BIOMATH 208 Week 10

Aidan Jan

March 10, 2025

Review

Geodesics

[FILL geodesic image] A geodesic is the shortest path between two points, p and q, on manifold $\mathcal{M}$.

- Riemannian exponential (and its inverse, Riemannian Log)
- states that if you walk along a curve for a unit of time, at a initial velocity, the endpoint is the result of the exponential.
- The log takes the endpoint and returns the initial velocity.

We looked at:

1. The scale group: The geodesic takes the form of $\exp(at)b$, where $a \in \mathbb{R}$, and $b \in \mathbb{R}'$
2. Positive definite matrices (DTI, covariance matrices): $b^{\frac{1}{2}} \exp(at) b^{\frac{1}{2}}$
   - The exp is the matrix exponential.
3. Probabilities: The geodesics took the form of $1/(1 + \exp(at + b))$

Averages

The sample mean is the (sum all elements / num elements). Weighted sample means weigh each element differently, (sum of all elements times their weights / total weight).

Frechet mean on a Riemannian manifold

[FILL Frechet image] The solution:

$$\exp_p(v_j) = q_j$$

Averages on Manifolds

Proof for Frechet Mean on Riemannian Manifolds

Let $f_i(p) = d^2(p, p_i)$ and $f(p) = \sum_{i=1}^{N} f_i(p)$ and work out

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} f_i(p + \epsilon \delta p)\bigg|_{\epsilon=0}$$

In a chart, let $q(t) + \epsilon \delta q(t)$ be a geodesic curve such that $q(1) = p_i$ and $q(0) = p + \epsilon \delta p$ (i.e., $\delta p = \delta q(0)$). Note this curve has a fixed endpoint, but not fixed start point.

First, plug in the definition of distance squared:

$$= \frac{\mathrm{d}}{\mathrm{d}\epsilon} \int_0^1 g_{ij}(q + \epsilon\delta q)(\dot{q}^i + \epsilon\delta\dot{q}^i)(\dot{q}^j + \epsilon\delta\dot{q}^j)\mathrm{d}t\bigg|_{\epsilon=0}$$

This is the same calculation we already did to derive the geodesic equation. But, there is an extra term that comes from integration by parts without fixed endpoints.

$$\begin{aligned} &= \int_0^1 \partial_k g_{ij}(q)\delta q^k \dot{q}^i\dot{q}^j + g_{ij}(q)\delta\dot{q}^i\dot{q}^j - g_{ij}(q)\dot{q}^i\delta\dot{q}^j\mathrm{d}t \\ &= 0 - g_{ij}(q)\delta q^i(0)\dot{q}^j(0) - g_{ij}q^i(0)\delta q^j(0) \end{aligned}$$

- The zero comes from setting the result to zero, and moving terms around.

Since $\delta q(0) = \delta p$, and $\dot{q}(0) = \log_p(p_i)$, we must have

$$\mathrm{d}f_i(p) = \log_p(p_i)$$

Therefore,

$$\mathrm{d}f(p) = \sum_{i=1}^{N}(-2\log_p(p_i)^\flat w_i)$$

and since the flat map is **always invertible** we recover our definition by setting the gradient to 0 for a stationary solution.

$$\sum_{i=1}^{N}\log_p(p_i) = 0$$

The Frechet Mean Algorithm (aka. Procrustes Analysis)

We start with an initial guess for the average point p_0 and iterate

1. $v_{k+1} = \frac{1}{\sum_{i=1}^N w_i}\sum_{i=1}^N w_i \log_{p_k}(y_i)$
2. $p_{k+1} = \exp_{p_k}(\epsilon v_{k+1})$

For some $\epsilon > 0$. Note when $\epsilon = 1$ we are updating our guess by replacing it with the average of our data in normal coordinates.

If $p_{k+1} = p_k$, then

$$\exp_p\left(\sum_{i=1}^{N} w_i \log_p(q_i)\right) = p$$

This is essentially stating that if we walk along a geodesic with speed 0, we stay where we are.

$$\sum_{i=1}^{N} w_i \log_p(q_i) = 0$$

The Frechet Mean Algorithm is Riemannian Gradient Descent.

