BIOMATH 208 Week 9

Aidan Jan

March 10, 2025

1 Review

[FILL]

Image Registration

- Minimize the sum of square error:
$$\text{SSE}(T \cdot I, J)$$

- T is a translator or affine transformation

- For T translator:
$$\mathrm{d}f(T) = \int (I(x - T) - J(x)) \cdot \mathrm{d}I(x - T)\mathrm{d}x$$

- For affine:
$$\mathrm{d}f(T) = \int (I(T^{-1}x) - J(x)) \cdot \mathrm{d}[I(T^{-1}x)]^T \cdot [T^{-1}x]^T \mathrm{d}x$$

Metric Manifolds (Riemannian Manifolds)

[FILL manifold drawing]

$$L(\gamma) = \int_0^1 g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))\mathrm{d}t$$

$$A(\gamma) = \int_0^1 \text{no square root}\mathrm{d}t$$

- Any minimizer of A is also a minimizer of L

- Any minimizer of A is constant speed

- Optimizing over A makes things easier (no square root) and gives a unique parameterization: constant speed geodesics

The Constant Speed Geodesic Equation

In a given coordinate chart, with components of a metric tensor field g written as g_{ij}, and its inverse written g^{ij}, constant speed geodesics are determined by

$$\ddot{q}^i + \frac{1}{2}g^{ij}(-\partial_j g_{kl} + \partial_k g_{jl} + \partial_l g_{kj})\dot{q}^k\dot{q}^l = 0$$

Proof

- We will consider a path $q(t)$ and a perturbation $q \mapsto q + \epsilon \delta q$.

- This denotes a corresponding perturbation in the components of velocity $\dot{q} \mapsto \dot{q} + \epsilon \delta \dot{q}$ (with $\delta \dot{q} = \frac{\mathrm{d}}{\mathrm{d}t} \delta q$).

- We seek to find stationary solutions via

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

  – e.g., find the directional derivative in the direction of δq.

First, we plug into our equation

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

We'll apply chain rule and get derivatives of g

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

Notice that the last two terms are quadratic in velocity.

Not trivial: We will apply integration by parts to write things as gradient dot direction for direction $= \delta q$. e.g., $\delta q(0) = \delta q(1) = 0$ (fixed endpoints). The boundary terms are zero in integration by parts.

$$= \int \partial_k g_{ij}(q(t)) \delta g^k(t) \dot{q}^i(t) \dot{q}^j(t) + g_{ij}(q(t)) \delta \dot{q}^i(t) \dot{q}^j(t) + g_{ij}(g(t)) \dot{q}^i(t) \delta \dot{q}^j(t) \mathrm{d}t$$

The first term is something acting linearly on the direction. The next two terms are not, since $\delta \dot{q}^i(t)$ is the time derivative of the direction.

$$= \int_0^1 \partial_k g_{ij} \dot{q} i(t) \dot{q}^j(t) \delta q^k(t) - \frac{\mathrm{d}}{\mathrm{d}t}(g_{ij}(q)\dot{q}^j)\delta q^i - \frac{\mathrm{d}}{\mathrm{d}t}(g_{ij}(q)\dot{q}^i)\delta q^j \mathrm{d}t$$

Note that δq has 3 different indices, but they are just dummy variables

$$= \int_0^1 \partial_k g_{ij} \dot{q}^i \dot{q}^j \delta q^k - \frac{\mathrm{d}}{\mathrm{d}t}(g_{kj}(q)\dot{q}^j)\delta q^k - \frac{\mathrm{d}}{\mathrm{d}t}(g_{ik}(q)\dot{q}^i)\delta q^k \mathrm{d}t$$

$$= \int partial_k g_{ij}(q)\dot{q}^i \dot{q}^j \delta q^k - \partial_i g_{kj}(q)\dot{q}^i \dot{q}^j \delta q^k - g_{kj}(q)\ddot{q} + \delta q^k$$
$$- \partial_j g_{ik}(q)\dot{q}^j \dot{q}^i \delta q^k - g_{ik}(q)\dot{q}^i \delta \dot{q}^k \mathrm{d}t$$

Now we simplify the equation. We want to simplify to the form

$$\int_0^1 \text{something}(t)^k \cdot q^k(t)\mathrm{d}t = 0$$

Since every term has a q^k, we can factor that out.

