INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Riemannian Framework for Tensor Computing

Xavier Pennec, Pierre Fillard, Nicholas Ayache

N° 5255

July 2004

Thème BIO

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

A Riemannian Framework for Tensor Computing

Xavier Pennec, Pierre Fillard, Nicholas Ayache

Thème BIO — Systèmes biologiques
Projet Epidaure

Rapport de recherche n° 5255 — July 2004 — 33 pages

Abstract: Positive definite symmetric matrices (so-called tensors in this article) are nowadays a common source of geometric information. In this paper, we propose to provide the tensor space with an affine-invariant Riemannian metric. We demonstrate that it leads to strong theoretical properties: the cone of positive definite symmetric matrices is replaced by a regular manifold of constant curvature without boundaries (null eigenvalues are at the infinity), the geodesic between two tensors and the mean of a set of tensors are uniquely defined, etc.

We have previously shown that the Riemannian metric provides a powerful framework for generalizing statistics to manifolds. In this paper, we show that it is also possible to generalize to tensor fields many important geometric data processing algorithms such as interpolation, filtering, diffusion and restoration of missing data. For instance, most interpolation schemes and Gaussian filtering can be tackled efficiently through a weighted mean computation. Linear and anisotropic diffusion schemes can be adapted to our Riemannian framework, through partial differential evolution equations, provided that the metric of the tensor space is taken into account. For that purpose, we provide intrinsic numerical schemes to compute the gradient and Laplacian operators. Finally, to enforce the fidelity to the data (either sparsely distributed tensors or complete tensors fields) we propose least-squares criteria based on our invariant Riemannian distance that are particularly simple and efficient to solve.

Key-words: Riemannian geometry, tensors fields, PDE, diffusion, restoration, invariance.

Unité de recherche INRIA Sophia Antipolis

Un cadre riemannien pour manipuler les tenseurs

Résumé : Les matrices définies positives, ici appelées tenseurs, sont des sources d'information géométrique de plus en plus disponibles. Dans cet article, nous proposons de munir l'espace des tenseurs d'une métrique riemannienne invariante par transformation affine. Cela mène à des propriétés théoriques fortes, puisque le cône des matrices définies positives (une variété plate mais à bords) est transformé en une variété régulière de courbure constante et surtout sans bord (les valeurs propres nulles sont à l'infini). De plus, les géodésiques entre deux tenseurs sont définies de manière unique, tout comme la moyenne d'un ensemble de tenseurs.

Nous avons auparavant montré que la géométrie riemannienne fournissait un cadre de travail puissant pour généraliser les statistiques aux variétés. Dans cet article, nous montrons qu'il est possible de généraliser aussi aux champs de tenseurs de nombreux algorithmes de traitement de données géométriques comme l'interpolation, le filtrage, la diffusion et la restauration de données manquantes. Par exemple, la plupart des schémas d'interpolation et le filtrage gaussien peuvent être considérés comme des moyennes pondérées. Les schémas de diffusion linéaire ou anisotropes peuvent être adaptés à notre cadre riemannien au travers d'équation d'évolution aux dérivées partielles, pourvu que la métrique de l'espace des tenseurs soit prise en compte. Nous fournissons pour cela des schémas numériques intrinsèques pour calculer les opérateurs gradient et laplacien. Pour finir, nous proposons des critères aux moindres carrés basés sur notre distance riemannienne invariante pour garantir une attache aux données, que ce soit pour des champs de tenseurs denses ou épars, qui s'avèrent être particulièrement simples et efficaces à résoudre.

Mots-clés : Géométrie riemannienne, champs de tenseurs, EDP, diffusion, restauration, invariance.

Contents

1 Introduction

Positive definite symmetric matrices (so-called tensors in this article) are nowadays a common source of geometric information, either as covariance matrices for characterizing statistics on deformations, or as an encoding of the principle diffusion directions in Diffusion Tensor Imaging (DTI). The measurements of these tensors is often noisy in real applications and we would like to perform estimation, smoothing and interpolation of fields of this type of features. The main problem is that the tensor space is a manifold that is not a vector space. As symmetric positive definite matrices constitute a convex half-cone in the vector space of matrices, many usual operations (like the mean) are stable in this space. However, problems arise when estimating tensors from data (in standard DTI, the estimated symmetric matrix could have negative eigenvalues), or when smoothing fields of tensors: the numerical schemes used to solve the Partial Differential Equation (PDE) may sometimes lead to negative eigenvalues if the time step is not small enough. Even when a SVD is performed to smooth independently the rotation (eigenvectors basis trihedron) and eigenvalues, there are continuity problem around equal eigenvalues.

In previous works [Pennec, 1996, Pennec and Ayache, 1998], we used invariance requirements to develop some basic probability tools on transformation groups and homogeneous manifolds. This statistical framework was then reorganized and extended in [Pennec, 1999, Pennec, 2004] for general Riemannian manifolds, invariance properties leading in some case to a natural choice for the metric. In this paper, we show how this theory can be applied to tensors, leading to a new intrinsic computing framework for these geometric features with many important theoretical properties as well as practical computing properties.

In the remaining of this section, we quickly investigate some connected works on tensors. Then, we summarize in Section 2 the main ideas of the statistical framework we developed on Riemannian manifolds. The aim is to exemplify the fact that choosing a Riemannian metric "automatically" determines a powerful framework to work on the manifold through the introduction of a few tools from differential geometry. In order to use this Riemannian framework on our tensor manifold, we propose in Section 3 an affine-invariant Riemannian distance on tensors. We demonstrate that it leads to very strong theoretical properties, as well as some important practical algorithms such as an intrinsic geodesic gradient descent. Section 4 focus on the application of this framework to an important geometric data processing problem: interpolation of tensor values. We show that this problem can be tackled efficiently through a weighted mean optimization. However, if weights are easy to define for regularly sampled tensors (e.g. for linear to tri-linear interpolation), the problem proved to be more difficult for irregularly sampled values.

With Section 5, we turn to tensors field computing, and more particularly filtering. If the Gaussian filtering may still be defined through weighted means, the partial differential equation (PDE) approach is slightly more complex. In particular, the metric of the tensor space has to be taken into account when computing the magnitude of the spatial gradient of the tensor field. Thanks to our Riemannian framework, we propose efficient numerical schemes for the computation of the gradient, its amplitude, and for the Laplacian used in linear diffusion. We also propose an adjustment of the Laplacian that realizes an anisotropic filtering. Finally, Section 6 focus on simple statistical approaches to regularize and restore missing values in tensor fields. Here, the use of the Riemannian distance inherited from the chosen metric is fundamental to define least-squares data attachment criteria for dense and sparsely distributed tensors fields that lead to simple implementation schemes in our intrinsic computing framework.

1.1 Related work

Quite an impressive literature has now been issued on the estimation and regularization of tensor fields, especially in the context of Diffusion Tensor Imaging (DTI) [Basser et al., 1994, Bihan et al., 2001, Westin et al., 2002]. Most of the works dealing with the geometric nature of the tensors has been perform for the discontinuity-

preserving regularization of the tensor fields using Partial Differential Equations (PDE) (see [Coulon et al., 2004] for a recent review). For instance, [Coulon et al., 2004] anisotropically restores the principal direction of the tensor, and uses the this regularized directions map as an input for the anisotropic regularization of the eigenvalues. A quite similar idea is adopted in [Tschumperlé, 2002], where a spectral decomposition $W(x) = U(x).D(x).U(x)^{\mathrm{T}}$ of the tensor field is performed at each points to independently regularize the eigenvalues and eigenvectors (orientations). This approach requires an additional reorientation step of the rotation matrices due to the non-uniqueness of the decomposition (each eigenvector is defined up its sign and there may be joint permutations of the eigenvectors and eigenvalues) in order to avoid the creation of artificial discontinuities. Another problem arises when two or more eigenvalues become equal: a whole subspace of unit eigenvectors is possible, and even a re-orientation becomes difficult. An intrinsic integration scheme for PDE that uses the exponential map has been added in [Chefd'hotel et al., 2002], and allows to perform PDE evolution on the considered manifold without re-projections. In essence, this is an infinitesimal version of the intrinsic gradient descent technique on manifolds we introduced in [Pennec, 1996, Pennec, 1999] for the computation of the mean.

In [Chefd'hotel et al., 2004], the same authors propose a rank-signature preserving flow that inherits its properties from the matrix exponential, without a clear reference to the metric used. Their derivation leads to an evolution equation similar to Eq. 2 that will define the geodesics of our metric in Section 3.3. They claim that this *rank/signature preserving flow tends to blend the orientation and diffusivity features (eigenvalue swelling effect)*. We do not observe such a behavior with our PDE evolution scheme of Sections 5 and 6. The explanation lies probably in the fact that we are not using the same metric when computing the driving gradient forces.

The affine-invariant Riemannian metric we detail in Section 3.3 may be traced back to the work of [Nomizu, 1954] on affine invariant connections on homogeneous spaces. It is implicitly hidden under very general theorems on symmetric spaces in many differential geometry textbooks [Gamkrelidze, 1991, Helgason, 1978, Kobayashi and Nomizu, 1969] and sometimes considered as a well known result as in [Bhatia, 2003]. In statistics, it has been introduced as the Fisher information metric [Skovgaard, 1984] to model geometry of the multivariate normal family. The idea of the invariant metric came to the mind of the first author during the IPMI conference in 2001 [Coulon et al., 2001, Batchelor et al., 2001], as an application to diffusion tensor imaging (DTI) of the statistical methodology on Riemannian manifolds previously developed (and summarized in the next Section). However, this idea was not exploited until the end of 2003, when the visit of P. Thompson (UCLA, USA) raised the need to interpolate tensors that represent the variability from specific locations on sulci to the whole volume[1]. The expertise of the second author on DTI [Fillard et al., 2003] provided an ideal alternative application field. During the writing of this paper, we discovered that the invariant metric has been independently proposed by [Förstner and Moonen, 1999] to deal with covariance matrices, and very recently by [Fletcher and Joshi, 2004] for the analysis of principal modes of sets of diffusion tensors. However, up to our knowledge, it has not been promoted to a complete computing framework, as we propose in this paper.

2 Statistics on geometric features

We summarize in this Section the theory of statistics on Riemannian manifolds developed in [Pennec, 1999, Pennec, 2004]. The aim is to exemplify the fact that choosing a Riemannian metric "automatically" determines a powerful framework to work on the manifold through the use of a few tools from differential geometry.

[1] A research report on the application of the theory developed here to that problem is currently in preparation.

In the geometric framework, one can specify the structure of a manifold $\mathcal{M}$ by a *Riemannian metric*. This is a continuous collection of dot products on the tangent space at each point of the manifold. Thus, if we consider a curve on the manifold, we can compute at each point its instantaneous speed vector and its norm, the instantaneous speed. To compute the length of the curve, we can proceed as usual by integrating this value along the curve. The distance between two points of a connected Riemannian manifold is the minimum length among the curves joining these points. The curves realizing this minimum for any two points of the manifold are called geodesics. The calculus of variations shows that geodesics are the solutions of a system of second order differential equations depending on the Riemannian metric. In the following, we assume that the manifold is *geodesically complete*, i.e. that the definition domain of all geodesics can be extended to $\mathbb{R}$. This means that the manifold has no boundary nor any singular point that we can reach in a finite time. As an important consequence, the Hopf-Rinow-De Rham theorem states that there always exists at least one minimizing geodesic between any two points of the manifold (i.e. whose length is the distance between the two points).

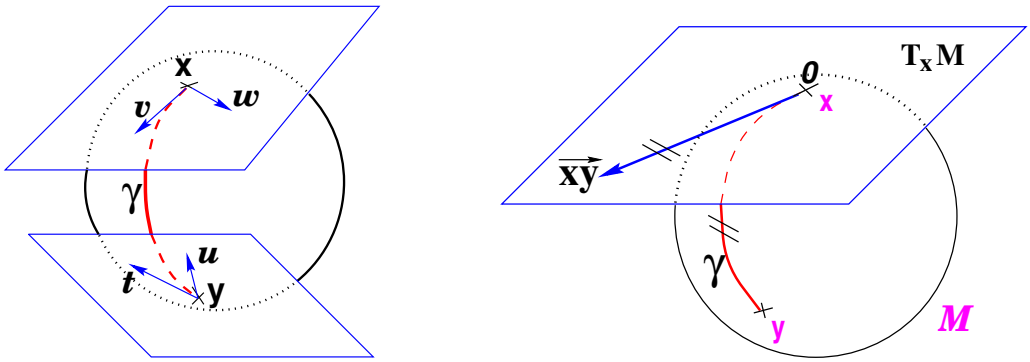

Figure 1: **Left:** The tangent planes at points x and y of the sphere $\mathcal{S}_2$ are different: the vectors v and w of $T_{\mathrm{x}}\mathcal{M}$ cannot be compared to the vectors t and u of $T_{\mathrm{y}}\mathcal{M}$. Thus, it is natural to define the dot product on each tangent plane. **Right:** The geodesics starting at x are straight lines in the exponential map and the distance along them is conserved.

2.1 Exponential chart

Let x be a point of the manifold that we consider as a local reference and $\overrightarrow{\mathrm{xy}}$ a vector of the tangent space $T_{\mathrm{x}}\mathcal{M}$ at that point. From the theory of second order differential equations, we know that there exits one and only one geodesic starting from that point with this tangent vector. This allows to develop the manifold in the tangent space along the geodesics (think of rolling a sphere along its tangent plane at a given point). The geodesics going through the reference point are transformed into straight lines and the distance along these geodesics is conserved (at least in a neighborhood of x).

The function that maps to each vector $\overrightarrow{\mathrm{xy}} \in T_{\mathrm{x}}\mathcal{M}$ the point y of the manifold that is reached after a unit time by the geodesic starting at x with this tangent vector is called the *exponential map*. This map is defined in the whole tangent space $T_{\mathrm{x}}\mathcal{M}$ (since the manifold is geodesically complete) but it is generally one-to-one only locally around 0 in the tangent space (i.e. around x in the manifold). In the sequel, we denote by $\overrightarrow{\mathrm{xy}} = \log_{\mathrm{x}}(\mathrm{y})$ the inverse of the exponential map: this is the smallest vector such that $\mathrm{y} = \exp_{\mathrm{x}}(\overrightarrow{\mathrm{xy}})$. If we look for the maximal definition domain, we find out that it is a star-shaped domain delimited by a continuous curve C_{x} called the *tangential cut-locus*. The image of C_{x} by the exponential map is the cut locus $\mathcal{C}_{\mathrm{x}}$ of point x. This is the closure of the set of points where several minimizing geodesics starting from x meet.

