

Manifold-valued image processing with SPD matrices

Xavier Pennec

* Inria and Université Côte d’Azur, Epione team, 2004 Route des Lucioles, BP93, 06902 Sophia-Antipolis Cedex, France
Corresponding: xavier.pennec@inria.fr

Abstract

Symmetric positive definite (SPD) matrices are geometric data that appear in many applications. They are used in particular in Diffusion tensor Imaging (DTI) as a simple model of the anisotropic diffusion of water in the tissues. This chapter extends the Riemannian computing statistical estimation framework of Chapter 1 and 2 to manifold-valued images with the example of SPD matrices. SPD matrices constitutes a smooth but non-complete manifold with the classical Euclidean metric on matrices. This creates important computational problems for image processing since one easily passes the boundaries to end-up with negative eigenvalues.

This chapter describes how we can give this space more interesting properties thanks to invariance: asking the Riemannian metric on covariance matrices to be invariant by affine changes of coordinates of the space leads to a one-parameter family of affine-invariant metrics that share the same affine connection. The geodesics are thus the same for all these metrics, even if the distance in the direction of scaling is different. With this structure, SPD matrices with null eigenvalues are at an infinite distance of any SPD matrix so that they cannot be reached in finite time. The space which is obtained is a typical example of a Hadamard space of non-positive curvature. It is also a symmetric Riemannian space of non-constant curvature. This can be seen thanks to the explicit computation of the Riemannian, sectional and Ricci curvatures.

Thanks to the relative simplicity of the geodesic equation, we can work out in detail the algorithms need for statistical estimation, for instance the Fréchet mean computation. Moreover, the knowledge of the Ricci curvature allows to see clearly what is controlling the difference between tangent PCA and Principal geodesic Analysis (PGA). Building on the reformulation of the mean value in manifolds as the minimization of an intrinsic functional, we also generalize many important image processing algorithms such as interpolation, filtering, diffusion and restoration of missing data to manifold-valued images.

We illustrate this framework on diffusion tensor image processing and on the modeling of the brain variability from a dataset of lines on the cerebral cortex. We also discuss alternative choices to the affine-invariant metrics. Log-Euclidean metrics, for instance, provide a very fast approximation when the data are concentrated with respect to the curvature. Other metrics give to the SPD space a positive curvature with a boundary at finite distance for rank-deficient matrices. Since the metric determines the main properties of the space, one should carefully review the assumptions that one can do on our data. Once the metric is chosen, the implementation of the geodesics (Exp and Log) provides once again the basis for all our manifold-valued image processing algorithms.

Chapter points

- Affine invariant metrics (Matrix exp and log, geodesics, exp and log maps)
- Interpolation and filtering as weighted means
- Harmonic and anisotropic diffusion
- Applications in DTI
- Other metrics on SPD matrices (log-Euclidean, Cholesky,...)

Manifold-valued image processing with SPD matrices	1
1. Introduction	3
2. Exponential, logarithm and square root of SPD matrices	6
2.1. <i>Differential of the matrix exponential</i>	7
2.2. <i>Differential of the matrix logarithm</i>	8
3. Affine invariant metrics	8
3.1. <i>Affine invariant distances</i>	9
3.2. <i>An invariant Riemannian metric</i>	9
3.3. <i>The one-parameter family of affine-invariant metrics</i>	12
3.4. <i>Curvature of affine-invariant metrics</i>	17
4. Basic statistical operations on SPD matrices	21
4.1. <i>Computing the mean and the covariance matrix</i>	21
4.2. <i>Tangent PCA and PGA of SPD matrices</i>	23
4.3. <i>Gaussian Distributions on SPD matrices</i>	23
5. Manifold-Valued Image Processing	27
5.1. <i>Interpolation</i>	27
5.2. <i>Gaussian and kernel-based filtering</i>	28
5.3. <i>Harmonic regularization</i>	29
5.4. <i>Anisotropic diffusion</i>	32
5.5. <i>Inpainting and extrapolation of sparse SPD fields</i>	32
6. Other metrics on SPD matrices	33
6.1. <i>Log Euclidean metrics</i>	33
6.2. <i>Cholesky metrics</i>	36
6.3. <i>Square root and Procrustes metrics</i>	36
6.4. <i>Extrinsic “distances”</i>	37
6.5. <i>Power Euclidean metrics</i>	37
6.6. <i>Which metric for which problem?</i>	38
7. Applications in Diffusion Tensor Imaging (DTI)	39
8. Learning Brain Variability from Sulcal Lines	44

1. Introduction

Symmetric positive definite (SPD) matrices are geometric data that appear in many applications. In medical imaging, SPD matrices are called tensors when they model the covariance matrix of the Brownian motion of water in Diffusion Tensor Imaging (DTI) [BMB94, LBMP⁺01]. They were also used to encode the joint variability at different places in shape analysis [FAP⁺05, FAP⁺07], and they are classically used in computer vision and image analysis to guide the segmentation, grouping and motion analysis [MLT00, WB02, BWBM06, WH06]. SPD matrices are also used as local metrics in numerical analysis to drive the size of adaptive meshes to solve PDEs in 3D more efficiently [BGM97]. In echo-Doppler or radar images, circular complex random processes with a null mean are characterized by Teoplitz Hermitian positive definite matrices [MSH05]. In brain-computer interfaces (BCI), the time-correlation of electro-encephalogram (EEG) signals are encoded through positive definite matrices and the choice of the metric on this space has been shown to drastically impact classification algorithms on these extremely low signal-to-noise ratio data [BBCJ12].

The space of SPD matrices is a smooth manifold which is not a vector space with the usual additive matrix structure. Indeed, a classical result in linear algebra states that P is SPD if and only if all its symmetric sub-matrices (including P itself) have non-negative determinant [Rao65]. Symmetric sub-matrices are obtained by removing at most $\dim(P) - 1$ rows and the corresponding columns. The space of SPD matrices is thus a subset of symmetric matrices delimited by multiple polynomial constraints on the coefficients. Since all these constraints are homogeneous (invariant under the multiplication by a positive scalar), this defines a convex half-cone in the vector space of symmetric matrices. Thus, convex operations like the mean are stable, but many other classical operations on SPD matrices are non convex and lead to matrices that are not positive definite. Gradient descent with the classical Froebenius (Euclidean) norm, for instance, amounts to evolve along a line in the vector space of symmetric matrices: one side of this line inevitably hit the boundary of the cone. When performing regularization of DTI images, there is thus almost inevitably a point in the image where the time step is not small enough and where we end-up with negative eigenvalues. A large part of the literature before 2006 was trying to get around these problems using the spectral decomposition of SPD matrices (e.g. [Tsc02, CTDF02]). However, processing independently the rotation (eigenvectors basis trihedron) and the eigenvalues is creating a continuity problem around equal eigenvalues.

To cope with that problem in the context of Diffusion Tensor Images (DTI), several authors proposed concurrently to consider Riemannian metrics on the space of SPD matrices which are invariant by affine change of the underlying space coordinates. The family of all affine-invariant metrics (up to a global scaling factor) induce the

Riemannian distance:

$$\text{dist}^2(P, Q) = \text{Tr}(L^2) + \beta \text{Tr}(L)^2 \quad \text{with} \quad L = \log(P^{-1/2}QP^{-1/2}), \quad (0.1)$$

where \log stands for the matrix logarithm and $\beta > -1/n$ is a free real parameter. The space of SPD matrices becomes a very regular Hadamard manifold structure (a space without cut-locus globally diffeomorphic to a Euclidean space). This considerably simplifies the computations compared to other non-invariant metrics. With this structure, symmetric matrices with null and infinite eigenvalues are both at an infinite distance of any SPD matrix: the cone of SPD matrices is mapped to a homogeneous space of non-constant curvature without boundaries. Moreover, there is a unique geodesic joining any two SPD matrices, the mean of a set of SPD matrices exists and is unique, and we can even define globally consistent orthonormal coordinate systems of tangent spaces. Thus, the structure obtained has many properties of Euclidean spaces even if it remains a manifold because of the curvature. The drawback with respect to the Euclidean case is the important increase of computational complexity due to the use matrix exponential and logarithms (computed through series or diagonalization of symmetric matrices) in the Riemannian distance / geodesics.

The affine-invariant metric with $\beta = 0$ has been put forward independently for several applications around 2005. [FJ07] used it for the analysis of principal modes of sets of diffusion tensors; [Moa05] analyzed its mathematical properties which were exploited in [BMA⁺05] for a new anisotropic DTI index; [PFA06] suggested it as the basis to develop the SPD matrix-valued images processing algorithms presented in this Chapter; [LRDF06] came to the same metric by looking for a natural metric on the space of Gaussian distributions for the segmentation of diffusion tensor images.

In statistics, this metric has been introduced in the 1980ies to model the geometry of the multivariate normal family [BR82, Sko84, CO91]. In this field, it is known as the Fisher-Rao metric. This metric is well known in other branches of mathematics [Bha03]. In computer vision, the metric was rediscovered to compute with covariance matrices [FM99]. The Levi-Civita connection of that metric was used in [HM94] to develop geometric integrators for flows of dynamic systems on the space of symmetric matrices. The geodesic walk along the geodesics of this metric was used for the anisotropic regularization of diffusion tensor images in [CTDF04] and [Bie04]. It is noticeable that so many different approaches lead to the same metric on the SPD space.

Although this affine-invariant Riemannian metric is often thought to be unique, a slightly different but still invariant metric was proposed in [LMO00] using a geometric embedding construction. It turns out that both metrics actually share the same affine invariant connection, so that there is an isometry between the two SPD spaces. The uniqueness of that affine invariant connection on SPD matrices is well known in differential geometry [Nom54, KN69, Hel78, Gam91]. However, the family of all affine invariant metrics was apparently not described before [Pen06b, Pen08].

Other metrics on SPD matrices with different properties can be constructed. By trying to put a Lie group structure on SPD matrices, Vincent Arsigny discovered the Log-Euclidean metrics [AFPA06, AFPA07, AFPA05]. These metrics give a vector space structure to the SPD matrix manifold while keeping most of the interesting affine-invariant properties (Section 6.1). This drastically simplifies the algorithms and speeds up computations. Other Riemannian or non Riemannian metrics were also proposed. We briefly present some of them for completeness in Section 6.

Chapter organization

In Section 2, we first describe the matrix exponential and logarithm function on symmetric matrices and their differential. Then we turn in Section 3 to the determination of Riemannian metrics that are invariant with respect to the natural action of the linear group on covariance matrices. We explicitly construct one affine-invariant Riemannian metric on SPD matrices, determine its geodesics and explicit formulas for the Riemannian exp and log maps. Then, we turn to the description of all the affine-invariant metrics. We show that there exists a one-parameter family of such metrics (up to a global scaling factor) which are sharing the same Levi-Civita connection. Establishing a globally orthonormal coordinate system for each of these metrics allows to compute explicitly the sectional, Ricci and scalar curvature. They are shown to be non-positive and bounded from below. Moreover, the expression of the curvatures is identical at all the points of the manifold although it is not constant in all directions. This is a feature of symmetric (and more generally homogeneous) spaces.

Section 4 illustrates how the statistical setting of Chapter 2 can be implemented on SPD matrices endowed with an affine-invariant metric. Because the space is Hadamard, the Fréchet mean is unique. However, the negative curvature has an important impact on the gradient descent algorithms and we illustrate cases where the time-step has to be adapted. The discussion of the generalized Gaussian distribution also illustrates how the Ricci curvature modifies the classical inverse relationship between the covariance matrix and the concentration parameter of the Gaussian.

We turn in Section 5 to the generalization of these statistical tools to more general manifold-valued image processing algorithms. Using weighted Fréchet means, we generalize many different types of interpolation, convolutions, isotropic and anisotropic filtering algorithms to manifold-valued images.

Because our manifold-valued image processing algorithms only depend on the chosen metric, they can also be used with other families of Riemannian and extrinsic metrics. We detail in Section 6 some of the metrics on SPD matrices that were proposed in the literature. Certain metrics like the log-Euclidean metrics do not display the full affine-invariance but conserve many of the good properties while providing a Euclidean structure to the space that considerably simplifies the computational frame-

work. Other metrics also give a symmetric space structure but of positive curvature: matrices with zero eigenvalues are then again at a finite distance of SPD matrices. There is indeed no canonical metric on SPD matrices that is suited to all applications: the choice of the metric should be guided by the invariance and computational properties that are important for the application.

Finally, the use of our manifold-valued image processing algorithms is illustrated in Section 7 with an application to the joint estimation and anisotropic smoothing of diffusion tensor images of the brain, and in Section 8 with the modeling of the variability of the brain anatomy.

2. Exponential, logarithm and square root of SPD matrices

Since we work with matrices in this chapter, we use capital letters like P and Q (instead of p and q or x and y in previous chapters) to denote the points on the manifold of $n \times n$ SPD matrices, and V and W for symmetric matrices (tangent vectors). The number of free parameters of symmetric and SPD matrices (the dimension of the manifold Sym_n^+) is $d = n(n+1)/2$.

In the sequel, we will extensively use the matrix exponential and logarithm functions. The exponential of a matrix W is defined through the convergent series $\exp(W) = \sum_{k=0}^{+\infty} \frac{W^k}{k!}$. We have drastic simplifications in the symmetric case thanks to the diagonalization $W = UDU^\top$, where U is an orthonormal matrix and $D = \text{DIAG}(d_i)$ is the diagonal matrix of eigenvalues. Indeed, powers of W can be written in the same basis: $W^k = UD^kU^\top$ and the rotation matrices can be factored out of the series. We are left with the exponential of each eigenvalue on the diagonal:

$$\exp(W) = \sum_{k=0}^{+\infty} \frac{W^k}{k!} = U \text{DIAG}(\exp(d_i)) U^\top. \quad (0.2)$$

For the inverse of the matrix exponential (the matrix logarithm), we may diagonalize any SPD matrix as $P = U \text{DIAG}(d_i) U^\top$, with (strictly) positive eigenvalues d_i . Thus, the function

$$\log(P) = U (\text{DIAG}(\log(d_i))) U^\top$$

is always well defined and realizes the inverse of the exponential of symmetric matrices. Moreover, the series defining the usual matrix log converges for small enough eigenvalues ($|d_i - 1| < 1$):

$$\log(P) = U \left(\text{DIAG} \left(\sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} (d_i - 1)^k \right) \right) U^\top = \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} (P - \text{Id})^k. \quad (0.3)$$

Classically, one defines the (left) square root of a matrix B as the set $\{B_L^{1/2}\} = \{A \in \text{GL}(n)/AA^\top = B\}$. One could also define the right square root: $\{B_R^{1/2}\} = \{A \in$

$GL(n)/A^\top A = B\}$. For SPD matrices, we define the symmetric square root as:

$$P^{1/2} = \{Q \in \text{Sym}_n^+ / Q^2 = P\}.$$

The symmetric square root of a SPD matrix is always defined and moreover unique: let $P = UD^2U^\top$ be a diagonalization (with positives values for the d_i 's). Then there exists one symmetric positive definite square root $Q = UDU^\top$. To prove that it is unique, let us consider two symmetric positive square roots Q_1 and Q_2 of P . Then, their square $Q_1^2 = P$ and $Q_2^2 = P$ obviously commute (because they are equal) and thus they can be diagonalized in the same basis: this means that the diagonal matrices D_1^2 and D_2^2 are equal in this common basis. As the elements of D_1 and D_2 are positive, they are also equal and $Q_1 = Q_2$.

More generally, one can define any power of a SPD matrix by taking the power of its eigenvalues or using the formula:

$$P^\alpha = \exp(\alpha(\log P)).$$

2.1. Differential of the matrix exponential

The matrix exponential and logarithm realize a one-to-one mapping between the space of symmetric matrices to the space of SPD matrices. Moreover, one can show that this mapping is diffeomorphic, since the differential has no singularities. Using the Taylor expansion of the matrix power $(W + \varepsilon V)^k = W^k + \varepsilon \sum_{i=0}^{k-1} W^i V W^{k-i-1} + O(\varepsilon^2)$ for $k \geq 1$, we obtain by identification the directional derivative $\partial_V \exp(W)$ by gathering the first order terms in ε in the series $\exp(W + \varepsilon V) = \sum_{k=0}^{+\infty} (W + \varepsilon V)^k / k!$:

$$\partial_V \exp(W) = (d \exp(W))(V) = \sum_{k=1}^{+\infty} \frac{1}{k!} \sum_{i=0}^{k-1} W^i V W^{k-i-1} \quad (0.4)$$

In order to simplify the formula, we insert the diagonalization $W = R S R^\top$ in the series to obtain:

$$\partial_V \exp(W) = R (\partial_{(R^\top V R)} \exp(S)) R^\top.$$

Thus, we are left with the computation of $\partial_V \exp(S)$ for S diagonal. As $[S^l V S^{k-l-1}]_{ij} = s_i^l v_{ij} s_j^{k-l-1}$, we have

$$[\partial_V \exp(S)]_{ij} = \left\{ \sum_{k=1}^{+\infty} \frac{1}{k!} \sum_{l=0}^{k-1} s_i^l s_j^{k-l-1} \right\} v_{ij} = v_{ij} \frac{\exp(s_i) - \exp(s_j)}{s_i - s_j}. \quad (0.5)$$

The value $a_{ij} = \frac{\exp(s_i) - \exp(s_j)}{s_i - s_j}$ is numerically unstable for almost equal eigenvalues s_i and s_j . However, we can rewrite this value as:

$$a_{ij} = \frac{\exp(s_i) - \exp(s_j)}{s_i - s_j} = \sum_{k=1}^{+\infty} \frac{1}{k!} \frac{s_i^k - s_j^k}{s_i - s_j} = \exp(s_j) \left(1 + \frac{(s_i - s_j)}{2} + \frac{(s_i - s_j)^2}{6} + O((s_i - s_j)^3) \right).$$

The Taylor expansion on the right hand side gives a numerically stable method to compute a_{ij} for equal eigenvalues. Moreover, since $\exp(s_i) > \exp(s_j)$ whenever $s_i > s_j$, we have $a_{ij} > 0$, so that we can conclude that $d \exp(S)$ is a diagonal linear form that is always invertible: the exponential of a symmetric matrix is a diffeomorphism.

2.2. Differential of the matrix logarithm

To compute the differential of the logarithm function, we can simply inverse the differential of the exponential as a linear form: as $\exp(\log(P)) = P$, we have $(d \log(P))(V) = (d \exp(\log(P)))^{-1} V$. Using $D = \exp(S)$, the inverse is easily expressed for a diagonal matrix: $[(d \exp(S))^{-1} V]_{ij} = v_{ij}/a_{ij}$. Thus we have:

$$[\partial_V \log(D)]_{ij} = v_{ij} \frac{\log(d_i) - \log(d_j)}{d_i - d_j}. \quad (0.6)$$

Like for the exponential, we can expand the value of a_{ij}^{-1} for close eigenvalues:

$$a_{ij}^{-1} = \frac{\log(d_i) - \log(d_j)}{d_i - d_j} = \frac{1}{d_j} \left(1 - \frac{d_i - d_j}{2d_j} + \frac{(d_i - d_j)^2}{3d_j^2} + O((d_i - d_j)^3) \right).$$

The Taylor expansion of the right hand side provides a numerically stable formulation for almost equal eigenvalues. Finally, using the identity $\log(P) = R^\top \log(RPR^\top) R$ for any rotation R , we have:

$$\partial_V \log(RDR^\top) = R (\partial_{R^\top V R} \log(D)) R^\top.$$

Using this formula, we may compute the differential at any point $P = RDR^\top$.

