BIOMATH 208 Week 7

Aidan Jan

February 26, 2025

Image Registration and Gradient Descent Optimization

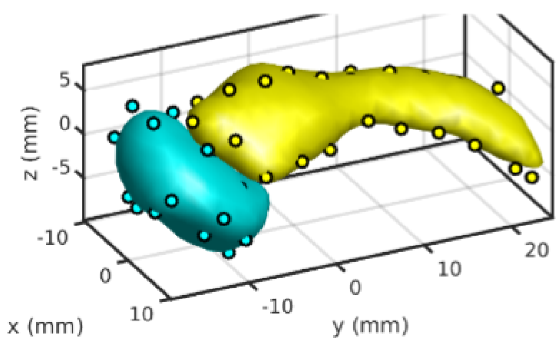
Motivation

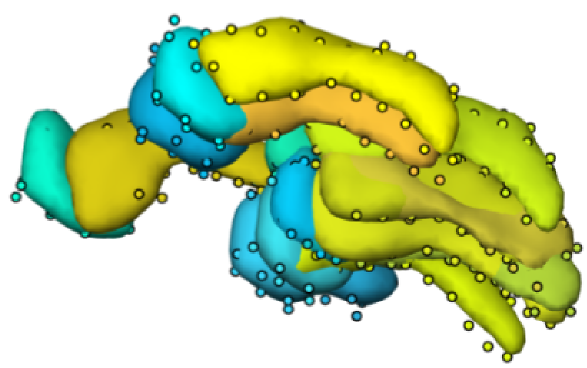
- Image registration consists of finding a transformation (often a member of a group) that spatially aligns a pair of images
- Useful for fusing multiple modalities or timepoints, or making comparisons across members of a population.
- We will use tools we developed to understand differentiation on manifolds to solve several problems analytically, and derive optimization algorithms to solve others.

Exact Landmark Based Registration

Pairs of "keypoints" or "landmark" points can be identified on imaging data, either manually or automatically (e.g., Scale Invariant Feature Transform)

We can align them by minimizing sum of square error, resulting in analytical solutions for many problems





- We will label each point, and find the same point in each image.

Notation

We will use matrix motation (Typically extrinsic coordinates), where Q is a $d \times N$ array such that Qij stores the ith component of the jth of N points. P is a set of corresponding points, and T is an element of some transformation group.

We align images by minimizing the cost

$$f(T) == \|T \cdot Q - P\|_F^2$$
$$= \mathrm{tr}\left[(T \cdot Q - P)(T \cdot Q - P)^T\right] \qquad = \sum_{ij} (T \cdot Q)_{ij}^2$$

- The $\cdot$ in this context means "group action" (e.g., matrix multiplication)
- This equation is the sum of square error cost.
- The subscript F is the Frobineous norm, sum of squares of traces in a matrix.

Aside: Properties of Traces:

- $\mathrm{tr}(A) = -\mathrm{tr}(A^T)$
- $\mathrm{tr}(ABC) = \mathrm{tr}(CBA)$: (invariant to cyclic permutations)
- Linear. $\mathrm{tr}(aA + bB) = a \cdot \mathrm{tr}(A) + b \cdot \mathrm{tr}(B)$

Translation

Consider T is a translation vector in $\mathbb{R}^d$, with $T \cdot Q = T\mathbb{1} + Q$, where $\mathbb{1}$ is a row vector of N ones.

The optimal translation aligns the center of mass of two point sets.

$$T = \frac{1}{N} \sum_{j=1}^{N} P_j - \frac{1}{N} \sum_{j=1}^{N} Q_j$$

Proof

Consider a path $\gamma(t) = T + t\delta T$ for δT an arbitrary translation, and consider the velocity

$$\frac{\mathrm{d}}{\mathrm{d}t} f(T + t\delta T)\bigg|_{t=0} \doteq v_{\gamma,t}(f).$$

- Note that $\dot{\gamma}(0) = \delta T$
- The $\dot{\gamma}$ is the derivative of γ.

Plugging our definition of f gives:

$$\begin{aligned}
&= \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{tr}(Q + T\mathbb{1} + t\delta T\mathbb{1} - P)(Q + T\mathbb{1} + t\delta T\mathbb{1} - p)^T\bigg|_{t=0} \\
&= \mathrm{tr}(\delta T\mathbb{1})(Q + T\mathbb{1} - P)^T + \mathrm{tr}(Q + T\mathbb{1} - P)(\delta T\mathbb{1})^T \\
&= 2\mathrm{tr}(Q + T\mathbb{1} - P)(\delta T\mathbb{1})^T \\
&= \mathrm{tr}(2(Q + T\mathbb{1} - P)\mathbb{1}^T)\delta T^T
\end{aligned}$$

- Note that $\delta T\mathbb{1}$ is linear, and $(Q + T\mathbb{1} - P)^T$ is constant.

We now set the derivative to zero.

$$0 = \mathrm{tr}(2(Q + T\mathbb{1} - P)\mathbb{1}^T)\delta T^T$$

Therefore,

$$2(Q + T\mathbb{1} - P)\mathbb{1}^T = 0$$

If we expand this, we get

$$Q\mathbb{1}^T + T\mathbb{1}\mathbb{1}^T - P\mathbb{1}^T = 0$$

- $\mathbb{1} \cdot \mathbb{1}^T = N$, since $\mathbb{1}$ is a row, $\mathbb{1}^T$ is a column

We can condense this into a summation:

$$\sum_{i=1}^{N} Q_i + NT - \sum_{i-1}^{N} P_i = 0 \Longrightarrow [FILL]$$

We can identify tangent and cotangent terms to define the gradient.