On the sphere $\mathcal{S}_2(1)$ for instance, the cut locus of a point x is its antipodal point and the tangential cut locus is the circle of radius π.

The exponential map within this domain realizes a chart called *the exponential chart*. It covers all the manifold except the cut locus of the development point, which has a null measure. In this chart, geodesics starting from x are straight lines. and the distance from the development point are conserved. This chart is somehow the "most linear" chart of the manifold with respect to the primitive x.

2.2 Practical implementation

In fact most of the usual operations using additions and subtractions may be reinterpreted in a Riemannian framework using the notion of *bipoint*, an antecedent of vector introduced during the 19th Century. Indeed, one defines vectors as equivalent classes of bipoint (oriented couples of points) in a Euclidean space. This is possible because we have a canonical way (the translation) to compare what happens at two different points. In a Riemannian manifold, we can still compare things locally (by parallel transportation), but not any more globally. This means that each "vector" has to remember at which point of the manifold it is attached, which comes back to a bipoint.

However, one can also see a vector $\overrightarrow{\mathrm{xy}}$ (attached at point x) as a vector of the tangent space at that point. Such a vector may be identified to a point on the manifold using the geodesic starting at x with tangent vector $\overrightarrow{\mathrm{xy}}$, i.e. using the exponential map: $\mathrm{y} = \exp_{\mathrm{x}}(\overrightarrow{\mathrm{xy}})$. Conversely, the logarithmic map may be used to map almost any bipoint (x, y) into a vector $\overrightarrow{\mathrm{xy}} = \log_{\mathrm{x}}(\mathrm{y})$ of $T_{\mathrm{x}}\mathcal{M}$.

Vector space	Riemannian manifold
$\overrightarrow{xy} = y - x$	$\overrightarrow{\mathrm{xy}} = \log_{\mathrm{x}}(\mathrm{y})$
$y = x + \overrightarrow{xy}$	$\mathrm{y} = \exp_{\mathrm{x}}(\overrightarrow{\mathrm{xy}})$
$\mathrm{dist}(x, y) = \|y - x\|$	$\mathrm{dist}(\mathrm{x}, \mathrm{y}) = \|\overrightarrow{\mathrm{xy}}\|_{\mathrm{x}}$

Table 1: Re-interpretation of addition and subtraction in a Riemannian manifold.

This reinterpretation of addition and subtraction using logarithmic and exponential maps is very powerful to generalize algorithms working on vector space to algorithms on Riemannian manifolds. It is also very powerful in terms of implementation since we can practically express all the geometric operations in these terms: the implementation of $\log_{\mathrm{x}}$ and $\exp_{\mathrm{x}}$ is the basis of any programming on Riemannian manifolds, as we will see in the following.

2.3 Basic statistical tools

The Riemannian metric induces an infinitesimal volume element on each tangent space, and thus a measure $d\mathcal{M}$ on the manifold that can be used to measure random events on the manifold and to define the probability density function (if it exists) of these random primitives. It is worth noticing that the induced measure $d\mathcal{M}$ represents the notion of *uniformity* according to the chosen Riemannian metric. This automatic derivation of the uniform measure from the metric gives a rather elegant solution to the Bertrand paradox for geometric probabilities [Poincaré, 1912, Kendall and Moran, 1963]. However, the problem is only shifted: which Riemannian metric do we have to choose ? We address this question in Section 3 for real positive definite symmetric matrices (tensors): it turns out that we will be able to require an invariance by the full linear group, which will lead to a very regular and convenient manifold structure.

Let us come back to the basic statistical tools we would like to develop. With the probability measure of a random primitive, we can integrate functions from the manifold to any vector space, thus defining the

expected value of this function. However, we generally cannot integrate manifold-valued functions. Thus, one cannot define the mean or expected "value" of a random primitive using a weighted sum or an integral as usual: we need to rely on distance-based variational formulation. The Fréchet or Karcher expected features basically minimize globally (or locally) the variance. As the mean is now defined through a minimization procedure, its existence and uniqueness are not ensured any more (except for distributions with a sufficiently small compact support). In practice, one mean value almost always exists, and it is unique as soon as the distribution is sufficiently peaked. The properties of the mean are very similar to those of the modes (that can be defined as central Karcher values of order 0) in the vectorial case.

To compute the mean value, we designed in [Pennec, 1999, Pennec, 2004] an original Gauss-Newton gradient descent algorithm that essentially alternates the computation of the barycenter in the exponential chart centered at the current estimation of the mean value, and a re-centering step of the chart at the point of the manifold that corresponds to the computed barycenter (geodesic marching step). To define higher moments of the distribution, we used the exponential chart at the mean point: the random feature is thus represented as a random vector with null mean in a star-shaped and symmetric domain. With this representation, there is no difficulty to define the covariance matrix and potentially higher order moments. Based on this covariance matrix, we may define a Mahalanobis distance between a random and a deterministic feature that basically weights the distance between the deterministic feature and the mean feature using the inverse of the covariance matrix. Interestingly, the expected Mahalanobis distance of a random primitive with itself is independent of the distribution and is equal to the dimension of the manifold, as in the vectorial case.

As for the mean, we chose in [Pennec, 1996, Pennec, 1999, Pennec, 2004] a variational approach to generalize the Normal Law: we define it as the distribution that minimizes the information knowing the mean and the covariance. Neglecting the cut-locus constraints, we show that this amounts to consider a Gaussian distribution on the exponential chart centered at the mean point that is truncated at the cut locus (if there is one). However, the relation between the concentration matrix (the "metric" used in the exponential of the probability density function) and the covariance matrix is slightly more complex that the simple inversion of the vectorial case, as it has to be corrected for the curvature of the manifold. Last but not least, using the Mahalanobis distance of a normally distributed random feature, we can generalize the χ^2 law: we were able to show that is has the same density as in the vectorial case up to an order 3 in σ. This opens the way to the generalization of many other statistical tests, as we may expect similarly simple approximations for sufficiently centered distributions.

3 Working on the Tensor space

Let us now focus on the space $\mathcal{S}ym_n^+$ of positive definite symmetric matrices (tensors). The goal is to find a Riemannian metric with interesting enough properties. It turns out that it is possible to require an invariance by the full linear group (Section 3.3). This leads to a very regular manifold structure where tensors with null and infinite eigenvalues are both at an infinite distance of any positive definite symmetric matrix: the cone of positive definite symmetric matrices is replaced by a space which has an infinite development in each of its $n(n+1)/2$ directions. Moreover, there is one and only one geodesic joining any two tensors, and we can even define globally consistent orthonormal coordinate systems of tangent spaces. Thus, the structure we obtain is very close to a vector space, except that the space is curved.

3.1 Exponential, logarithm and square root of tensors

In the following, we will make an extensive use of a few functions on symmetric matrices. The exponential of any matrix can be defined using the series $\exp(A) = \sum_{k=0}^{+\infty} \frac{A^k}{k!}$. In the case of tensors, we have some important simplifications. Let $\Sigma = U\ D\ U^{\mathrm{T}}$ be a diagonalization, where U is an orthogonal matrix, and

$D = \text{DIAG}(d_i)$ is the diagonal matrix of strictly positives eigenvalues. We can write any power of Σ in the same basis: $\Sigma^k = U\, D^k\, U^{\mathrm{T}}$. This means that we may factor out the rotation matrices in the series and map the exponential individually to each eigenvalue:

$$\exp(\Sigma) = \sum_{k=0}^{+\infty} \frac{\Sigma^k}{k!} = U\, \text{DIAG}(\exp(d_i))\, U^{\mathrm{T}}$$

The series defining the exponential function converges for any matrix argument, but this is generally not the case for the series defining its inverse function: the logarithm. However, in our case, the rotations in the series can be factored out just as above, and we end up with the diagonal matrix of the logarithm of the eigenvalues, which are always well defined since the matrix is positive definite:

$$\log(\Sigma) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k} (\Sigma - \text{ Id})^k = U \left(\text{DIAG} \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{k} (d_i - 1)^k \right) \right) U^{\mathrm{T}} = U\, (\text{DIAG}(\log(d_i)))\, U^{\mathrm{T}}$$

Classically, one defines the (left) square root of a matrix B as the set $\{B_L^{1/2}\} = \{A \in GL_n / A\, A^{\mathrm{T}} = B\}$. One could also define the right square root: $\{B_R^{1/2}\} = \{A \in GL_n \,/\, A^{\mathrm{T}}\, A = B\}$. For tensors, we define the square root as:

$$\Sigma^{1/2} = \{\Lambda \in \mathcal{S}ym_n^+ \,/\, \Lambda^2 = \Sigma\}$$

The square root is always defined and moreover unique: let $\Sigma = U\, D^2\, U^{\mathrm{T}}$ be a diagonalization (with positives values for the d_i's). Then $\Lambda = U\, D\, U^{\mathrm{T}}$ is of course a square root of Σ, which proves the existence. For the uniqueness, let us consider two symmetric and positive square roots Λ_1 and Λ_2 of Σ. Then, $\Lambda_1^2 = \Sigma$ and $\Lambda_2^2 = \Sigma$ obviously commute 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. As the elements of D_1 and D_2 are positive, they are also equal and $\Lambda_1 = \Lambda_2$. Last but not least, we have the property that

$$\Sigma^{1/2} = \exp\left(\frac{1}{2}(\log \Sigma)\right)$$

3.2 An affine invariant distance

Let us consider the following action of the linear group GL_n on the tensor space $\mathcal{S}ym_n^+$:

$$A \star \Sigma = A \Sigma A^{\mathrm{T}} \qquad \forall A \in GL_n \qquad \text{and} \qquad \Sigma \in \mathcal{S}ym_n^+$$

This group action corresponds for instance to the standard action of the affine group on the covariance matrix of a random variables $\boldsymbol{x}$ in $\mathbb{R}^n$: if $\boldsymbol{y} = A\boldsymbol{x} + t$, then $\Sigma_{\boldsymbol{yy}} = \boldsymbol{E}[\boldsymbol{y}\, \boldsymbol{y}^{\mathrm{T}}] = A \Sigma_{\boldsymbol{xx}} A^{\mathrm{T}}$.

This action is naturally extended to tangent vectors is the same way: if $\Gamma(t) = \Sigma + t\, W + O(t^2)$ is a curve passing at Σ with tangent vector W, then the curve $A \star \Gamma(t) = A\, \Sigma\, A^{\mathrm{T}} + t\, A\, W A^{\mathrm{T}} + O(t^2)$ passes through $A \star \Sigma$ with tangent vector $A \star W$.

Following [Pennec and Ayache, 1998], any invariant distance on $\mathcal{S}ym_n^+$ verifies $\text{dist}(A \star \Sigma_1, A \star \Sigma_2) = \text{dist}(\Sigma_1, \Sigma_2)$. Choosing $A = \Sigma_1^{-1/2}$, we can reduce this to a pseudo-norm, or distance to the identity:

$$\text{dist}(\Sigma_1, \Sigma_2) = \text{ dist}\left(\text{Id}, \Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} \right) = N\left(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} \right)$$

Moreover, as the invariance has to hold for any transformation, N should be invariant under the action of the isotropy group $\mathcal{H}(\text{ Id}) = \mathcal{O}_n = \{U \in GL_n \,/\, UU^{\mathrm{T}} = \text{ Id}\}$:

$$\forall U \in \mathcal{O}_n, \qquad N(U\, \Sigma\, U^{\mathrm{T}}) = N(\Sigma)$$

Using the spectral decomposition $\Sigma = U D^2 U^{\mathrm{T}}$, it is easy to see that $N(\Sigma)$ has to be a symmetric function of the eigenvalues. Moreover, the symmetry of the distance $\mathrm{dist}(\Sigma,\ \mathrm{Id}) = \mathrm{dist}(\ \mathrm{Id}, \Sigma)$ imposes that $N(\Sigma) = N(\Sigma^{(-1)})$. Thus, a good candidate is the sum of the squared logarithms of the eigenvalues:

$$N(\Sigma)^2 = \| \log(\Sigma) \|^2 = \sum_{i=1}^{n} (\log(\sigma_i))^2 \tag{1}$$

This "norm" verifies by construction the symmetry and positiveness. $N(\Sigma) = 0$ implies that $\sigma_i = 1$ (and conversely), so that the separation axiom is verified. What is much more difficult to show is the triangle inequality, which should read $N(\Sigma_1) + N(\Sigma_2) \geq N(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})$. If we can verify it experimentally, its direct theoretical proof is, up to our knowledge, unknown (see e.g. [Förstner and Moonen, 1999]).

3.3 An invariant Riemannian metric

Another way to determine the invariant distance is through the Riemannian metric. Let us take the most simple dot product on the tangent space at the identity matrix: if W_1 and W_2 are tangent vectors (i.e. symmetric matrices, not necessarily definite nor positive), we define the dot product to be the standard matrix dot product $\langle\ W_1\ |\ W_2\ \rangle = \mathrm{Tr}(W_1^{\mathrm{T}} W_2)$. This dot product if obviously invariant by the isotropy group $\mathcal{O}_n$. Now, if W_1 and W_2 are two tangent vector at Σ, we require their dot product to be invariant by the action of any transformation: $\langle\ W_1\ |\ W_2 \rangle_{\Sigma} = \langle\ A \star W_1\ |\ A \star W_2 \rangle_{A \star \Sigma}$. This should be true in particular for $A = \Sigma^{-1/2}$, which allows us to define the dot product at any Σ from the dot product at the identity:

$$\langle\ W_1\ |\ W_2 \rangle_{\Sigma} = \left\langle\ \Sigma^{-\frac{1}{2}} W_1 \Sigma^{-\frac{1}{2}}\ \middle|\ \Sigma^{-\frac{1}{2}} W_2 \Sigma^{-\frac{1}{2}}\ \right\rangle_{\mathrm{Id}} = \mathrm{Tr}\left(\Sigma^{-\frac{1}{2}} W_1 \Sigma^{-1} W_2 \Sigma^{-\frac{1}{2}}\right)$$

One can easily verify that this definition is left unchanged if we use any other transformation $A = U\ \Sigma^{-1/2}$ (where U is a free orthogonal matrix) that transports Σ to the identity: $A \star \Sigma = A\ \Sigma\ A^{\mathrm{T}} = U\ U^{\mathrm{T}} = \mathrm{Id}$.