$$0 = \partial_k g_{ij}(q)\dot{q}^i \dot{q}^j - \partial_i g_{kj}(q)\dot{q}^i \ddot{q}^j - g_{kj}(q)\ddot{q}^j - \partial_j g_{ik}(g)\dot{q}^j \dot{q}^k - g_{ik}(q)\ddot{q}^i$$

Now, we can (1) combine the pairs of items that look the same, and (2) act with the matrix inverse of g (sharp map.)

$$0 = g_{ik}(q)\ddot{q}^i + \frac{1}{2}(-\partial_k g_{ij}(q) + \partial_i g_{kj}(q) + \partial_j g_{ik}(q))\dot{q}^i \dot{q}^j$$

$$0 = \ddot{q}^l + \frac{1}{2}(g^{-1})^{kl}(-\partial_k g_{ij}(q) + \partial_i g_{kj}(q) + \partial_j g_{ik}(q))\dot{q}^i \dot{q}^j$$

[FILL]

Christoffel Symbols

We simplify by introducing the notation $\ddot{q}^k + \Gamma^k_{ij}\dot{q}^i\dot{q}^j = 0$. **Definition: Christoffel Symbol of the first kind**

$$\Delta_{kij} = \frac{1}{2}(-\partial_k g_{ij} + \partial_i g_{kj} + \partial_j g_{ik})$$

Definition: Christoffel symbol of the second kind

$$\Gamma^l_{ij} = g^{lk}\Gamma_{kij}$$
$$= \frac{1}{2}g^{lk}(-\partial_k g_{ij} + \partial_i g_{kj} + \partial_j g_{ik})$$

Note they are symmetric in the last two indices.

Geodesics in Euclidean Space

If g is constant everywhere in some chart, then the resulting geodesics are linear equations in this chart

$$q^i(t) = a^i t + b^i$$

Proof:

If g is constant then $\partial_k g_{ij} = 0$ for all i, j, k. The equation becomes

$$\ddot{q}^i = 0$$

Integrating once gives

$$\dot{q}^i = a^i$$

for some a. Integrating again gives the solution for arbitrary b.

Geodesics in polar coordinates

Earlier we showed that the Euclidean dot product in polar coordinates is

$$g_{ij}(r, \theta) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}_{ij}$$

Geodesics in this coordinate system are given by

$$\ddot{r} - r\dot{\theta}^2 = 0$$
$$\ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} = 0$$

Proof

[FILL]

Now, start calculating partial derivatives. Note they are only nonzero when we take the derivative of the $\theta - \theta$ component $(i = 1, j = 1)$, with respect to r.

$$[FILL]$$

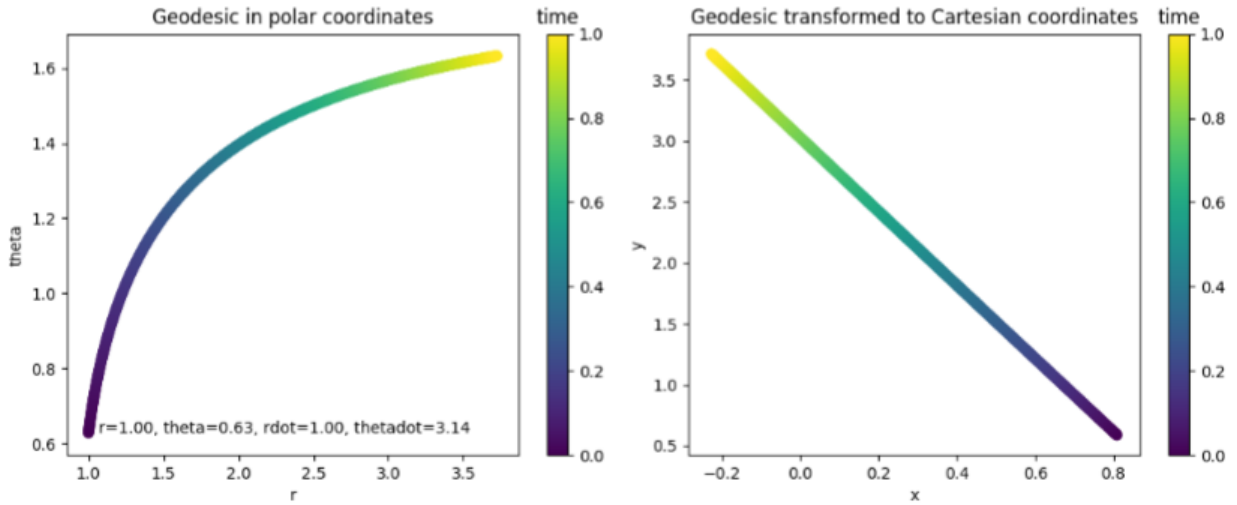Now Christoffel symbols of the first kind.