From the expression above, we can establish two very useful identities implying the differential of the log:

$$\partial_{\log(P)} \log(P) = P^{-1} \log(P) = \log(P) P^{-1} \quad (0.7)$$

$$\langle \partial_V \log(P), W \rangle = \langle \partial_W \log(P), V \rangle \quad (0.8)$$

3. Affine invariant metrics

Let us consider the action $(A, t) \cdot x = Ax + t$ of an element (A, t) of the affine group $\text{Aff}(n) = \text{GL}(n) \ltimes \mathbb{R}^n$ on a point $x \in \mathbb{R}^n$. The symbol \ltimes denotes a semi-direct product since the linear part of the transformation interferes with the translation in the composition of two affine transformations: $(A_1, t_1) \circ (A_2, t_2) = (A_1 A_2, A_1 t_2 + t_1)$. Now, if x is a random variable with mean \bar{x} and covariance matrix Σ_{xx} , then $y = Ax + t$ is a random variable with mean $\bar{y} = A\bar{x} + t$ and covariance matrix $\Sigma_{yy} = \mathbf{E}[(y - \bar{y})(y - \bar{y})^\top] = A \Sigma_{xx} A^\top$. Thus, the action of a linear transformation $A \in \text{GL}(n)$ on a covariance matrices $P \in \text{Sym}_n^+$ is:

$$A \cdot P = A P A^\top \quad \forall A \in \text{GL}(n) \quad \text{and} \quad P \in \text{Sym}_n^+.$$

3.1. Affine invariant distances

Following [PA98], any invariant distance on Sym_n^+ verifies $\text{dist}(A \cdot P_1, A \cdot P_2) = \text{dist}(P_1, P_2)$. Choosing $A = P_1^{-1/2}$, we can reduce this to the distance to the identity:

$$\text{dist}(P_1, P_2) = \text{dist}\left(\text{Id}, P_1^{-\frac{1}{2}} P_2 P_1^{-\frac{1}{2}}\right) = N\left(P_1^{-\frac{1}{2}} P_2 P_1^{-\frac{1}{2}}\right).$$

However, this should hold true for all transformations. In particular, the distance to identity N should be invariant by transformations that leaves the identity unchanged, i.e. the isotropy group of the identity $\mathcal{H}(\text{Id}) = \text{O}(n) = \{U \in \text{GL}(n) / UU^\top = \text{Id}\}$:

$$\forall U \in \text{O}(n), \quad N(UPU^\top) = N(P).$$

Using the spectral decomposition $P = UDU^\top$, we can conclude that $N(P)$ has to be a symmetric function of the eigenvalues $d_1 \geq d_2 \geq \dots d_n > 0$. Moreover, the symmetry of the invariant distance $\text{dist}(P, \text{Id}) = \text{dist}(\text{Id}, P^{-1})$ implies that $N(P) = N(P^{-1})$.

The sum of the squared logarithms of the eigenvalues is a candidate that satisfies these constraints:

$$N(P)^2 = \|\log(P)\|^2 = \sum_{i=1}^n (\log(d_i))^2. \quad (0.9)$$

By construction, $N\left(P_1^{-\frac{1}{2}} P_2 P_1^{-\frac{1}{2}}\right)$ fulfills the symmetry and definiteness axioms of the distance. Because $N(P) = 0$ implies that the eigenvalues are $d_i = 1$ (and conversely), the separation axiom is also verified. However, if the triangular inequality $N(P_1) + N(P_2) \geq N(P_1^{-1/2} P_2 P_1^{-1/2})$ can be verified numerically (see e.g. [FM99]), a formal proof is quite difficult to establish. Moreover, many other functions N may give rise to other valid distances.

3.2. An invariant Riemannian metric

Another way to determine the invariant distance is through a Riemannian metric. For that, we need to define the differential structure of the manifold.

3.2.1. Tangent vectors

Tangent vectors to SPD matrices are simply symmetric matrices with no constrain on the eigenvalues: if $\Gamma(t) = P + tW + O(t^2)$ is a curve on the SPD space, the tangent vector W is obviously symmetric, and there is no other constraint as symmetric and SPD matrices both have the same dimension $d = n(n+1)/2$.

Our group action naturally extends to tangent vectors: if $\Gamma(t) = P + tW + O(t^2)$ is a curve passing at P with tangent vector W , then the curve $A \cdot \Gamma(t) = APA^\top + tAWA^\top + O(t^2)$ passes through $A \cdot P = APA^\top$ with tangent vector $A \cdot W = AWA^\top$. The tangent space $T_P \text{Sym}_n^+$ at the point P is thus identified with the space of symmetric matrices.

3.2.2. Riemannian metric

On the tangent space at the identity matrix, we can now choose one of the most simple scalar product: if W_1 and W_2 are tangent vectors (i.e. symmetric matrices, not necessarily definite nor positive), the standard Froebenius scalar product on matrices is $\langle W_1, W_2 \rangle = \text{Tr}(W_1^T W_2)$. This scalar product is moreover invariant by the isotropy group $O(n)$. Now, if $W_1, W_2 \in T_P \text{Sym}_n^+$ are two tangent vectors at P , the invariance of the Riemannian metric under the action of $A \in GL(n)$ means that $\langle W_1, W_2 \rangle_P = \langle A \cdot W_1, A \cdot W_2 \rangle_{A \cdot P}$. Using $A = P^{-1/2}$, we see that we can define the metric at P from the metric at the identity:

$$\langle W_1, W_2 \rangle_P = \left\langle P^{-\frac{1}{2}} W_1 P^{-\frac{1}{2}}, P^{-\frac{1}{2}} W_2 P^{-\frac{1}{2}} \right\rangle_{\text{Id}} = \text{Tr}(P^{-1} W_1 P^{-1} W_2).$$

One can easily verify that using any other transformation $A = UP^{-1/2}$ (where U is a free orthonormal matrix) that transports P to the identity does not change the metric since $A \cdot P = APA^T = UU^T = \text{Id}$ and $A \cdot W_i = UP^{-1/2} W_i P^{-1/2} U^T$.

3.2.3. A Symmetric space structures

The invariant metric construction considers the space of SPD matrices as the quotient $\text{Sym}_n^+ = GL^+(n)/SO(n)$ (we take here the connected components of positive determinant to simplify), where $SO(n)$ is the isotropy group of the identity matrix taken as the origin of Sym_n^+ . In this homogeneous space, each SPD matrix P is seen as the equivalent class (a coset) of all invertible matrices $A \in GL^+(n)$ satisfying $A \text{Id} A^T = P$, i.e. $A = P^{1/2} V$ for $V \in SO(n)$. This polar decomposition of the positive linear group into the isotropy group times the SPD manifold can be expressed infinitesimally as the lie algebra decomposition $\mathfrak{gl}^+(n) = \mathfrak{so}(n) \times \mathfrak{sim}(n)$, where the Lie algebra of the isotropy subgroup $SO(n)$ is the algebra $\mathfrak{so}(n)$ of skew symmetric matrices (vertical vectors), and $\mathfrak{sim}(n)$ is a complementary space of horizontal vectors that we can choose to be the algebra of symmetric matrices, so that it is $Ad(SO(n))$ -invariant. It is important to realize that many different choices are possible for this complement, but the Ad -invariance and the invariant metric turn the homogeneous space into a naturally reductive homogeneous space for which geodesics are induced by one-parameter subgroups (see below).

In fact, we have in our case an even stronger structure: a symmetric space. Let us consider the mapping $s_P(Q) = P Q^{-1} P$. We have $s_P^{-1} = s_P$ and $s_P(P + \epsilon W) = P - \epsilon W + O(\epsilon^2)$ for $W \in T_P \text{Sym}_n^+$. Thus s_P is a smooth involution of Sym_n^+ that fixes P and acts on the tangent space at P by minus the identity. Moreover, it transforms a tangent vector W at Q into the tangent vector $W' = ds_P|_Q W = -P Q^{-1} W Q^{-1} P$ at $Q' = s_P(Q) = P Q^{-1} P$. Thus, we see that our Riemannian metric is actually invariant under this symmetry:

$$\langle W'_1, W'_2 \rangle_{Q'} = \text{Tr}((Q')^{-1} W'_1 (Q')^{-1} W'_2) = \text{Tr}(Q^{-1} W_1 Q^{-1} W_2) = \langle W_1, W_2 \rangle_Q.$$

Our symmetry is thus an isometry. These properties establish that the space of SPD matrices with the above metric is a Riemannian symmetric space, with more properties than general Riemannian manifolds. For instance, the curvature tensor is covariantly constant.

3.2.4. Geodesics

The determination of geodesics usually relies on the computation of the Christoffel symbols in general Riemannian manifolds. However, we may use more powerful tools since we have a naturally reductive homogeneous space. In these spaces, geodesics going through a point are generated by the action of the horizontal one-parameter subgroups of the acting Lie group on that point [Gam91, Hel78, KN69]. In our case, the one parameter subgroups of the linear group are the curves $\exp(tA)$ for $t \in \mathbb{R}$ and the horizontal vectors are the symmetric matrices $A = A^\top \in \text{sim}(n)$. Thus, the geodesic going through Id with tangent vector W need to have the form: $\Gamma_{(\text{Id}, W)}(t) = \exp(tA)$. $\text{Id} = \exp(tA) \exp(tA)^\top$ with $W = A + A^\top$. Because A is symmetric (horizontal), we find that:

$$\Gamma_{(\text{Id}, W)}(t) = \exp(tW). \quad (0.10)$$

We obtain the geodesics starting at another point P thanks to the invariance: the geodesic starting at $A \cdot P$ with tangent vector $A \cdot W$ is the transport by the group action of the geodesic starting at P with tangent vector W . Taking $A = P^{-1/2}$ gives $P^{-1/2} \cdot \Gamma_{(P, W)}(t) = \Gamma_{(\text{Id}, P^{-1/2} \cdot W)}(t)$, or in short:

$$\Gamma_{(P, W)}(t) = P^{\frac{1}{2}} \exp\left(tP^{-\frac{1}{2}}WP^{-\frac{1}{2}}\right)P^{\frac{1}{2}} \quad (0.11)$$

A different expression is appearing in some works. It can be related to this one thanks to the identity $\exp(WP^{-1})P = P \exp(P^{-1}W)$ which can be verified using the series expansion of the matrix exponential.

We now want to compute the distance by integrating the length along geodesics. Let $W = U \text{DIAG}(w_i) U^\top$ be a tangent vector at the identity. The tangent vector at time t to the geodesics $\Gamma_{(\text{Id}, W)}(t) = \exp(tW)$ starting from the identity along W is:

$$\frac{d\Gamma(t)}{dt} = U \text{DIAG}(w_i \exp(tw_i)) U^\top = \Gamma(t)^{\frac{1}{2}} W \Gamma(t)^{\frac{1}{2}} = \Gamma(t)^{\frac{1}{2}} \cdot W.$$

This is the transport of the initial tangent vector by the group action. Thus, thanks to our invariant metric, the norm of this vector is constant: $\|\Gamma(t)^{1/2} \cdot W\|_{\Gamma(t)^{1/2}, \text{Id}}^2 = \|W\|_{\text{Id}}^2$. This was expected since geodesics are parametrized by arc-length. Thus, the length of the curve between time 0 and 1 is

$$\mathcal{L} = \int_0^1 \left\| \frac{d\Gamma(t)}{dt} \right\|_{\Gamma(t)}^2 dt = \|W\|_{\text{Id}}^2.$$

Solving for $\Gamma_{(\text{Id}, W)}(1) = P$, we obtain the distance to identity $N(P) = \text{dist}^2(P, \text{Id}) = \|\log(P)\|_{\text{Id}}^2 = \sum_i (\log d_i)^2$ of Eq.(0.9). To obtain the distance from another point, we use the invariance properties of our metric, as in Section 3.1:

$$\text{dist}(P, Q) = \text{dist}\left(\text{Id}, P^{-\frac{1}{2}}QP^{-\frac{1}{2}}\right) = \left\|\log\left(P^{-1/2}QP^{-1/2}\right)\right\|_{\text{Id}}^2.$$

With the geodesic equation and the distance formula, we clearly see that SPD matrices with null eigenvalues are the limit at infinity of geodesics with positive eigenvalues everywhere. The other side of this geodesic has the corresponding eigenvalues going to infinity. Symmetric matrices with infinite and null eigenvalues are thus as far from any SPD matrix that they could be: at infinity. This contrasts with the flat Euclidean metric where zero eigenvalues PSD matrices are at finite distance: the cone of SPD matrices is not geodesically complete nor metrically complete with the Euclidean metric. From a computational point of view, the geodesic completeness of the affine-invariant structure ensures that any finite geodesic walking scheme will remain in the space of SPD matrices, contrarily to the Euclidean case.

3.2.5. Riemannian Exponential and Log Maps

As in all geodesically complete Riemannian manifolds, geodesics map the tangent space at P to a neighborhood of P in the manifold: $\Gamma_{(P, W)}(1) = \text{Exp}_P(W)$ associates to each tangent vector $W \in T_P \text{Sym}_n^+$ a point of the manifold. This mapping is a local diffeomorphism called the exponential map, because it corresponds to the usual exponential in some matrix groups. This is exactly our case for the exponential map around the identity:

$$\text{Exp}_{\text{Id}}(UDU^\top) = \exp(UDU^\top) = U \text{DIAG}(\exp(d_i)) U^\top.$$

However, the Riemannian exponential map associated to our invariant metric has a more complex expression at other SPD matrices:

$$\text{Exp}_P(W) = P^{\frac{1}{2}} \exp\left(P^{-\frac{1}{2}}WP^{-\frac{1}{2}}\right) P^{\frac{1}{2}}. \quad (0.12)$$

In our case, this diffeomorphism is global, and we can uniquely define the inverse mapping everywhere:

$$\overrightarrow{PQ} = \text{Log}_P(Q) = P^{\frac{1}{2}} \log\left(P^{-\frac{1}{2}}QP^{-\frac{1}{2}}\right) P^{\frac{1}{2}}. \quad (0.13)$$

Thus, Exp_P gives us a collection of one-to-one and complete charts of the manifold centered at any point P .

3.3. The one-parameter family of affine-invariant metrics

One can question the uniqueness of the previous affine-invariant Riemannian metric. Indeed, the previous construction uses one particular scalar product at the identity

(the Froebenius one), but other scalar products could also work. It was shown in [Pen06b, Pen08] that there is actually a one-parameter family of such affine-invariant Riemannian metrics on SPD matrices that all share the same connection. This is the affine invariant connection on homogeneous spaces of [Nom54] which is well known in symmetric spaces [KN69, Hel78, Gam91].

3.3.1. $GL(n)$ -invariant metrics

A $GL(n)$ -invariant scalar product has to verify $\langle V, W \rangle_P = \langle AVA^\top, AWA^\top \rangle_{APA^\top}$. In particular, this should be true for the isotropy group of the identity (the linear transformations that leave the identity matrix unchanged: the rotation matrices). All the rotationally invariant scalar products on symmetric matrices are given (up to a constant global multiplicative factor) by:

$$\langle V, W \rangle_{\text{Id}} = \text{Tr}(VW) + \beta \text{Tr}(V)\text{Tr}(W) \quad \text{with} \quad \beta > -\frac{1}{n}$$

where n is the dimension of the space (see for instance [Pen06b, Pen08]). The sketch of the proof is the following: these scalar products are derived from rotationally invariant norms $\|W\|^2$, which are quadratic forms on (symmetric) matrices. By isotropy, such forms can only depend on the matrix invariants $\text{Tr}(W)$, $\text{Tr}(W^2)$, $\text{Tr}(W^3)$, \dots . However, as the form is quadratic in W , we are left only with $\text{Tr}(W)^2$ and $\text{Tr}(W^2)$ that can be weighted by α and β . One easily verifies that $\beta > -\alpha/n$ is a necessary and sufficient condition to ensure positive definiteness. This metric at the identity can then be transported at any point by the group action using the (symmetric or any other) square root $P^{1/2}$ considered as a group element:

$$\begin{aligned} \langle V, W \rangle_P &= \langle P^{-1/2}VP^{-1/2}, P^{-1/2}WP^{-1/2} \rangle_{\text{Id}} \\ &= \alpha \text{Tr}(VP^{-1}WP^{-1}) + \beta \text{Tr}(VP^{-1})\text{Tr}(WP^{-1}) \end{aligned}$$

Theorem 1 (Family of affine-invariant metrics on SPD matrices [Pen06b, Pen08]). *All the metrics on the space of SPD matrices Sym_n^+ that are invariant under the $GL(n)$ action $A \cdot P = APA^\top$ are given by the 1-parameter family (up to a global scaling factor):*

$$\langle V, W \rangle_P = \text{Tr}(VP^{-1}WP^{-1}) + \beta \text{Tr}(VP^{-1})\text{Tr}(WP^{-1}) \quad \text{with} \quad \beta > -\frac{1}{n}. \quad (0.14)$$

For $\beta = 0$, we retrieve the affine-invariant metric that was proposed in [Sko84, FM99, Bha03, FJ07, LRDF06, PFA06]. Up to our knowledge, the only case of a different invariant metric was proposed by [LMO00] with $\beta = -1/(n+1)$. This metric was obtained by embedding the space of SPD matrices of dimension n into the space of $n+1$ square matrices using homogeneous coordinates and by quotienting out $n+1$ dimensional rotations. This type of embedding is interesting as it allows to represent also the mean of a Gaussian distributions in addition to its covariance matrix. The

embedding can be done in spaces of square matrices of dimension $n + k$ ($k \geq 1$), in which case one would obtain the invariant metric with $\beta = -1/(n + k)$. Interestingly, the metric proposed by [LMO00] with $-1/\beta = n + 1$ is the first authorized integer to obtain a proper metric.

Although many people favor the value $\beta = 0$, there is no invariance reason to do so. Indeed, the general linear group $GL(n)$ can be seen as the direct product $GL(n) = SL(n) \otimes \mathbb{R}^*$ where $SL(n)$ is the unimodular group (unit determinant matrices) and \mathbb{R}^* is the multiplicative group of real numbers. Likewise, SPD matrices can be decomposed into unimodular SPD matrices $\mathcal{USym}_n^+ = \{P \in \text{Sym}_n^+, \det(P) = 1\}$ and their determinant in \mathbb{R}_*^+ : $\text{Sym}_n^+ = \mathcal{USym}_n^+ \otimes \mathbb{R}_*^+$. The split decomposes an element of the tangent space at identity $W \in T_{\text{Id}}\text{Sym}_n^+$ into the traceless part $W_0 = W - \text{Tr}(W)/n \text{Id}$ and the trace part $\text{Tr}(W) \in \mathbb{R}$. Since the group action splits into two independent group actions, we can choose an invariant metric on each part and recombine them afterward. For the unimodular part \mathcal{USym}_n^+ , the $SL(n)$ -invariant metric at $P \in \mathcal{USym}_n^+$ is uniquely defined by $\langle V_0, W_0 \rangle_P = \text{Tr}(V_0 P^{-1} W_0 P^{-1})$ (up to a scale factor), since the trace part of each tangent vector vanishes. Likewise, the multiplicative-invariant metric on the scale is unique (up to a scale factor). Now, when recombining both parts, we can independently scale the two parts with the direct product, which exactly describes all the $GL(n)$ -invariant metrics on SPD matrices.

In the limit case $\beta = -1/n$, the bilinear form becomes degenerate in the direction of the trace of the matrix at the identity: $\|V - \text{Tr}(V) \text{Id}\|_{\text{Id}} = \|V\|_{\text{Id}}$. This direction in the tangent space corresponds to the scalar multiplication of SPD matrices: the SPD matrix sP (for any $s > 0$) is at null distance of the SPD matrix P with this metric. Thus, this is still defining a metric on the space of unimodular SPD matrices, but not on the full SPD matrix space. For $\beta < -1/n$, the bilinear form has a negative eigenvalue in the trace direction and defines a semi-Riemannian metric.

3.3.2. Different metrics for a unique affine connection

The Koszul formula below is the key step to establish the uniqueness of the Levi-Civita connection of a Riemannian metric [Car92, p.55]:

$$2 \langle \nabla_V W, U \rangle = \partial_V \langle W, U \rangle + \partial_W \langle V, U \rangle - \partial_U \langle V, W \rangle \\ + \langle [V, W], U \rangle - \langle [V, U], W \rangle - \langle [W, U], V \rangle.$$

In our case, expanding the terms in the Koszul formula by chain rule and using $\partial_X \Sigma^{-1} = -\Sigma^{-1}(\partial_X \Sigma) \Sigma^{-1}$, we obtain that all the affine-invariant metrics have the same Levi-Civita connection:

$$\nabla_V W = \partial_V W - \frac{1}{2}(VP^{-1}W + WP^{-1}V). \quad (0.15)$$

This formula is in agreement with the connection computed for $\beta = 0$ by [Sko84] (save that $\partial_V W$ is omitted in this work).