Instead of parameters in a straight line along a chart, we update them along a geodesic on the manifold. Instead of:

$$p^i \mapsto p^i - \epsilon g^{-1}(p)^{ij}\mathrm{d}f_j(p)$$

we use

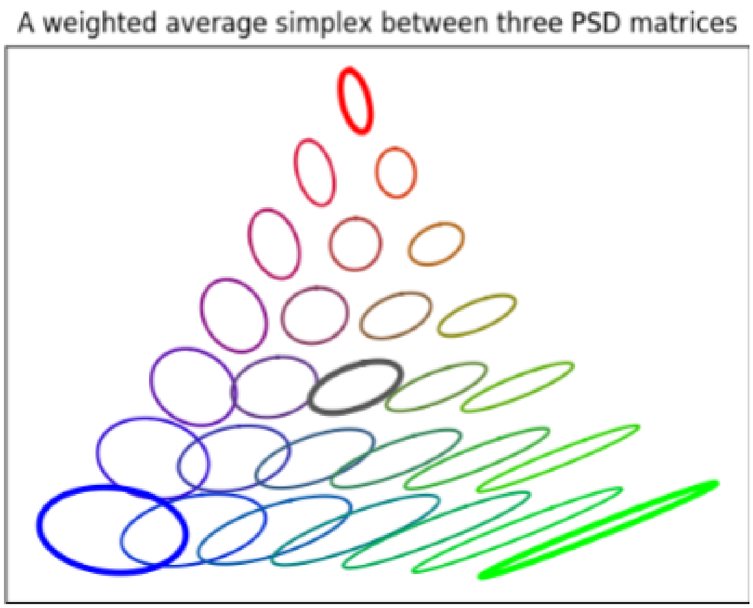
$$p \mapsto \exp_p(-\epsilon\mathrm{d}f^\sharp(p))$$

- We use a sharp map rather than a value in a straight line in a chart. This is because the line may not necessarily be "straight". Using a sharp map allows generalization to any chart.

This makes our gradient descent algorithm coordinate independent.

- Additionally, when ϵ (the learning rate or step size) is small, then the two equations are similar by Taylor expansion.