$$\Gamma_{kij} = -\frac{1}{2}(-\partial_k g_{ij} + \partial_i g_{kj} + \partial_j g_{ik})$$
$$\Gamma_{000} = 0$$
$$\Gamma_{010} = 0$$
$$\Gamma_{100} = 0$$
$$\Gamma_{110} = \Gamma_{101} = r\Gamma_{001} \qquad\qquad = [FILL]\Gamma_{011} = \Gamma_{101} \qquad = \Gamma_{111} =$$

Now, Christoffel symbols of the second kind

$$\Gamma^l_{ij} = g^{lk}\Gamma_{kij}$$
$$\Gamma^0_{00} = g^{00}\Gamma_{000} + g^{01}\Gamma_{100} = 0$$
$$\Gamma^0_{01} = g^{00}\Gamma_{001} + g^{01}\Gamma_{101} = 0$$
$$\Gamma^0_{10} = g^{00}\Gamma_{001} + g^{01}\Gamma_{101} = 0$$
$$\Gamma^0_{10} = g^{00}\Gamma_{010} + g^{01}\Gamma_{110} = -r$$
$$\Gamma^1_{00} = g^{10}\Gamma_{000} + g^{11}\Gamma_{100} = 0$$
$$\Gamma^1_{01} = g^{10}\Gamma_{001} + g^{11}\Gamma_{101} = y_r$$
$$\Gamma^1_{10} = g^{10}\Gamma_{001} + g^{11}\Gamma_{101} = y_r$$
$$\Gamma^1_{10} = g^{10}\Gamma_{010} + g^{11}\Gamma_{110} = 0$$

An example of geodesic curves are shown below:

Christoffel Symbols are not Tensors

Why? Because a tensor that is 0 in one chart induced basis will be 0 in all others.

- Change of basis involved multiplying by Jacobians, these are always invertible matrices (smoothly compatible atlas).

Size Data

Size data is very common in medical imaging, where it is often called volumetry or morphometry. It gives a means to quantify data in structural images.

Geodesics on size Consider $\mathcal{M}$ the space of sizes (or the scale group), with the "natural" coordinate chart (i.e., a number in $\mathbb{R}^+$). Let $g(q) = \frac{1}{q^2}$, and so $g^{-1}(q) = q^2$. Note this metric is left invariant. To see this, just push forward vectors with q^{-1}.

$$g(q)(u,v) = \frac{1}{q^2}uv = g(I)(\frac{1}{q}u, \frac{1}{q}v)$$

- e.g., we push forward, back to identity, with q^{-1}.

For $a, b \in \mathbb{R}$, geodesics are given by

$$q(t) = \exp(at+b)$$

Proof

First calculate g^{-1}

$$\left(\frac{1}{q^2}^{-1}\right) = q^2 \qquad \rightarrow \text{ sharp map}$$

Now, calculate the derivative of g.

$$\partial_0 g_{00} = \frac{\mathrm{d}}{\mathrm{d}q}\frac{1}{q^2} = -2q^{-3}$$

Now calculate the Christoffel symbol of the first kind.

$$\Gamma_{000} = \frac{1}{2}(-2q^{-3})$$

Note, in 1D, two of the terms cancel out because of a minus sign.

Now, calculate the Christoffel symbol of the second kind. [FILL]
   Write the geodesic equation:

$$\ddot{q} - \frac{1}{q}\dot{q}^2 = 0$$

Notice that the coefficients are the Christoffel symbols.

Show exponentials are a solution:

$$q(t) = \exp(at + b)$$
$$\dot{q}(t) = a \exp(at + b)$$
$$\ddot{q}(t) = a^2 \exp(at + b)$$

Plugging in,

$$a^2 \exp(at + b) - \frac{1}{\exp(at + b)}(a \exp(at + b))^2 = 0$$

Distances on Size Data

The Riemannian distance between two points $0 < a < b$ is given by the log of their ration

$$d(a, b) = \log(b/a) = \log(b) - \log(a)$$

Note the distance between sa and sb is the same as that between a and b (left invariance).

