# Manifold Statistics: Fréchet Mean

## Geometry of Data

September 29, 2022

This lecture is about statistical analysis of data on a **known manifold**.

Such data often arises from nonlinear constraints.

For example, if your data has unit length (such as directional data), it lives on the unit sphere, $S^d$.

**Theorem.** Let $f : M \to N$ be a smooth map between two manifolds of dimension $m \geq n$. If $y \in N$ is a regular value, then the set $f^{-1}(y)$ is a smooth manifold of dimension $m - n$.

# Manifold Examples

▶ **Euclidean Space:** $\mathbb{R}^d$

$\mathrm{id} : \mathbb{R}^d \to \mathbb{R}^d$ is a global coordinate chart
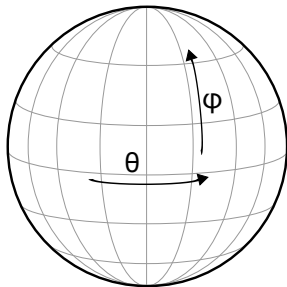
# Manifold Examples

▶ **Euclidean Space:** $\mathbb{R}^d$

  $\mathrm{id} : \mathbb{R}^d \to \mathbb{R}^d$ is a global coordinate chart

▶ **The Sphere:** $S^d$

  ▶ Local coordinate chart for $S^2$:

$$(-\pi, \pi) \times (0, 2\pi) \to S^2$$

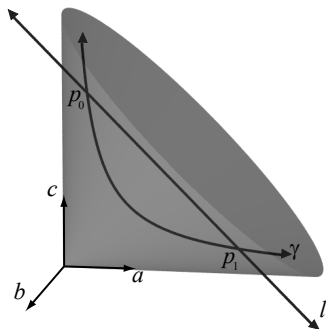$$(\theta, \phi) \mapsto (\cos(\theta)\cos(\phi), \cos(\theta)\sin(\phi), \sin(\theta))$$

# Manifold Examples: Matrix Groups

- **General Linear Group:** $GL(n)$
  - Space of nonsingular $n \times n$ matrices
  - Open set of $\mathbb{R}^{n \times n}$

# Manifold Examples: Matrix Groups

► **General Linear Group:** $GL(n)$
  ► Space of nonsingular $n \times n$ matrices
  ► Open set of $\mathbb{R}^{n \times n}$
► **Special Orthogonal Group:** $SO(n)$
  ► Rotations of $\mathbb{R}^n$
  ► All matrices $R \in GL(n)$ such that $RR^T = I$ and $\det(R) = 1$

# Manifold Examples: Positive-Definite Tensors



$A \in \mathrm{PD}(2)$ is of the form

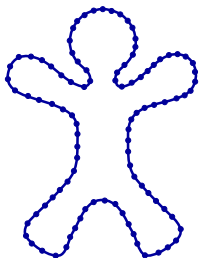$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

$$ac - b^2 > 0, \quad a > 0.$$

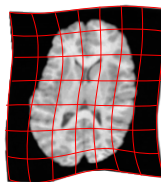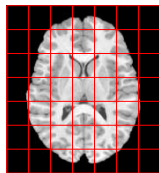Similar situation for $\mathrm{PD}(3)$ (6-dimensional) or $\mathrm{PD}(d)$.
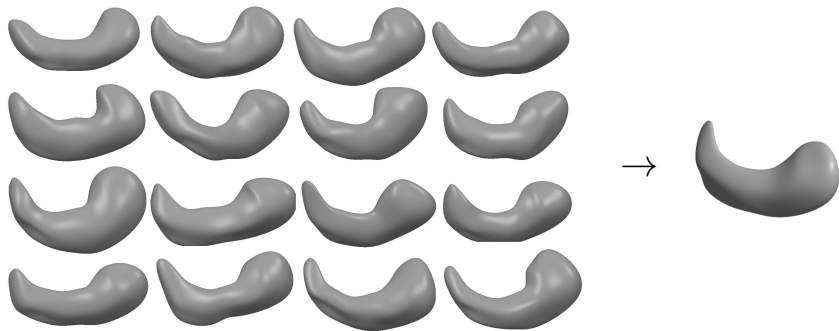
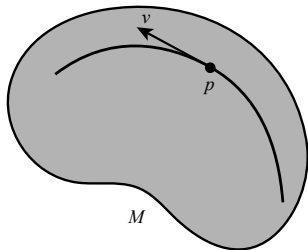# Manifold Examples: Shape Spaces

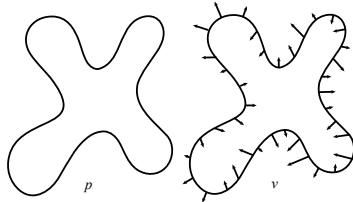**Kendall's Shape Space**

**Space of Diffeomorphisms**

# Shape Statistics: Averages

# Review: Tangent Spaces



Infinitesimal change in shape:

A **tangent vector** is the velocity of a curve on $M$.

# Review: Riemannian Metrics

A **Riemannian metric** is a smoothly varying inner product on the tangent spaces, denoted $\langle v, w \rangle_p$ for $v, w \in T_p M$.

