuroradiology 46 (2004), pp. 339–350, Available at <http://dx.doi.org/10.1007/s00234-003-1114-x>.
- [36] R. Tibshirani, *Regression shrinkage and selection via the lasso: a retrospective*, J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (2011), pp. 273–282, Available at <http://dx.doi.org/10.1111/j.1467-9868.2011.00771.x>.
- [37] D.S. Tuch, T.G. Reese, M.R. Wiegell, N. Makris, J.W. Belliveau, and V.J. Wedeen, *High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity*, Magnetic Resonance in Medicine 48 (2002), pp. 577–582, Available at <http://dx.doi.org/10.1002/mrm.10268>.
- [38] Z. Wang and B. Vemuri, *Tensor field segmentation using region based active contour model*, in *Computer Vision - ECCV 2004*, T. Pajdla and J. Matas, eds., Lecture Notes in Computer Science, Vol. 3024, Springer Berlin Heidelberg, 2004, pp. 304–315, Available at http://dx.doi.org/10.1007/978-3-540-24673-2_25.

- [39] C.A. Wheeler-Kingshott, G.J. Barker, S.C.A. Steens, and M.A. van Buchem, *D: The diffusion of water*, in *Quantitative MRI of the Brain: Measuring Changes Caused by Disease*, P. Tofts, ed., John Wiley & Sons, Ltd. Chichester, 2004, pp. 203–256, Available at <http://dx.doi.org/10.1002/0470869526.ch7>.
- [40] M.R. Wiegell, D.S. Tuch, H.B. Larsson, and V.J. Wedeen, *Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging*, *NeuroImage* 19 (2003), pp. 391 – 401, Available at <http://www.sciencedirect.com/science/article/pii/S1053811903000442>.
- [41] D. Zhou, *Statistical analysis for diffusion tensor imaging*, Ph.D. thesis, University of Nottingham, 2010.
- [42] D. Zhou, I.L. Dryden, A.A. Koloydenko, and B. Li, *A Bayesian method with reparameterization for diffusion tensor imaging*, in *Proceedings of SPIE Medical Imaging 2008: Image Processing*, Vol. 6914, Available at <http://dx.doi.org/10.1117/12.771697>, 2008, p. 69142J.
- [43] D. Zhou, I.L. Dryden, A.A. Koloydenko, and L. Bai, *Procrustes analysis of diffusion tensor data*, in *Proceedings of 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, Vol. 17, 2009, p. 3584.
- [44] D. Zhou, I.L. Dryden, A.A. Koloydenko, and L. Bai, *Procrustes analysis for diffusion tensor image processing*, *International Journal of Computer Theory and Engineering* 5 (2013), pp. 108–113.
- [45] L. Zhukov, K. Museth, D. Breen, A. Barr, and R. Whitaker, *Level set segmentation and modeling of DT-MRI human brain data*, *Journal of Electronic Imaging* 12 (2003).