To find the geodesic without going though the computation of Christoffel symbols, we may rely on a result from differential geometry [Gamkrelidze, 1991, Helgason, 1978, Kobayashi and Nomizu, 1969] which says that the geodesics for the invariant metrics on affine symmetric spaces are generated by the action of the one-parameter subgroups of the acting Lie group[2]. Since the one-parameter subgroups of the linear group are given by the matrix exponential $\exp(t\ A)$, geodesics on our tensor manifold going through Σ with tangent vector W should have the following form:

$$\Gamma_{(\Sigma,W)}(t) = \exp(t\ A)\ \Sigma\ \exp(t\ A)^{\mathrm{T}} \qquad \text{with} \qquad W = A\ \Sigma + \Sigma\ A^{\mathrm{T}} \tag{2}$$

This expression was used directly in [Chefd'hotel et al., 2004]. For our purpose, we need to relate explicitly the geodesic to the tangent vector in order to define the exponential chart. Since Σ is a symmetric matrix, there is hopefully an explicitly solution to the Sylvester equation $W = A\ \Sigma + \Sigma\ A^{\mathrm{T}}$. We get $A = \frac{1}{2}\left(W\ \Sigma^{(-1)} + \Sigma^{1/2}\ Z\ \Sigma^{-1/2}\right)$, where Z is a free skew-symmetric matrix. However, introducing this solution into the equation of geodesics (Eq. 2) does not lead to a very tractable expression. Let us look at an alternative solution.

Since our metric (and thus the geodesics) is invariant under the action of the group, we can focus on the geodesics going through the origin (the identity). In that case, a symmetric solution of the Sylvester equation is $A = \frac{1}{2} W$, which gives the following equation for the geodesic going through the identity with tangent vector W:

$$\Gamma_{(\ \mathrm{Id},W)}(t) = \exp\left(\frac{t}{2}\ W\right) \exp\left(\frac{t}{2}\ W\right)^{\mathrm{T}} = \exp(t\ W).$$

[2] To be mathematically correct, we should consider the quotient space $\mathcal{S}ym_n^+ = GL_n^+ / SO_n$ instead of $\mathcal{S}ym_n^+ = GL_n / O_n$ so that all spaces are simply connected.

We may observe that the tangent vector along this curve is the parallel transportation of the initial tangent vector. If $W = U\ \text{DIAG}(w_i)\ U^{\text{T}}$,

$$\frac{d\Gamma(t)}{dt} = U\ \text{DIAG}\left(w_i \exp(t\, w_i)\right)\ U^{\text{T}} = \Gamma(t)^{\frac{1}{2}}\ W\ \Gamma(t)^{\frac{1}{2}} = \Gamma(t)^{\frac{1}{2}} \star W$$

By definition of our invariant metric, the norm of this vector is constant: $\|\Gamma(t)^{1/2} \star W\|^2_{\Gamma(t)^{1/2} \star \text{Id}} = \|W\|^2_{\text{Id}} = \|W\|^2_2$. This was expected since geodesics are parameterized by arc-length. Thus, the length of the curve between time 0 and 1 is

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

Solving for $\Gamma_{(\text{Id},W)}(1) = \Sigma$, we obtain the "norm" $N(\Sigma)$ of Eq.(1). Using the invariance of our metric, we easily obtain the geodesic starting from any other point of the manifold using our group action:

$$\Gamma_{(\Sigma,W)}(t) = \Sigma^{\frac{1}{2}} \star \Gamma_{\left(\text{Id},\Sigma^{-1/2} \star W\right)}(t) = \Sigma^{\frac{1}{2}} \exp\left(t\, \Sigma^{-\frac{1}{2}} W \Sigma^{-\frac{1}{2}}\right) \Sigma^{\frac{1}{2}}$$

Coming back to the distance $\text{dist}^2(\Sigma,\ \text{Id}) = \sum_i (\log \sigma_i)^2$, it is worth noticing that tensors with null eigenvalues are located as far from the identity as tensors with infinite eigenvalues: at the infinity. Thanks to the invariance by the Linear group, this property holds for the distance to any (positive definite) tensor of the manifold. Thus, the original cone of positive definite symmetric matrices (a manifold with a flat metric but with boundaries) has been changed into a regular manifold of constant curvature with an infinite development in each of its $n(n+1)/2$ directions.

3.4 Exponential and logarithm maps

As a general property of Riemannian manifolds, geodesics realize a local diffeomorphism from the tangent space at a given point of the manifold to the manifold: $\Gamma_{(\Sigma,W)}(1) = \exp_\Sigma(W)$ associates to each tangent vector $W \in T_\Sigma \mathcal{S}ym_n^+$ a point of the manifold. This mapping is 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:

$$\exp_{\text{Id}}(UDU^{\text{T}}) = \exp(UDU^{\text{T}}) = U\ \text{DIAG}\left(\exp(d_i)\right)\ U^{\text{T}}$$

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

$$\exp_\Sigma(W) = \Sigma^{\frac{1}{2}} \exp\left(\Sigma^{-\frac{1}{2}} W \Sigma^{-\frac{1}{2}}\right) \Sigma^{\frac{1}{2}}$$

As we have no cut locus, this diffeomorphism is moreover global, and we can uniquely define the inverse mapping everywhere:

$$\log_\Sigma(\Lambda) = \Sigma^{\frac{1}{2}} \log\left(\Sigma^{-\frac{1}{2}} \Lambda \Sigma^{-\frac{1}{2}}\right) \Sigma^{\frac{1}{2}}$$

Thus, $\exp_\Sigma$ gives us a collection of one-to-one and complete maps of the manifold, centered at any point Σ. As explained in Section 2.1, these charts can be viewed as the development of the manifold onto the tangent space along the geodesics. Moreover, since there is no cut-locus, the statistical properties detailed in [Pennec, 2004] hold in their most general form. For instance, we have the existence and uniqueness of the mean of any distribution with a compact support.

3.5 Induced and orthonormal coordinate systems

One has to be careful because the coordinate system of all these charts is not orthonormal. Indeed, the coordinate system of each chart is induced by the standard coordinate system (here the matrix coefficients), so that the vector $\overrightarrow{\Sigma\Lambda}$ corresponds to the standard derivative in the vector space of matrices: we have $\Lambda = \Sigma + \overrightarrow{\Sigma\Lambda} + O(\|\overrightarrow{\Sigma\Lambda}\|^2)$. Even if this basis is orthonormal at some points of the manifold (such as at the identity for our tensors), it has to be corrected for the Riemannian metric at other places due to the manifold curvature.

From the expression of the metric, one can observe that

$$\|\overrightarrow{\Sigma\Lambda}\|_\Sigma^2 = \|\log_\Sigma(\Lambda)\|_\Sigma^2 = \|\Sigma^{-\frac{1}{2}}\log_\Sigma(\Lambda)\Sigma^{-\frac{1}{2}}\|_{\mathrm{Id}}^2 = \|\log(\Sigma^{-\frac{1}{2}} \star \Lambda)\|_2^2,$$

which shows that $\overrightarrow{\Sigma\Lambda}_\perp = \log(\Sigma^{-\frac{1}{2}} \star \Lambda) \in T_\Sigma\mathcal{S}ym_n^+$ is the expression of the vector $\overrightarrow{\Sigma\Lambda}$ in an orthonormal basis. In our case, the transformation $\Sigma^{1/2} \in GL_n$ is moreover uniquely defined (as a positive square root) and is a smooth function of Σ over the complete tensor manifold. Thus, $\overrightarrow{\Sigma\Lambda}_\perp$ realizes an atlas of orthonormal exponential charts which is globally smooth with respect to the development point[3]. This group action approach was chosen in earlier works [Pennec, 1996, Pennec and Thirion, 1997, Pennec and Ayache, 1998] with what we called the placement function.

For some statistical operations, we need to use a minimal representation (e.g. 6 parameters for 3×3 tensors) in a (locally) orthogonal basis. 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 $i\,n + j$st element $\mathrm{Vec}(A)_{i\,n+j}$ of a $n \times n$ dimensional vector $\mathrm{Vec}(A)$. Since we are working with symmetric matrices, we have only $n(n+1)/2$ independent coefficients (say the upper triangular part). Moreover, the off-diagonal coefficients are counted twice in the L_2 norm at the identity: $\|W\|_2^2 = \sum_{i=1}^n w_{i,i}^2 + 2\ \sum_{i<j\leq n} w_{i,j}^2$. The corresponding projection finally gives us:

$$\mathrm{Vec}_{\mathrm{Id}}(W) = \left(w_{1,1}, \sqrt{2}\,w_{1,2}, w_{2,2}, \sqrt{2}\,w_{1,3}, \sqrt{2}\,w_{2,3}, w_{3,3}, \ldots \sqrt{2}\,w_{1,n}, \ldots \sqrt{2}\,w_{(n-1),n}, w_{n,n}\right)^{\mathrm{T}}$$

Now, for a vector $\overrightarrow{\Sigma\Lambda} \in\in T_\Sigma\mathcal{S}ym_n^+$, we define its minimal representation in the orthonormal coordinate system as:

$$\mathrm{Vec}_\Sigma(\overrightarrow{\Sigma\Lambda}) = \mathrm{Vec}_{\mathrm{Id}}(\overrightarrow{\Sigma\Lambda}_\perp) = \mathrm{Vec}_{\mathrm{Id}}\left(\Sigma^{-\frac{1}{2}}\,\overrightarrow{\Sigma\Lambda}\,\Sigma^{-\frac{1}{2}}\right) = \mathrm{Vec}_{\mathrm{Id}}\left(\log(\Sigma^{-\frac{1}{2}} \star \Lambda)\right)$$

The mapping Vec_Σ realizes an explicit isomorphism between $T_\Sigma\mathcal{S}ym_n^+$ and $\mathbb{R}^{n(n+1)/2}$ with the canonical metric. The reverse mappings will be denoted by $\mathrm{Vec}_\Sigma^{(-1)}$.

3.6 Gradient descent and PDE evolution: an intrinsic scheme

Let $f(\Sigma)$ be an objective function to minimize, Σ_t the current estimation of Σ, and $W_t = \partial f/\partial\Sigma = [\partial f/\partial\sigma_{ij}]$ its matrix derivative at that point, which is of course symmetric. The principle of a first order gradient descent is to go toward the steepest descent, in the direction opposite of the gradient for a short time-step ε, and iterate the process. However, the standard operator $\Sigma_{t+1} = \Sigma_t - \varepsilon W_t$ is only valid for very short time-steps, and we could easily go out of the manifold of positive definite tensors. A much more interesting numerical operator is given by following the geodesic backward starting at Σ with tangent vector W_t during a time ε. This intrinsic gradient descent ensures that we cannot leave the manifold and can easily be expressed using the exponential map:

$$\Sigma_{t+1} = \Gamma_{(\Sigma_t, W_t)}(-\varepsilon) = \exp_{\Sigma_t}(-\varepsilon W_t) = \Sigma^{\frac{1}{2}}\exp(-\varepsilon\Sigma^{-\frac{1}{2}}W_t\Sigma^{-\frac{1}{2}})\Sigma^{\frac{1}{2}}$$

[3] On most homogeneous manifolds, this can only be realized locally. For instance, on the sphere, there is a singularity at the antipodal point of the chosen origin for any otherwise smooth placement function.

This intrinsic scheme is trivially generalized to partial differential evolution equations (PDEs) on tensor fields such as $\partial\Sigma(x,t)/\partial t = -W(x,t)$. We obtain $\Sigma(x,t+dt) = \exp_{\Sigma(x,t)}(-dt\, W(x,t))$.

3.7 Example with the mean value

Let $\Sigma_1 \dots \Sigma_N$ be a set of measures of the same Tensor. The Karcher or Fréchet mean is the set of tensors minimizing the sum of squared distance: $C(\Sigma) = \sum_{i=1}^{N} \text{dist}^2(\Sigma, \Sigma_i)$. In the case of tensors, there is not cut locus, so that there is one and only one mean value $\bar{\Sigma}$ [Pennec, 2004]. Moreover, a necessary and sufficient condition for an optimum is a null gradient of the criterion. Differentiating one step further, we obtain a constant Hessian matrix. Thus, the intrinsic second order Newton gradient descent algorithm gives the following mean value at estimation step $t+1$:

$$\bar{\Sigma}_{t+1} = \exp_{\bar{\Sigma}_t}\left(\frac{1}{N}\sum_{i=1}^{N}\log_{\bar{\Sigma}_t}(\Sigma_i)\right) = \bar{\Sigma}_t^{\frac{1}{2}} \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log\left(\bar{\Sigma}_t^{-\frac{1}{2}}\Sigma_i\bar{\Sigma}_t^{-\frac{1}{2}}\right)\right)\bar{\Sigma}_t^{\frac{1}{2}} \tag{3}$$

Notice that we cannot easily simplify more this expression as in general the data Σ_i and the mean value $\bar{\Sigma}_t$ cannot be diagonalized in a common basis. However, this gradient descent algorithm usually converges very fast (about 10 iterations, see Fig. 2 below).