This metric now gives us the **norm** of a tangent vector:

$$\|v\|_p = \sqrt{\langle v, v \rangle_p}.$$
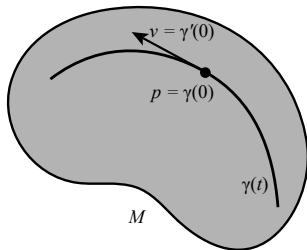
# Review: Geodesics

A **geodesic** is a curve $\gamma \in M$ that locally minimizes

$$E(\gamma) = \int_0^1 \|\gamma'(t)\|^2 dt.$$

Turns out it also locally minimizes arc-length,

$$L(\gamma) = \int_0^1 \|\gamma'(t)\| dt.$$

# Review: Directional Derivatives



Tangent vector $v \in T_p M$ is directional derivative operator on smooth functions, $f : M \to \mathbb{R}$

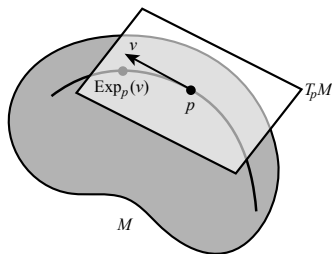Notation: $vf(p) = \frac{d}{dt} f(\gamma(t))|_{t=0}$

# Gradient on a Manifold

Given $f : M \to \mathbb{R}$, define gradient $\mathrm{grad} f(p)$ as vector in $T_p M$ such that:

$$vf(p) = \langle \mathrm{grad} f(p), v \rangle,$$

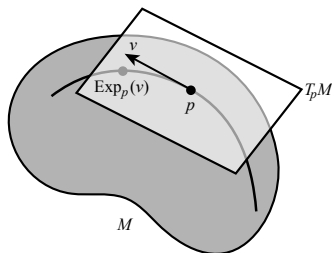for all tangent vectors $v \in T_p M$.

# The Exponential Map



Notation: $\mathrm{Exp}_p(v)$

▶ $p \in M$: starting point

▶ $v \in T_pM$: initial velocity

▶ Output: endpoint of geodesic segment, starting at $p$, with velocity $v$, with same length as $\|v\|$

# The Log Map



Notation: $\mathrm{Log}_p(q)$

- ▶ Inverse of $\mathrm{Exp}$
- ▶ $p, q$: two points in $M$
- ▶ Output: tangent vector at $p$, such that $\mathrm{Exp}_p(\mathrm{Log}_p(q)) = q$
- ▶ Gives distance between points: $d(p, q) = \|\mathrm{Log}_p(q)\|$.

# Intrinsic Means (Fréchet)

The *intrinsic mean* of a collection of points $x_1, \ldots, x_N$ in a metric space $M$ is

$$\mu = \arg \min_{x \in M} \sum_{i=1}^{N} d(x, x_i)^2,$$

where $d(\cdot, \cdot)$ denotes distance in $M$.
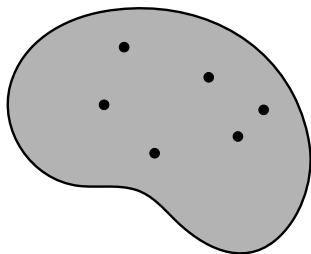
# Gradient of the Geodesic Distance

The gradient of the Riemannian distance function is

$$\mathrm{grad}_x d(x, y)^2 = -2 \, \mathrm{Log}_x(y).$$

So, gradient of the sum-of-squared distance function is

$$\mathrm{grad}_x \sum_{i=1}^{N} d(x, x_i)^2 = -2 \sum_{i=1}^{N} \mathrm{Log}_x(x_i).$$

# Computing Means



**Gradient Descent Algorithm:**

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**
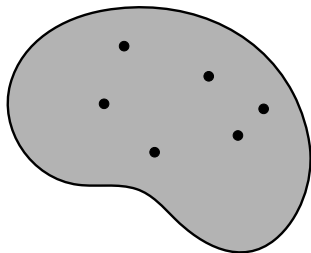
Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**
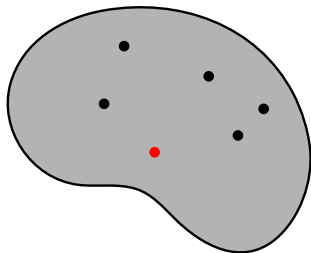
Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**

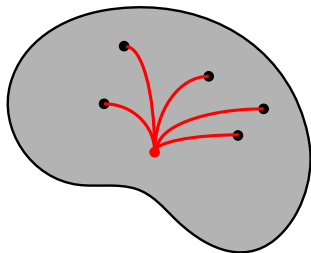Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**

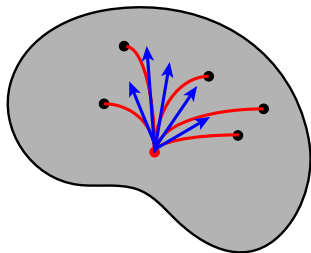Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**

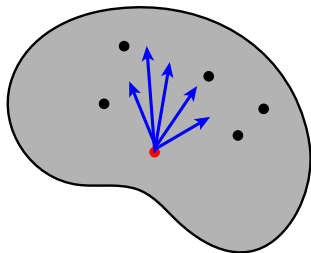Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**