Weighted Averages of 2D PDS Matrices



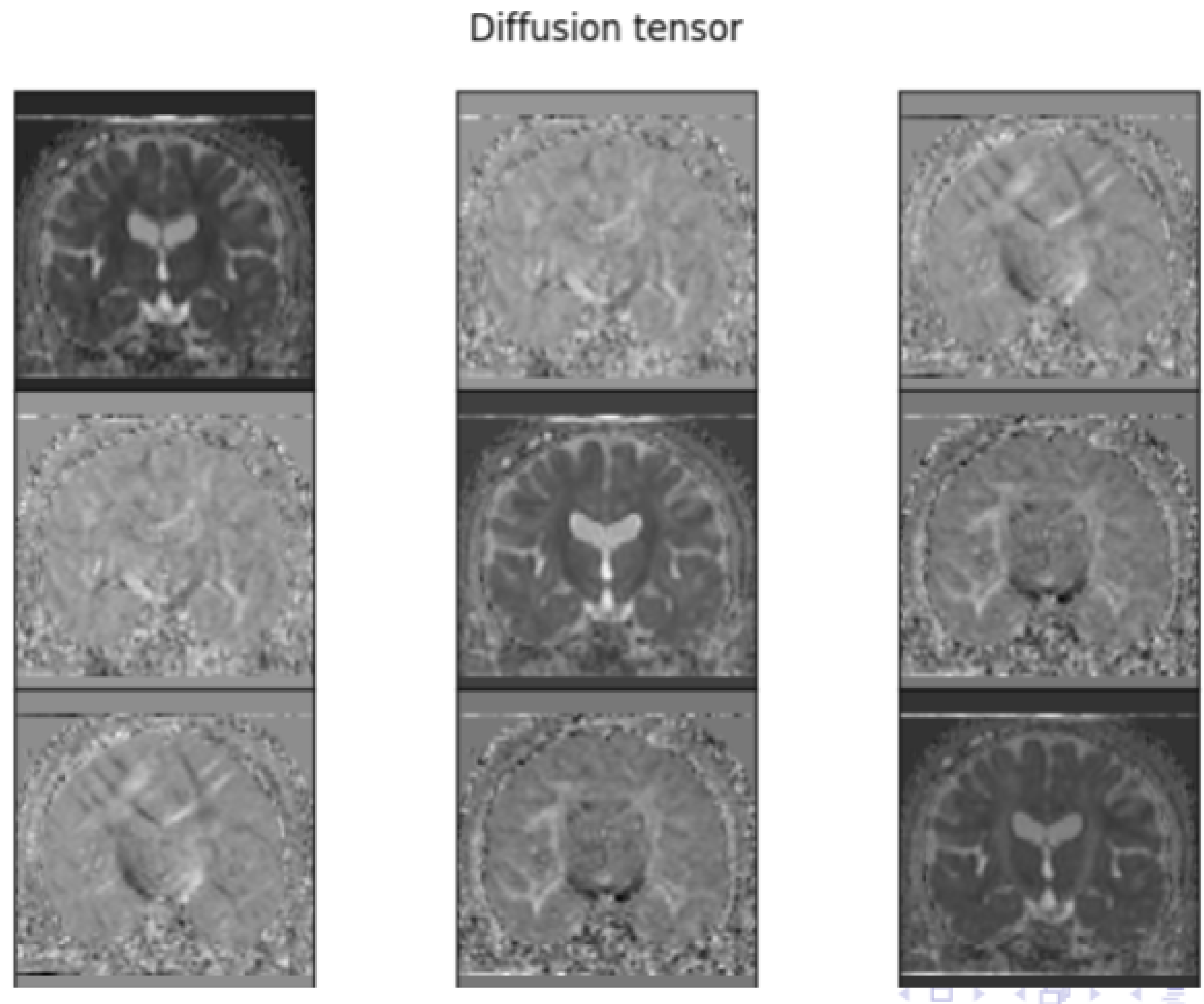
- In the above image, we have three weights (the three vertices of the triangle, R, G, B.)
- In the middle, $w_1 = w_2 = w_3$, where all weights are equal, and each ellipse is interpolated the same amount.
- Weights are added when you move away from the center
- An edge of the triangle (connecting two of the vertices) is a geodesic; it is the shortest path between the two interpolations.

Filtering

- In a vector space, each pixel is replaced with some linear combination of its neighbors. (aka. convolution)
- This can be interpreted as a Frechet mean algorithm, but note that if weights are negative there may not be a solution.
- We will demonstrate this approach applied to DTI images.

Diffusion Tensor Images (DTI)

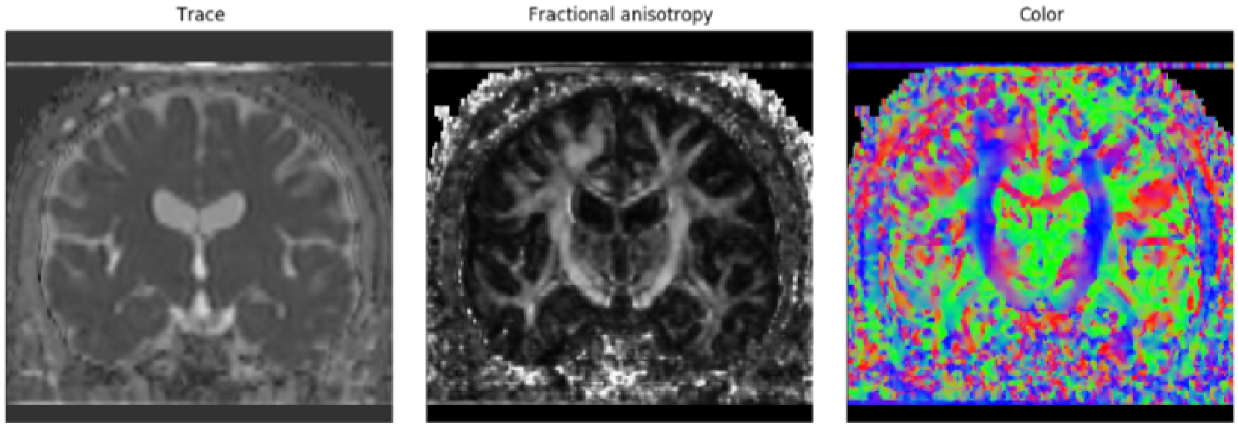
A DTI image measures the 6 components of a PDS diffusion tensor for water through tissue.