Proof

Let $q(t)$ be any increasing function with $q(0) = a$ and $q(1) = b$. Then,

$$d(a, b) = \int_0^1 \sqrt{\frac{\dot{q}^2(t)}{q^2(t)}}\mathrm{d}t$$
$$[FILL]$$

Normal coordinates on sizes

We chose a base of point 1, and use a chart where

$$x(p) = d(1, p)\mathrm{sign}(p - 1) = \log(p)$$

This chart has the property that the Euclidean distance from 1 to p in the chart, is the Riemnanian distance on the manifold.

This property defines "normal coordinates" on a manifold.

Probability data

Probability data shows up in image segmentation/classification tasks, where an image or a pixel is assigned to a given category with some probability (e.g., foreground vs. background)

- Consider $\mathcal{M}$ the space of probabilities, with the "natural" coordinate chart (i.e., a number in $(0, 1)$). Let's choose the metric tensor $g(q) = \frac{1}{q^2(1-q)^2}$ which says that the parameter changes near the endpoints are larger than those near the middle. Note, g is invariant under $q \mapsto 1 - q$.

- For $a, b \in \mathcal{R}$, geodesics are given by the logistic function (sigmoid)

$$q(t) = \frac{1}{1 + \exp(at + b)}$$

Review

Geodesics: Shortest path between two endpoints on a Riemannian manifold (metric tensor at every $p \in \mathcal{M}$) [FILL drawing/equations]

$$A[\gamma] = \int_0^1 \text{no square root} dt$$

Minimizers of A also minimize L_1 and are constant speed. We optimized the action integral by taking a directional derivative (in the space of curves), and setting the result to 0.

For a curve $g(t)$ in a chart, $\Rightarrow \ddot{q}^i(t) + \Gamma_i$ [FILL]

Examples of Geodesics

- Euclidean space, linear functions of time

- Polar coordinates, it was hard, but we did get the same answer.

- Scale group, left invariant metric $g_q(x, y) = \frac{1}{q^2} xy$

- A solution:

$$q(t) = \exp(at + b)$$

  - a and b describe initial velocity and position (although they are not exactly initial velocity and position)

$$\text{distance}(p, q) = \log(q/p) \qquad \{q > p\}$$

Probability data

Probability data shows up in image segmentation/classification tasks, where an image or a pixel is assigned a to a given category with some probability (e.g. foreground versus background).

Consider $\mathcal{M}$ in the space of probabilities, with the "natural" coordinate chart (i.e., a number in $(0, 1)$). Let's choose the metric tensor $g(q) = \frac{1}{q^2(1-q)^2}$, which says that marameter changes near the endpoints are larger than those near the middle. Note, g is invariant under $q \mapsto 1 - q$.

For $a, b \in \mathcal{R}$, geodesics are given by the logistic function (sigmoid)

$$q(t) = \frac{1}{1 + \exp(at + b)}$$

Proof:

[FILL]

Distances on Probabilities

For $a < b \in (0,1)$, the distance between two probabilities is the difference between their log of odds

$$d(a,b) = \log\left(\frac{b}{1-b}\right) - \log\left(\frac{a}{1-a}\right)$$

- Each term here is called a "log of odds", aka a "logit"

Proof:

Note $\log(q/(1-q))$ is the inverse of the logistic function (function inverse, not matrix inverse). Let $q(t) = 1/(1 + \exp(ct + d))$ (for $c < 0$).

Consider the sigmoid curve:

$$d(a,b) = \int \sqrt{\frac{\dot{q}^2(t)}{q(t)(1 - q(t))}} dt$$

If we define:

$$\dot{q} = cq(1 - q)$$
$$\dot{q}^2 = c^2 q^2 (1 - q^2)$$

then, we can simplify:

$$= \int_0^1 \sqrt{c^2} dt$$

$$= \int_0^1 c \, dt$$

Note that c must be negative for an increasing curve. (Take the correct sign when doing a square root)

Now, work out c as a function of a and b.