3.8 Simple statistical operations on tensors

As described in [Pennec, 2004], we may generalize most of the usual statistical methods by using the exponential chart at the mean point. For instance, the empirical covariance matrix of a set of N tensor Σ_i of mean $\bar{\Sigma}$ will be: $\frac{1}{N-1}\sum_{i=1}^{n} \overrightarrow{\bar{\Sigma}\Sigma_i} \otimes \overrightarrow{\bar{\Sigma}\Sigma_i}$. Using our *Vec* mapping, we may come back to more usual matrix notations and write its expression in a minimal representation with an orthonormal coordinate system:

$$\text{Cov} = \frac{1}{N-1}\sum_{i=1}^{N} \text{Vec}_{\bar{\Sigma}}\left(\overrightarrow{\bar{\Sigma}\Sigma_i}\right)\ \text{Vec}_{\bar{\Sigma}}\left(\overrightarrow{\bar{\Sigma}\Sigma_i}\right)^{\text{T}}$$

One may also define the Mahalanobis distance

$$\mu^2_{(\bar{\Sigma},\text{Cov})}(\Sigma) = \text{Vec}_{\bar{\Sigma}}\left(\overrightarrow{\bar{\Sigma}\Sigma}\right)^{\text{T}}\ \text{Cov}^{(-1)}\ \text{Vec}_{\bar{\Sigma}}\left(\overrightarrow{\bar{\Sigma}\Sigma}\right)$$

Looking for the probability density function that minimizes the information with a constrained mean and covariance, we obtain a generalization of the Gaussian distribution of the form:

$$N_{\bar{\Sigma},\Gamma}(\Sigma) = k\ \exp\left(-\frac{1}{2}\mu^2_{\bar{\Sigma},\Gamma}(\Sigma)\right)$$

The main difference with a Euclidean space is that we have a curvature to take into account: the invariant measure induced on the manifold by our metric is linked to the usual matrix measure by $d\mathcal{M}(\Sigma) = d\Sigma/\det(\Sigma)$. Likewise, the curvature slightly modifies the usual relation between the covariance matrix, the concentration matrix Γ and the normalization parameter k of the Gaussian distribution [Pennec, 2004]. These differences have an impact on the calculations using continuous probability density functions. However, from a practical point of view, we only deal with a discrete sample set of measurements, so that the measure-induced corrections are hidden. For instance, we can generate a random (generalized) Gaussian tensor using the following procedure: we sample $n(n+1)/2$ independent and normalized real Gaussian samples, multiply the corresponding vector by the square root of the desired covariance matrix (expressed

in our Vec coordinate system), and come back the the tensor manifold using the $\text{Vec}^{(-1)}$ mapping. That way, we can easily generate noisy measurements of known tensors (see e.g. Fig. 7).

To verify the implementation of our charts and geodesic marching algorithms, we have generated 10000 random Gaussian tensors around a random tensor $\bar{\Sigma}$ with a variance of $\gamma = 1$. We computed the mean using the algorithm of Eq. 3. The convergence is clearly very fast (Fig. 2, left). Then, we computed statistics on the Mahalanobis distance to the mean (since we used a unit variance for generating our random tensors, this corresponds to the squared distance): the distribution closely follows a χ^2_6 distribution, as expected, with an empirical mean of 6.031 and a variance of 12.38 (expected values are 6 and 12).

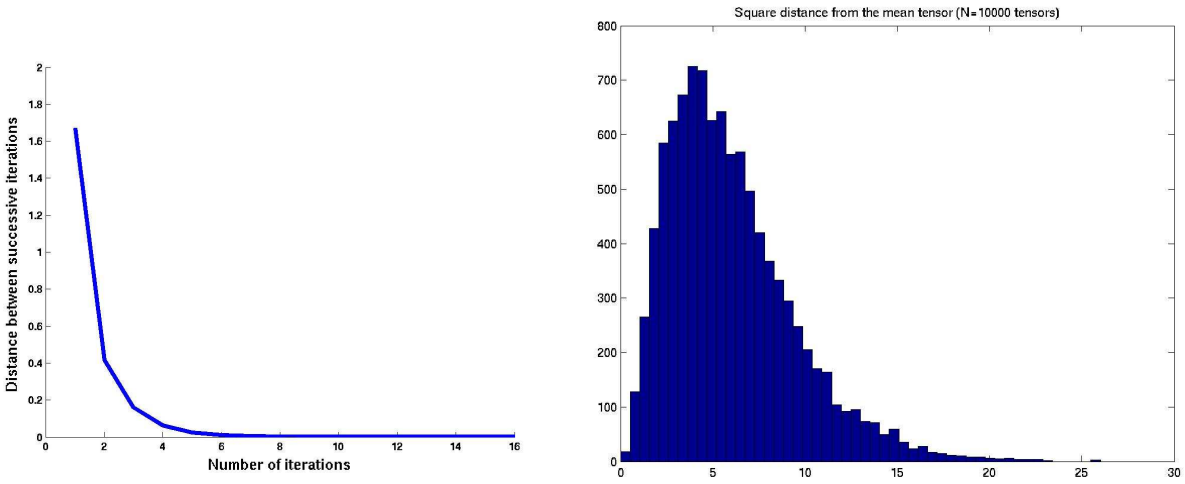

Figure 2: Mean of 10000 random Gaussian tensors. **Left:** evolution of the distance between successive iterations. The convergence is clearly very fast. **Right:** Histogram of the squared-distance to the computed mean. Since we used a unit mean for the variance, this is also an histogram of the Mahalanobis distance.

4 Tensor Interpolation

One of the important operations in geometric data processing is to interpolate values between known measurements. In 3D image processing, (tri-) linear interpolation is often used thanks to its very low computational load and comparatively much better results than nearest neighbor interpolation. Other popular methods include the cubic and, more generally, spline interpolations [Thévenaz et al., 2000, Meijering, 2002].

The standard way to define an interpolation on a regular lattice of dimension d is to consider that the interpolated function $f(x)$ is a linear combination of samples f_k at integer (lattice) coordinates $k \in \mathbb{Z}^d$: $f(x) = \sum_k w(x-k)\, f_k$. To realize an interpolation, the "sample weight" function w has to vanish at all integer coordinates except 0 where it has to be one. A typical example where the convolution kernel has an infinite support is the sinus cardinal interpolation. With the nearest-neighbor, linear (or tri-linear in 3D), and higher order spline interpolations, the kernel is piecewise polynomial, and limited to a few neighboring points in the lattice.

When it comes to an irregular sampling (i.e. a set of measurements f_k at positions x_k), interpolation may still be defined using a weighted mean: $f(x) = \sum_{k=1}^{N} w_k(x)\, f_k$. To ensure that this is an interpolating function, one has to require that $w_i(x_j) = \delta_{ij}$ (where δ_{ij} is the Kronecker symbol). Moreover, the coordinates are usually normalized so that $\sum_{i=1^N} w_k(x) = 1$ for all position x within the domain of interest. Typical examples in triangulations or tetrahedrizations are barycentric and natural neighbor coordinates [Sibson, 1981] (see Section 4.4 below).

4.1 Interpolation through Weighted mean

To generalize interpolation methods defined using weighted means to our tensor manifolds, let us assume that the sample weights $w_k(x)$ are defined as above in $\mathbb{R}^d$. Thanks to their normalization, the value $f(x)$ interpolated from vectors f_k verifies $\sum_{i=1}^N w_i(x)\ (f_i - f(x)) = 0$. Thus, similarly to the Fréchet mean, we can define the interpolated value $\Sigma(x)$ on our tensor manifold as the tensor that minimizes the weighted sum of squared distances to the measurements Σ_i: $C(\Sigma(x)) = \sum_{i=1}^N w_i(x)\ \text{dist}^2(\Sigma_i, \Sigma(x))$. Of course, we loose in general the existence and uniqueness properties. However, for positive weights, the existence and uniqueness theorems for the Karcher mean can be adapted. In practice, this means that we have a unique tensor that verifies $\sum_{i=1}^N w_i(x)\overrightarrow{\Sigma(x)\Sigma_i} = 0$. To reach this solution, it is easy to adapt the Gauss-Newton scheme proposed for the Karcher mean. The algorithm becomes:

$$\Sigma_{t+1}(x) = \exp_{\Sigma_t(x)}\left(\sum_{i=1}^N w_i(x)\ \log_{\Sigma_t(x)}(\Sigma_i)\right) \tag{4}$$

$$= \Sigma_t(x)^{\frac{1}{2}} \exp\left(\sum_{i=1}^N w_i(x)\ \log\left(\Sigma_t(x)^{-\frac{1}{2}}\Sigma_i\Sigma_t(x)^{-\frac{1}{2}}\right)\right)\Sigma_t(x)^{\frac{1}{2}} \tag{5}$$

Once again, this expression cannot be easily simplified, but the convergence is very fast (usually less than 10 iterations as for the mean).

4.2 Example of the linear interpolation

The linear interpolation is somehow simple as this a walk along the geodesic joining the two tensors. We have the closed-form expression: $\Sigma(t) = \exp_{\Sigma_1}(t\ \log_{\Sigma_1}(\Sigma_2)) = \exp_{\Sigma_2}((1-t)\ \log_{\Sigma_2}(\Sigma_1))$ for $t \in [0;1]$. To compare, the equivalent interpolation in the standard matrix space would give $\Sigma'(t) = (1-t)\,\Sigma_1 + t\,\Sigma_2$. We displayed in Fig. 3 the flat and the Riemannian interpolations between 2D tensors of eigenvalues (5,1) horizontally and (1,50) at 45 degrees, along with the evolution of the eigenvalues, their mean (i.e. trace of the matrix) and product (i.e. determinant of the matrix or volume of the ellipsoid).

With the standard matrix coefficient interpolation, the evolution of the trace is perfectly linear (which was expected since this is a linear function of the coefficients), and the principal eigenvalue regularly grows almost linearly, while the smallest eigenvalue slightly grows toward a local maxima before lowering. What is much more annoying is that the determinant (i.e. the volume) does not grow regularly in between the two tensors, but goes through a maximum. If we interpret our tensors as covariance matrices of Gaussian distributions, this means that the probability of a random point to be accepted as a realization of our distribution is larger in between than at the measurement points themselves! On the contrary, one can clearly see a regular evolution of the eigenvalues and of their product with the interpolation in our Riemannian space. Moreover, there is a much smoother rotation of the eigenvectors than with the standard interpolation.

4.3 Tri-linear interpolation

The bi- and tri-linear interpolation of tensors on a regular grid in 2D or 3D are almost as simple, except that we do not have any longer an explicit solution using geodesics since there are more than two reference points. After computing the (bi-) tri-linear weights with respect to the neighboring sites of the point we want to evaluate, we now have to go through the iterative optimization of the weighted mean (Eq. 4) to compute the interpolated tensor. We display an example in Figure 4. One can see that the volume of the tensors is much more important with the classical than with the Riemannian interpolation. We also get a much smoother interpolation of the principal directions with our method.

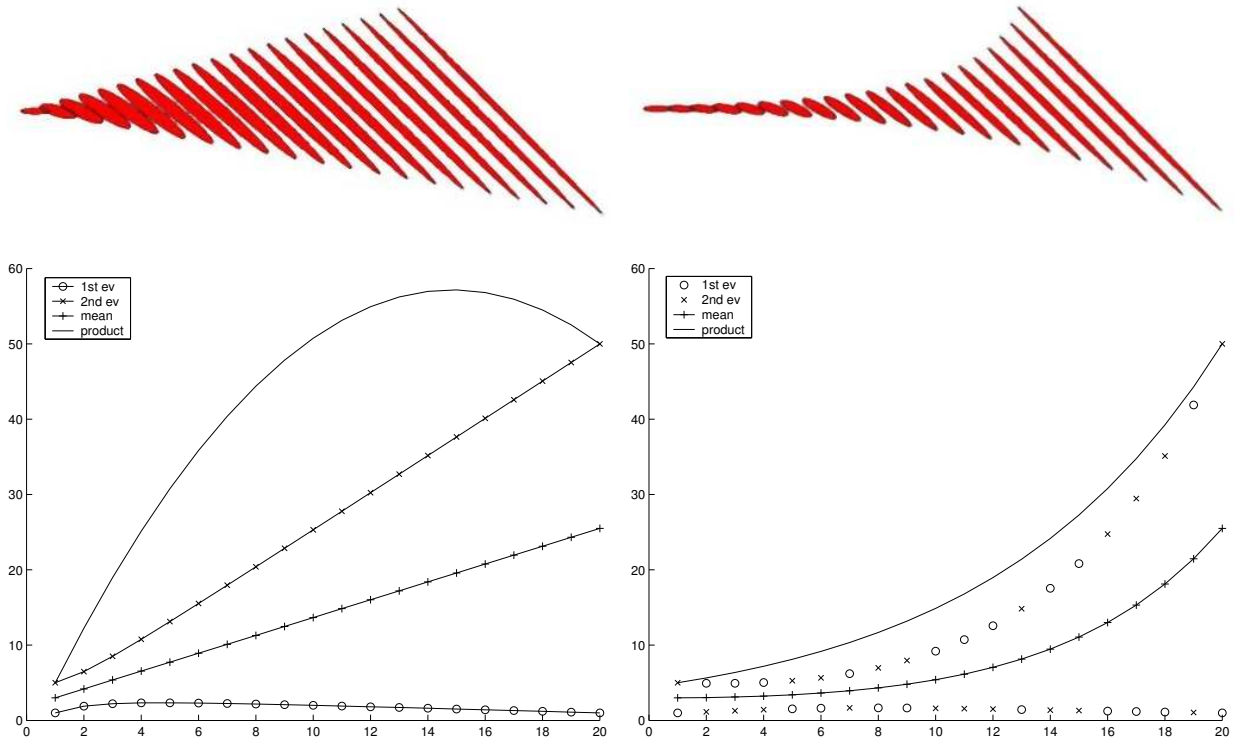

Figure 3: **Top:** Linear interpolation between 2D tensors of eigenvalues (5,1) horizontally and (1,50) at 45 degrees. **Left:** interpolation in the standard matrix space (interpolation of the coefficients), and **right:** in our Riemannian space. **Bottom:** evolution of the eigenvalues, their mean (i.e. trace of the matrix) and product (i.e. determinant of the matrix or volume of the ellipsoid).

4.4 Interpolation of non regular measurements

When tensors are not measured on a regular grid but "randomly" localized in space, defining neighbors becomes an issue. One solution, proposed by [Sibson, 1981] and later used for surfaces by [Cazals and Boissonnat, 2001], is the natural neighbor interpolation. For any point x, its natural neighbors are the the points of $\{x_i\}$ whose Voronoi cells are chopped off upon insertion of x into the Voronoi diagram. The weight w_i of each natural neighbor x_i is the proportion of the new cell that is taken away by x to x_i in the new Voronoï diagram. One important restriction of these interesting coordinates is that they are limited to the convex hull of the point set (otherwise the volume or surface of the cell is infinite).

Another idea is to rely on radial-basis functions to define the relative influence of each measurement point. For instance, a Gaussian influence would give a weight $w_i(x) = G_\sigma(x - x_i)$ to the measurement Σ_i located at x_i. Since weights need to be renormalized in our setup, this would lead to the following evolution equation:

$$\Sigma_{t+1}(x) = \exp_{\Sigma_t(x)}\left(\frac{\sum_{i=1}^{N} G_\sigma(x - x_i)\,\overrightarrow{\Sigma_t(x)\Sigma_i}}{\sum_{i=1}^{N} G_\sigma(x - x_i)}\right) \tag{6}$$

The initialization could be the (normalized) Gaussian mean in the matrix space. An example of the result of this evolution scheme is provided on top of Figure 10. However, this algorithm does not lead to an interpolation, but rather to an approximation, since the weights are not zero at other measurement points. Moreover, we have little control on the quality of this approximation. It is only at the limit where σ goes to zero that we end-up with a (non-continuous) closest point interpolation.