Since all the invariant metrics have the same Levi-Civita connection, they share the same geodesics and the Riemannian exp and log maps at each point:

$$\begin{aligned}\text{Exp}_P(W) &= P^{1/2} \exp(P^{-1/2} W P^{-1/2}) P^{1/2} \\ \overrightarrow{PQ} = \text{Log}_P(Q) &= P^{1/2} \log(P^{-1/2} Q P^{-1/2}) P^{1/2}\end{aligned}$$

However, one should be careful that the orthonormal bases are different for each metric which means that distances along the geodesics are different. The Riemannian distance is obtained as before by integration, or more easily by the norm of the initial tangent vector of the geodesic joining the two points:

$$\begin{aligned}\text{dist}^2(P, Q) &= \|\text{Log}_P(Q)\|_P^2 = \|P^{-1/2} \text{Log}_P(Q) P^{-1/2}\|_{\text{Id}}^2 \\ &= \text{Tr}(\log(Q P^{-1})^2) + \beta \text{Tr}(\log(Q P^{-1}))^2.\end{aligned}$$

3.3.3. Orthonormal coordinate systems

For many computations, it is convenient to use a minimal representation (e.g. 6 parameters for 3×3 SPD matrices) in an orthonormal basis. In classical Euclidean matrix spaces, this can be realized through the classical “Vec” operator that maps the element $a_{i,j}$ of a $n \times n$ matrix A to the $(in + j)$ th element $\text{Vec}(A)_{in+j}$ of a n^2 dimensional vector $\text{Vec}(A)$. Since we are working with symmetric matrices, we have only $d = n(n + 1)/2$ independent coefficients, say the upper triangular part of the symmetric matrix $W \in T_P \text{Sym}_n^+$. This corresponds to take the basis vectors $F_{ii} = e_i e_i^\top$ and $F_{ij} = e_i e_j^\top + e_j e_i^\top$ ($1 \leq i < j \leq n$) for the space of symmetric matrices, where the e_i are the standard basis vectors of \mathbb{R}^n . However, this basis is not orthonormal at the identity: a direct computation of the Gram matrix

$$g_{ij,kl} = \langle F_{ij}, F_{kl} \rangle_{\text{Id}} = \text{Tr}(F_{ij} F_{kl}) + \beta \text{Tr}(F_{ij})(F_{kl})$$

shows that the non-zero terms are $g_{ii,ii} = \|F_{ii}\|_{\text{Id}}^2 = 1 + \beta$, $g_{ii,jj} = \beta$ for $j \neq i$ and $g_{ij,ij} = \|F_{ij}\|_{\text{Id}}^2 = 2$ for $i < j$.

A field of orthonormal bases for $\beta = 0$

The previous non-orthonormal basis can be easily corrected for $\beta = 0$ by normalizing the basis vectors for the off-diagonal coefficient: an orthonormal basis of $T_{\text{Id}} \text{Sym}_n^+$ for the affine-invariant metric with $\beta = 0$ is given by the vectors:

$$E_{ij}^0 = \begin{cases} e_i e_i^\top & (1 \leq i = j \leq n), \\ (e_i e_j^\top + e_j e_i^\top) / \sqrt{2} & (1 \leq i < j \leq n). \end{cases}$$

The vector of coordinates in this basis is:

$$\text{Vec}_{\text{Id}}^0(W) = (w_{1,1}, \dots, w_{n,n}, \sqrt{2} w_{1,2}, \dots, \sqrt{2} w_{(n-1)n})^\top.$$

This realizes an explicit isomorphism between $T_{\text{Id}}\text{Sym}_n^+$ endowed with the Froebenius metric and \mathbb{R}^d with the L_2 metric.

It is important to notice that the above basis is orthonormal only at the identity and not at other places due to the curvature of the Riemannian manifold: a field of orthonormal bases usually depends on the base-point P of the tangent space $T_P\text{Sym}_n^+$. Such a frame-field can be obtained by the group action since we are dealing with a naturally reductive homogeneous space: $E_{ij}^0|_P = P^{\frac{1}{2}} E_{ij}^0 P^{\frac{1}{2}}$ with $1 \leq i \leq j \leq n$. Because there are no closed geodesics, $E_{ij}^0|_P$ is a smooth vector field over Sym_n^+ , and the set of vectors fields $\{E_{ij}^0|_P\}_{1 \leq i \leq j \leq n}$ constitute a smooth global frame-field. Moreover, the vector of coordinates $\text{Vec}_P^0(W) = \text{Vec}_{\text{Id}}^0(P^{-1/2} \cdot W) = \text{Vec}_{\text{Id}}^0(P^{-1/2} W P^{-1/2})$ in the frame at point P also realizes an explicit isomorphism between $T_P\text{Sym}_n^+$ with the metric $\langle U, V \rangle_P^0 = \text{Tr}(U P^{-1} V P^{-1})$ and \mathbb{R}^d with the canonical L_2 metric.

A field of orthonormal bases for $\beta \neq 0$

To obtain an orthonormal basis for the affine-invariant metric with $\beta \neq 0$, we can build an isomorphism of Riemannian manifolds by modifying the trace part. First, we observe that $A_\alpha(V) = V - \alpha \text{Tr}(V) \text{Id}$ is a linear map of the space of symmetric matrices identified to $T_{\text{Id}}\text{Sym}_n^+$. This map leaves the off-diagonal coefficients unchanged and transforms the diagonal basis vectors according to:

$$E_{ii}^\beta = A_\alpha(E_{ii}^0) = E_{ii}^0 - \alpha \sum_{j=1}^n E_{jj}^0.$$

Thus, when restricted to the basis $(E_{11}^0, \dots, E_{nn}^0)$ this mapping can be written in matrix form $A_\alpha = \text{Id}_n - \alpha \mathbb{1}_n \mathbb{1}_n^\top$ with $\mathbb{1}_n^\top = (1, \dots, 1)$ (we should be careful that this matrix just operate on the diagonal of symmetric matrices and not on the standard Euclidean space \mathbb{R}^n). Its determinant is $\det(A_\alpha) = \det(\text{Id}_n - \alpha \mathbb{1}_n \mathbb{1}_n^\top) = 1 - \alpha n$, so that the linear map has positive determinant if $\alpha < 1/n$. Thus, we can invert it by Sherman-Morrison formula to get: $A_\alpha^{-1} = \text{Id} + \alpha \mathbb{1}_n \mathbb{1}_n^\top / (1 - \alpha n)$.

Now we can look for the value of α such that the mapping A_α transforms the affine-invariant metric $\langle \cdot, \cdot \rangle_P$ of Eq. (0.14) with $\beta \neq 0$ to the previous affine invariant metric with $\beta = 0$. This equation reads $\langle A_\alpha(V), A_\alpha(W) \rangle_P = \langle A_0(V), A_0(W) \rangle_P$, or:

$$\text{Tr}(A_\alpha(V) P^{-1} A_\alpha(W) P^{-1}) + \beta \text{Tr}(A_\alpha(V) P^{-1}) \text{Tr}(A_\alpha(W) P^{-1}) = \text{Tr}(V P^{-1} W P^{-1}).$$

Thanks to the invariance, this is verified for all P if this holds for $P = \text{Id}$. Plugging $A_\alpha(V) = V - \alpha \text{Tr}(V) \text{Id}$ in the above equation with $P = \text{Id}$ leads to

$$\text{Tr}((V - \alpha \text{Tr}(V) \text{Id})(W - \alpha \text{Tr}(W) \text{Id})) + \beta(1 - n\alpha)^2 \text{Tr}(V) \text{Tr}(W) = \text{Tr}(VW).$$

Thus, we have to solve $n\alpha^2 - 2\alpha + \beta(1 - n\alpha)^2 = 0$. This leads to the second order equation $n\alpha^2 - 2\alpha + \beta/(1 + n\beta) = 0$. Because $\beta > -1/n$, it has two solutions and the one that gives a positive determinant for the linear mapping A_α is $\alpha = \frac{1}{n} \left(1 - 1/\sqrt{1 + n\beta} \right)$. One can now verify by direct computation of the Gram matrix that the vectors

$$E_{ij} = \begin{cases} e_i e_i^\top - \alpha \text{Id} & (1 \leq i = j \leq n), \\ (e_i e_j^\top + e_j e_i^\top) / \sqrt{2} & (1 \leq i < j \leq n), \end{cases} \quad \text{with} \quad \alpha = \frac{1}{n} \left(1 - \frac{1}{\sqrt{1 + n\beta}} \right)$$

constitute an orthonormal basis of $T_{\text{Id}} \text{Sym}_n^+$ with the affine-invariant metric with any value of $\beta > -1/n$. By invariance, the vectors $\{E_{ij}|_P = P^{1/2} E_{ij} P^{1/2}\}_{1 \leq i \leq j \leq n}$ form an orthonormal basis of $T_P \text{Sym}_n^+$ for the affine-invariant metric.

To obtain a mapping $\text{Vec}_P(W)$ that realizes an explicit isomorphism between $T_P \text{Sym}_n^+$ with the general affine-invariant metric and \mathbb{R}^d with the canonical L_2 metric, we have first to transport W to the tangent space at identity using the action of $P^{-1/2}$ and then use the inverse mapping A_α^{-1} on the diagonal coefficients before using the previous mapping Vec_{Id}^0 :

$$\text{Vec}_P(W) = \text{Vec}_{\text{Id}}(P^{-\frac{1}{2}} W P^{-\frac{1}{2}}) = \text{Vec}_{\text{Id}}^0(A_\alpha^{-1}(P^{-\frac{1}{2}} W P^{-\frac{1}{2}})).$$

Since $A_\alpha^{-1}(V) = V + \alpha \text{Tr}(V) \text{Id} / (1 - \alpha n) = V + \frac{1}{n} (\sqrt{1 + \beta n} - 1) \text{Tr}(V) \text{Id}$, we get the following theorem.

Theorem 2 (Orthonormal field for the affine invariant metric). *The $d = n(n - 1)/2$ vectors*

$$E_{ij}|_P = \begin{cases} P^{\frac{1}{2}} e_i e_i^\top P^{\frac{1}{2}} - \alpha P & (1 \leq i = j \leq n), \\ P^{\frac{1}{2}} (e_i e_j^\top + e_j e_i^\top) P^{\frac{1}{2}} / \sqrt{2} & (1 \leq i < j \leq n), \end{cases} \quad \text{with} \quad \alpha = \frac{1}{n} \left(1 - \frac{1}{\sqrt{1 + n\beta}} \right),$$

form an orthonormal basis of $T_P \text{Sym}_n^+$ with the affine-invariant metric for any value of $\beta > -1/n$. Moreover, the mapping $\text{Vec}_P(W) = \text{Vec}_{\text{Id}}(P^{-\frac{1}{2}} W P^{-\frac{1}{2}})$ with

$$\text{Vec}_{\text{Id}}(W) = (w_{1,1} + \delta, \dots, w_{n,n} + \delta, \sqrt{2}w_{1,2}, \dots, \sqrt{2}w_{(n-1),n})^\top \quad \text{and} \quad \delta = \frac{1}{n} (\sqrt{1 + \beta n} - 1) \text{Tr}(W)$$

realizes an explicit isomorphism between $T_P \text{Sym}_n^+$ with the general affine-invariant metric and \mathbb{R}^d with the canonical L_2 metric.

3.4. Curvature of affine-invariant metrics

Thanks to the explicit expression of the connection, one can compute the Riemannian curvature tensor $R(V, W)U = \nabla_V \nabla_W U - \nabla_W \nabla_V U - \nabla_{[V, W]} U$. We find:

$$R(V, W)U = \frac{1}{4} (WP^{-1}VP^{-1}U + UP^{-1}VP^{-1}W - VP^{-1}WP^{-1}U - UP^{-1}WP^{-1}V).$$

Equivalently the $(0, 4)$ -Riemannian curvature tensor is given by

$$R(U, V, W, Z) = \langle R(U, V)Z, W \rangle = \frac{1}{2} \text{Tr} \left(UP^{-1}VP^{-1}WP^{-1}ZP^{-1} - UP^{-1}VP^{-1}ZP^{-1}WP^{-1} \right).$$

The conventions used here are the one of Section 1.4.4 of Chapter 1. Notice that there is a factor -2 with respect to the formula of [Sko84] which is due to a global scaling of $1/2$ of their Riemannian metric with respect to ours when $\beta = 0$ and an opposite convention for the $(0, 4)$ Riemannian tensor.

3.4.1. Sectional Curvature

From this tensor, one gets the sectional curvature $\kappa(U, V)|_P$ in the subspaces spanned by the vector fields U and V at P (see Eq.?? in Chapter ??). We recall that this measures the Gauss curvature of the 2D geodesic surface generated by linear combinations of the two vectors. This is the function:

$$\kappa(U, V) = \frac{\langle R(U, V)U, V \rangle}{\langle U, U \rangle \langle V, V \rangle - \langle U, V \rangle^2} = \frac{R(U, V, U, V)}{\langle U, U \rangle \langle V, V \rangle - \langle U, V \rangle^2}.$$

This sectional curvature can be computed easily in our previous orthonormal basis $E_{ij}|_P$. We find that the only non-zero terms are independent of P :

$$\begin{aligned} \kappa(E_{ii}, E_{ij}) &= \kappa(E_{ii}|_P, E_{ij}|_P) = \kappa(E_{ij}|_P, E_{jj}|_P) = -\frac{1}{4} \quad (j \neq i), \\ \kappa(E_{ij}, E_{ik}) &= \kappa(E_{ij}|_P, E_{ik}|_P) = \kappa(E_{ij}|_P, E_{kj}|_P) = -\frac{1}{8} \quad (i \neq j \neq k \neq i). \end{aligned}$$

Other 2-subspaces are obtained by rotating the vectors $U|_P$ and $V|_P$ with respect to our orthonormal basis vectors. The sectional curvature is thus in between $\kappa_{\min} = -1/4$ and $\kappa_{\max} = 0$. In consequence, the manifold of SPD matrices with any affine-invariant metric has bounded non-positive curvature and is a Hadamard manifold.

Thanks to our well chosen orthonormal basis, it is immediately clear that the sectional curvature is “the same” at every point of the manifold. This is *not* a constant curvature since it varies depending on the chosen 2-subspace within the tangent space, but it is comparable at every point up to a rotation. This is a feature of homogeneous manifolds: there exists a way to compare the geometry at every point thanks to the group action. The fact that the sectional curvature does not depend on β was expected since all the affine-invariant Riemannian spaces are isomorphic to each-other.

To compare our results to those of [Sko84] for $\beta = 0$, we have to compute the curvature in the planes spanned by two vectors of the basis $\{F_{ij}\}$. We recall that this basis is orthogonal but not orthonormal at Id , and not orthogonal at other values of P . In our case, we find $\kappa(F_{ii}, F_{ij})|_P = -1/4$. Because of the non-orthogonality of this basis at $P \neq \text{Id}$, we also get the extra non-vanishing term $\kappa(F_{ii}, F_{ij})|_P = -\frac{1}{2}\bar{\rho}_{ij}^2/(1 + \bar{\rho}_{ij}^2)$, where $\bar{\rho}_{ij}^2 = (e_i^\top P^{-1}e_j)^2/(e_i^\top P^{-1}e_i e_j^\top P^{-1}e_j)$ is the ij partial correlation coefficient of P . These results are consistent with the ones of [Sko84], up to a factor 2 due the

global scaling of the metric. Surprisingly, the non-vanishing term $\kappa(E_{ij}, E_{ik}) = -1/8$ seems to be forgotten in their work.

3.4.2. Ricci Curvature

The Ricci curvature is the trace of the linear operator $U \rightarrow R(U, V)Z$. In coordinates, this writes $\text{Ric}_{ab} = R_{acb}^c = g^{dc} R_{dacb} = \sum_c R(e_c, e_a, e_c, e_b)$, where the right-hand-side is only valid in an orthonormal basis $\{e_c\}$. Because this is a bilinear form, it is entirely described by the quadratic form $\text{Ric}(V)|_P = \text{Ric}(V, V)|_P$ for all the unit vectors in the tangent space at P . $\text{Ric}(V)$ is thus the average (up to a scale factor n) of the sectional curvatures $\kappa(V, W)$ of 2-planes of the tangent space containing the vector V , assuming that W follows a uniform distribution on the unit sphere in the tangent space. The scalar curvature is the metric trace of the Ricci curvature matrix: $R = \text{Tr}_g(\text{Ric}) = g^{ab} R_{ab}$. The scalar curvature describes how the volume of geodesic ball in a manifold deviates from the volume of standard euclidean ball with the same radius.

In dimension 2 and 3, the Ricci curvature determines the Riemannian curvature tensor, like the sectional curvature. In dimension greater than 3, the sectional curvature continues to determine the full curvature tensor but the Ricci curvature generally contains less information. However, it encodes important information about the diffusion on the manifold (using the Laplace-Beltrami operator) since $\text{Ric}(V)|_P$ encodes how the volume of an infinitesimal neighborhood of a point evolves when it is transported along the geodesic starting at P in the direction V . In a normal coordinate system at P , we have indeed the Taylor expansion or the Riemannian measure:

$$d\mathcal{M}(\text{Exp}_P(V)) = \left(1 - \frac{1}{6}\text{Ric}(V)|_P + O(\|V\|_P^3)\right) dV.$$

A related formulation is developed in the excellent paper of Yann Ollivier *a visual introduction to curvature* [Oll13]: if C is a small neighborhood of P , then the volume of the neighborhood $C_V = \{\exp_Q(\Pi_P^Q V), Q \in C\}$ transported along the geodesics in the direction V evolves at second order in V with the multiplicative factor $(1 - \frac{1}{6}\text{Ric}(V)|_P)$. Thus, a negative Ricci curvature indicates a volume expansion of the neighborhood while a positive curvature indicates a volume contraction in the direction V .

Coming back to SPD matrices with the affine-invariant metric, we want to compute the Ricci curvature in our orthonormal basis $\{E_{ij}|_P\}$ at any point P . Using $E_{ij}|_P = P^{1/2} E_{ij} P^{1/2}$ and the above expression of the $(0, 4)$ Riemannian tensor, we get that:

$$R(E_{ab}|_P, E_{ij}|_P, E_{ab}|_P, E_{kl}|_P) = \frac{1}{2} \text{Tr} \left(E_{ab} E_{ij} E_{ab} E_{kl} - E_{ab} E_{ij} E_{kl} E_{ab} \right).$$

Thus, we see that the Ricci curvature matrix at any point is the same than at identity

in our specific coordinate system:

$$\text{Ric}(E_{ij}, E_{kl}) = \sum_{1 \leq a \leq b \leq n} R(E_{ab}, E_{ij}, E_{ab}, E_{kl}) = \frac{1}{2} \sum_{1 \leq a \leq b \leq n} \text{Tr}(E_{ab} E_{ij} E_{ab} E_{kl} - E_{ab} E_{ij} E_{kl} E_{ab}).$$

Lengthy computations verified by computer show that only the following terms do not vanish: $\text{Ric}(E_{ii}, E_{ii}) = -\frac{n-1}{4}$, $\text{Ric}(E_{ij}, E_{ij}) = -\frac{n}{4}$ for $i \neq j$, and $\text{Ric}(E_{ii}, E_{jj}) = \frac{1}{4}$. Ordering the n diagonal basis vectors E_{ii} before the $n(n-1)/2$ off-diagonal basis vectors E_{ij} ($i < j$), we obtain the following diagonal by bloc Ricci curvature matrix:

$$\text{Ric} = -\frac{n}{4} \begin{bmatrix} \text{Id}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\top & 0 \\ 0 & \text{Id}_{n(n-1)/2} \end{bmatrix}.$$

We easily see that this matrix has one eigenvector $(\mathbb{1}_n^\top, 0_{n(n-1)/2})^\top$ with eigenvalue 0 and $d-1$ orthogonal eigenvectors associated to the multiple eigenvalue $-n/4$. The null eigenvalue along the trace part in the tangent space of SPD matrices at identity (which corresponds to a scalar multiplication of SPD matrices when it is exponentiated) means that the volume of a neighborhood remains constant when we shoot in the direction of SPD matrix rescaling. In all other directions, the Ricci curvature is negative, which indicate that a neighborhood is volume increasing.