- If $t = 0$, then $q(0) = \frac{1}{1 + \exp(d)} = a \Rightarrow d = \log(a/1 - q)$.
- If $t = 1$, then $q(t) = \frac{1}{1 + \exp(c+d)} = b$

Therefore,

$$c + d = \log\left(\frac{b}{1-b}\right)$$

$$-c = d - (c + d) = \log\left(\frac{b}{1-b}\right) - \left(\frac{a}{1-a}\right)$$

Normal Coordinates on Probabilities

Choose a base point of 0.5 and write

$$x(p) = \text{sign}(p - 0.5)\text{d}(0.5, p) = \log\left(\frac{p}{1-p}\right)$$

These are normal coordinates for this 1D manifold. A statistics, this is called the "canonical parameterization" (of a Bernoulli random variable)

- In normal coordinates, every straight line moving away from the base point is a constant speed geodesic

Positive Definite Symmetric Matrix Data

Diffusion tensor imaging outputs a positive definite symmetric matrix at each pixel. This helps us understand local orientation of tissue microstructure, which is crucial for identifying white matter tracts in the brain, or fiber orientations in muscle such as heart.

With A a square matrix, we denote its exponentail as $\exp(A)$. It can be defined using a Taylor series

$$\exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$$

or with a "compound interest formula"

$$\exp(A) = \lim_{n \to \infty} \left(I + \frac{1}{n} A \right)^n$$

Properties of the Matrix Exponential

1. $\exp(A)A = A\exp(A)$

2. $\frac{\mathrm{d}}{\mathrm{d}t}\exp(At) = A\exp(At) = \exp(At)A$

3. If A and B commute, then

$$\exp(A + B) = \exp(A)\exp(B) = \exp(B)\exp(A)$$

Positive Definite Symmetric Matrix Data

The matrix logarithm is the inverse of the matrix exponential, i.e., $\exp(\log(A)) = A$. It can also be defined by a Taylor series

$$\log(I + A) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{N} A^n$$

Geodesics on Positive Definite Symmetric Matrices

Consider $\mathcal{M}$ the set of positive definite symmetric matrices, with the metric

$$g_A(U, V) = \mathrm{tr}[(A^{-\frac{1}{2}} U A^{-\frac{1}{2}})(A^{-\frac{1}{2}} V A^{-\frac{1}{2}})]$$
$$= \mathrm{tr}[(A^{-1}U)(A^{-1}V)]$$

Here U, V are symmetric, but not necessarily positive definite. Geodesics take the form $A(t) = b^{\frac{1}{2}} \exp(at) b^{\frac{1}{2}}$, where a is symmetric and b is symmetric positive definite.

Proof

We will work in <u>extrinsic coordinates</u>, and find stationary paths of the action integral directly. Consider a path $A(t)$, perturbed to $A(t) + \epsilon \delta A(t)$. Here, δA is symmetric, with $\delta A(0) = \delta A(1) = 0$. Take the derivative of our action integral

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \int_0^1 \mathrm{tr}\left[(A + \epsilon \delta A)^{-1} (\dot{A} + \epsilon \delta \dot{A})(A + \epsilon \delta A)^{-1}(\dot{A} + \epsilon \delta \dot{A}) \right] \mathrm{d}t \Bigg|_{\epsilon=0}$$
$$= \int \mathrm{tr}\left[-A^{-1}\delta A A^{-1}\dot{A}A^{-1}\dot{A} \right] + \mathrm{tr}\left[A^{-1}\delta \dot{A}A^{-1}\dot{A} \right] + \mathrm{tr}\left[-A^{-1}\dot{A}A^{-1}\delta A A^{-1}\dot{A} \right] + \mathrm{tr}\left[A^{-1}\dot{A}A^{-1}\delta \dot{A} \right] \mathrm{d}t$$
$$= \int_0^1 \partial \mathrm{tr}[-A^{-1}\dot{A}A^{-1}\dot{A}A^{-1}\delta A + A^{-1}\dot{A}A^{-1}\delta \dot{A}] \mathrm{d}t$$

By integration by parts:

$$= \int_0^1 \partial \mathrm{tr} \left[(-A^{-1}\dot{A}A^{-1}\dot{A}A^{-1} - \frac{\mathrm{d}}{\mathrm{d}t}A^{-1}\dot{A}A^{-1})\delta A \right] \mathrm{d}t$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}\left(A^{-1}\dot{A}A^{-1} \right)$$

$$= -A^{-1}\dot{A}A^{-1}\dot{A}A^{-1} + A^{-1}\ddot{A}A^{-1} - A^{-1}\dot{A}A^{-1}\dot{A}A^{-1}$$

Now, we can rewrite the whole expression and set it to 0.