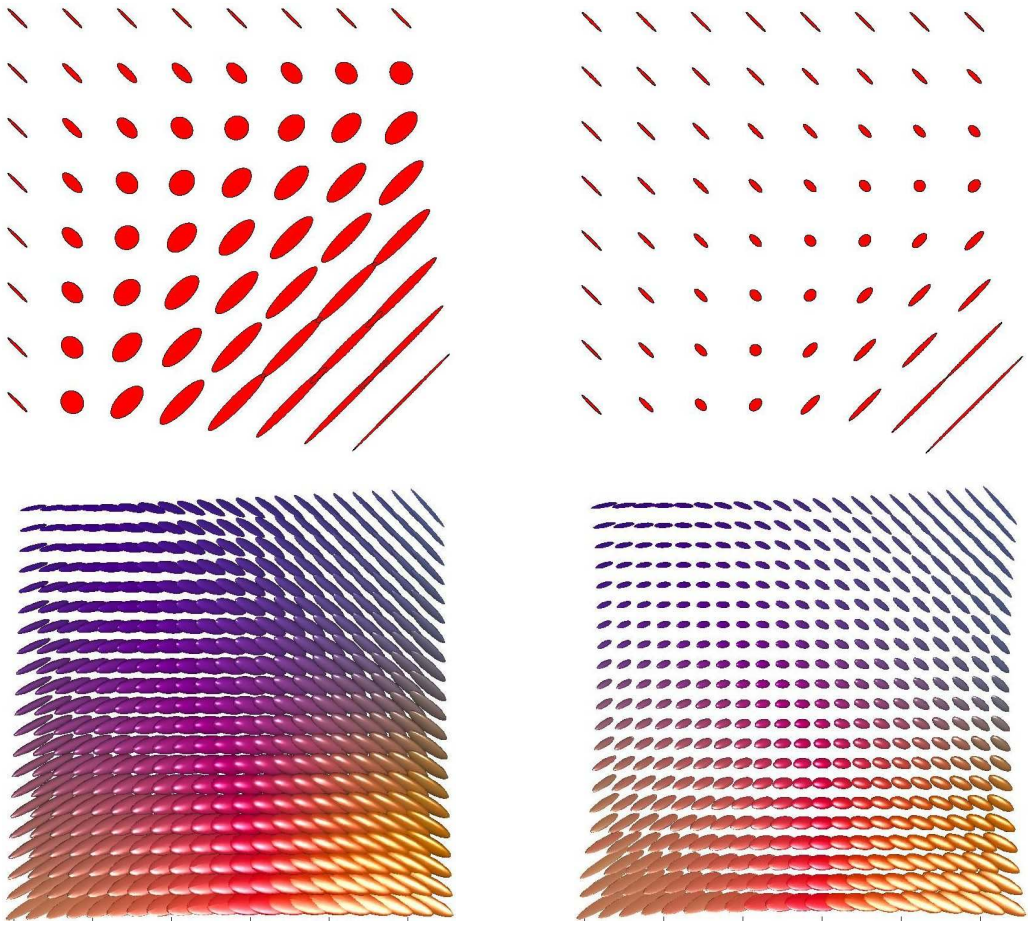

Figure 4: **Top:** Bi-linear interpolation between the 4 2D tensors at the corners. **Bottom:** A slice of the tri-linear interpolation between 3D tensors. **Left:** interpolation in the standard matrix space (interpolation of the coefficients), and **right:** in our Riemannian space.

We will describe in Section 6.3 a last alternative that performs the interpolation and extrapolation of sparsely distributed tensor measurements using diffusion.

5 Filtering tensor fields

Let us now consider that we have a tensor field, for instance like in Diffusion Tensor Imaging (DTI) [Bihan et al., 2001], where the tensor is a first order approximation of the anisotropic diffusion of the water molecules at each point of the images tissues. In the brain, the diffusion is much favored in the direction of oriented structures (fibers of axons). One of the goal of DTI is to retrieve the main tracts along these fibers. However, the tensor field obtained from the images is noisy and needs to be regularized before being further analyzed. A naive but simple and often efficient regularization on signal or images is the convolution by a Gaussian. The generalization to Tensor fields is quite straightforward using once again weighted means (Section 5.1 below). An alternative is to consider a regularization using possibly anisotropic diffusion. This will be the subject of Section 5.3.

5.1 Gaussian Filtering

In the continuous setting, the convolution of a vector field $F_0(x)$ by a Gaussian is:

$$F(x) = \int_y G_\sigma(y-x)\, F_0(y)\, dy$$

In the discrete setting, coefficients are renormalized since the neighborhood $\mathcal{V}$ is usually limited to points within one to three times the standard deviation:

$$F(x) = \frac{\sum_{u\in\mathcal{V}(x)} G_\sigma(u)\, F_0(x+u)}{\sum_{u\in\mathcal{V}(x)} G_\sigma(u)} = \arg\min_F \sum_{u\in\mathcal{V}(x)} G_\sigma(u)\, \|F_0(x+u) - F\|^2$$

Like previously, this weighted mean can be solved on our manifold using our intrinsic gradient descent scheme. Starting from the measured tensor field $\Sigma_0(x)$, the evolution equation is

$$\Sigma_{t+1}(x) = \exp_{\Sigma_t(x)}\left(\frac{\sum_{u\in\mathcal{V}} G_\sigma(u)\, \overrightarrow{\Sigma_t(x)\Sigma_t(x+u)}}{\sum_{u\in\mathcal{V}} G_\sigma(u)}\right)$$

We illustrate in Fig. 5 the comparative Gaussian filtering of a slice of a DT MR image using the flat metric on the coefficient (since weights are positive, a weighted sum of positive definite matrices is still positive definite) and our invariant Riemannian metric. Figure 9 displays closeups around the ventricles to compare the different regularization methods (including the anisotropic filters of Section 5.3.4). One can see a more important blurring of the corpus callosum fiber tracts using the flat metric. However, the integration of this filtering scheme into a complete fiber tracking system would be necessary to fully evaluate the pros and cons of each metric.

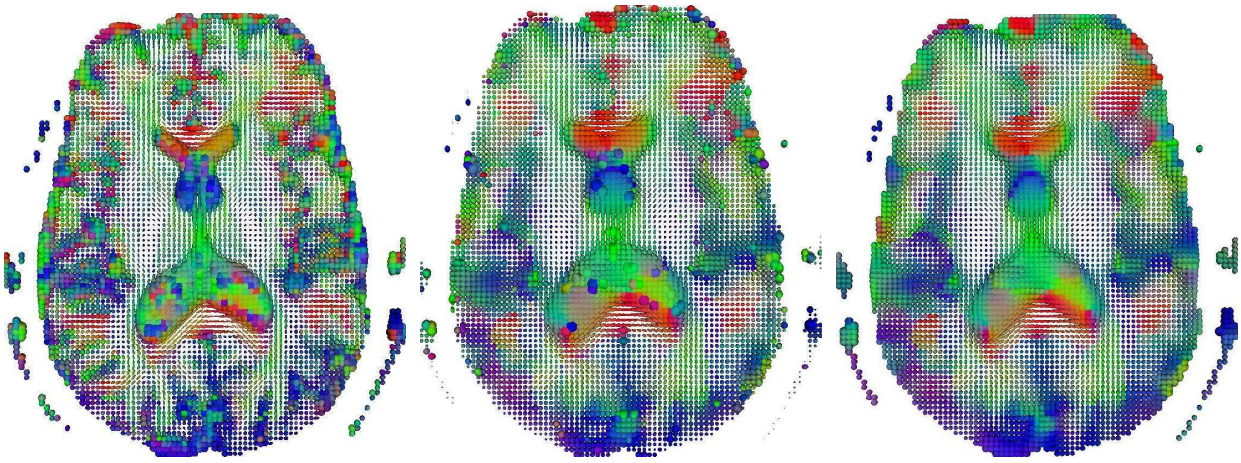

Figure 5: Regularization of a DTI slice around the corpus callosum by isotropic Gaussian filtering. **On the left:** raw estimation of the tensors. The color codes for the direction of the principal eigenvector (red: left/right, green anterior/posterior, blue: top/bottom). **On the middle:** Gaussian filtering of the coefficients (5x5 window, $\sigma = 2.0$). **On the right:** equivalent filtering (same parameters) using the Riemannian metric.

5.2 Spatial gradient of Tensor fields

On a n-dimensional vector field $F(x) = (f_1(x1, \dots x_d), \dots f_n(x1, \dots x_d))^{\mathrm{T}}$ over $\mathbb{R}^d$, one may express the spatial gradient in an orthonormal basis as:

$$\nabla F^{\mathrm{T}} = \left(\frac{\partial F}{\partial x}\right) = [\partial_{x_1} F, \dots \partial_{x_d} F] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}, & \dots & \frac{\partial f_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}, & \dots & \frac{\partial f_n}{\partial x_d} \end{bmatrix}$$

Notice the linearity of the derivatives implies that we could use directional derivatives in more than the d orthogonal directions. This is especially well adapted to stabilize the discrete computations: the finite difference estimation of the directional derivative is $\partial_u F(x) = F(x+u) - F(x)$. By definition, the spatial gradient is related to the directional derivatives through $\nabla F^{\mathrm{T}}\, u = \partial_u F(x)$. Thus, we may compute ∇F as the matrix that best approximate (in the least-square sense) the directional derivatives in the neighborhood $\mathcal{V}$ (e.g. 6, 18 or 26 connectivity in 3D):

$$\begin{aligned} \nabla F(x) &= \arg\min_G \sum_{u\in\mathcal{V}} \|G^{\mathrm{T}}\, u - \partial_u F(x)\|^2 = \left(\sum_{u\in\mathcal{V}} u\, u^{\mathrm{T}}\right)^{(-1)} \left(\sum_{u\in\mathcal{V}} u\, \partial_u F(x)^{\mathrm{T}}\right) \\ &\simeq \left(\sum_{u\in\mathcal{V}} u\, u^{\mathrm{T}}\right)^{(-1)} \left(\sum_{u\in\mathcal{V}} u\, (F(x+u) - F(x))^{\mathrm{T}}\right) \end{aligned}$$

We experimentally found in other applications (e.g. to compute the Jacobian of a deformation field in non-rigid registration) that this gradient approximation scheme was more stable and much faster than computing all derivatives using convolutions, for instance by the derivative of the Gaussian.

To quantify the local amount of variability independently of the space direction, one usually takes the norm of the gradient: $\|\nabla F(x)\|^2 = \sum_{i=1}^d \|\partial_{x_i} F(x)\|^2$. Once again, this can be approximated using all directional derivatives in the neighborhood

$$\|\nabla F(x)\|^2 \simeq \frac{d}{Card(\mathcal{V})} \sum_{u\in\mathcal{V}} \frac{\|F(x+u) - F(x)\|^2}{\|u\|^2} \tag{7}$$

Notice that this approximation is consistent with the previous one only if the directions u are normalized to unity.

For a manifold valued field $\Sigma(x)$ define on $\mathbb{R}^d$, we can proceed similarly, except that the directional derivatives $\partial_{x_i}\Sigma(x)$ are now tangent vectors of $T_{\Sigma(x)}\mathcal{M}$. They can be approximated just like above using finite "differences" in our exponential chart:

$$\partial_u \Sigma(x) \simeq \log_{\Sigma(x)}(\Sigma(x+u)) = \overrightarrow{\Sigma(\mathrm{x})\,\Sigma(\mathrm{x}+\mathrm{u})} = \Sigma(x)^{\frac{1}{2}}\, \log\left(\Sigma(x)^{-\frac{1}{2}}\, \Sigma(x+u)\, \Sigma(x)^{-\frac{1}{2}}\right)\, \Sigma(x)^{\frac{1}{2}} \tag{8}$$

As observed in Section 3.5, we must be careful that this directional derivative is expressed in the standard matrix coordinate system (coefficients). Thus, the basis is not orthonormal: to quantify the local amount of variation, we have to take the metric at the point $\Sigma(x)$ into account, so that:

$$\|\nabla\Sigma(x)\|^2_{\Sigma(x)} = \sum_{i=1}^d \|\partial_{x_i}\Sigma(x)\|^2_{\Sigma(x)} \simeq \frac{d}{Card(\mathcal{V})} \sum_{u\in\mathcal{V}} \frac{\left\|\log\left(\Sigma(x)^{-\frac{1}{2}}\, \Sigma(x+u)\, \Sigma(x)^{-\frac{1}{2}}\right)\right\|_2^2}{\|u\|^2} \tag{9}$$

Figure 6: Norm of the gradient of the tensor field. **On the left:** computed on the coefficients with Eq. 7 (with the flat metric). **On the middle:** we computed the directional derivatives with the exponential map (Eq. 8), but the norm is taken without correcting for the metric. As this should be very close to the flat gradient norm, we only display the difference image. The main differences are located on very sharp boundaries, where the curvature of our metric has the most important impact. However, the relative differences remains small (less than 10%), which shows the stability of both the gradient and the $\log / \exp$ computation schemes. **On the right:** Riemannian norm of the Riemannian gradient (Eq. 9). One can see much more detailed structures within the brain, which will now be preserved during an anisotropic regularization step.

5.3 Filtering using PDE

Regularizing a scalar, vector or tensor field F aims at reducing the amount of its spatial variations. The first order measure of such variations is the spatial gradient ∇F that we dealt with in the previous section. To obtain a regularity criterion over the domain Ω, we just have to integrate: $Reg(F) = \int_\Omega \|\nabla F(x)\|^2 \, dx$. Starting from an initial field $F_0(x)$, the goal is to find at each step a field $F_t(x)$ that minimizes the regularity criterion by gradient descent in the space of (sufficiently smooth and square integrable) functions.