Theorem 3 (Curvature of affine-invariant metrics). *The space Sym_n^+ of SPD matrix of dimension $n \geq 2$ endowed with the affine invariant metric:*

$$\langle V, W \rangle_P = \text{Tr}(VP^{-1}WP^{-1}) + \beta \text{Tr}(VP^{-1}) \text{Tr}(WP^{-1}) \quad \text{with } \beta > -\frac{1}{n},$$

has Riemannian curvature tensor:

$$R(U, V, W, Z) = \langle R(U, V)Z, W \rangle = \frac{1}{2} \text{Tr}(UP^{-1}VP^{-1}WP^{-1}ZP^{-1} - UP^{-1}VP^{-1}ZP^{-1}WP^{-1}). \quad (0.16)$$

The sectional curvature is:

$$\kappa(E_{ij}, E_{ik}) = \kappa(E_{ij}, E_{ki}) = \begin{cases} -\frac{1}{8} & \text{if } i \neq j \neq k \neq i, \\ -\frac{1}{4} & \text{if } j = i \neq k \text{ or } k = i \neq j, \\ 0 & \text{otherwise.} \end{cases} \quad (0.17)$$

The Ricci curvature matrix is bloc-diagonal:

$$\text{Ric} = -\frac{n}{4} \begin{bmatrix} \text{Id}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\top & 0 \\ 0 & \text{Id}_{n(n-1)/2} \end{bmatrix}. \quad (0.18)$$

It has one vanishing eigenvalue along the trace-part and $(d-1)$ orthogonal eigenvectors associated to the multiple eigenvalue $-n/4$. Finally, the scalar curvature is:

$$R = -\frac{n(n-1)(n+2)}{8}. \quad (0.19)$$

The sectional, Ricci and scalar curvatures are non-positive.

4. Basic statistical operations on SPD matrices

Now that we have the atomic Exp and Log maps for SPD matrices and that we have reviewed the main geometrical properties of the affine-invariant structure on SPD matrices, let us turn to statistical inference methods in this space. As the manifold has a non-positive curvature and there is no cut locus (the injection radius is infinite), the statistical properties detailed in previous chapters hold in their most general form.

4.1. Computing the mean and the covariance matrix

Let $P_1 \dots P_m$ be a set of measurements of SPD matrices. It can be seen as a distribution using a mixture of delta Diracs at the sample points: $\mu = \frac{1}{m} \sum_{i=1}^m \delta_{P_i}$. We recall from Chapter 2 that the empirical Fréchet mean is the set of SPD matrices minimizing the sum of squared distances $\text{Var}(P) = \frac{1}{m} \sum_{i=1}^m \text{dist}^2(P, P_i)$. Since the manifold has no cut-locus, the variance is everywhere smooth.

The first order moment of the sample distribution is the contravariant vector field defined all over the SPD matrix manifold by:

$$\mathfrak{M}_1(P) = \int_{\mathcal{M}} \overrightarrow{PQ} \mu(dQ) = \frac{1}{m} \sum_{i=1}^m \overrightarrow{PP_i} = \frac{1}{m} \sum_{i=1}^m P^{\frac{1}{2}} \log(P^{-\frac{1}{2}} P_i P^{-\frac{1}{2}}) P^{\frac{1}{2}}. \quad (0.20)$$

Note that this expression cannot be further simplified in general when the data P_i and the matrix P do not commute. The gradient of the variance is related very simply to this first order moment by $\nabla \text{Var}(P) = -2 \mathfrak{M}_1(P)$, and the critical points of the variance (exponential barycenters) are the zeros of this field. Because SPD matrices have a non-positive curvature and an infinite injection radius with an affine invariant metric, there is one and only one global minimum: the Fréchet mean \bar{P} is unique [Ken90].

Algorithm 1 Intrinsic Gradient Descent to Compute the Mean SPD Matrix

Initialize

Set current estimate of the mean with one of the data points, e.g. $\bar{P}_0 = P_1$.

Set error $\epsilon_0 = \|\mathfrak{M}_1(\bar{P}_0)\|_{\bar{P}_0}$, step-size $\tau = 1/2$ and time $t = 0$.

Iterate

$t = t + 1$.

$\bar{P}_t = \text{Exp}_{\bar{P}_{t-1}}(2\tau \mathfrak{M}_1(\bar{P}_{t-1}))$.

$\epsilon_t = \|\mathfrak{M}_1(\bar{P}_t)\|_{\bar{P}_t}$.

if $\epsilon_t > \epsilon_{t-1}$ then $\tau = \tau/2$ and $\bar{P}_t = \bar{P}_{t-1}$.

Until $\epsilon_t < \epsilon$.

To compute this mean, we can use the gradient descent algorithm on the space of SPD matrices described in Algorithm 1. The idea is to follow the flow of the gradient of the variance for a certain time step τ :

$$\bar{P}_{t+1} = \text{Exp}_{\bar{P}_t}(-\tau \nabla \text{Var}(\bar{P}_t)) = \text{Exp}_{\bar{P}_t}(2\tau \mathfrak{M}_1(\bar{P}_t)). \quad (0.21)$$

With the time-step $\tau = 1/2$, this iterative algorithm corresponds to the classical Gauss-Newton gradient descent algorithm on manifolds, which was shown to be globally convergent when the data are sufficiently concentrated in Riemannian symmetric spaces and compact Lie groups [Le01, Man04]. In practice, the Gauss-Newton algorithm converges very quickly when the data points are sufficiently close to each other.

However, the algorithm may diverge when the SPD matrices are too far away from each other, as noted in [FJ07]. Sufficient conditions on the time-step τ for the convergence of the algorithm were established in [BI13] based on the condition numbers (ratio of minimal and maximal eigenvalues) of $\bar{P}_t^{-1/2} P_i \bar{P}_t^{-1/2}$. When the matrices $P_1 \dots P_n$ commute, then the choice of $\tau = 1/2$ is actually optimal and leads to a super-linear (quadratic) convergence. When these matrices are close to each other, a fast convergence is still expected, although it is not quadratic any more. However, when the matrices are far away from each other, [BI13] showed that the Gauss-Newton algorithm may diverge even if the matrices almost commute. For example, computing the mean of the three SPD matrices $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and $\begin{pmatrix} x & 1 \\ 1 & 2 \end{pmatrix}$ converges for $x = 1000$ but diverges for $x = 10000$. Thus, it is generally necessary to use an adaptive time-step τ in the Riemannian gradient descent. One of the simplest adaptive algorithm is to start with $\tau = 1/2$ and decrease it instead of updating \bar{P}_{t+1} when the variance increases during the optimization [FJ07], as described in Algorithm 1.

The divergence of the Gauss-Newton scheme with a fixed time-step is due to the curvature of the manifold. This can be illustrated in constant curvature spaces where we can explicitly compute the eigenvalues of the Hessian of the square Riemannian distance (see for instance [Pen18]). On the sphere (positive curvature), the eigenvalues are lower than 2, and can become negative when data leave the Kendall/Karcher conditions of uniqueness. Because the inverse of the Hessian is used to modulate the gradient in Newton methods, $\tau = 1/2$ is a conservative bound which guaranties the convergence, with potentially a suboptimal rate. On the contrary, the Hessian has eigenvalues larger than 2 in the hyperbolic space (negative curvature) so that a time-step of $\tau = 1/2$ can be too large. Computing explicit tight bounds on the spectrum of the Hessian of the variance is possible in connected locally symmetric spaces [FXCB13], and this enables the use of Newton methods with a quadratic convergence rate.

The covariance, or second order moment of the sample distribution, is the 2-

contravariant field defined all over the SPD matrix manifold by the tensor product:

$$\mathfrak{M}_2(P) = \int_{\mathcal{M}} \overrightarrow{PQ} \otimes \overrightarrow{PQ} \mu(dQ) = \frac{1}{m} \sum_{i=1}^m \overrightarrow{PP_i} \otimes \overrightarrow{PP_i}.$$

The covariance matrix is the value of this field at the Fréchet mean \bar{P} . This tensor can be expressed as a matrix in our orthonormal basis using the *Vec* mapping:

$$\Sigma = \frac{1}{m} \sum_{i=1}^m \text{Vec}_{\bar{P}}(\overrightarrow{\bar{P}P_i}) \text{Vec}_{\bar{P}}(\overrightarrow{\bar{P}P_i})^\top.$$

4.2. Tangent PCA and PGA of SPD matrices

Assume that we have computed the sample Fréchet mean \bar{P} and the sample covariance matrix Σ of a set of SPD matrices. We may want to identify subspaces of low dimensions that best approximate our SPD data. In Euclidean spaces, this is usually done using Principal Component Analysis (PCA). As explained in Chapter 2, we have several choices in a manifold. We can first maximize the explained variance, which corresponds to choosing the subspace generated by the eigenvectors of the largest eigenvalues of the covariance matrix. This is called tangent PCA (tPCA). We can also minimize the unexplained variance, which is measured by the sum of square of residues, which are the tangent vectors pointing from each data point to the closest point on the geodesic subspace generated by k modes at the mean (Principal Geodesic Analysis, or PGA).

Figure 0.1, adapted from [FJ07], presents an example of the difference between PCA on SPD matrices with the Euclidean metric and tPCA with one of the affine invariant metric. One of the main differences is that any matrix generated by the principal geodesic subspace remains positive-definite, contrarily to the Euclidean PCA.

4.3. Gaussian Distributions on SPD matrices

Several definitions have been proposed to extend the notion of Gaussian distribution to Riemannian manifolds. The natural definition from the stochastic point of view is the heat kernel $p_t(x, y)$, which is the transition density of the Brownian motion (see Chapter 4 of this book). The heat kernel is the smallest positive fundamental solution to the heat equation $\frac{\partial f}{\partial t} - \Delta f = 0$, where $\Delta f = \text{Tr}_g(\text{Hess } f) = g^{ij}(\text{Hess } f)_{ij} = \nabla^i \nabla_i f$ is the Laplace-Beltrami operator. However, the heat kernel has a non-linear dependency in time which makes it difficult to use in statistics as a classical Gaussian distribution.

To obtain more tractable formulas, the wrapped Gaussian distribution was proposed in several domains [Gre63, MJ00]. The idea is to take a Gaussian distribution in the tangent space at the mean value and to consider the push-forward distribution on the manifold that wraps the distribution along the closed geodesics. The proper

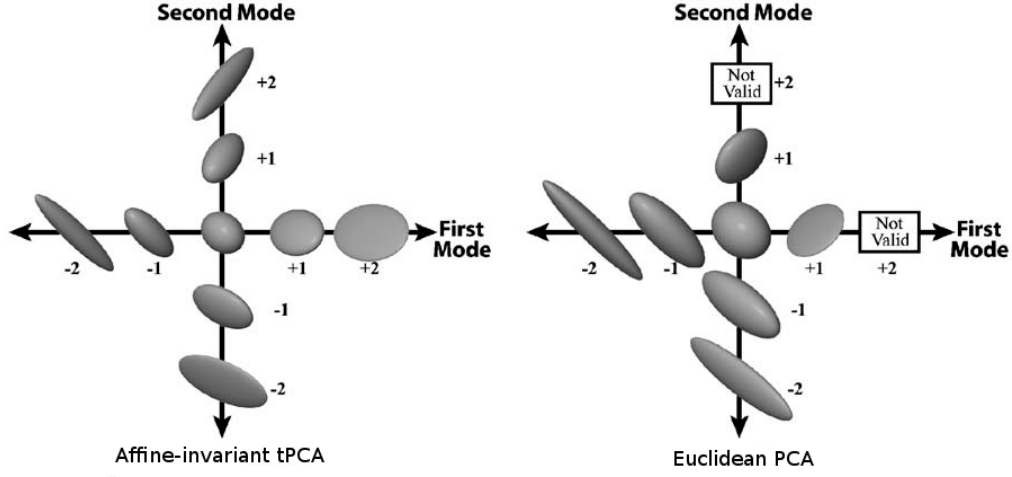


Figure 0.1 The first two modes of variation of the simulated data: (left) using the affine-invariant tPCA, and (right) using linear PCA. Units are in standard deviations. The boxes labeled “Not Valid” indicate that the tensor was not positive-definite, i.e., it had negative eigenvalues. [Adapted from [FJ07]]

renormalization of this push-forward distribution is an issue since the renormalization constant includes infinite series along these closed geodesics. The wrapped Gaussian distribution naturally corresponds to the infinitesimal heat kernel for small time steps and tends towards the Dirac mass distribution if the variance goes to zero. In the compact case, it also tends towards the Riemannian uniform distribution for a large variance. [OC95] considered an extension of this definition with non-centered Gaussian distributions on the tangent spaces of the manifold in order to tackle the asymptotic properties of estimators. In this case, the mean value is generally not any more simply linked to the Gaussian parameters.

A third approach, detailed below, considers the probability density function (pdf) that minimizes the information (or equivalently maximizes the entropy) knowing the mean and the covariance matrix. This approach was taken for example in [Sou70] for dynamical Lie groups and in [Pen96, Pen06a] for Riemannian manifolds. This leads to another family of exponential distributions that are cut at the tangential cut-locus. Obviously, wrapped and cut Gaussians in tangent spaces are the same in Hadamard spaces since the cut-locus is at infinity in all directions.

To define the entropy, we consider a probability $\mu = \rho d\mathcal{M}$ that is absolutely continuous with respect to the Riemannian measure so that it has a pdf ρ that is an integrable non-negative function on \mathcal{M} . With the affine-invariant metric, the Riemannian mea-

sure on the space of SPD matrices can be expressed as

$$d\mathcal{M}(P) = 2^{n(n-1)/4} \sqrt{(1 + \beta n)} \det(P)^{-1} dP \propto \det(P)^{-1} dP,$$

with respect to the standard Lebesgue measure on the upper triangular coefficients of the SPD matrix P . This expression is obtained by expressing the formula of the Riemannian measure $d\mathcal{M}(P) = \sqrt{\det(g|_P)} dP$ (see Chapter 1, Section 4.3) in the chart of upper triangular coefficients F_{ij} of section 3.3.3. The change of coordinates from the orthonormal basis $E_{ij}|_P$ to the basis $E_{ij}|_{\text{Id}}$ at identity is responsible for the term $\det(P)^{-1}$. The mapping $A_\alpha^{-1} = \text{Id} + \alpha \mathbb{1}_n \mathbb{1}_n^\top / (1 - \alpha n)$ transforming to the basis $E_{ij}^0|_{\text{Id}}$ has determinant: $\det(A_\alpha^{-1}) = 1/(1 - \alpha n) = \sqrt{(1 + \beta n)}$. Finally, the rescaling of the $n(n-1)/2$ off diagonal terms by $\sqrt{2}$ to obtain the L_2 metric accounts for the factor $2^{n(n-1)/4}$. Changing for the Lebesgue measure on the full matrix coefficients only changes this last constant multiplicative factor. The entropy of an absolutely continuous distribution $\mu = \rho d\mathcal{M}$ is then:

$$H(\mu) = - \int_{\mathcal{M}} \log(\rho(P)) \rho(P) d\mathcal{M}(P) = - \int_{\text{Sym}_n^+} \log(\rho(P)) \frac{\rho(P)}{\det(P)} dP.$$

Its negative is called the information, or negentropy. One verifies that the pdf maximizing the entropy in a compact set is the uniform distribution over that set with respect to the Riemannian measure. Maximizing the entropy is thus consistent with the Riemannian framework.

Applying the maximum entropy principle to characterize Gaussian distributions, we can look for the pdf maximizing the entropy with a fixed mean and covariance matrix. The constraints are thus:

- the normalization $\int_{\mathcal{M}} \rho(Q) d\mathcal{M}(Q) = 1$,
- the prescribed mean value $\int_{\mathcal{M}} \vec{PQ} \rho(Q) d\mathcal{M}(Q) = 0$,
- and the prescribed covariance $\int_{\mathcal{M}} \vec{PQ} \otimes \vec{PQ} \rho(Q) d\mathcal{M}(Q) = \Sigma$.

These constraints can be expressed in a normal coordinate system at the mean and using the convexity of the real function $-x \log(x)$, one can show that the maximum entropy is attained by the distributions of density $\rho(v) = k \exp\left(-\langle \beta, v \rangle_x - \frac{1}{2} \Gamma(v, v)\right)$, where v is a tangent vector at the mean, provided that there exists a constant k , a vector β and a bilinear form Γ on the tangent space at the mean such that our constraints are fulfilled. Moreover, when the tangential cut-locus at the mean is symmetric, we find that $\beta = 0$ satisfies the prescribed mean.

Theorem 4 (Normal density on a Riemannian manifold [Pen06a]). *In a complete Riemannian manifold with a symmetric tangential cut-locus at P , the density maximizing the entropy with a prescribed mean P and covariance Σ has the form of a Gaussian in*

the tangent space at P truncated at the cut-locus: $G(Q) = k \exp(-\frac{1}{2} \vec{PQ}^\top \Gamma \vec{PQ})$. The exponential of the squared distance $G(Q) = \exp(-\frac{\tau}{2} \text{dist}^2(P, Q))$ used in Chapter 2 is a particular case obtained with an isotropic concentration matrix $\Gamma = \tau \text{Id}$.

Let $r = i(\mathcal{M}, P)$ be the injectivity radius at the mean point (by convention $r = +\infty$ if there is no cut-locus). Assuming a finite variance for any concentration matrix Γ , we have the following approximations of the normalization constant and concentration matrix for a covariance matrix Σ of small variance $\sigma^2 = \text{Tr}(\Sigma)$:

$$k = \frac{1 + O(\sigma^3) + \epsilon\left(\frac{\sigma}{r}\right)}{\sqrt{(2\pi)^n \det(\Sigma)}} \quad \text{and} \quad \Gamma = \Sigma^{-1} - \frac{1}{3} \text{Ric} + O(\sigma) + \epsilon\left(\frac{\sigma}{r}\right)$$

Here, $\epsilon(x)$ is a function that is a $O(x^k)$ for any positive k , with the convention that $\epsilon\left(\frac{\sigma}{+\infty}\right) = \epsilon(0) = 0$.

In Riemannian symmetric spaces, the tangential cut-locus is symmetric at all points. Indeed, the geodesic symmetry with respect to a point P reads $\vec{PQ} \rightarrow -\vec{PQ}$ in the log-map at P . Since this is an isometry, the distance to the tangential cut-locus is the same in opposite directions. In the case of our SPD matrices space, this distance is moreover infinite and the conditions involving the injection radius can be waved since we are in a Hadamard manifold. Thus, we obtain a generalization of the Gaussian distribution on the space of SPD matrices which is a standard Gaussian in the tangent space at the mean:

$$G_{(\bar{P}, \Gamma)}(Q) = k \exp\left(-\frac{1}{2} \vec{PQ}^\top \Gamma \vec{PQ}\right).$$

However, contrarily to the Euclidean case where the concentration matrix Γ is simply the precision matrix Σ^{-1} , there is a correction term for the curvature which is due to the change of the volume of infinitesimal geodesic balls with respect to Euclidean spaces. It is thus natural to see the Ricci curvature appearing as a correction term. In order to understand better what this means, let us consider a Gaussian distribution at $\bar{P} = \text{Id}$ (by invariance we are not loosing any generality) with covariance matrix Σ and concentration matrix Γ that are jointly diagonal with the Ricci curvature matrix. According to Theorem 3, the Ricci curvature matrix is block-diagonal in the basis $\{E_{ij}\}$. Let Γ_{ij} and Σ_{ij} be the coefficients of the diagonal concentration / covariance matrices. Theorem 4 states that $\Gamma_{ij} = \Sigma_{ij}^{-1} - \frac{1}{3} \text{Ric}(E_{ij}) + O(\sigma)$. Thus, we see that the Ricci curvature acts as the inverse of a typical length separating two different behaviors: when the variance is small ($\Sigma_{ij} \ll \text{Ric}(E_{ij})^{-1}$), then the curvature correction term can be neglected and the usual Gaussian in the tangent space is a good model. This is also the regime where tangent PCA (tPCA) will work without having to worry about curvature. On the contrary, if the variance is large with respect to the curva-

ture ($\Sigma_{ij} \gg \text{Ric}(E_{ij})^{-1}$), then the curvature has to be taken into account and tPCA and principal geodesic analysis (PGA) will give different results.

5. Manifold-Valued Image Processing

We turn in this section to the generalization of image processing algorithms like interpolation and diffusion to manifold-valued images. We show that most interpolation and filtering methods can be reformulated using weighted Fréchet means. The linear and non-linear diffusion schemes can be adapted to Manifolds through PDEs, provided that we take into account the variations of the metric. For details, we refer the reader to [PFA06]. More elaborate methods were developed since then using total variation [WDS14, GS16] or second order [BBSW16] regularization, as well as non-local mean denoising of manifold-valued images [LNPS17].