$2(Q + T\mathbb{1} - P)\mathbb{1}^T$ acts linearly on our velocity δT, to give a real number, therefore it is the gradient covector, $\mathrm{d}f(T)$ (e.g., derivative at the point T).

$$\dot{\partial}^i \left(\frac{\partial}{\partial_q}\right) [FILL]$$

Summary

In summary, we

1. Wrote down the loss function.
2. Consider a path, passing through T with direction δT.
3. The components of δT are the components of our velocity.
4. We worked out the time derivative, and found "something" acting linearly on δT.
5. Set expression to zero to find a stationary point.
6. Argued that the "something" $=0$, if this is stationary for any δT.
7. The "something" is the gradient of our function, $\mathrm{d}f$.

Linear Alignment

Consider T an element of the general linear matrix group, with $T \cdot Q = TQ$ (matrix multiplication). Note that if Q has a 0 center of mass, then so does TQ, so we often "preprocess" by translating the center of mass to the origin.

The optimal transformation takes the form of "corss covariance times inverse variance":

$$T = \Sigma_{PQ}\Sigma_{QQ}^{-1}$$

where $\Sigma_{PQ} = \frac{PQ^T}{N}$, $\Sigma_{QQ} = \frac{QQ^T}{N}$

- In this case, Σ_{QQ} is a 3x3 covariance matrix.
- Σ_{PQ} is a 3x3 cross covariance matrix.

Proof

Consider a path $\gamma(T) = T + t\delta T$, where δT is an arbitrary matrix. δT need not be invertible, as long as $\gamma(t)$ is. This may require a small t, but we can always find one (why?). Consider the velocity

$$\frac{\mathrm{d}}{\mathrm{d}t} f(T + t\delta T)\bigg|_{t=0} = v_{\gamma,t}(f)$$

Note that $\dot{\gamma}(0) = \delta T$.

Plugging in our definition of f gives:

$$
\begin{aligned}
&= \mathrm{tr}(TQ + t\delta TQ - P)(TQ + t\delta TQ - P)^T \Big|_{t=0} \\
&= \mathrm{tr}\delta TQ(TQ - P) + \mathrm{tr}(TQ - P)(\delta TQ)^T \\
&= 2 \cdot \mathrm{tr}(TQ - P)(\delta TQ)^T \qquad = \mathrm{tr}\left[2(TQ - P)Q^T\right] \delta T^T
\end{aligned}
$$

In this case, the quantity in square brackets is our gradient. For this entire equation to be 0 for all δT, we need the gradient to equal 0.

$$
\begin{aligned}
0 &= 2(TQ - P)Q^T \\
&\Rightarrow TQQ^T = PQ^T \\
&\Rightarrow T(\frac{1}{N}QQ^T) = \frac{1}{N}PQ^T \\
&\Rightarrow T\Sigma_{QQ} = \Sigma_{PQ} \\
&\Rightarrow T = \Sigma_P Q \Sigma_{QQ}^{-1}
\end{aligned}
$$

We predict P using a linear transform of Q, and minimize δSE.

- P = data (observed)
- Q = covariants (design)

Linear Alignment

We can identify tangent and cotangent terms to define the gradient

- "tangent" $\rightarrow \delta T_{ij}$
- "cotangent" $\rightarrow \mathrm{d}f_{ij} = [2(T \cdot Q - P)Q^T]_{ij}$

[FILL]

Affine Alignemnt

Consider T an element of the affine transformation group in homogeneous coordinates. If Q and P are also in homogeneous coordinates, then we consider $T \cdot Q = TQ$ (matrix multiplication).

The optimal affine transform takes the form of "cross covariance times inverse variance":

$$T = \Sigma_{PQ}\Sigma_{QQ}^{-1}$$

where Σ_{PQ}, ΣQQ are $(d+1) \times (d+1)$ matrices.

- Bottom row of P and Q is all 1's.
- Typical regression problems are affine in your covariants.

Proof

Consider a path $\gamma(t) = T + t\delta T$, where δT is a matrix whose bottom row is all 0's.

The proof is exactly the same as above, but

$$\mathrm{d}f(\delta T) = \mathrm{tr}(\mathrm{d}f \delta T^T) = 0$$

This is because of the row of zeros. We get no information about the sum of the components of $\mathrm{d}f$, because we are extrinsic coordinates. We use 16 numbers to deal with a 12-parameter Lie group.

- Ignore the bottom row of $\mathrm{d}f$ or set it to 0. [FILL]

Rotations

We can choose a parameterization of the rotation group, and identify an optimal rotation iteratively by gradient descent. But, in practice, we can find an exact solution using singular value decompositions.

Gradient Descent

Objective functions other than sum of square error (e.g., distance between curves and surfaces), cannot by solved analytically. For these problems we require an iterative method that approaches an optimal solution. Usually these methods are based on gradient calculations.

Naive Gradient Descent in Vector Spaces

Given a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, the naive gradient descent algorithm procedes as follows:

1.
2.
3.
4.
5.