To compute the first order variation, we write a Taylor expansion for an incremental step in the direction of the field H. Notice that $H(x)$ is a tangent vector at $F(x)$:

$$Reg(F + \varepsilon\, H) = Reg(F) + 2\,\varepsilon \int_\Omega \langle\, \nabla F(x) \mid \nabla H(x) \,\rangle \; dx + O(\varepsilon^2).$$

We get the directional (or Gâteau) derivative: $\partial_H Reg(F) = 2 \int_\Omega \langle\, \nabla F(x) \mid \nabla H(x) \,\rangle \; dx$. To compute the steepest descent, we now have to find the gradient $\nabla Reg(F)$ such that for all variation H, we have $\partial_H Reg(F) = \int_\Omega \langle\, \nabla Reg(F)(x) \mid H(x) \,\rangle_{F(x)} \; dx$. Notice that $\nabla Reg(F)(x)$ and $H(x)$ are elements of the tangent space at $F(x)$, so that the dot product should be taken at $F(x)$ for a Tensor field.

5.3.1 The case of a scalar field

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a scalar field. Our regularization criterion is $Reg(f) = \int_\Omega \|\nabla f(x)\|^2 \; dx$. Let us introduce the contravariant derivative $\text{div}(.) = \langle\, \nabla \mid . \,\rangle$ and the Laplacian operator $\Delta f = \text{div}(\nabla f)$. The divergence is usually written $\nabla^{\mathrm{T}} = (\partial/\partial x_1, \ldots, \partial/\partial x_d)$, so that in an orthonormal coordinate system we have $\Delta f = \langle\, \nabla \mid \nabla f \,\rangle = \sum_{i=1}^d \partial^2_{x_i} f$. Using the standard differentiation rules, we have:

$$\text{div}(h\, \nabla f) = \langle\, \nabla \mid h\, \nabla f \,\rangle = h\, \Delta f + \langle\, \nabla h \mid \nabla f \,\rangle$$

Now, thanks to the Green's formula (see e.g. [Gallot et al., 1993]), we know that the flux going out of the boundaries of a (sufficiently smooth) region Ω is equal to the integral of the divergence inside this region. If we denote by n the normal pointing outward at a boundary point, we have:

$$\int_{\partial\Omega} \langle\, h\, \nabla f \mid n \,\rangle \; dn = \int_\Omega \text{div}(h\, \nabla f) = \int_\Omega h\, \Delta f + \int_\Omega \langle\, \nabla h \mid \nabla f \,\rangle$$

This result can also be interpreted as an integration by part in $\mathbb{R}^d$. Assuming homogeneous Neumann boundary conditions (gradient orthogonal to the normal on $\partial\Omega$: $\langle\, \nabla f \mid n \,\rangle = 0$), the flow across the boundary vanishes, and we are left with:

$$\partial_h Reg(f)(x) = 2 \int_\Omega \langle\, \nabla f(x) \mid \nabla h(x) \,\rangle \, dx = -2 \int_\Omega h(x)\, \Delta f(x)\, dx$$

Since this last formula is no more than the dot product on the space $L_2(\Omega, \mathbb{R})$ of square integrable functions, we end-up with the classical Euler-Lagrange equation: $\nabla Reg(f) = -2\Delta f(x)$. The evolution equation used to filter the data is thus

$$f_{t+1}(x) = f_t(x) - \varepsilon \nabla Reg(f)(x) = f_t(x) + 2\,\varepsilon \Delta f_t(x)$$

5.3.2 The vector case

Let us decompose our vector field $F(x)$ into its n scalar components $f_i(x)$. Likewise, we can decompose the $d \times n$ gradient ∇F into the gradient of the n scalar components $\nabla f_i(x)$ (columns). Thus, choosing an orthonormal coordinate system on the space $\mathbb{R}^n$, our regularization criterion is decoupled into n independent scalar regularization problems:

$$Reg(F)(x) = \sum_{i=1}^{n} \int_\Omega \|\nabla f_i(x)\|^2 \, dx = \sum_{i=1}^{n} Reg(f_i)$$

Thus, each component f_i has to be independently regularized with the Euler-Lagrange equation: $\nabla Reg(f_i) = -2\Delta f_i$. With the convention that the Laplacian is applied component-wise (so that we still have $\Delta F = \text{div}(\nabla F) = \nabla^{\mathrm{T}} \nabla F = (\Delta f_1, \dots \Delta f_n)^{\mathrm{T}}$), we end-up with the vectorial equation:

$$\nabla Reg(F) = -2\Delta F \qquad \text{for} \qquad Reg(F) = \int_\Omega \|\nabla F(x)\| \, dx$$

The associated evolution equation is $F_{t+1}(x) = F_t(x) + 2\,\varepsilon \Delta F_t(x)$.

5.3.3 Tensor fields

For a tensor field $\Sigma(x)$, the procedure appears to be more complex. However, a tangent vector to a tensor field (e.g. $\partial_u \Sigma(x)$) is simply a vector field that maps to each point $x \in \mathbb{R}^d$ a vector of $T_{\Sigma(x)}\mathcal{S}ym_n^+$. Thus, we may simply apply the above framework. Our regularization criterion is:

$$Reg(\Sigma) = \int_\Omega \|\nabla \Sigma(x)\|^2_{\Sigma(x)} \, dx = \sum_{i=1}^{d} \int_\Omega \|\partial_{x_i} \Sigma(x)\|^2_{\Sigma(x)} \, dx = \sum_{i=1}^{d} \int_\Omega \left\| \Sigma^{-\frac{1}{2}}\, \partial_{x_i} \Sigma(x)\, \Sigma^{-\frac{1}{2}} \right\|_2^2 dx \quad (10)$$

and its gradient is simply:

$$\nabla Reg(\Sigma)(x) = -2\Delta\Sigma(x) = -2 \sum_{i=1}^{d} \partial^2_{x_i} \Sigma(x)$$

In the above formula, is should be noticed that the second order spatial derivative $\partial^2_{x_i} \Sigma(x)$ is the derivative of a tangent vector $\partial_{x_i} \Sigma(x)$. Thus, $\nabla Reg(\Sigma)(x)$ is a vector of $T_{\Sigma(x)}\mathcal{S}ym_n^+$ that is expressed in the same

coordinate system as the tangent vector $\partial_{x_i}\Sigma(x)$. Finally, the gradient descent on the regularization criterion with the intrinsic geodesic scheme of Section 3.6 leads to:

$$\Sigma_{t+1}(x) = \exp_{\Sigma_t(x)}\left(-\varepsilon\,\nabla Reg(\Sigma)(x)\right) = \exp_{\Sigma_t(x)}\left(2\,\varepsilon\,\Delta\Sigma(x)\right) \tag{11}$$

For the numerical computation of the Laplacian, let us observe that, from the Taylor expansion of a vector field F at x, we have: $\big(F(x+u)-F(x)\big)+\big(F(x-u)-F(x)\big)\big) = \partial_u^2 F(x) + O(\|u\|^3)$. Thus, we may approximate the second order tensor derivative by

$$\partial_u^2\Sigma(x) \simeq \overrightarrow{\Sigma(\mathrm{x})\Sigma(\mathrm{x}+\mathrm{u})} + \overrightarrow{\Sigma(\mathrm{x})\Sigma(\mathrm{x}-\mathrm{u})}. \tag{12}$$

In this formula, all vectors belong to $T_{\Sigma(x)}\mathcal{S}ym_n^+$ so that we do not have any problem with the coordinate system used. Finally, like for the computation of the gradient, we may improve the computation of the Laplacian by using second order derivatives in all possible directions in the neighborhood $\mathcal{V}$. Assuming a symmetric neighborhood (i.e. both u and $-u$ belong to $\mathcal{V}$), this can be further simplified into:

$$\Delta\Sigma(x) = \frac{d}{Card(\mathcal{V})}\sum_{u\in\mathcal{V}}\frac{\partial_u^2\Sigma(x)}{\|u\|^2} \simeq \frac{2\,d}{Card(\mathcal{V})}\sum_{u\in\mathcal{V}}\frac{\overrightarrow{\Sigma(\mathrm{x})\Sigma(\mathrm{x}+\mathrm{u})}}{\|u\|^2} \tag{13}$$

5.3.4 Anisotropic filtering

In practice, we would like to filter within the homogeneous regions, but not across their boundaries. The basic idea is to penalize the smoothing in the directions where the derivative is important [Perona and Malik, 1990, Gerig et al., 1992]. If $c(.)$ is a weighting function decreasing from $c(0)=1$ to $c(+\infty)=0$, this can be realized directly in the discrete implementation of the Laplacian (Eq. 13): the contribution of $\partial_u^2\Sigma$ is weighted by $c(\|\partial_u\Sigma\|/\|u\|)$. With our finite difference approximations, this leads to the following modified Laplacian:

$$\begin{aligned}
\Delta_{aniso}\Sigma(x) &= \frac{d}{Card(\mathcal{V})}\sum_{u\in\mathcal{V}} c\left(\frac{\partial_u\Sigma(x)}{\|u\|}\right)\frac{\partial_u^2\Sigma(x)}{\|u\|^2} \\
&\simeq \frac{2\,d}{Card(\mathcal{V})}\sum_{u\in\mathcal{V}} c\left(\frac{\left\|\overrightarrow{\Sigma(\mathrm{x})\Sigma(\mathrm{x}+\mathrm{u})}\right\|_{\Sigma(x)}}{\|u\|}\right)\frac{\overrightarrow{\Sigma(\mathrm{x})\Sigma(\mathrm{x}+\mathrm{u})}}{\|u\|^2}
\end{aligned}$$

Figures 7 and 8 present example results of this very simple anisotropic filtering scheme on synthetic and real DTI images. We used the function $c(x) = \exp\left(-x^2/\kappa^2\right)$, where the threshold κ controls the amount of local regularization. For both synthetic and real data, the histogram of the gradient norm is very clearly bimodal so that the threshold κ is easily determined.

In Fig. 7, we generated a tensor field with a discontinuity, and add independent Gaussian noises according to Section 3.8. The anisotropic smoothing perfectly preserves the discontinuity while completely smoothing each region. In this synthetic experiment, we retrieve tensor values that are very close to the initial tensor field. This could be expected since the two regions are perfectly homogeneous. After enough regularization steps, each region is a constant field equal to the mean of the 48 initially noisy tensors of the region: the regularized tensors should be roughly 7 times more accurate than the noisy ones.

In Figure 8, we display the evolution of (a slice of) the tensors, the norm of the gradient and the fraction anisotropy (FA) at different steps of the anisotropic filtering of a 3D DTI. The FA is based on the normalized variance of the eigenvalues. It shows the differences between an isotropic diffusion in the brain (where the diffusion tensor is represented by a sphere, FA=0) and a highly directional diffusion (cigar-shaped ellipsoid,

Figure 7: **Left:** 3D synthetic tensor field with a clear discontinuity. **Middle:** The field has been corrupted by a Gaussian noise (in the Riemannian sense). **Right:** result of the regularization after 30 iterations (time step $\varepsilon = 0.01$).

FA=1). Consequently, the bright regions in the image are the potential areas where nervous fibers are located. On can see that the tensors are regularized in "homogeneous" regions (ventricles, temporal areas), while the main tracts are left unchanged. It is worth noticing that the fractional anisotropy is very well regularized even though this measure has almost nothing in common with our invariant tensor metric.

Figure 9 displays closeups around the ventricles to compare the different regularization methods developed so far. 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.

6 Regularization and restoration of tensor fields

The pure diffusion is efficient to reduce the noise in the data, but it also reduces the amount of information. Moreover, the amount of smoothing is controlled by the time of diffusion (time step ε times the number of iterations), which is not an easy parameter to tune. At an infinite diffusion time, the tensor field will be completely homogeneous (or homogeneous by part for some anisotropic diffusion schemes), with a value corresponding to the mean of the measurements over the region (with Neumann boundary conditions). Thus, the absolute minimum of our regularization criterion alone is of little interest.

To keep close to the measured tensor field $\Sigma_0(x)$ while still regularizing, a more theoretically grounded approach is to consider an optimization problem with a competition between a data attachment term and a possibly non-linear anisotropic regularization term:

$$C(\Sigma) = Sim(\Sigma, \Sigma_0) + \lambda\, Reg(\Sigma)$$

Like before, the intrinsic evolution equation leading to a local minimum is:

$$\Sigma_{t+1}(x) = \exp_{\Sigma_t(x)}\left(-\varepsilon\ \left(\nabla Sim(\Sigma, \Sigma_0) + \lambda\, \nabla Reg(\Sigma)(x)\right)\right)$$

6.1 The regularization term

As we saw in the previous section, the simplest regularization criterion is the norm of the gradient of the field $Reg(F) = \int_\Omega \|\nabla F(x)\|^2\, dx$. To preserve the discontinuities, the gradient of this criterion (the Laplacian) may be tailored to prevent the smoothing across them, as we have done in Section 5.3.4. However, there is no more convergence guaranty, since this anisotropic regularization "force" may not derive from a well-posed criterion (energy). Following the pioneer work of [Perona and Malik, 1990], there has been quite an extensive amount of work to propose well posed PDE for the non-linear, anisotropic and non-stationary