- The top left is the 00 component
- The top right is the 02 component
- The bottom right is the 22 component.

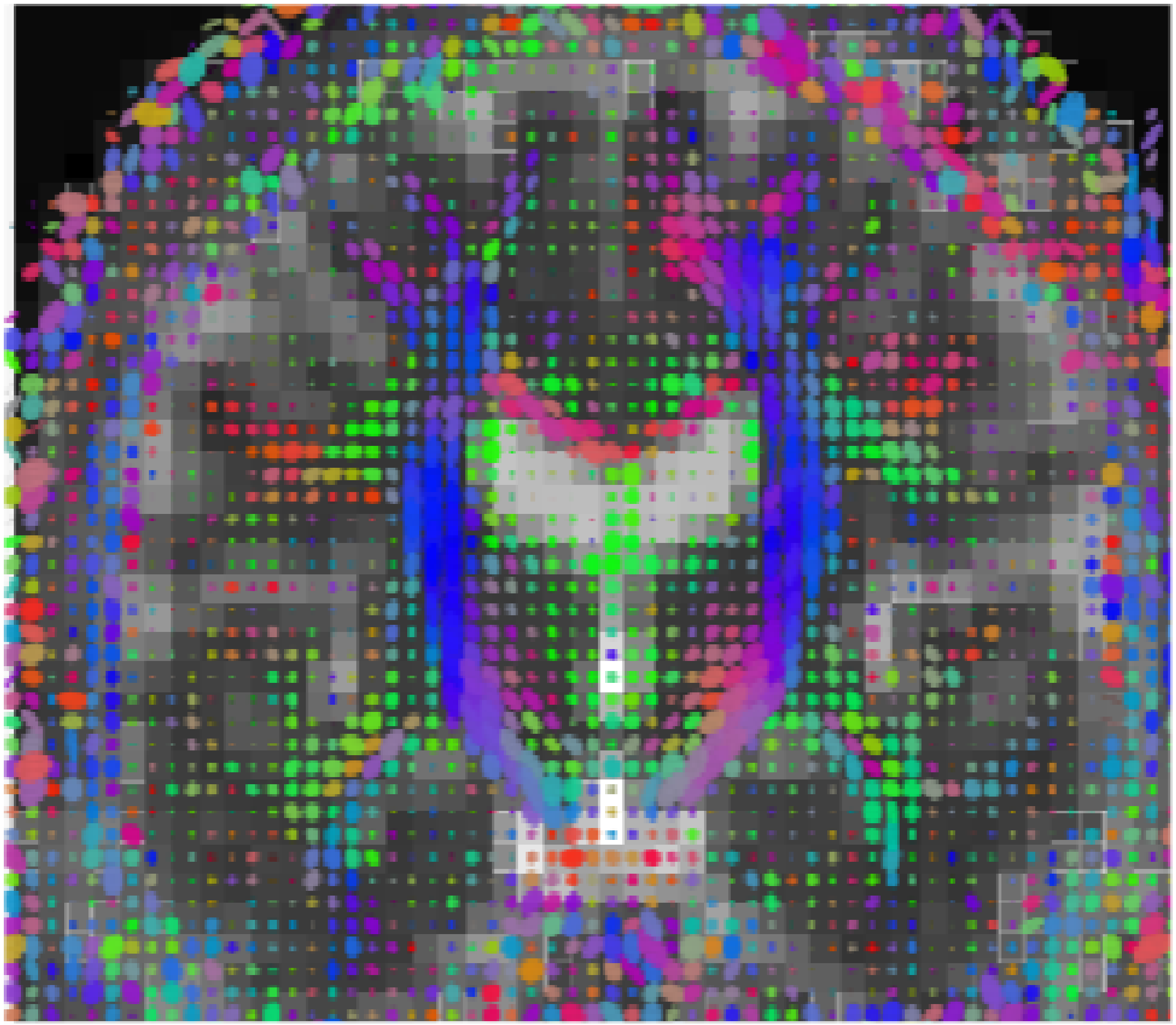
Visualizing DTI Images

We often visualize these images using rotationally invariant measures derived from the tensors. Or we use color to show the direction of the largest eigenvector: $(x, y, z) \mapsto (R, G, B)$.



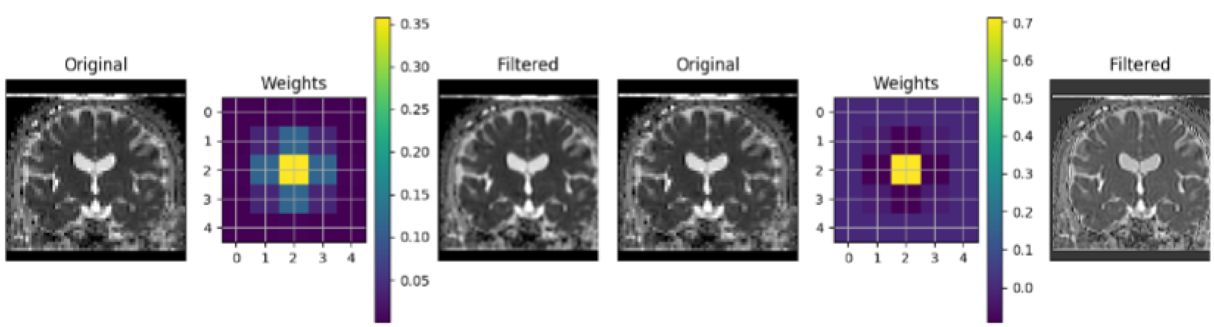
- The fractional anisotropy is like a standard deviation
- It represents the difference between the smallest and largest eigenvector.
- For the color image, the blue part is where water is diffusing up/down. The red is left/right, and green is front/back.

It is also common to show these all together. Below shows a trace image, with colored ellpisoids overlayed, whose size is given by the fractional anisotropy.



Filtering a Trace Image

- Recall: traces are scalor, and vector-valued.



- Left: lowpass, image is less noisy
- Right: highpass, image is clearer, but more noisy
- Noise and resolution tends to be a tradeoff.

Filtering a PDS Image

Left: original, center: lowpass, right: highpass

- Middle: same filter as weights in our Frechet Means Algorithm
- Right: [FILL]

Image Interpolation

Our original definition of image interpolation used a kernel function and the definition of plus and times:

$$I(x) = \sum_{ijk} I[i, j, k] K(x - x[i, j, k])$$

We can now extend this definition to manifold valued images by performing averaging using the Kernel function as weights.

- Now any time you see a linear combination, just replace with [FILL]

Riemannian Regression

Suppose we are given data in pairs: (t_i, y_i) for $t \in \mathbb{R}$ and $y \in \mathcal{M}$, and $w \in \mathbb{R}^+$.

We extend our objective function to

$$f(\theta) = \sum_{i=1}^{N} w_i d^2(\gamma_\theta(t_i), y_i)$$

where γ_θ is a family of curves indexed by θ. e.g., in a vector space we would write $\gamma_\theta(t) = \theta_0 t \theta_1$ for linear regression.

On a manifold we could choose $\gamma_\theta(t) = \exp_{\theta_i}(\theta_0 t)$.

This objective can be optimized using gradient based methods.