In this section, $P(x)$ is a $n \times n$ SPD matrix-valued image with Euclidean coordinates $x \in \mathbb{R}^m$. Because we are in a Hadamard space where there exists a global coordinate system (for instance $E_{ij}(P(x))$), this SPD image can be represented as a d -dimensional vector image, with $d = n(n+1)/2$. This would not be possible for an image of orientations belonging to a sphere, for instance. However, we will see below that processing manifold-valued images is in both cases different from processing vector images. In practice the image coordinates are sampled at the points x_k of a grid where $k = \{k_1 \dots k_m\}$ is the index of the signal element ($m = 1$), the pixel ($m = 2$) or the voxel ($m = 3$). The SPD matrix-valued image is thus encoded by the m -dimensional array of values $P_k = P(x_k)$ at the voxels x_k .

5.1. Interpolation

One of the important operations in geometric data processing is to interpolate values between known measurements. Beyond the nearest neighbor interpolation, which is not continuous, one of the simplest interpolation for 1D signal processing is linear interpolation. In 2D and 3D image processing, this generalizes to bilinear and trilinear interpolation, which are particularly popular due to their very low computational complexity. Higher order methods include quadratic, cubic and higher order spline interpolation [TBU00, Mei02].

Assuming that we have signal values f_k sampled on a regular lattice with integer coordinates $x_k \in \mathbb{Z}^m$. It is usual to interpolate the value at a non-integer coordinates x using a linear combination of the signal value $f(x) = \sum_k w(x - x_k) f_k$ weighted by a kernel w normalized with $w(0) = 1$ and $w(k) = 0$ at other integer values so that the interpolated signal $f(x_k) = f_k$ is equal to the sampled measure at the lattice points. A typical example of an interpolation kernel with infinite support is the sinus cardinal. Classical finite support kernels are the nearest-neighbor, linear (or tri-linear in 3D) and piecewise polynomial kernels realizing spline interpolations [TBU00, Mei02].

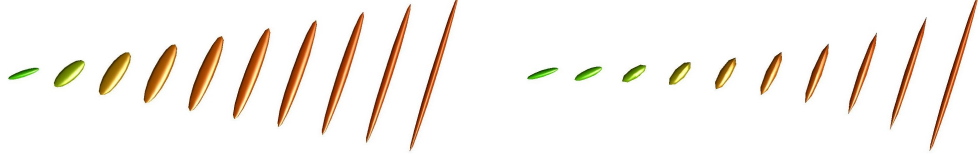


Figure 0.2 Linear interpolation of SPD matrices. **Left:** linear interpolation on coefficients. **Right:** affine-invariant interpolation. We can notice the characteristic swelling effect observed in the Euclidean case, which is not present in the Riemannian framework.

When the weights are summing up to one, this can be seen as a weighted Fréchet mean. Thus, interpolation of manifold-valued images can be formulated as an optimization problem: the interpolated value is then defined as:

$$P(x) = \arg \min_{Q \in \mathcal{M}} \left(\sum_k w(x - x_k) \text{dist}^2(P_k, Q) \right).$$

The cost to pay is that the value has to be determined by gradient descent, for instance using the extension of algorithm 1 with weights. Linear interpolation between two points $P(0) = P$ and $P(1) = Q$ can be written explicitly since it is a simple geodesic walking scheme: $P(t) = \text{Exp}_P(t\vec{PQ}) = \text{Exp}_Q((1-t)\vec{QP})$.

For our SPD matrices example, this gives the following interpolation with the standard Euclidean and affine-invariant metrics:

$$\begin{aligned} P_{Euc}(t) &= (1-t)P + tQ \\ P_{Aff}(t) &= P^{1/2} \exp\left(t \log\left(P^{-1/2}QP^{-1/2}\right)\right) P^{1/2} \end{aligned}$$

For a Euclidean metric, the trace is linearly interpolated. With an affine invariant metric, the trace is not linear anymore but the determinant is geometrically interpolated and its logarithm is linearly interpolated [AFPA07]. This is illustrated in Fig. 0.2 for the linear interpolation and in Fig. 0.3 for the bilinear interpolation with the Euclidean and affine-invariant metric.

5.2. Gaussian and kernel-based filtering

Considering weighted means allows to generalize many other image processing operations. A kernel convolution, for instance, can be viewed as the average value of the signal at neighboring points weighted by the respective kernel value. For a translation invariant kernel $k(x, y) = k(y - x)$ in a Euclidean space, the convolution $\hat{f}(x) = \int_{\mathbb{R}^n} k(u) f(x + u) du$ is indeed the minimizer of the criterion $C(\hat{f}) = \int_{\mathbb{R}^n} k(u) \text{dist}^2(f(x + u), \hat{f}(x)) du$. Here, the kernel might be a discrete measure, for instance when sample points are localized at discrete nodes of grid.

This variational formulation is still valid in Riemannian manifolds. However, in

general, the minimum may not exist nor be unique. For Hadamard spaces (thus for SPD matrices with an affine-invariant metric), the existence and uniqueness are ensured for a non-negative kernel. In general, there is no closed form for the minimizer of this cost function. Thus, we need to rely on an iterative gradient descent algorithm. Assuming that the kernel is normalized ($\int_{\mathbb{R}^n} k(u) du = 1$), we can use:

$$\hat{P}_{t+1}(x) = \text{Exp}_{P_t} \left(\tau \int_{\mathbb{R}^n} k(u) \text{Log}_{\hat{P}_t(x)}(P(x+u)) du \right), \quad (0.22)$$

where the time-step τ is updated like in algorithm 1. Fig. 0.4 illustrates the Gaussian filtering of a SPD image (here a diffusion tensor image) with the flat and affine-invariant metrics. One can visualize the blurring that occurs on the corpus callosum fiber tracts using the flat metric.

5.3. Harmonic regularization

In order to filter an image, one can also want to minimize the norm of the derivative of the field in the direction u . The directional derivative $\partial_u P(x)$ is a tangent vector of $T_{P(x)}\mathcal{M}$ which can be evaluated using finite differences: $\partial_u P(x) \simeq \text{Log}_{P(x)}(P(x+u)) + O(\|u\|^2)$. The norm of this vector has to be measured with the Riemannian metric at the foot-point: $\|\partial_u P(x)\|_{P(x)}$. However, penalizing the derivative only in one direction of the space is not sufficient. The spatial differential of the field $P(x)$ is the linear form that maps to any spatial direction u the directional derivative $dP(x)u = \partial_u P(x)$. it can be expressed as a matrix with the column coefficients related to the spatial basis (say e_i) and row coefficients related to a basis of the $T_{P(x)}\text{Sym}_n^+$ (for instance $E_{jk}(P(x))$). The dual of that $d \times m$ matrix with respect to the spatial metric (this is the transpose of the matrix in an orthonormal coordinate system) has columns that gives the direction of

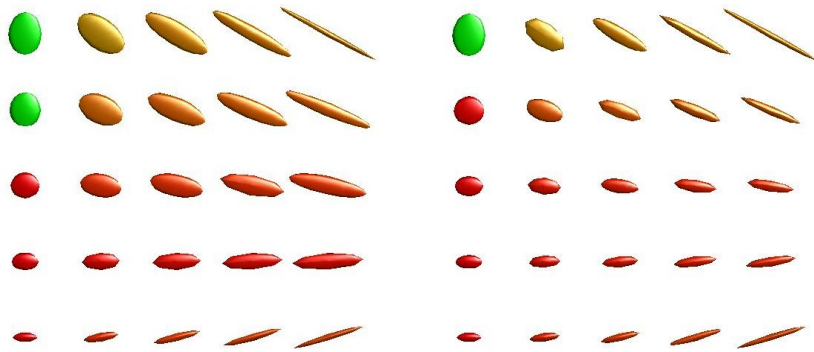


Figure 0.3 Bilinear interpolation of SPD matrices. **Left:** linear interpolation on coefficients. **Right:** affine-invariant interpolation. We can notice once again the characteristic swelling effect observed in the Euclidean case, which is not present in the Riemannian framework. [Figure adapted from [AFPA07]]

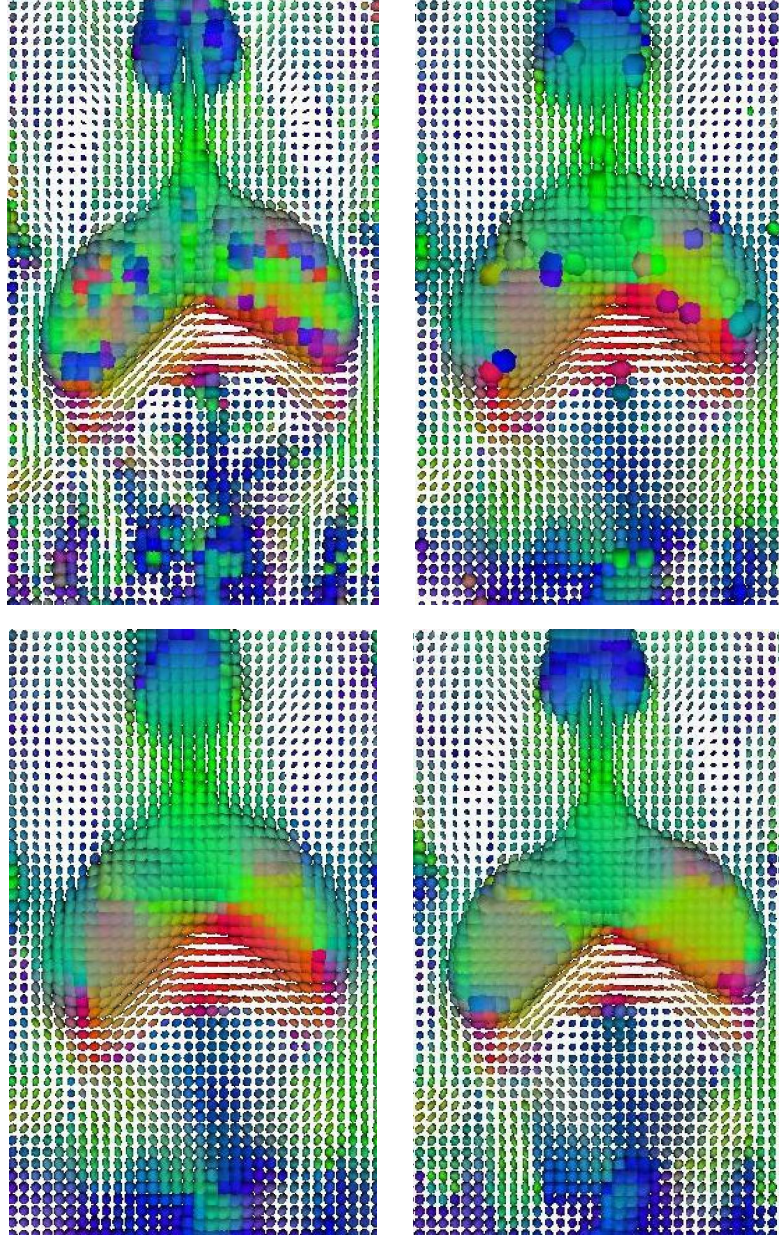


Figure 0.4 Results of Euclidean Gaussian, Riemannian Gaussian and Riemannian anisotropic filtering on a 3D DTI of the brain (Closeup around the splenium of the corpus callosum). The color codes for the direction of the principal eigenvector (red: left-right, green: posterior-anterior, blue: inferior-superior). **Top left:** Original image. **Top right:** Gaussian filtering using the flat metric (5x5 window, $\sigma = 2.0$). This metric gives too much weight to SPD matrices with large eigenvalues, thus leading to clear outliers in the ventricles or in the middle of the splenium tract. **Bottom left:** Gaussian filtering using the Riemannian metric (5x5 window, $\sigma = 2.0$). Outliers disappeared, but the discontinuities are not well preserved, for instance in the ventricles at the level of the cortico-spinal tracts (upper-middle part of the images). **Bottom right:** Anisotropic filtering in the Riemannian framework (time step 0.01, 50 iterations). The ventricles boundary is very well conserved with an anisotropic filter and both isotropic (ventricles) and anisotropic (splenium) regions are regularized. Note that the U-shaped tracts at the boundary of the grey/white matter (lower left and right corners of each image) are preserved with an anisotropic filter and not with a Gaussian filter. [Figure reproduced from [PFA06]]

the spatial domain in which each component of the SPD matrix field evolve the fastest: we denote this gradient by $\nabla_x P(x) = dP(x)^\top$ to differentiate it from the connection ∇ .

As there are more directional derivatives in classical 6, 18 or 26 neighborhoods (in 3D) than the three needed to compute the full gradient matrix, the gradient may be approximated by minimizing the least square error:

$$\nabla_x P(x) = \arg \min_{A \in M_{m \times d}} \sum_i \|A^\top u_i - \text{Log}_{P(x)}(P(x + u_i))\|_{P(x)}^2,$$

where A is a matrix with 3 rows and 6 columns in the case of a 3D image of 3D SPD matrices. The simplest regularization criterion based on the spatial gradient is the harmonic energy:

$$\text{Reg}(P) = \frac{1}{2} \int_{\Omega} \|\nabla_x P(x)\|_{P(x)}^2 dx = \frac{1}{2} \sum_{i=1}^m \int_{\Omega} \|\partial_{x_i} P(x)\|_{P(x)}^2 dx.$$

The expression on the right assumes an orthonormal coordinate system $\partial_{x_1}, \dots, \partial_{x_m}$ of the image domain. The Euler-Lagrange equation $\nabla_P \text{Reg}(P)(x) = -\Delta P(x)$ of this harmonic regularization criterion with Neumann boundary conditions involves the Laplace-Beltrami operator $\Delta P(x)$ on the manifold. In addition to summing the flat Euclidean second order directional derivatives $\partial_{x_i}^2 P(x)$ in a locally orthogonal system, this operator has an additional term encoding the curvature of the manifold that distorts the ortho-normality of this coordinate system from one point to another in the neighborhood. However, we need not compute the full Riemannian curvature operator to compute the Laplace-Beltrami operator. Indeed, we only need to access to the second-order derivative along geodesics starting with orthonormal vectors at a point. Because the Christoffel symbols and their radial derivatives (along the geodesics starting from the foot-point P) vanish in a normal coordinate system at P , computing the standard Laplacian in this specific coordinate system actually already includes the correction for the curvature.

This gives rise to the following very general and efficient scheme for the second order derivative in the spatial direction u [PFA06]:

$$\Delta_u P = \text{Log}_{P(x)}(P(x + u)) + \text{Log}_{P(x)}(P(x - u)) + O(\|u\|^4).$$

Averaging over all the spatial directions in a spatial neighborhood \mathcal{V} finally gives a robust and efficient estimation scheme:

$$\Delta P(x) \propto \sum_{u \in \mathcal{V}} \frac{1}{\|u\|^2} \text{Log}_{P(x)}(P(x + u)) \quad (0.23)$$

The optimization of the harmonic energy can be performed as previously using a first order gradient descent technique $P^{t+1}(x) = \text{Exp}_{P^t(x)}(-\tau \Delta P^t(x))$ that iteratively shoot for an adaptive time-step τ in the (opposite) direction of the regularization criterion

gradient.

5.4. Anisotropic diffusion

In order to filter in homogeneous regions and to keep edges sharp, one can modulate the regularization depending on the direction u . A large body of work has been focusing on the inhomogeneous and anisotropic filtering of classical grey-valued images (see e.g. [Wei97, Wei98]). A comparatively much smaller part of this early literature has been devoted to anisotropic diffusion on very specific manifold-valued images, like circular data [TSC00]. The case of SPD-valued images illustrated below was investigated in [PFA06] and generalizes easily to generic Riemannian-valued images.

A first method consists in smoothing the image in the direction u if the directional derivative $\partial_u P(x)$ is small, and penalize the smoothing whenever we cross an edge as measured by a large directional derivative [PM90, GKKJ92]. This can be realized directly in the discrete implementation of the Laplacian by weighting the directional contribution $\Delta_u P = \text{Log}_{P(x)}(P(x+u))/\|u\|$ to the Laplacian with a decreasing function of the norm $\|\partial_u P\|_P$. Thus, a natural generalization of the weighted Laplacian to manifold-valued images is $\Delta_u^\phi P = \sum_u \phi(\|\partial_u P\|_P) \Delta_u P$, for instance with $\phi(x) = \exp(-x^2/\kappa^2)$ [PFA06].

One of the key problem of the anisotropically weighted Laplacian is that its evolution equation is not guarantied to converge, even with Euclidean images, since the anisotropic regularization “forces” may not derive from a well-posed energy. An alternative is to construct a variational formulation with a weighted regularization criterion that penalizes the spatial variations of the field in the homogeneous areas. It is classical to take for instance a robust M-estimator of the Riemannian norm: $\text{Reg}_\phi(P) = \frac{1}{2} \int_\Omega \phi(\|\nabla_x P(x)\|_{P(x)}) dx$. By choosing an adequate ϕ -function (for instance $\phi(s) = 2\sqrt{1 + s^2/\kappa^2} - 2$), one can give to the regularization an isotropic or an anisotropic behavior [AK01]. The main difference with a classical Euclidean calculation is the use of the Laplace-Beltrami operator and the Riemannian norm [FAAP05]. Using $\Psi(x) = \phi'(x)/x$, we get the following gradient:

$$\nabla_P \text{Reg}_\phi(P) = -\Psi(\|\nabla_x P\|_P) \Delta P - \sum_{i=1}^d \partial_{x_i} \Psi(\|\nabla_x P\|_P) \partial_{x_i} P.$$

The evolution equation is thus $\partial_t P = -\nabla_P \text{Reg}_\phi(P)$. An illustration of the anisotropic diffusion of an SPD image (here a diffusion tensor image) is given in Fig. 0.4. One can notice that some important U-shaped tracts at the boundary of the grey/white matter are preserved with our Riemannian anisotropic filter and not with a Gaussian filter.

5.5. Inpainting and extrapolation of sparse SPD fields

The harmonic or anisotropic diffusion schemes developed above reduce the noise on the image but also the information that it contains. Indeed, an infinite diffusion time

leads to a completely homogeneous field. Thus, it is interesting to consider adding a data attachment term (a data likelihood) in addition to the regularization which constitutes then a spatial prior. One usually assumes an independent Gaussian noise at each observed point of the image so that the maximum likelihood corresponds to a least-square criterion. This is still the case up to the first order with our definition of the Gaussian on manifolds of Section 4.3 when the variance is small enough. However, correction terms should be considered for covariances that are larger than the Ricci curvature (see also Chapter 2 on regression on manifolds).

When we have a dense data field $Q(x)$ as previously, the natural similarity criterion is the classical sum of square differences $\text{Sim}(P) = \int_{\Omega} \text{dist}^2(P(x), Q(x)) dx$, expressed with the Riemannian distance. This criterion adds a geodesic spring (whose strength depends on the weighting of each criterion) $\nabla_P \text{dist}^2(P, Q) = -2\overrightarrow{PQ}$ to the gradient of the regularization which prevents from getting too much away from the data.

For sparse delta-Dirac measurements Q_i at points x_i , we need to restrict the integral in the above similarity criterion to the points where we have measurements. Unfortunately, these mass distributions induce singularities in the gradients. One solution is to regularize the data attachment term with Gaussian convolutions:

$$\text{Sim}(P(x)) = \int_{\Omega} \sum_{i=1}^n G_{\sigma}(x - x_i) \text{dist}^2(P(x), Q_i) dx.$$

This leads to the regularized derivative [PFA06]:

$$\nabla_P \text{Sim}(x) = -2 \sum_{i=1}^n G_{\sigma}(x - x_i) \overrightarrow{P(x)P_i}.$$

An example of extrapolation (inpainting) of a whole image from 4 SPD matrix measurements is illustrated in Figure 0.5.

6. Other metrics on SPD matrices

Affine-invariant and Euclidean metrics are two families of metrics that endow the space of SPD matrices with very different properties. There are actually quite a few other families of metrics which are interesting to know.