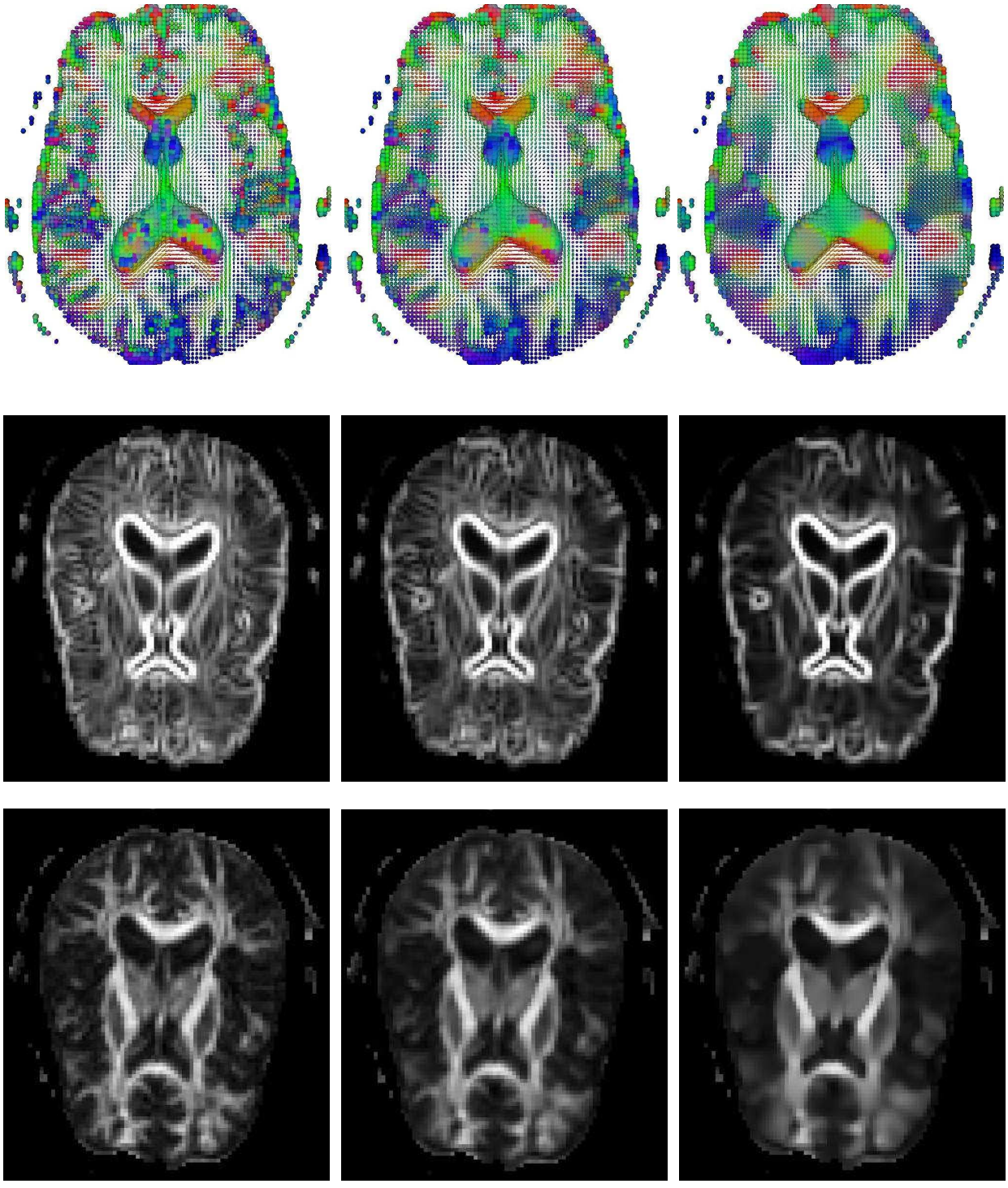

Figure 8: Anisotropic filtering of a DTI slice (time step 0.01, $\kappa = 0.046$). From left to right: at the beginning, after 10 and after 50 iterations. **Top:** A 3D view of the tensors as ellipsoids. The color codes for the direction of the principal eigenvector. The results could be compared with the isotropic Gaussian filtering displayed in Figure 5. **Middle:** norm of the Riemannian gradient. **Bottom:** fractional anisotropy.

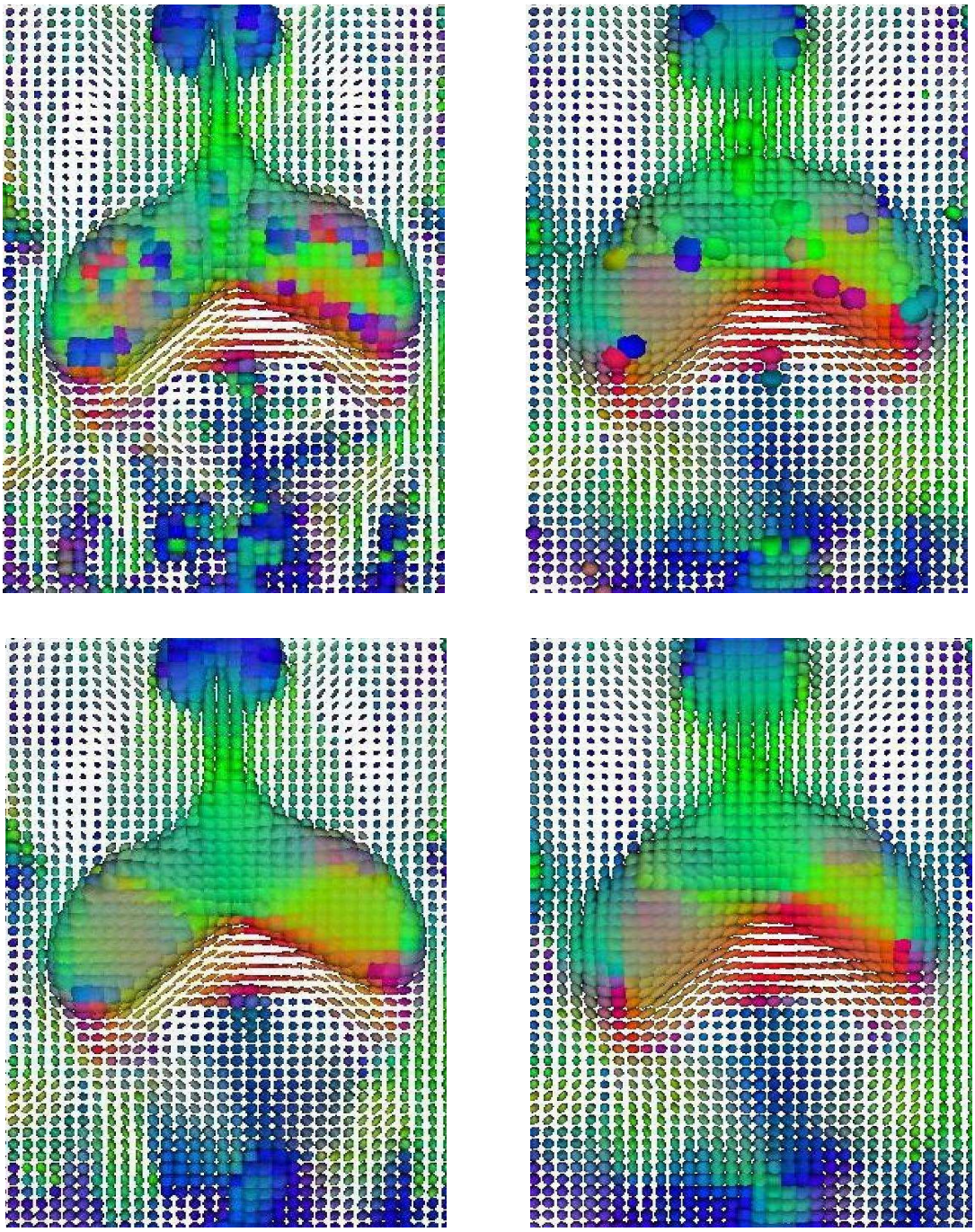

Figure 9: Closeup on the results of the different filtering methods around the splenium of the corpus callosum. The color codes for the direction of the principal eigenvector (red: left-right, green: posterior-anterior, blue: inferior-superior). **Upper left:** Original image. **Upper right:** Gaussian filtering using the flat metric (5x5 window, $\sigma = 2.0$). This metric gives too much weight to tensors with large eigenvalues, thus leading to clear outliers in the ventricles or in the middle of the splenium tract. **Lower right:** 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). **Lower left:** 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.

regularization of vector fields (see e.g. [Weickert, 1998, Sapiro, 2001] to cite only a few recent books). Some of these techniques were recently adapted to work on some manifolds [Tschumperle and Deriche, 2002, Chefd'hotel et al., 2004].

One of the main idea is to replace the usual simple regularization term $Reg(F) = \int_\Omega \|\nabla F(x)\|^2\, dx$ by an increasing function Φ of the norm of the spatial gradient: $Reg(F) = \int_\Omega \Phi(\|\nabla F(x)\|)\, dx$. With some regularity conditions on the Φ-function, one can redo the previous derivations with this Φ-function, and we end-up with [Aubert and Kornprobst, 2001]:

$$\nabla Reg(F)(x) = -2 \text{ div} \left(\frac{\Phi'(\|\nabla F\|)}{\|\nabla F\|} \nabla F \right) = -2 \sum_{i=1}^{d} \partial_{x_i} \left(\frac{\Phi'(\|\nabla F\|)}{\|\nabla F\|} \partial_{x_i} F \right)$$

This continuous scheme can be adapted to the Riemannian framework using the proper gradient norm. However, designing an efficient discrete computation scheme is more difficult. We may compute the directional derivatives using finite differences in the flat matrix space and use the intrinsic evolution scheme, but we believe that there are more efficient ways to do it using the exponential map. We are still investigating that aspect. In the following, we keep the isotropic regularization based on the squared amplitude of the gradient

6.2 A least-squares attachment term

Usually, one consider that the data (e.g. a scalar image or a displacement vector field $F_0(x)$) are corrupted by a uniform (isotropic) Gaussian noise independent at each space position. With a maximum likelihood approach, this amounts to consider a least-squares criterion $Sim(F) = \int_\Omega \|F(x) - F_0(x)\|^2\, dx$. Like in the previous section, we compute the first order variation by writing the Taylor expansion

$$Sim(F + \varepsilon\, H) = Sim(F) + 2\, \varepsilon \int_\Omega \langle\, H(x) \mid F(x) - F_0(x) \,\rangle\; dx + O(\varepsilon^2).$$

This time, the directional derivative $\partial_H Sim(F)$ is directly expressed using a dot product with H in the proper functional space, so that the steepest ascent direction is $\nabla Sim(F) = 2\,(F(x) - F_0(x))$.

On the tensor manifold, assuming a uniform (generalized) Gaussian noise independent at each position also leads to a least-squares criterion thought a maximum likelihood approach. The only difference is that is uses our Riemannian distance:

$$Sim(\Sigma) = \int_\Omega \text{dist}^2\,(\Sigma(x)\,,\, \Sigma_0(x))\; dx = \int_\Omega \left\| \overrightarrow{\Sigma(\mathrm{x})\Sigma_0(\mathrm{x})} \right\|^2_{\Sigma(x)} dx$$

Thanks to the properties of the exponential map, one can show that the gradient of the squared distance is: $\nabla_\Sigma \text{dist}^2(\Sigma\,,\, \Sigma_0) = -2\, \overrightarrow{\Sigma\Sigma_0}$ [Pennec, 2004]. One can verify that this is a tangent vector at Σ whereas $\overrightarrow{\Sigma_0\Sigma}$ is not. Finally, we obtain a steepest ascent direction of our criterion which is very close to the vector case:

$$\nabla Sim(\Sigma)(x) = -2\, \overrightarrow{\Sigma(\mathrm{x})\Sigma_0(\mathrm{x})} \tag{14}$$

6.3 A least-squares attachment term for sparsely distributed tensors

Now, let us consider the case where we do not have a dense measure of our tensor field, but only N measures Σ_i at irregularly distributed sample points x_i. Assuming a uniform Gaussian noise independent at each position still leads to a least-squares criterion:

$$Sim(\Sigma) = \sum_{i=1}^{N} \text{dist}^2\,(\Sigma(x_i)\,,\, \Sigma_i) = \int_\Omega \sum_{i=1}^{N} \text{dist}^2\,(\Sigma(x)\,,\, \Sigma_i)\; \delta(x - x_i)\, dx$$

In this criterion, the tensor field $\Sigma(x)$ is related to the data only at the measures points x_i though the Dirac distributions $\delta(x - x_i)$. If the introduction of distributions may be dealt with for the theoretical differentiation of the criterion with respect to the continuous tensor field Σ, it is a real problem for the numerical implementation. In order to regularize the problem, we consider the Dirac distribution as the limit of the Gaussian function G_σ when σ goes to zero. Using that scheme, our criterion becomes the limit case $\sigma = 0$ of:

$$Sim_\sigma(\Sigma) = \int_\Omega \sum_{i=1}^{N} \text{dist}^2\left(\Sigma(x)\ ,\ \Sigma_i\right)\ G_\sigma(x - x_i)\, dx \tag{15}$$

From a practical point of view, we need to use a value of σ which is of the order of the spatial resolution of the grid on which $\Sigma(x)$ is evaluated, so that all measures can at least influence the neighboring nodes.

Now that we came back to a smooth criterion, we may differentiate it exactly as we did for the dense measurement setup. The first order variation is:

$$Sim_\sigma(\Sigma + \varepsilon\Lambda) = Sim_\sigma(\Sigma) - 2\,\varepsilon \int_\Omega \left\langle\ \Lambda(x)\ \middle|\ \sum_{i=1}^{N} G_\sigma(x - x_i)\ \overrightarrow{\Sigma(\mathrm{x})\Sigma_\mathrm{i}}\ \right\rangle\ dx + O(\varepsilon^2),$$

so that we get:

$$\nabla Sim_\sigma(x) = -2\ \sum_{i=1}^{N} G_\sigma(x - x_i)\ \overrightarrow{\Sigma(\mathrm{x})\Sigma_\mathrm{i}} \tag{16}$$

6.3.1 Interpolation through diffusion

With the sparse data attachment term (16) and the isotropic first order regularization term (10), we are looking for a tensor field that minimizes its spatial variations while interpolating (or more precisely approximating at the desired precision) the measurement values:

$$C(\Sigma) = \sum_{i=1}^{N} G_\sigma(x - x_i)\ \text{dist}^2\left(\Sigma(x_i)\ ,\ \Sigma_i\right) + \lambda \int_\Omega \|\nabla\Sigma(x)\|^2_{\Sigma(x)}\, dx$$

According to the previous sections, the gradient of this criterion is

$$\nabla C(\Sigma)(x) = -2\ \sum_{i=1}^{N} G_\sigma(x - x_i)\ \overrightarrow{\Sigma(\mathrm{x})\Sigma_\mathrm{i}} - 2\,\lambda\,\Delta\Sigma(x)$$

Using our finite difference approximation scheme (Eq. 13), the intrinsic geodesic gradient descent scheme (Sec. 3.6) is finally:

$$\Sigma_{t+1}(x) = \exp_{\Sigma_t(x)}\left(\varepsilon\left\{\sum_{i=1}^{N} G_\sigma(x - x_i)\ \overrightarrow{\Sigma(\mathrm{x})\Sigma_\mathrm{i}} + \lambda' \sum_{u\in\mathcal{V}} \frac{\overrightarrow{\Sigma(\mathrm{x})\Sigma(\mathrm{x}+\mathrm{u})}}{\|u\|^2}\right\}\right) \tag{17}$$

Last but not least, we need an initialization of the tensor field $\Sigma_0(x)$ to obtain a fully operational algorithm. This is easily done with any radial basis function approximation, for instance the renormalized Gaussian scheme that we investigated in Section 4.4. Figure 10 displays the result of this algorithm on the interpolation between 4 tensors. On can see that the soft closest point approximation is well regularized into a constant field equal to the mean of the four tensors if data attachment term is neglected. On the contrary, a very small value of λ is sufficient for regularizing the field between known tensors (as soon as σ is much smaller than the typical spatial distance between two measures).