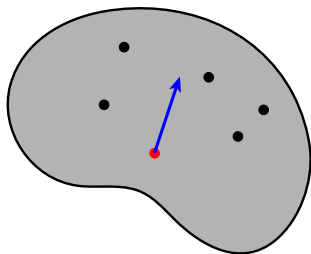Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**

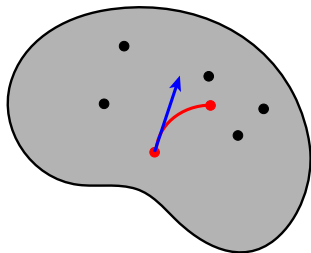Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Computing Means



**Gradient Descent Algorithm:**

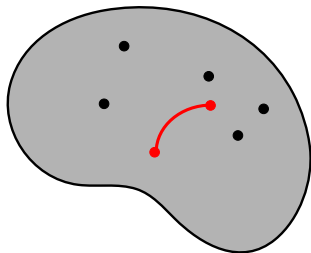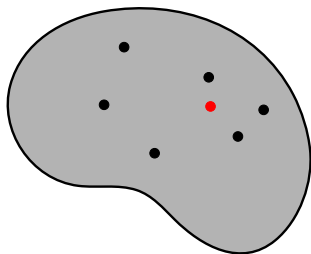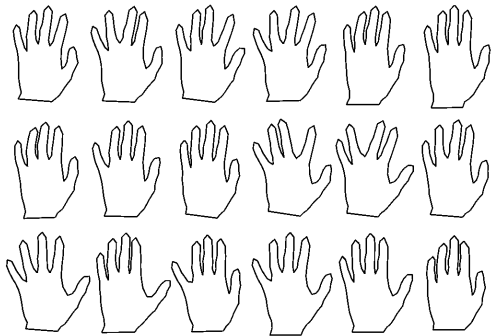Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N}\sum_{i=1}^{N}\mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$
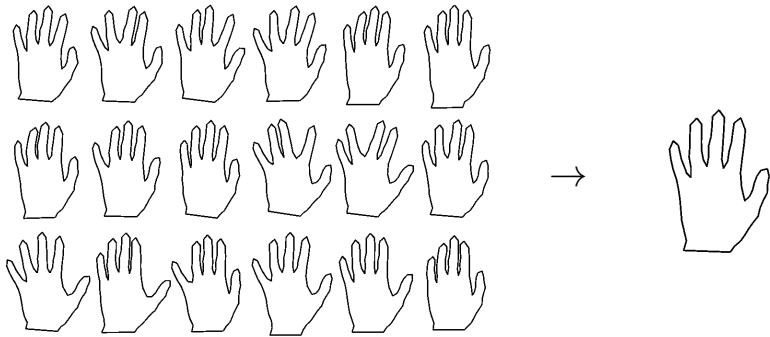
$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

# Example of Mean on Kendall Shape Space



Hand data from Tim Cootes

# Example of Mean on Kendall Shape Space



Hand data from Tim Cootes

# Least Squares and Maximum Likelihood

**Geometric:** Least squares

$$\min_{\text{model}} \sum_{i=1}^{N} d(\text{model}, y_i)^2$$

**Probabilistic:** Maximum likelihood

$$\max_{\text{model}} \prod_{i=1}^{N} p(y_i; \text{model})$$

How about this "Gaussian" likelihood?

$$p(y_i; \text{model}) \propto \exp\left(-\tau d(\text{model}, y_i)^2\right)$$

# A Riemannian Normal Distribution

For a simple model with Fréchet mean:

$$p(y; \mu, \tau) = \frac{1}{C(\mu, \tau)} \exp\left(-\tau d(\mu, y)^2\right)$$

Notation: $y \sim N_M(\mu, \tau^{-1})$

**Problem:** Normalizing constant may depend on $\mu$:

$$\ln p(y; \mu, \tau) = -\ln C(\mu, \tau) - \tau d(\mu, y)^2$$

Note: not a problem in $\mathbb{R}^d$ because $C(\mu, \tau) \propto \tau^{-d/2}$.

# Riemannian Homogeneous Spaces

**Definition:** A Riemannian manifold $M$ is called a **Riemannian homogeneous space** if its isometry group $G$ acts transitively.

**Theorem:** If $M$ is a homogeneous space, the normalizing constant for a normal distribution on $M$ does not depend on $\mu$.

Fletcher, IJCV 2013

# Examples of Homogeneous Spaces

▶ **Constant curvature spaces:** Euclidean spaces, spheres, hyperbolic spaces

▶ **Lie groups:** $SO(n)$ (rotations), $SE(n)$ (rigid transforms), $GL(n)$ (non-singular matrices), $Aff(n)$ (affine transforms), etc.

▶ **Stiefel manifolds:** space of orthonormal $k$-frames in $\mathbb{R}^n$

▶ **Grassmann manifolds:** space of $k$-dimensional subspaces in $\mathbb{R}^n$

▶ **Positive-definite symmetric matrices**