6.1. Log Euclidean metrics

In 2006, soon after the affine-invariant metrics were proposed, [AFPA06, AFPA07] realized that the matrix exponential was a global diffeomorphism from the space of symmetric matrices to the space of SPD matrices (a proof of that fact is given in Section 2 of this chapter), a fact which is mathematically well-known in Hadamard manifolds. However, it was apparently not used previously to transfer the vector space structure of symmetric matrices to SPD matrices. In particular, it is possible to endow

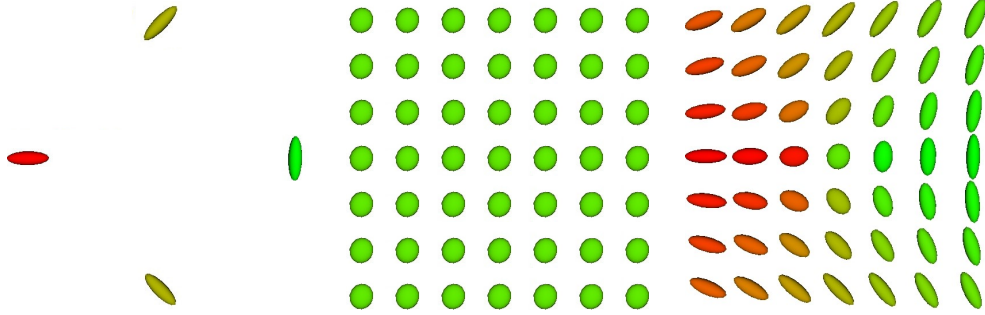


Figure 0.5 Extrapolation of SPD matrix values from four measurements using harmonic diffusion. **left:** The four initial SPD matrix values. **Middle:** result of the diffusion without the data attachment term. The field of SPD matrix converges to a uniform image with the Fréchet mean value of the 4 SPD matrices. **Right:** result of the diffusion with a strong data attachment term. The diffusion smoothly extrapolates outside the measurements with a very small smoothing towards the mean at each of the original measurements. Notice that the center point is a multiple of the identity matrix around which it is not possible to consistently orient the principal eigenvector.

SPD matrices with a commutative Lie group structure where the composition is defined by $P_1 \diamond P_2 = \exp(\log(P_1) + \log(P_2))$ (the log-product). With the logarithmic scalar multiplication $\lambda \cdot P = \exp(\lambda \log(P)) = P^\lambda$ in addition, this endows the SPD space with a vector space structure. To finish, any Euclidean metric on symmetric matrices is also transformed into a bi-invariant Riemannian metric on SPD matrices thanks to the differential of the matrix logarithm:

$$\langle V, W \rangle_P^{LE} = \langle d \log(P)(V), d \log(P)(W) \rangle_{Euc} = \langle \partial_V \log(P), \partial_W \log(P) \rangle_{Euc}.$$

We recall that the differential of the matrix logarithm was explicitly determined in Section 2.2. It is well defined and invertible everywhere on the space of SPD matrices.

Using the fact that Euclidean geodesics are straight lines in the space of symmetric matrices, the expression of the Exp, Log and distance maps metric are easily determined for the Log-Euclidean metric:

$$\begin{aligned} \text{Exp}_P^{LE}(W) &= \exp(\log(P) + \partial_W \log(P)). \\ \text{Log}_P^{LE}(Q) &= d \exp(\log(P))(\log(Q) - \log(P)) = \partial_{\log(Q) - \log(P)} \exp(\log(P)). \\ \text{dist}_{LE}^2(P_1, P_2) &= \|\log(P_1) - \log(P_2)\|_{Euc}^2. \end{aligned}$$

Because the differential of the matrix log at the identity is the identity, we can see that log-Euclidean geodesics through the identity are the same as the affine-invariant geodesics. However, this is not true at other points of the SPD manifold [AFPA07].

By construction, log-Euclidean metrics are invariant by inversion and by a change of scale of the space. If we choose in addition a Euclidean scalar product which is invariant by rotation (we already know that they are all of the form $\|W\|_{Euc}^2 = \text{Tr}(W^2) +$

$\beta \text{Tr}(W)^2$ with $\beta > -1/n$, we conclude that all the similarity invariant log-Euclidean Riemannian distances are:

$$\begin{aligned} \text{dist}_{LE}^2(P, Q) &= \|\log(P) - \log(Q)\|_{Euc}^2 \\ &= \text{Tr}\left((\log(P) - \log(Q))^2\right) + \beta \text{Tr}(\log(P) - \log(Q))^2. \end{aligned}$$

The relative complexity of the exp/log formulas with respect to the affine-invariant case is due to the use of matrix exponential and logarithm differentials. However, log-Euclidean exp and log are nothing else than the transport of the addition and subtraction through the exponential of symmetric matrices. In practice, the log-Euclidean framework consists in taking the logarithm of the SPD matrices, computing like usual in the Euclidean space of symmetric matrices, and coming back at the end to the SPD matrix space using the exponential [AFPA06, AFPA05]. The vector space structure gives a closed-form expression to many of the operations that were defined in this chapter through optimization. For instance, the log-Euclidean mean is simply [AFPA05]:

$$\bar{P}_{LE} = \exp\left(\frac{1}{n} \sum_i^n \log(P_i)\right),$$

while the affine-invariant mean has to be obtained through the iterative algorithm 1 of Section 4.1.

This shows that we can have very different flat structures on the space of SPD matrices. In fact, the log-Euclidean structure is much closer to the affine-invariant structure than to the Euclidean structure: log-Euclidean geodesics are complete (never leaving the space) and are identical to the affine-invariant geodesics at the identity, contrarily to Euclidean geodesics. The log-Euclidean and affine-invariant means are identical if the mean commutes with all the data. When they are not equal, one can show that (close enough to the identity) the log-Euclidean mean is slightly more anisotropic [AFPA07]. A careful comparison of both metrics in practical applications [AFPA05, AFPA06] showed that there was very few differences on the results (of the order of 1%) on real diffusion tensor images (see next section), with a gain of computation time from 4 to 10 folds for the log-Euclidean. The difference is actually due to the curvature induced by the affine-invariant metric (the log-Euclidean structure is flat since it is a vector space): when the distance of the data to the identity is less than the typical Ricci curvature, one can conclude that the log-Euclidean framework is a good first order approximation of the affine-invariant computations. When there is another reference point \bar{Q} which is more central than the identity matrix Id for the data, one can also use the log-Euclidean structure induced by $\bar{Q}^{1/2} \log(\bar{Q}^{-1/2} P \bar{Q}^{-1/2}) \bar{Q}^{1/2}$ instead of $\log(P)$.

For other types of applications where the variability or the anisotropy of the SPD matrices is higher than the affine-invariant Ricci curvature, affine-invariant and log-

Euclidean results might change a lot. This is for instance the case in adaptive remeshing [BGM97]. However, initializing the iterative optimizations of affine-invariant algorithms with the log-Euclidean result drastically speeds-up the convergence.

6.2. Cholesky metrics

Other families of metrics were also proposed to work with SPD matrices. For instance, [WVCM04] proposed to parametrize tensors by their Cholesky decomposition $P = LL^\top$ where L is upper triangular with positive diagonal entries. Taking the standard flat Euclidean metric on the (positive) upper diagonal matrices leads to straight line geodesics in that space: $L_t = L + t\dot{L}$ is the geodesic starting at L with (upper-triangular) tangent vector \dot{L} . It can be transported to a Cholesky geodesic in SPD matrix space using the product $P_t = L_t L_t^\top = P + t(\dot{L}L^\top + L\dot{L}^\top) + t^2\dot{L}\dot{L}^\top$. We see that the matrix product $L \rightarrow LL^\top$ plays here the role that was taken by the matrix exponential in the log-Euclidean framework. This mapping is a diffeomorphism between positive definite upper triangular and SPD matrices. However, contrarily to the log-Euclidean case, symmetric matrices with null eigenvalues are at a finite distance of any SPD matrix with the Cholesky metric, like for the Euclidean case.

6.3. Square root and Procrustes metrics

Since $P = (LR)(LR)^\top$ is also a valid decomposition of P for any rotation R , other definitions of square roots of SPD matrices can be used. For instance, the symmetric square root $P^{1/2} = (P^{1/2})^\top$ lead to a well defined metric on tensors which has similar properties to the Cholesky metric above, yet having different geodesics. The fact that the rotation R can be freely chosen to compute the square root led [DKZ09] to propose a distance measuring the shortest extrinsic distance between all the square roots $L_1 R_1$ of P_1 and $L_2 R_2$ of P_2 . The minimal extrinsic distance is realized by the Procrustes match of the square roots:

$$\text{dist}(P_1, P_2) = \min_{R \in O(n)} \|L_2 - L_1 R\|_{\text{Eucl}},$$

and the optimal rotation $\hat{R} = UV^\top$ is obtained thanks to the singular value decomposition of $L_2^\top L_1 = USV^\top$. This distance is in fact the standard Kendall structure on the reflection size-and-shape space of $n + 1$ points in dimension n [DKZ09, DM98, Sma96], which geometry is well known. For instance, the minimal geodesic joining P_1 to P_2 is given by

$$P(t) = \left((1-t)L_1 + tL_2\hat{R} \right) \left((1-t)L_1 + tL_2\hat{R} \right)^\top.$$

From the equation of the geodesics, one can derive the Riemannian exp and log map and proceed with the general computing framework. However, one must be careful that this space is not complete and has singularities when the matrix P has rank $n - 2$,

i.e. when 2 eigenvalues are going to zero [LK93]. The curvature of this space positive, which makes an important difference with the affine-invariant / log-Euclidean setting.

6.4. Extrinsic “distances”

The symmetrized Kullback Leibler divergence (J-divergence) was proposed as a “distance” on the SPD matrix space in [WV05] (it is specified in the paper that triangular inequality might not be verified):

$$\text{dist}_J^2(P_1, P_2) = \text{Tr}(P_1 P_2^{-1} + P_2 P_1^{-1}) - 2n.$$

This J-distance has interesting properties: it is affine invariant, and the Fréchet mean value of a set of tensors P_i has a closed form solution:

$$\bar{P} = B^{-1/2} \left(B^{1/2} A B^{1/2} \right)^{1/2} B^{-1/2},$$

with $A = \sum_i P_i$ and $B = \sum_i P_i^{-1}$. However, this is not a Riemannian distance as a Taylor expansion

$$\text{dist}_J^2(P, P + \epsilon V) = \frac{\epsilon^2}{2} \text{Tr}(P^{-1} V P^{-1} V) + O(\epsilon^3)$$

shows that the underlying infinitesimal dot product is the usual affine invariant metric $\langle V, W \rangle_P = \frac{1}{2} \text{Tr}(P^{-1} V P^{-1} W)$. In fact, this divergence is probably an extrinsic distance (whose triangular inequality remains to be shown) and it would be quite interesting to determine the underlying embedding. In any case, the algorithms based on this symmetric divergence should have results close to the affine-invariant ones when the data are sufficiently concentrated, provided that these algorithms can accommodate an extrinsic distance without direct correspondence to geodesics.

6.5. Power Euclidean metrics

In between a Euclidean metric and its log-Euclidean counterpart, we can design a family of Riemannian metrics based of power of the SPD matrices [DPP10]. The basic idea is to take the Euclidean distance after the power transformation $\frac{1}{\alpha} P^\alpha$ for $\alpha \neq 0$, so that the distance is:

$$\text{dist}_\alpha^2(P, Q) = \frac{1}{\alpha} \|P^\alpha - Q^\alpha\|_{Euc}^2.$$

The mapping $\exp(\alpha(\log P))/\alpha$ has a smooth invertible differential $d \exp|_{\alpha(\log P)} d \log|_P$ which realizes an isometry between Sym_n^+ and a subset of symmetric matrices with a Euclidean metric. SPD matrices with power-Euclidean metrics are thus flat spaces.

We recall from Section 2 that the power α of a SPD matrix is obtained by taking the power of the eigenvalues in the eigen-decomposition $P = U \text{DIAG}(d_i) U^\top$, using the formula $P^\alpha = U \text{DIAG}(d_i^\alpha) U^\top = \exp(\alpha(\log P))$. Since $\lim_{\alpha \rightarrow 0} (x^\alpha - 1)/\alpha =$

$\log(x)$, we see that the family tends to the log-Euclidean metric for $\alpha = 0$, and comprises the Euclidean metric for $\alpha = 1$ and the Euclidean metric on the precision matrix for $\alpha = -1$. If $\alpha \leq 0$ the symmetric matrices must be positive definite or have infinite eigenvalues, but if $\alpha > 0$ then we can also compare symmetric positive semi-definite (PSD) matrices.

The question of how to choose the value of α in practice for a specific dataset was investigated in [DPP10] using statistical inference to find the optimal value for which the data are as Gaussian as possible after transformation, similarly to a Box-Cox transformation. Preliminary tests on canine diffusion tensor image data suggested that an intermediate value between Euclidean and log-Euclidean $\alpha = 1/2$ (a square-root metric) could lead to a good description of the data in this specific case.

6.6. Which metric for which problem?

The zoo of intrinsic Riemannian and extrinsic metrics that we just described raise the problem of the choice of the metric, which has to be data dependent. One can think of optimizing the best suited metric, as described in [DPP10] for the power-Euclidean metrics. Taking into account all the metrics above would require finding a larger parametric family comprising all of them. This is an interesting challenge. A simpler method is to list the main properties of these metrics in order to identify which of them make sense for the data under investigation.

The space of SPD matrices endowed with affine-invariant and log-Euclidean metrics is geodesically and metrically complete. This is a desirable feature when we want to have matrices with null eigenvalues at an infinite distance of any SPD matrix and cannot be reached in finite time in practice. This is an important feature for gradient descent algorithms. In both the Hadamard and flat log-Euclidean case, the mean always exists and is unique, which simplifies many algorithms. For diffusion tensor data, the negative curvature of the affine-invariant metric seems to be small enough so that algorithms can be made much more efficient using the flat log-Euclidean metric.

For many other flat or positively curved metrics (e.g. power Euclidean or Cholesky), the space of SPD matrices is open and has a boundary including some rank-deficient matrices that can be reached in finite time. This means that positive semi-definite (PSD) matrices should make sense for the investigated data and that most of the algorithms developed so far need to be adapted to manifold with boundaries. Moreover, the potential multiplicity of the mean value in positively curved (or in flat but non-convex) manifolds raises other algorithmic challenges.

An example of PSD matrix data is found in computer vision with the structure tensor of an image [KWA11]. This is the Gaussian convolution of the gradient of the image tensored by itself $S_\sigma = G_\sigma \star (\nabla I \nabla I^T)$. The structure tensor field reveals structural information like edges and corners which are used to guide the anisotropic

smoothing of the original image. High values of σ favor smooth structure tensor fields while smaller values can help extracting low level features in images but produces a lot of noise on the field. Smoothing anisotropically this tensor field may help regularizing homogeneous regions while preserving edges. Here, filtering the coefficients only (using the Euclidean metric) produces negative eigenvalues, while filtering with the affine-invariant or log-Euclidean metrics forbids being close to null eigenvalues and gives the same importance to small differences in small tensors as to large differences in large ones [FAAP05]. This means that the anisotropic diffusion enhance these small details as much as the large scale ones. In fact, we should realize that structure tensors with one vanishing eigenvalue in 2D (or 2 vanishing eigenvalues in 3D) represents a perfect infinite edge in the image. This type of PSD matrices should thus be naturally reachable (i.e. at a finite distance) while negative eigenvalues should be forbidden. In this case, metrics with flat or positive curvature like Cholesky or Procrustes could be better suited.

In order to understand the behavior of each type of metric, Fig. 0.6 illustrates the geodesic shooting from a SPD matrix with the same tangent vector with the Euclidean, log-Euclidean metric (affine-invariant is very similar), and Cholesky metrics. The Euclidean geodesic quickly reaches the boundary of the space (a matrix with zero eigenvalues) after which it is not defined anymore. The log-Euclidean geodesic reaches null eigenvalues asymptotically at infinity, while the Cholesky geodesic reaches a null eigenvalue in finite time but bounce back and becomes positive anew after that point. There is thus no universal metric for SPD matrices, but many different families of metrics sharing similar or distinct characteristics that one has to investigate to choose the most adapted to the application needs.



Figure 0.6 Geodesic shooting with several metrics from the SPD matrix $P = \text{DIAG}(4, 1)$ with the tangent vector $\dot{P} = -4P$. **Left:** Along the Euclidean geodesic, the geodesic is $P_t = (1 - 4t)P$, so that both eigenvalues vanish at $t = 1/4$ and the geodesic cannot be continued after in Sym_n^+ . One can see the linear height decrease. **Middle:** The Cholesky decomposition of P gives the diagonal (thus upper diagonal) matrix $L = \sqrt{P} = \text{DIAG}(2, 1)$ and the tangent vector $\dot{L} = -2L$. The Cholesky geodesic is thus $P_t = (L + t\dot{L})^2 = (1 - 2t)^2P$. Both eigenvalues vanish at $t = 1/2$ and regrow symmetrically on the other side. Notice the quadratic height evolution. **Right:** along the log-Euclidean geodesic, matrices are always SPD. The height decrease is exponential.

7. Applications in Diffusion Tensor Imaging (DTI)

Diffusion Tensor Imaging (DTI) is a Magnetic Resonance Imaging (MRI) modality which was developed to measure in vivo oriented structures within tissues thanks to

the directional measure of the water diffusion. The measure of intravoxel incoherent motions within each voxel was proposed as early as 1986 [LBBL⁺86]. Specific gradient pulses produce a spin-echo attenuation that can be related to the diffusion in the gradient direction. Thus, one can estimate an Apparent Diffusion Coefficient (ADC) from a reference B_0 image (without gradient) and a few Diffusion Weighted Images (DWI) with gradients in different directions. In 1993, the single isotropic model of diffusion evolved into a matrix field that is allowing the measure of the anisotropic behavior of diffusion [BMB94], in particular within the white matter. The diffusion matrix is symmetric and can be interpreted as the covariance matrix of the Brownian motion of water in the tissues. This symmetric matrix image model was named Diffusion Tensor Imaging (DTI), and symmetric positive definite matrices are now classically referred to as tensors in medical image analysis. The Stejskal-Tanner equation relates the diffusion tensor D to noise-free DWI B_i acquired with the encoding gradient g_i and diffusion factor b :

$$B_i = B_0 \exp(-b g_i^\top D g_i). \quad (0.24)$$

Taking the logarithm of this equation leads to a very simple linear system which can be solved in the least-square sense using algebraic methods (see e.g. [WMM⁺02]).

In an ideal fibrous tissue, the principal eigenvector of the diffusion tensor is aligned with the fiber orientation. This gives an information about the direction of the neural tracts at each point of the image, whose global shape can be reconstructed into fiber bundles using tractography. Thus, diffusion imaging provides in vivo imaging of the white matter architecture of nervous fibers (axons) and allows to get an insight on the brain’s information highways! Fiber tracking is providing a massive amount of detailed information about the macroscopic structures of the brain and is used in many neuroscience studies, even if the validity of the anatomical interpretation of tractographies at the microscopic level remains under discussion. Higher order models of diffusion are nowadays developed (Diffusion Spectrum Imaging (DSI), High Angular Resolution Diffusion Imaging (HARDI), Q-ball, etc [TRW⁺02]) both to increase the orientation accuracy and to determine the different compartments at the microstructure level within each voxel. In all cases, good images require a good scanner quality (in particular very high and fast gradients) and a relatively long acquisition time.

On the other side of the application spectrum, DTI is finding clinical applications in brain tumors, spinal chord pathologies, epilepsy, diffuse axonal injury, multiple Sclerosis, Alzheimer Disease and ischemic stroke [LMK⁺14]. Here, DTI as a quantitative tool for medical diagnosis is hampered by the lower scanners quality in clinical environments and by the limited time for a clinical MRI acquisition. This results into images with a quite low signal-to-noise ratio (SNR) and a limited number of encoding gradients. Because the estimation of the tensor field from DWIs is quite noise sensitive, fiber tracking is often difficult with clinical DTI.

Simple methods estimate the tensor image with the above linear method and then spatially regularize the tensor field. From the signal processing point of view, this amounts to assume a log-Gaussian noise on the images, while it is rather Rician in MRI as we take the amplitude of a complex Gaussian signal in the k-space [SdD04]. The Gaussian or log-Gaussian noise assumption is valid only for high SNR images and it biases the estimation for low SNR clinical DTI. Rician noise removal on DWI images has been investigated [BFW06]. However, smoothing independently each DWI before estimating the tensor may blur the transitions between distinct fiber tracts which delimit anatomical and functional brain regions since the transitions may not be seen correctly in all gradient directions.