$$0 = A^{-1}\ddot{A}A^{-1} - A^{-1}\dot{A}A^{-1}\dot{A}A^{-1}$$

We can multiply all the terms by $A^{-\frac{1}{2}}$ to further simplify:

$$0 = A^{-\frac{1}{2}}\ddot{A}A^{-\frac{1}{2}} - (A^{-\frac{1}{2}}\dot{A}A^{-\frac{1}{2}})(A^{-\frac{1}{2}}\dot{A}A^{-\frac{1}{2}})$$

$$0 = \ddot{A} - \dot{A}A^{-1}\dot{A}$$

Now we show our solution satisfies this equation:

$$A(t) = b^{\frac{1}{2}}\exp(at)b^{\frac{1}{2}}$$

$$\dot{A}(t) = b^{\frac{1}{2}}a\exp(at)b^{\frac{1}{2}} - b^{\frac{1}{2}}a^{\frac{1}{2}}\exp(at)a^{\frac{1}{2}}b^{\frac{1}{2}}$$

$$\ddot{A} = b^{\frac{1}{2}}[FILL]$$

Distances on Positive Definite Matrices

If A and B are two positive definite symmetric matrices, then the Riemannian distance between them is

$$d(A,B) = \sqrt{\mathrm{tr}\left[\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}) \right]}$$

Proof

Consider the path $C(t) = A^{\frac{1}{2}}\exp(\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})t)$. Note that

$$C(0) = [FILL]$$

and that

$$C(1) = [FILL]$$

The initial velocity is then

$$\dot{C}(0) = A^{\frac{1}{2}}\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})A^{\frac{1}{2}}$$

Its initial speed squared is then:

- the velocity times A^{-1}, then product with itself, then trace.

$$\mathrm{tr}\left[A^{-\frac{1}{2}}\dot{C}(0)A^{-\frac{1}{2}}A^{-\frac{1}{2}}\dot{C}(0)A^{-\frac{1}{2}} \right]$$

To find distance squared from A to B, it is $\dot{C}(0)$. Constant speed, integrating for 1 unit of time, then initial speed squared is equal to distance squared.

The Riemannian Exponential

We have seen two examples where geodesics turn out to be exponentials. This motivates the use of this language in other situations. We define the Riemannian exponential as

$$\exp_p(v)$$

for $p \in \mathcal{M}$ and $v \in T_p\mathcal{M}$ to be the endpoint of a constant speed geodesic starting at p with initial velocity v. Note that the square distance between endpoints is given by

$$d^2(p, \exp_p(v)) = g_p(v, v)$$

Normal Coordinates

We can use the matrix exponential to build a coordinate system. Given a base point p_0, we chose a coordinate map

$$x^i(p) = \mathrm{d}x^i(\log_{p_0}(p))$$

With this chart all straight lines through the origin are geodesics!

- Note that $\log_{p_0}(p)$ is the inverse of the Riemannian exponential

Averaging, Filtering, and Regression

Motivation

- Here we show how the Riemannian exponential and logarithm can be used to define averages by minimizing sum of square distance problems

- These averages will be used to define means, interpolation and extrapolation, and regression.

Sample Mean in Vector Spaces

For x_i a set of N points in a vector space V, and using the definitions of plus and times, we define the sample mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

or the weighted sample mean

$$\bar{x}_w = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$$

for weights $w_i \in \mathbb{R}^+$. (Note subscript does not refer to covector components here).

This is the common definition of 'mean' that was taught in high school.

Locally weighted averages

In imaging, locally weighted averages are used for processing such as denoising or antialiasing, for example

$$\bar{I}[i,j] = \frac{1}{5}(I[i,j] + I[i+1,j] + I[i-1,j] + I[i,j+1] + I[i,j-1])$$