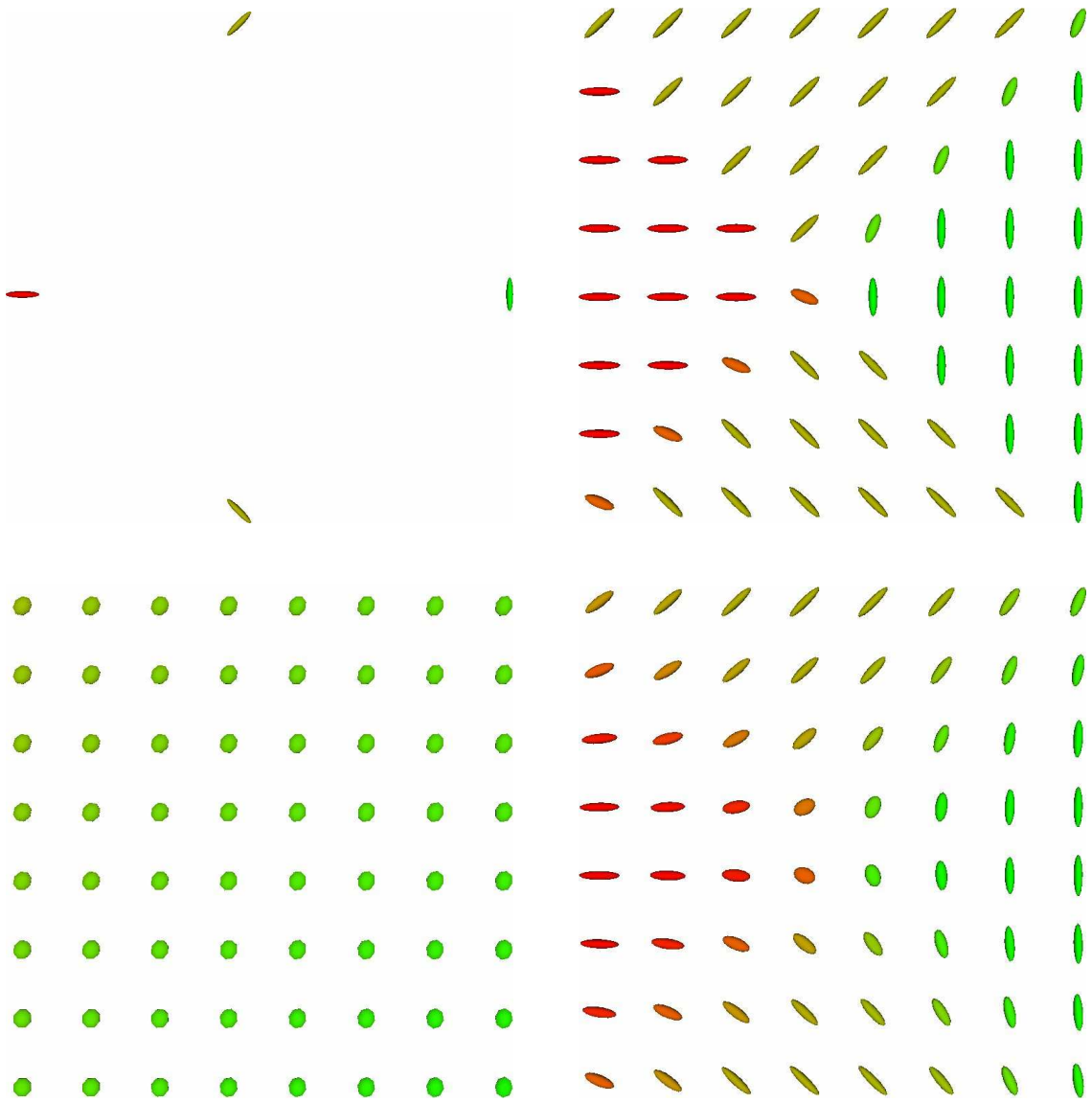

Figure 10: Interpolation and extrapolation of tensor values from four measurements using diffusion. **Top left:** The four initial tensor measurements. **Top right:** Initialization of the tensor field using a soft closest point interpolation (mean of the four tensors with a renormalized spatial Gaussian influence). **Bottom left:** result of the diffusion without the data attachment term (1000 iterations, time-step $\varepsilon = 1$, $\lambda = +\infty$). **Bottom right:** result of the diffusion with an attachment term after (1000 iterations, time-step $\varepsilon = 1$, $\lambda = 0.01$, $\sigma = 1$ pixel of the reconstruction grid). The algorithm did in fact converge in about 100 iterations.

7 Conclusion

We propose in this paper an affine invariant metric that gives to the space of positive define symmetric matrices (tensors) a very regular manifold structure. In particular, tensors with null and infinite eigenvalues are both at an infinite distance of any positive definite symmetric matrix: the cone of positive definite symmetric matrices is replaced by a space which has an infinite development in each of its $n(n+1)/2$ directions. Moreover, there is one and only one geodesic joining any two tensors, and we can even define globally consistent orthonormal coordinate systems of tangent space. Thus, the structure we obtain is very close to a vector space, except that the space is curved. We exemplify some the the good metric properties for some simple statistical operations. For instance, the Karcher mean in Riemannian manifolds has to be defined through a distance-based variational formulation. With our invariant metric on tensor, the existence and uniqueness is insured, which is generally not the case.

A second contribution of the paper is the application of this framework to important geometric data processing problem such as interpolation, filtering, diffusion and restoration of tensor fields. We show that interpolation and Gaussian filtering can be tackled efficiently through a weighted mean computation. However, if weights are easy to define for regularly sampled tensors (e.g. for linear to tri-linear interpolation), the problem proved to be more difficult for irregularly sampled values. The solution we propose is to consider this type of interpolation as a statistical restoration problem where we want to retrieve a regular tensor field between (possibly noisy) measured tensors values at sparse points. This type of problem is usually solved using a PDE evolution equation. We show that the usual linear regularization (minimizing the magnitude of the gradient) and some anisotropic diffusion schemes can be adapted to our Riemannian framework, provided that the metric of the tensor space is taken into account. We also provide intrinsic numerical schemes for the computation of the gradient and Laplacian operators. Finally, simple statistical considerations led us to propose least-squares data attachment criteria for dense and sparsely distributed tensors fields. The differentiation of these criterion is particularly efficient thanks to the use of the Riemannian distance inherited from the chosen metric.

From a theoretical points of view, this paper is a striking illustration of the general framework we are developing since [Pennec, 1996] to work properly with geometric objects. This framework is based on the choice of a Riemannian metric on one side, which leads to powerful differential geometry tools such as the the exponential maps and geodesic marching techniques, and on the transformation of definitions based on linear combination or integrals into minimization problems on the other side. The Karcher mean and the generalized Gaussian distribution are a typical example that we have previously investigated [Pennec, 2004]. In the present paper, we provide new examples with interpolation, filtering and PDE on Riemannian-valued fields.

Many research avenues are still left open, in particular concerning the choice of the metric to use. In a more practical domain, we believe that investigating new intrinsic numerical schemes to compute the derivatives in the PDEs could lead to important gains in accuracy and efficiency. Last but not least, all the results presented in this paper still need to be confronted to other existing methods and validated in the context of medical DTI applications. We are currently investigating another very interesting application field in collaboration with P. Thompson and A. Toga at UCLA: the analysis and the modeling of the variability of brain.

References

[Aubert and Kornprobst, 2001] Aubert, G. and Kornprobst, P. (2001). *Mathematical problems in image processing*, volume 147 of *Applied Mathematical Sciences*. Springer.

[Basser et al., 1994] Basser, P., Mattiello, J., and Bihan, D. L. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267.

[Batchelor et al., 2001] Batchelor, P., Hill, D., Calamante, F., and Atkinson, D. (2001). Study of the connectivity in the brain using the full diffusion tensor from MRI. In Insana, M. and Leahy, R., editors, *Proc. of the 17th Int. Conf. on Information Processing in Medical Imaging (IPMI 2001)*, volume LNCS 2082, pages 121–133. Springer Verlag.

[Bhatia, 2003] Bhatia, R. (2003). On the exponential metric increasing property. *Linear Algebra and its Applications*, 375:211–220.

[Bihan et al., 2001] Bihan, D. L., Manguin, J.-F., Poupon, C., Clark, C., Pappata, S., Molko, N., and Chabriat, H. (2001). Diffusion tensor imaging: Concepts and applications. *Journal Magnetic Resonance Imaging*, 13(4):534–546.

[Cazals and Boissonnat, 2001] Cazals, F. and Boissonnat, J.-D. (2001). Natural coordinates of points on a surface. *Comp. Geometry Theory and Applications*, 19:155–173.

[Chefd'hotel et al., 2002] Chefd'hotel, C., Tschumperlé, D., Deriche, R., and Faugeras, O. (2002). Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. In et al, A. H., editor, *Proc. of ECCV 2002*, number LNCS 2350, pages 251–265. Springer Verlag.

[Chefd'hotel et al., 2004] Chefd'hotel, C., Tschumperlé, D., Deriche, R., and Faugeras, O. (2004). Regularizing flows for constrained matrix-valued images. *J. Math. Imaging and Vision*, 20(1-2):147–162.

[Coulon et al., 2001] Coulon, O., Alexander, D., and Arridge, S. (2001). A regularization scheme for diffusion tensor magnetic resonance images. In Insana, M. and Leahy, R., editors, *Proc. of the 17th Int. Conf. on Information Processing in Medical Imaging (IPMI 2001)*, volume LNCS 2082, pages 92–105. Springer Verlag.

[Coulon et al., 2004] Coulon, O., Alexander, D., and Arridge, S. (2004). Diffusion tensor magnetic resonance image regularization. *Medical Image Analysis*, 8(1):47–67.

[Fillard et al., 2003] Fillard, P., Gilmore, J., Piven, J., Lin, W., and Gerig, G. (2003). Quantitative analysis of white matter fiber properties along geodesic paths. In Ellis, R. E. and Peters, T. M., editors, *Proc. of MICCAI'03, Part II*, volume 2879 of *LNCS*, pages 16–23, Montreal. Springer Verlag.

[Fletcher and Joshi, 2004] Fletcher, P. and Joshi, S. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proc. of the workshop on Computer Vision Approaches to Medical Image Analuy (CVAMIA 2004)*.

[Förstner and Moonen, 1999] Förstner, W. and Moonen, B. (1999). A metric for covariance matrices. In Krumm, F. and Schwarze, V. S., 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.

[Gallot et al., 1993] Gallot, S., Hulin, D., and Lafontaine, J. (1993). *Riemannian Geometry*. Springer Verlag, 2nd edition edition.

[Gamkrelidze, 1991] Gamkrelidze, R., editor (1991). *Geometry I*, volume 28 of *Encyclopaedia of Mathematical Sciences*. Springer Verlag.

[Gerig et al., 1992] Gerig, G., Kikinis, R., Kübler, O., and Jolesz, F. (1992). Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232.

[Helgason, 1978] Helgason, S. (1978). *Differential Geometry, Lie groups, and Symmetric Spaces*. Academic Press.

[Kendall and Moran, 1963] Kendall, M. and Moran, P. (1963). *Geometrical probability*. Number 10 in Griffin's statistical monographs and courses. Charles Griffin & Co. Ltd.

[Kobayashi and Nomizu, 1969] Kobayashi, S. and Nomizu, K. (1969). *Foundations of differential geometry*, volume II of *Interscience tracts in pure and applied mathematics*. John Whiley & Sons.

[Meijering, 2002] Meijering, E. (2002). A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proceedings of the IEEE*, 90(3):319–342.

[Nomizu, 1954] Nomizu, K. (1954). Invariant affine connections on homogeneous spaces. *American J. of Math.*, 76:33–65.

[Pennec, 1996] Pennec, X. (1996). *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).

[Pennec, 1999] Pennec, X. (1999). Probabilities and statistics on riemannian manifolds: Basic tools for geometric measurements. In Cetin, A., Akarun, L., Ertuzun, A., Gurcan, M., and Yardimci, Y., editors, *Proc. of Nonlinear Signal and Image Processing (NSIP'99)*, volume 1, pages 194–198, June 20-23, Antalya, Turkey. IEEE-EURASIP.

[Pennec, 2004] Pennec, X. (2004). Probabilities and statistics on riemannian manifolds: A geometric approach. Research Report 5093, INRIA.

[Pennec and Ayache, 1998] Pennec, X. and Ayache, N. (1998). Uniform distribution, distance and expectation problems for geometric features processing. *Journal of Mathematical Imaging and Vision*, 9(1):49–67.

[Pennec and Thirion, 1997] Pennec, X. and Thirion, J.-P. (1997). A framework for uncertainty and validation of 3D registration methods based on points and frames. *Int. Journal of Computer Vision*, 25(3):203–229.

[Perona and Malik, 1990] Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 12(7):629–639.

[Poincaré, 1912] Poincaré, H. (1912). *Calcul des probabilités*. 2nd edition, Paris.

[Sapiro, 2001] Sapiro, G. (2001). *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press.

[Sibson, 1981] Sibson, R. (1981). A brief description of natural neighbour interpolation. In Barnet, V., editor, *Interpreting Multivariate Data*, pages 21–36. John Wiley & Sons, Chichester.

[Skovgaard, 1984] Skovgaard, L. (1984). A riemannian geometry of the multivariate normal model. *Scand. J. Statistics*, 11:211–223.

[Thévenaz et al., 2000] Thévenaz, P., Blu, T., and Unser, M. (2000). Interpolation revisited. *IEEE Transactions on Medical Imaging*, 19(7):739–758.

[Tschumperlé, 2002] Tschumperlé, D. (2002). *PDE-Based Regularization of Multivalued Images and Applications*. PhD thesis, University of Nice-Sophia Antipolis.

[Tschumperle and Deriche, 2002] Tschumperle, D. and Deriche, R. (2002). Orthonormal vector sets regularization with PDE's and applications. *International Journal on Computer Vision*, 50(3):237–252.

[Weickert, 1998] Weickert, J. (1998). *Anisotropic Diffusion in Image Processing*. Teubner-Verlag.

[Westin et al., 2002] Westin, C., Maier, S., Mamata, H., Nabavi, A., Jolesz, F., and Kikinis, R. (2002). Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6(2):93–108.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399