Instead of separating the tensor estimation from the DWI and the DTI regularization, a statistically more interesting model is to consider the spatial regularization as a prior distribution in a Bayesian tensor estimation with a Rician DWI likelihood [FAPA06, FAPA07]. For instance, a maximum a posteriori (MAP) algorithm jointly (rather than sequentially) performs the estimation and the regularization of the tensor field. Such an optimization is very easily developed in the log-Euclidean framework. One could also formulate it with an affine-invariant metric but calculations are slightly more complex. Let us consider a Gaussian DWI noise model to start. The tensor image $D(x) = \exp(W(x))$ is parametrized by an unconstrained symmetric matrix $W(x)$. In the following, we skip the position x in the image when not necessary. The log-likelihood for a Gaussian noise is $\text{Sim}_{\text{Gauss}}(W) = \sum_i (\hat{B}_i - B_i(W))^2$, where \hat{B}_i is the DWI observed with the gradient direction g_i and $B_i(W) = B_0 \exp(-b g_i^\top \exp(W) g_i)$ is the value modeled from the tensor parameter W using the Stejskal-Tanner equation. The derivative of this criterion with respect to W is:

$$\nabla_W \text{Sim}_{\text{Gauss}}(W) = 2b \sum_i (\hat{B}_i - B_i(W)) \partial_W B_i(W) \text{ with } \partial_W B_i(W) = B_i(W) \partial_{g_i g_i^\top} \exp(W).$$

The partial derivative of the matrix exponential was computed in Equation 0.4.

In order to take into account the native Gaussian noise in the k-space, [WVCM04] developed an estimation criterion on the complex DWI signal with a computationally grounded optimization framework based on the Cholesky decomposition. In clinical images, the phase is often discarded to conserve only the amplitude. For a Rician noise of variance σ^2 on the data, the probability density of the measured signal \hat{B} knowing the expected signal B is in that case [SdDSD98]:

$$\rho(\hat{B}|B) = \frac{\hat{B}}{\sigma^2} \exp\left(-\frac{\hat{B}^2 + B^2}{2\sigma^2}\right) I_0\left(\frac{B\hat{B}}{\sigma^2}\right), \quad (0.25)$$

where I_0 is the modified 0 order Bessel function of the first kind. The Rician noise induces a signal-dependent bias of the order of $\sigma^2/2B$ on the DWI signal [SdDSD98]. As a consequence, tensors are under-estimated with the least-squares estimation because the signal is systematically overestimated.

Under a Rician noise, the log-likelihood of DWI images with independent voxels is then: $\text{Sim}_{\text{Rice}}(W) = -\sum_i \log(\rho(\hat{B}_i|B_i(W)))$. The derivative of this criterion with respect to the tensor parameter W is quite simple [FAPA06, FAPA07]:

$$\nabla_W \text{Sim}_{\text{Rice}}(W) = -\frac{2b}{\sigma^2} \sum \left(B_i(W) - \alpha \left(\frac{\hat{B}_i B_i(W)}{\sigma^2} \right) \hat{B}_i \right) \partial_W B_i(W) \quad \text{with} \quad \alpha(x) = \frac{I'_0(x)}{I_0(x)}.$$

It is very similar to the previous Gaussian case up to a correction factor α that depends on the signal and the noise variance. The variance of the noise is usually estimated on the background of the image where there is no signal.

The Markovian prior $p(P(x+dx)|P(x)) \propto \exp\left(-\frac{1}{\lambda} \|\partial_{dx} P(x)\|_{P(x)}\right)$ is the discrete version of the harmonic energy of Section 5.3 for the spatial regularity. For preserving discontinuities, we may use a redescending M-estimator such as $\phi(s) = 2\sqrt{1+s^2/\kappa^2} - 2$ [FAPA07]. The ϕ -function helps preserving the edges of the tensor field while promoting the smoothing of homogeneous regions. The prior becomes: $\text{Reg}(W) = \int_{\Omega} \phi(\|\nabla W\|)$. In order to adapt the previous maximum-likelihood (ML) gradient descent into a maximum a-posteriori (MAP) estimation, we simply need to add the derivative of this prior to the ML criterion gradient:

$$\nabla_W \text{Reg}(W) = -\psi(\|\nabla W\|) \Delta W - \sum_i \partial_i (\psi(\|\nabla W\|)) \partial_i W \quad \text{where} \quad \psi(s) = \phi'(s)/s.$$

Directional derivatives, gradient and Laplacian can be estimated with finite differences like for scalar images (see [FAPA06, FAPA07] for details).

Experiments on synthetic data with a Rician noise showed that the above MAP technique was correctly avoiding the negative eigenvalues which inevitably appear in the standard linear estimation technique. The results of the ML (without regularization) and the MAP (with regularization) estimation methods with a log-Euclidean parametrization of the tensor image showed that the volume of the tensors was underestimated by 20% when we assume a log-Gaussian noise. Assuming a Gaussian noise model leads to an even higher underestimation of 30%. The estimation of the volume was within 5% with the Rician ML and MAP methods [FAPA07].

Results on two clinical acquisitions are presented in (Fig. 0.7). A first dataset is a brain image with a very low SNR, and a second dataset is an experimental acquisition of a tumor in the spinal chord, both with 7 gradient directions [FOF⁺05]. This last type of acquisition is sometimes difficult to perform because the position is uncomfortable due to the tumor and the coil cannot be perfectly adapted to the body as it is for the head. Consequently, spinal chord images are noisier than brain ones.

As for synthetic data, the negative tensor eigenvalues of the standard method disappear with the ML or MAP estimation. Several biomarkers such as the fractional anisotropy (FA) and the volume of the diffusion tensors in the ventricles/corpus callosum were used to assess the influence of the noise model used. Results using the Rician ML estimation showed an increase of the tensor volume and the ADC of about

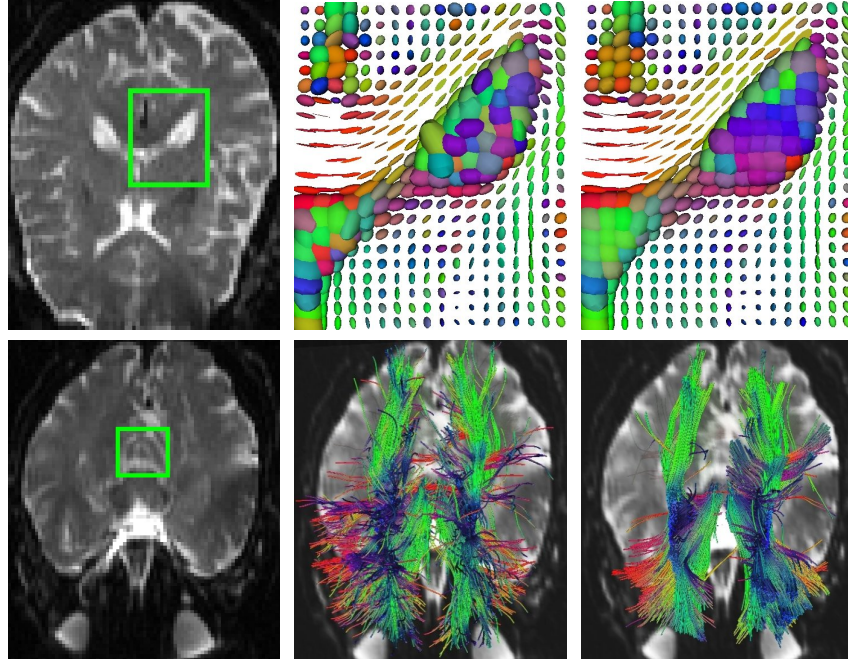


Figure 0.7 Tensor field estimation of a brain (top row) and improvement of the fiber tracking (bottom row). **Top Left:** A slice of the B_0 image. **Top Middle:** The classic log-Gaussian estimation on the ROI. The color codes for the principal direction of tensors: **red:** left-right, **green:** anterior-posterior, **blue:** inferior-superior. Missing tensors in the splenium region are non-positive. **Top Right:** The MAP estimation of the same region. **Bottom Left:** ROI where the tracking is initiated. **Bottom middle:** The cortico-spinal tract reconstructed after a classic estimation. **Bottom Right:** Same tract reconstructed after the MAP estimation. [Image reproduced from [Pen08]].

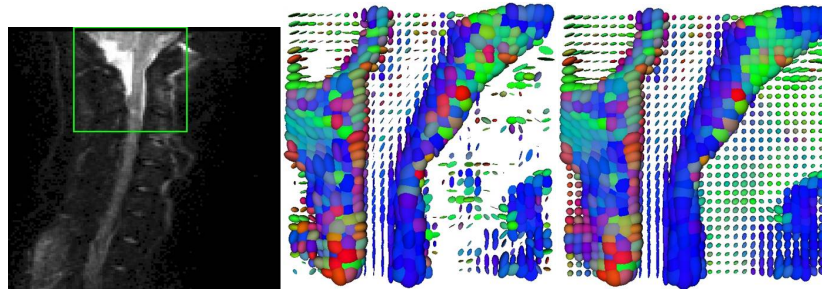


Figure 0.8 Tensor field estimation of the spinal cord. **Left:** A slice of the B_0 image with the ROI squared in green. **Middle:** Classic log-Gaussian ML tensor estimation. There are many missing (non-positive) tensors around and in the spinal cord. **Right:** Rician MAP tensor estimation: tensors are all positive and the field is much more regular while preserving discontinuities. Original DWI are courtesy of D. Ducreux, MD. [Image reproduced from [Pen08]].

10% in isotropic brain regions, 1 to 2% in anisotropic brain regions, and about 30% in the spinal chord, without modifying the FA [FAPA07]. These values are in line with the expected correction of the shrinking effect due to the log-Gaussian noise model.

In MAP methods where spatial regularization is added, Gaussian and log-Gaussian Rician noise models exhibited a more severe decrease of the FA than the Rician noise model. Moreover, the effect is much stronger in homogeneous regions (15% FA decrease in the ventricles with Rician vs 30% with log-Gaussian) than in anisotropic regions (3% FA decrease in the corpus callosum with Rician vs 11% with log-Gaussian). Thus, the reproducibility of the biomarkers measurements is higher with the MAP Rician tensor reconstruction. The tractography results with the MAP Rician estimation showed more numerous, longer, smoother and less dispersed fibers. In fact, tracts that were previously stopped because of the noise seem to be fully reconstructed. The radiologists found these results anatomically more meaningful.

8. Learning Brain Variability from Sulcal Lines

Modeling the statistical variability of the brain shape in a population from 3D images is a second interesting application of SPD-matrix-valued image processing [FAP⁺05, FAP⁺07]. In Chapters 4 and 5 of this book, the inter-individual shape variability will be modeled by a diffeomorphic deformation from a template to all the subject images. In this Section, we assume a simpler setting where we identify corresponding anatomical features (points or more generally lines or surfaces) among the anatomy of individuals (structural homologies). The statistical analysis is then performed on the displacement field between the template and the individuals. The covariance matrix of each point of the template independently could be seen as a first order estimation of the spatial Riemannian metric that we should use for inter-subject registration. Because structural variations are larger along certain directions [TMW⁺01], we cannot simplify this SPD matrix field into a simpler scalar variance function. Since the structural homologies can only be sparsely estimated, we actually have to extrapolate a SPD matrix field from sparse measurements, similarly to Section 5.5.

Sulcal lines are low dimensional structures easily identified by neuroscientists that consistently appear in the normal brain anatomy. The main sulcal lines are used to subdivide the cortex into major lobes and gyri [MRC⁺04]. In [FAP⁺05, FAP⁺07], 72 sulcal curves that consistently appear in all subjects were selected. The curves were manually delineated in 98 subjects by expert neuroanatomists according to a precise protocol with formal rules governing the handling of branching patterns, breaks in sulci, and doubling of specific sulci (see above references for details). The inter- and intra-rater error (reliability) is less than 2mm (in r.m.s.) everywhere, and in most regions less than 1mm, far less than the inter-subject anatomical variance. To determine the point correspondences between different instances of each sulcal line, a classi-

cal method is to alternate the computation of the closest point from the mean curve to each instance and the re-estimation of the mean curve from the updated correspondences (this is the max-max algorithm of Chapter 9). A constraint imposing monotonic matches along the curves was added with dynamic programming and a global affine transformation per subject was also removed. This matching obviously underestimates the variability in the tangent direction to the sulcal lines except maybe at the endpoints that can be considered as landmarks. The mean sulcal lines are illustrated in Figure 0.9.

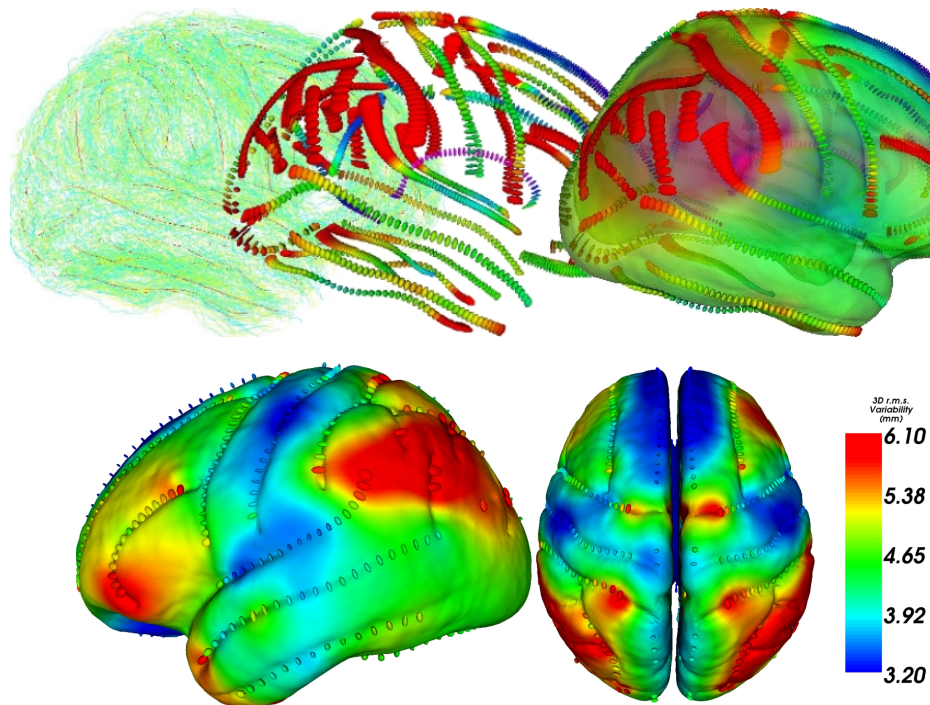


Figure 0.9 Top: From sulcal lines to the brain variability. Left : sulcal lines of 80 subjects in green with the mean sulcal lines in red. Middle: variability measured along the mean sulcal lines (covariance matrix at one sigma). Right: the color encodes the amount of variability everywhere on the cortex after the extrapolation of the covariance onto the whole 3D space. **Bottom:** The 366 SPD matrices selected along each sulcal lines, and the trace of the extrapolated covariance matrix. Lateral view from the left (on the left) and from above (right). Images courtesy of P. Fillard.

The correspondences from the mean sulcal lines to all subjects are then summa-

rized in a covariance matrix along each sulcal line that encodes the point-wise variability up to second order (since the mean is zero by construction of the mean sulcal lines). These covariance matrices are highly correlated along each line. In order to retain only the informative ones, the subset of SPD matrices optimizing a prescribed maximal distance (0.2 in the affine-invariant metric sense) between interpolated and measured tensors along the line is computed. The selection of 366 covariance matrices was found to be sufficient to encode the variability of the 72 sulci with a low RMS error. The resulting sparse field of SPD matrices was then extrapolated to the whole space using the framework described in Section 5.5 (see Fig. 0.9).

The dense map of covariance matrices which is obtained was qualitatively agreeing with the previously reported brain variability results [FAP⁺07]: areas that are highly specialized and lateralized such as the planum parietale and the temporo-parietal areas exhibit the largest variability. Phylogenetically older areas (e.g. orbitofrontal cortex) and primary cortices that myelinate earliest during development (e.g., primary somatosensory and auditory cortex) display the minimum of variability.

However, the 3D covariance matrix field contains more information than a scalar evaluation of the variability: it also contains directions that could be used to better guide the registration for instance. Even more interestingly, it is possible to study with this approach the co-variability of different areas of the brain: instead of constructing the covariance matrix of the inter-subject displacements at each point, we may construct the joint variance-covariance matrix of the inter-subject displacements at two points x and y of the space. For prescribe initial points x , this amount to extrapolate an image of 6×6 covariance matrices (the point y being the spatial coordinate of the image).

The total covariance matrix $\Lambda(x, y) = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix}$ can be analyzed using Canonical Correlation Analysis (CCA): the correlation matrix $\Gamma(x, y) = \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2}$ is decomposed using SVD to find the correlation coefficients between the x and y variables. A chi-square test allows to state if the correlation matrix is significantly different from zero, (i.e. if there exists at least an axis in which there is some correlation), in which case we may report the maximal correlation coefficient and the related axes in which this correlation happens at both points. In this process, the components that are tangential to the sulcal lines may be questioned since they are known to be notably underestimated (this is called the aperture problem). They can be removed completely from the test at this stage to avoid biases. Since this test should be corrected for multiple comparisons, it is also important to limit the number of candidate pairs of points that are tested.

In [FPTA07], three reference positions (beginning, middle, and end point) were selected along two important sulci: the Central Sulcus (CS) and the Inferior Temporal Sulcus (ITS). These sulci lie in different lobes, develop at different times during

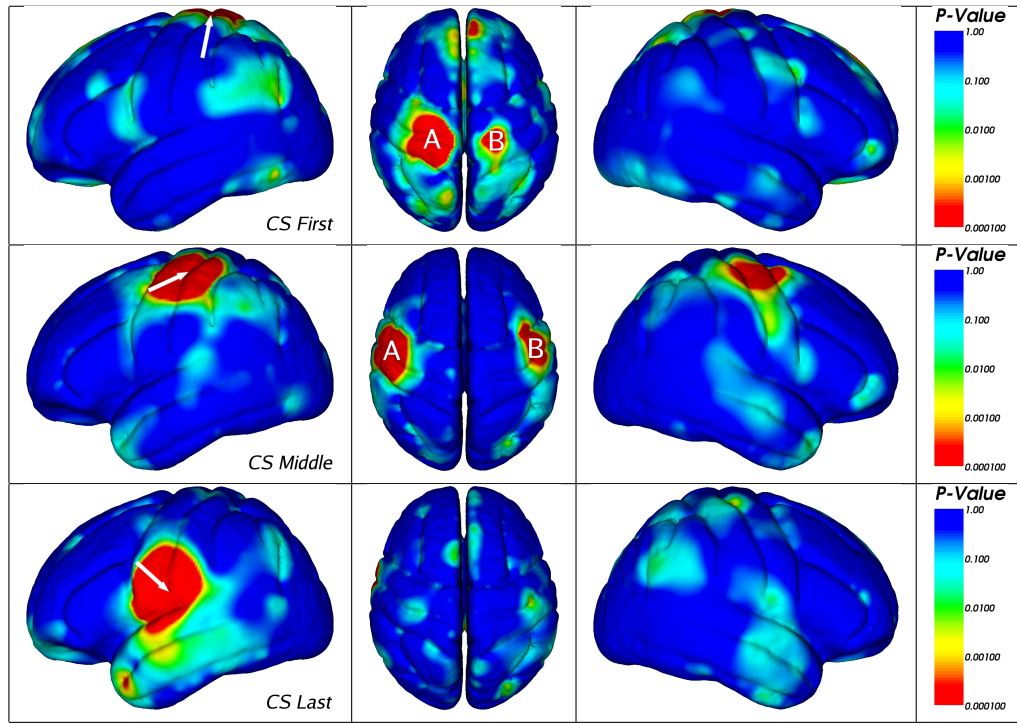


Figure 0.10 Correlation Maps between the Central Sulcus and Other Brain Regions. A white arrow in each row indicates a reference landmark; correlations with the reference landmark are plotted. Correlations for 3 reference landmarks on the CS are shown: the first (top row), the middle (second row), and the last, i.e. most inferior, position (third row) on the sulcal trace. Corresponding regions in the opposite hemisphere are highly correlated for the top and middle points (marked A and B). The lower end of the sulcus, however, exhibits low correlation with its symmetric contralateral counterpart. It is interesting to notice an unexpected long-range correlation between the variability of the top point of the left CS and a point of the right frontal lobe.

gyrogenesis (CS developing earlier) and are distant in terms of fiber and functional connectivity: we expect a priori a very low correlation between them. Thus, they are good candidates for assessing inter-structure correlation. The correlation maps over the brain surface are displayed in Figure 0.10 and Figure 0.11 colored by their p-value. The threshold for statistical significance of the CCA was set to 0.0001 to correct for multiple comparisons. A large area around the reference points shows high p-values: points that are anatomically close to the reference are likely to have a correlated distribution among individuals. More interestingly, corresponding brain regions in each hemisphere are highly correlated, except for regions including Wernicke’s and Broca’s areas, which are known to be functionally specialized in one hemisphere.