This is a weighted average where weights are 0 for pixels away from (i,j). These types of local averages are usually called "filters".
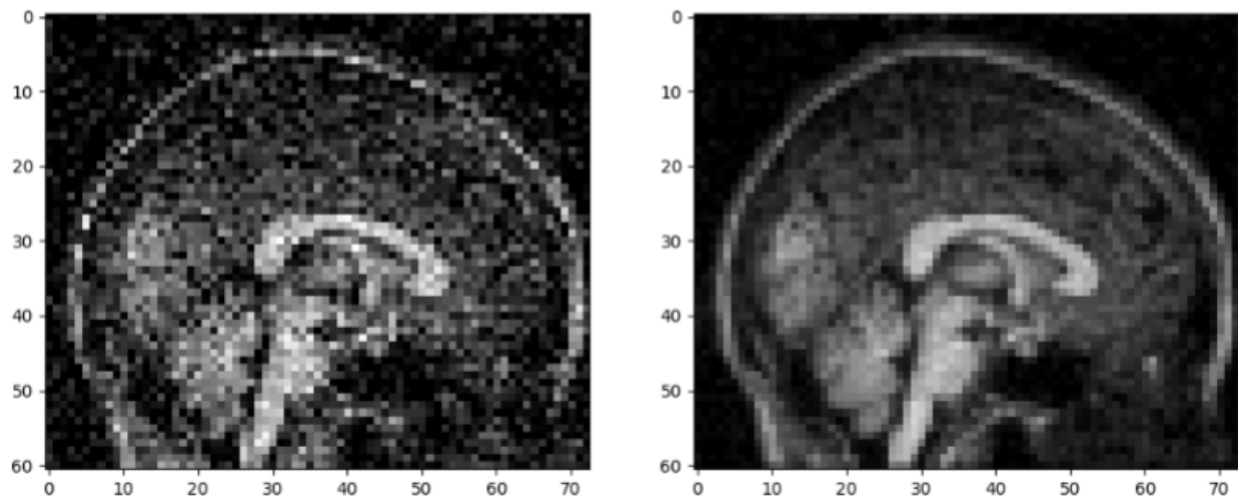
Denoising

Here is an example of the above filter applied to denoising
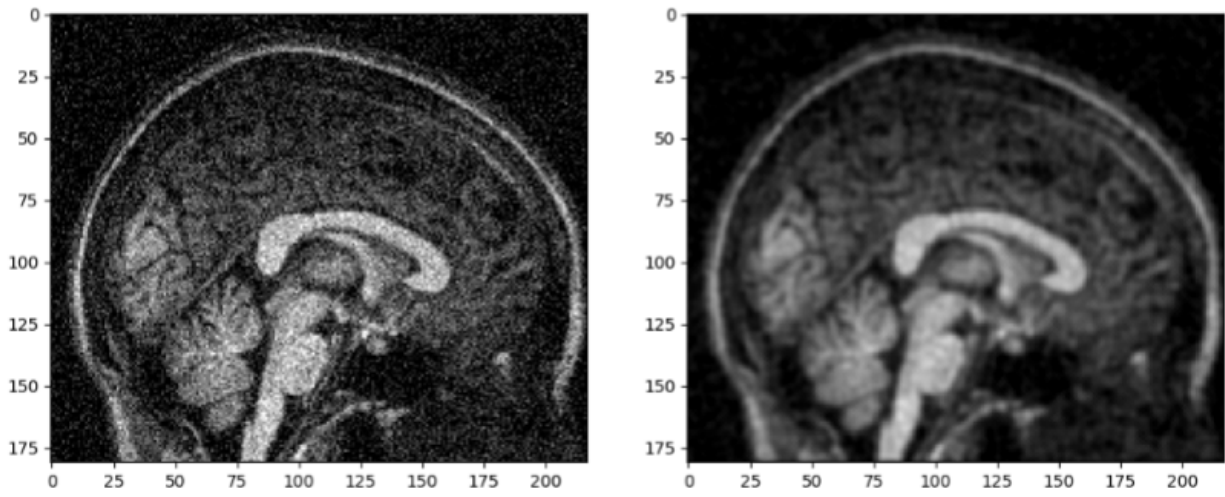
Antialiasing

Filtering before downsampling is critical, otherwise we will end up with noise aliasing.

The Frechet Mean in Metric Spaces

A metric space is one where we can define a distance between any two points. An alternative definition is to find a point that minimizes the sum of square distances to our sample

$$\bar{x} = \arg\min_{y} \sum_{i=1}^{N} w_i d^2(y, x_i)$$

The Frechet Mean in Vector Spaces

If data lies in a vector space with the Euclidian metric, this definition is equivalent to the previous.

Proof.

Search for a stationary point:

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \sum_{i=1}^{N} w_i |y + \epsilon\delta y - x_i|^2 \bigg|_{\epsilon=0}$$

$$= \sum_{i=1}^{N} w_i (y - x_i)^T \delta y$$

The part before the δy acts linearly on the velocity to give the directional derivative

Next, we find a stationary point by setting it to 0

$$= \sum_{i=1}^{N} w_i y = \sum_{i=1}^{N} [FILL]$$

Averages on Manifolds

On a Riemannian Manifold we can't apply the first definition because we have no plus or times, but we can apply the second: **Frechet Mean on Riemannian Manifolds:**

For N samples, $p_i \in \mathcal{M}$, the Frechet mean is characterized by

$$\sum_{i=1}^{N} w_i \log_{\bar{p}_w}(p_i) = 0$$

That is, we find a point such that our samples are 0 mean in normal coordinates.