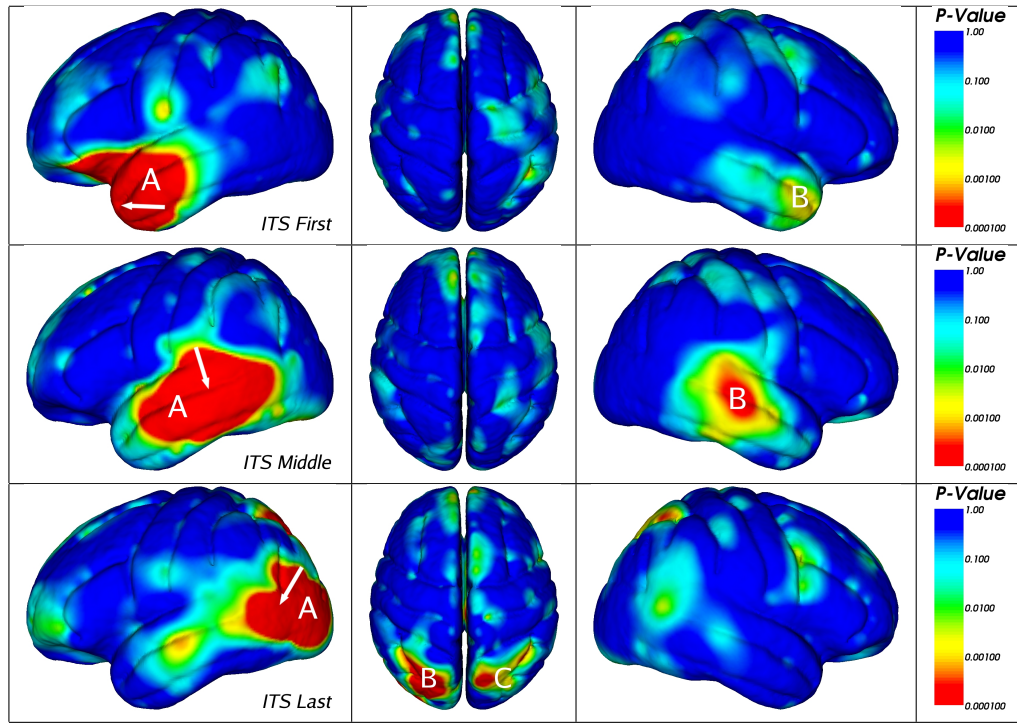


Figure 0.11 Correlation Maps between the Inferior Temporal Sulcus (ITS) and Other Brain Regions. A white arrow in each row indicates a reference landmark; correlations with the reference landmark are plotted. Correlations for 3 reference landmarks on the ITS are shown: the first (top row), the middle (second row), and the last, i.e. most inferior, position (third row) on the sulcal trace. The first and middle positions are symmetrically correlated (marks A and B). The last position (third row) correlates less with its opposite hemisphere counterpart, than with the intra-parietal sulci (marked B and C). This could be expected because the intersubject variability is very low and reaching the inter-rater reliability. [Figure reproduced from [FPTA07]]

The posterior tip of the ITS is also very loosely correlated to its opposite hemisphere counterpart. As this structure is highly variable and is specialized for understanding the semantics of language in the left hemisphere and the prosodic aspects of language in the right hemisphere, this may suggest partially independent developmental programs. The long-range correlation between the back of the ITS and the left and right intra-parietal sulci is in itself an interesting neuroscience finding: the planum temporale and planum parietale are the two distinct areas most widely studied in neuroscience for their very high hemispheric asymmetry. Such long-range correlations may reflect common factors driving programmed asymmetries for both regions.

Acknowledgments

The author would like to thank Yann Thanwerdas for the careful proofreading of the chapter and for checking the lengthy computations of the connection and curvatures. This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant G-Statistics agreement No 786854).

- [AFPA05] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In J. Duncan and G. Gerig, editors, *Proceedings of the 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005, Part I*, volume 3749 of *LNCS*, pages 115–122, Palm Springs, CA, USA, October 26–29, 2005. Springer. PMID: 16685836.
- [AFPA06] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, August 2006. PMID: 16788917.
- [AFPA07] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007.
- [AK01] Gilles Aubert and Pierre Kornprobst. *Mathematical problems in image processing - Partial differential equations and the calculus of variations*, volume 147 of *Applied Mathematical Sciences*. Springer, 2001.
- [BBCJ12] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass BrainComputer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, April 2012.
- [BBSW16] Miroslav Bak, Ronny Bergmann, Gabriele Steidl, and Andreas Weinmann. A Second Order Nonsmooth Variational Model for Restoring Manifold-Valued Images. *SIAM Journal on Scientific Computing*, 38(1):A567–A597, January 2016.
- [BFW06] Saurav Basu, Thomas Fletcher, and Ross Whitaker. Rician Noise Removal in Diffusion Tensor MRI. In Rasmus Larsen, Mads Nielsen, and Jon Sporring, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2006*, pages 117–125, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [BGM97] Houman Borouchaki, Paul-Louis George, and Bijan Mohammadi. Delaunay mesh generation governed by metric specifications. Part II: applications. *Finite Elements in Analysis and Design*, pages 85–109, 1997.
- [Bha03] Rajendra Bhatia. On the exponential metric increasing property. *Linear Algebra and its Applications*, 375:211–220, 2003.
- [BI13] Dario A. Bini and Bruno Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. *Linear Algebra and its Applications*, 438(4):1700–1710, February 2013.
- [Bie04] Joris Bierkens. Geometric methods in diffusion tensor regularization. Master’s thesis, Technische Universiteit Eindhoven, Dept. of Math and Comp. Sci., 2004.
- [BMA⁺05] Philip Batchelor, Maher Moakher, David Atkinson, Fernando Calamante, and Alan Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53:221–225, 2005.
- [BMB94] Peter J. Basser, James Mattiello, and Denis Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
- [BR82] Jacob Burbea and C. Radhakrishna Rao. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.
- [BWBM06] Thomas Brox, Johachim Weickert, Bernhard Burgeth, and Pavel Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006.
- [Car92] Manfredo P. do Carmo. *Riemannian geometry*. Mathematics. Theory & applications. Birkhäuser, Boston, 1992.
- [CO91] Miquel Calvo and Josep M. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*, 9:119–138, 1991.
- [CTDF02] Christophe Chefd’hotel, David Tschumperlé, Rachid Deriche, and Olivier Faugeras. Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proc. of ECCV 2002*, volume 2350 of *LNCS*, pages 251–265. Springer Verlag., 2002.
- [CTDF04] Christophe Chefd’hotel, David Tschumperlé, Rachid Deriche, and Olivier Faugeras. Regularizing flows for constrained matrix-valued images. *J. Math. Imaging and Vision*, 20(1-

- 2):147–162, January - March 2004.
- [DKZ09] Ian L. Dryden, Alexey Koloydenko, and Diwei Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3(3):1102–1123, September 2009.
- [DM98] Ian L. Dryden and Kanti V. Mardia. *Statistical Shape Analysis*. John Wiley, Chichester, 1998.
- [DPP10] Ian L. Dryden, Xavier Pennec, and Jean-Marc Peyrat. Power Euclidean metrics for covariance matrices with application to diffusion tensor imaging. *arXiv:1009.3045 [stat]*, September 2010. arXiv: 1009.3045.
- [FAAP05] Pierre Fillard, Vincent Arsigny, Nicholas Ayache, and Xavier Pennec. A Riemannian framework for the processing of tensor-valued images. In Ole Fogh Olsen, Luc Florak, and Arjan Kuijper, editors, *Deep Structure, Singularities, and Computer Vision (DSSCV)*, LNCS, pages 112–123. Springer, June 2005.
- [FAP⁺05] Pierre Fillard, Vincent Arsigny, Xavier Pennec, Paul M. Thompson, and Nicholas Ayache. Extrapolation of sparse tensor fields: Application to the modeling of brain variability. In Gary Christensen and Milan Sonka, editors, *Proc. of Information Processing in Medical Imaging 2005 (IPMI’05)*, volume 3565 of LNCS, pages 27–38, Glenwood springs, Colorado, USA, July 2005. Springer. PMID: 17354682.
- [FAP⁺07] Pierre Fillard, Vincent Arsigny, Xavier Pennec, Kiralee M. Hayashi, Paul M. Thompson, and Nicholas Ayache. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*, 34(2):639–650, January 2007. Also as INRIA Research Report 5887, April 2006. PMID: 17113311.
- [FAPA06] Pierre Fillard, Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI 2006)*, pages 786–789, Crystal Gateway Marriott, Arlington, Virginia, USA, April 2006.
- [FAPA07] Pierre Fillard, Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, 26(11):1472–1482, November 2007. PMID: 18041263.
- [FJ07] P. Thomas Fletcher and Sarang Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, February 2007.
- [FM99] Wolfgang Förstner and Boudewijn Moonen. A metric for covariance matrices. In F. Krumm and V. S. Schwarze, editors, *Qua vadis geodesia...? Festschrift for Erik W. Grafarend on the occasion of his 60th birthday*, number 1999.6 in Tech. Report of the Dpt of Geodesy and Geoinformatics, pages 113–128. Stuttgart University, 1999.
- [FOF⁺05] David Facon, Augustin Ozanne, Pierre Fillard, Jean-François Lepeintre, Caroline Tournoux-Facon, and Denis Ducreux. MR diffusion tensor imaging and fiber tracking in spinal cord compression. *American Journal of Neuroradiology (AJNR)*, 26:1587–1594, 2005.
- [FPTA07] Pierre Fillard, Xavier Pennec, Paul Thompson, and Nicholas Ayache. Evaluating Brain Anatomical Correlations via Canonical Correlation Analysis of Sulcal Lines. In *MICCAI’07 Workshop on Statistical Registration: Pair-wise and Group-wise Alignment and Atlas Formation*, Brisbane, Australia, Australia, 2007. <https://hal.inria.fr/inria-00616033>.
- [FXCB13] Ricardo Ferreira, Joao Xavier, Joao P. Costeira, and Victor Barroso. Newton Algorithms for Riemannian Distance Related Problems on Connected Locally Symmetric Manifolds. *IEEE Journal of Selected Topics in Signal Processing*, 7(4):634–645, August 2013.
- [Gam91] R.V. Gamkrelidze, editor. *Geometry I*, volume 28 of *Encyclopaedia of Mathematical Sciences*. Springer Verlag, 1991.
- [GKKJ92] Guido Gerig, Ron Kikinis, Olaf Kübler, and Ferenc A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, June 1992.
- [Gre63] Ulf Grenander. *Probabilities on Algebraic Structures*. Wiley, 1963.
- [GS16] Philipp Grohs and Markus Sprechler. Total variation regularization on Riemannian manifolds by iteratively reweighted minimization. *Information and Inference*, 5(4):353–378, December 2016.

- [Hel78] Sigurdur Helgason. *Differential Geometry, Lie groups, and Symmetric Spaces*. Academic Press, 1978.
- [HM94] Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. Communication and Control Engineering Series. Springer, 1994.
- [Ken90] Wilfrid S. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.
- [KN69] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry, vol. II*. Number 15 in Interscience tracts in pure and applied mathematics. John Wiley & Sons, 1969.
- [KWA11] Hans Knutsson, Carl-Fredrik Westin, and Mats Andersson. Representing Local Structure Using Tensors. In Anders Heyden and Fredrik Kahl, editors, *Image Analysis*, pages 545–556, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [LBBL⁺86] Denis Le Bihan, Eric Breton, D Lallemant, P Grenier, E Cabanis, and M Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407, November 1986.
- [LBMP⁺01] Denis Le Bihan, Jean-François Mangin, Cyril Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriet. Diffusion tensor imaging: Concepts and applications. *Journal Magnetic Resonance Imaging*, 13(4):534–546, 2001.
- [Le01] Huiling Le. Locating Frchet Means with Application to Shape Spaces. *Advances in Applied Probability*, 33(2):324–338, 2001.
- [LK93] Huiling Le and David G. Kendall. The Riemannian structure of Euclidean shape space: a novel environment for statistics. *Annals of Statistics*, 21:1225–1271, 1993.
- [LMK⁺14] Alexander Lerner, Monique A. Mogensen, Paul E. Kim, Mark S. Shiroishi, Darryl H. Hwang, and Meng Law. Clinical Applications of Diffusion Tensor Imaging. *World Neurosurgery*, 82(1-2):96–109, July 2014.
- [LMO00] Miroslav Lovrić and Maung Min-Oo. Multivariate normal distributions parametrized as a Riemannian symmetric space. *Journal of Multivariate Analysis*, 74(1):36–48, 2000.
- [LNPS17] Friederike Laus, Mila Nikolova, Johannes Persch, and Gabriele Steidl. A Nonlocal Denoising Algorithm for Manifold-Valued Images Using Second Order Statistics. *SIAM Journal on Imaging Sciences*, 10(1):416–448, January 2017.
- [LRDF06] Christophe Lenglet, Michael Rousson, Rachid Deriche, and Olivier Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision*, 25(3):423–444, October 2006.
- [Man04] Jonathan H. Manton. A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups. In *Proc of ICARCV 2004 8th Control, Automation Robotics and Vision Conference*, volume 3, pages 2211–2216. IEEE, 2004.
- [Mei02] Eric Meijering. A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proceedings of the IEEE*, 90(3):319–342, March 2002.
- [MJ00] Kanti V. Mardia and Peter E. Jupp. *Directional statistics*. Wiley, Chichester, 2000.
- [MLT00] Gérard Medioni, Mi-Suen Lee, and Chi-Keung Tang. *A Computational Framework for Segmentation and Grouping*. Elsevier, 2000.
- [Moa05] Maher Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal of Matrix Analysis and Applications*, 26(3):735–747, 2005.
- [MRC⁺04] Jean-François Mangin, Denis Rivière, Arnaud Cachia, Edouard Duchesnay, Yves Cointepas, Dimitri Papadopoulos-Orfanos, Paola Scifo, T. Ochiai, Francis Brunelle, and Jean Régis. A framework to study the cortical folding patterns. *NeuroImage*, 23(Supplement 1):S129–S138, 2004.
- [MSH05] Bill Moran, Sofia Suvorova, and Stephen Howard. Sensor management for radar: a tutorial. In *Advances in Sensing with Security Applications 17–30 July 2005, Il Ciocco, Italy*. NATO Advanced Study Institute, 2005.
- [Nom54] Katsumi Nomizu. Invariant affine connections on homogeneous spaces. *American J. of*

- Math.*, 76:33–65, 1954.
- [OC95] Josep M. Oller and José Manuel Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, 23(5):1562–1581, 1995.
 - [Oll13] Yann Ollivier. A visual introduction to Riemannian curvatures and some discrete generalizations. In Alina Stancu Galia Dafni, Robert McCann, editor, *Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the 50th Séminaire de Mathématiques Supérieures (SMS), Montréal, 2011*, pages 197–219. AMS, 2013.
 - [PA98] Xavier Pennec and Nicholas Ayache. Uniform distribution, distance and expectation problems for geometric features processing. *Journal of Mathematical Imaging and Vision*, 9(1):49–67, July 1998. A preliminary version appeared as INRIA Research Report 2820, March 1996.
 - [Pen96] Xavier Pennec. *L’incertitude dans les problèmes de reconnaissance et de recalage – Applications en imagerie médicale et biologie moléculaire*. Thèse de sciences (phd thesis), Ecole Polytechnique, Palaiseau (France), December 1996.
 - [Pen06a] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. A preliminary appeared as INRIA RR-5093, January 2004.
 - [Pen06b] Xavier Pennec. *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, December 2006.
 - [Pen08] Xavier Pennec. Statistical computing on manifolds: from Riemannian geometry to computational anatomy. In Frank Nielsen, editor, *Emerging Trends in Visual Computing*, volume 5416 of *LNCS*, pages 347–386. Springer, 2008.
 - [Pen18] Xavier Pennec. Barycentric subspace analysis on manifolds. *The Annals of Statistics*, 46(6A):2711–2746, 2018.
 - [PFA06] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, January 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.
 - [PM90] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 12(7):629–639, 1990.
 - [Rao65] Callyampudi Radhakrishna Rao, editor. *Linear Statistical Inference and its Applications*. John Wiley and Sons, 1965.
 - [SdD04] Jan Sijbers and Arnold J. den Dekker. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magnetic Resonance in Medicine*, 51(3):586–594, March 2004.
 - [SdDSD98] Jan Sijbers, Arnold J. den Dekker, Paul Scheunders, and Dirk Van Dyck. Maximum likelihood estimation of Rician distribution parameters. *TMI*, 17(3), June 1998.
 - [Sko84] Lene Theil Skovgaard. A Riemannian geometry of the multivariate normal model. *Scand. J. Statistics*, 11:211–223, 1984.
 - [Sma96] Christopher G. Small. *The Statistical Theory of Shapes*. Springer series in statistics. Springer, 1996.
 - [Sou70] Jean-Marie Souriau. *Structure des systmes Dynamiques*. Dunod, 1970.
 - [TBU00] Philippe Thévenaz, Thierry Blu, and Michael Unser. Interpolation revisited. *IEEE Transactions on Medical Imaging*, 19(7):739–758, July 2000.
 - [TMW⁺01] Paul M. Thompson, Michael S. Mega, Roger P. Woods, Chris I. Zoumalan, Chris J. Lindshield, Rebecca E. Blanton, Jacob Moussai, Colin J. Holmes, Jeffrey L. Cummings, and Arthur W. Toga. Cortical change in Alzheimer’s disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex*, 11(1):1–16, January 2001.
 - [TRW⁺02] David S. Tuch, Timothy G. Reese, Mette R. Wiegell, Nikos Makris, John W. Belliveau, and Van J. Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48(4):577–582, October 2002.
 - [TSC00] Bei Tang, Guillermo Sapiro, and Vicent Caselles. Diffusion of General Data on Non-Flat Manifolds via Harmonic Maps Theory: The Direction Diffusion Case. *International Journal of Computer Vision*, 36(2):149–161, 2000.

- [Tsc02] David Tschumperlé. *PDE-Based Regularization of Multivalued Images and Applications*. PhD thesis, University of Nice-Sophia Antipolis, dec 2002.
- [WB02] Joachim Weickert and Thomas Brox. Diffusion and regularization of vector- and matrix-valued images. In M.Z. Nashed and O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging.*, volume 313 of *Contemporary Mathematics*, pages 251–268, Providence, 2002. AMS.
- [WDS14] Andreas Weinmann, Laurent Demaret, and Martin Storath. Total Variation Regularization for Manifold-Valued Data. *SIAM Journal on Imaging Sciences*, 7(4):2226–2257, January 2014.
- [Wei97] Joachim Weickert. A review of nonlinear diffusion filtering. In Gerhard Goos, Juris Hartmanis, Jan Leeuwen, Bart Haar Romeny, Luc Florack, Jan Koenderink, and Max Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252, pages 1–28. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.
- [Wei98] Joachim Weickert. *Anisotropic Diffusion in Image Processing*. ECMI Series. Teubner-Verlag, Stuttgart, Germany, 1998.
- [WH06] Joachim Weickert and Hans Hagen, editors. *Visualization and Processing of Tensor Fields*. Mathematics and Visualization. Springer, 2006.
- [WMM⁺02] Carl-Frederic Westin, Stephan E. Maier, Hatsuho Mamata, Arya Nabavi, Ferenc A. Jolesz, and Ron Kikinis. Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6(2):93–108, June 2002.
- [WV05] Zhizhou Wang and Baba C. Vemuri. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Trans. on Medical Imaging*, 24(10):1267–1277, October 2005.
- [WVCM04] Zhizhou Wang, Baba C. Vemuri, Yunmei Chen, and Thomas H. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Trans. on Medical Imaging*, 23(8), 2004.