Proof

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 a fixed start point.

Plug in the definition of distance

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

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. We can continue to simplify:

$$= \int_0^1 \partial_k g_{ij}(q)\delta q^k \dot{q}^i \dot{q}^j + g_{ij}(q)\delta\dot{q}^i \dot{q}^j + g_{ij}(q)\dot{q}^i \delta\dot{q}^j \mathrm{d}t$$

$$= 0 - g_{ij}(q)\delta q^i(0)\dot{q}^j(0) - g_{ij}(q)q^i(0)\delta q^j(0)$$

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

Therefore,

$$\mathrm{d}f(p) = [FILL]$$

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

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

Interpolation by Weighted Averaging

When we have only two datapoints, this equation is very easy to solve. With $w_i \in [0,1]$ and $w_0 = 1 - w_1$, this corresponds to linear interpolation between the two points.

When $w_1 = 1$ this will give p_1, when $w_1 = 0$ this will give p_0. If w_1 is outside of $[0, 1]$ this corresponds to extrapolation.

The weighted average interpolation between two points p_0 and p_1 is found by minimizing

$$(1 - w_1)d^2(\bar{p}_w, p_0) + w_1 d^2(\bar{p}_w, p_1)$$

is given by a point a fraction w_1 along the geodesic from p_0 to p_1.

Proof

With two points, the above solution gives

$$(1 - w_1)\log_{\bar{p}_w}(p_0) + w_1 \log_{\bar{p}_w}(p_1) = 0$$

This means the initial velocities are colinear and correspond to a single geodesic connecting p_0 to p_1.

Also, the distance from $\bar{p}_w$ to p_0 is $w_1/(1 - w_1)$ times as long as the distance from $\bar{p}_w$ to p_1. This is true when $\bar{p}_w$ is a fraction w_1 of the way from p_0 to p_1.

Weighted avverages in a chart will leave the manifold

Here we show an interpolation between two 2D rotation matrices, through their action on an image. Simple weighted averages in a chart leave the manifold.

Weighted average interpolation

Riemannian interpolation
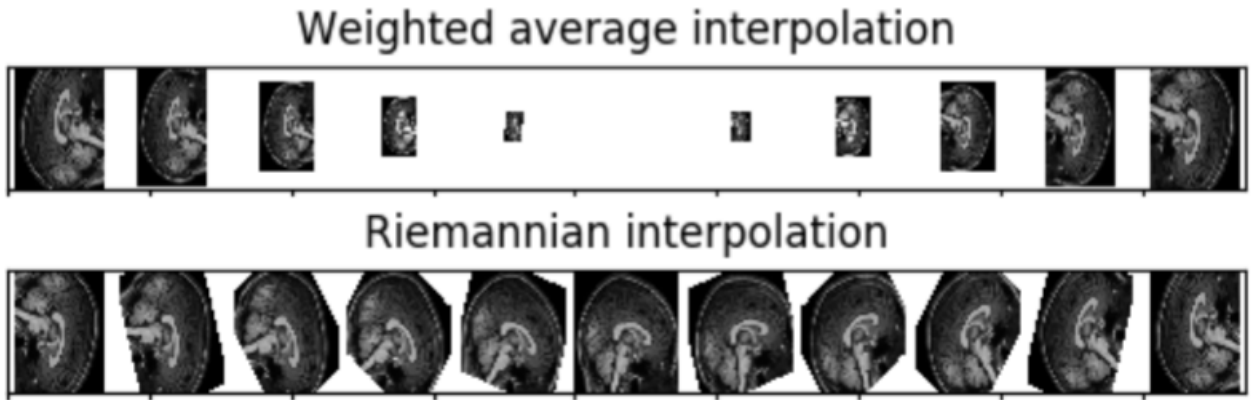
Interpolation between PDS Matrices

If A is PDS, then $x^T A^{-1} x = 1$ defines an ellipse(oid) to be visualized. Note $A^{\frac{1}{2}}(\cos(\theta), \sin(\theta))^T$ is a solution for all θ.

Interpolation by weighted average

Riemannian interpolation