UNIVERSITY OF COPENHAGEN

Welcome to Copenhagen!

Schedule:

	Monday	Tuesday	Wednesday	Thursday	Friday
8	Registration and welcome				
9	Crash course on Differential and Riemannian Geometry 1.1 (Feragen)	Crash course on Differential and Riemannian Geometry 3 (Lauze)	Introduction to Information Geometry 3.1 (Amari)	Information Geometry & Stochastic Optimization 1.1 (Hansen)	Information Geometry & Stochastic Optimization in Discrete Domains 1.1 (Màlago)
10	Crash course on Differential and Riemannian Geometry 1.2 (Feragen)	Tutorial on numerics for Riemannian geometry 1.1 (Sommer)	Introduction to Information Geometry 3.2 (Amari)	Information Geometry & Stochastic Optimization 1.2 (Hansen)	Information Geometry & Stochastic Optimization in Discrete Domains 1.2 (Màlago)
11	Crash course on Differential and Riemannian Geometry 1.3 (Feragen)	Tutorial on numerics for Riemannian geometry 1.2 (Sommer)	Introduction to Information Geometry 3.3 (Amari)	Information Geometry & Stochastic Optimization 1.3 (Hansen)	Information Geometry & Stochastic Optimization in Discrete Domains 1.3 (Màlago)
12	Lunch	Lunch	Lunch	Lunch	Lunch
13	Crash course on Differential and Riemannian Geometry 2.1 (Lauze)	Introduction to Information Geometry 2.1 (Amari)	Information Geometry & Reinforcement Learning 1.1 (Peters)	Information Geometry & Stochastic Optimization 1.4 (Hansen)	Information Geometry & Cognitive Systems 1.1 (Ay)
14	Crash course on Differential and Riemannian Geometry 2.2 (Lauze)	Introduction to Information Geometry 2.2 (Amari)	Information Geometry & Reinforcement Learning 1.2 (Peters)	Information Geometry & Stochastic Optimization 1.5 (Hansen)	Information Geometry & Cognitive Systems 1.2 (Ay)
15	Introduction to Information Geometry 1.1 (Amari)	Introduction to Information Geometry 2.3 (Amari)	Information Geometry & Reinforcement Learning 1.3 (Peters)	Stochastic Optimization in Practice 1.1 (Hansen)	Information Geometry & Cognitive Systems 1.3 (Ay)
16	Introduction to Information Geometry 1.2 (Amari)	Introduction to Information Geometry 2.4 (Amari)	Social activity/ Networking event	Stochastic Optimization in Practice 1.2 (Hansen)	Information Geometry & Cognitive Systems 1.3 (Ay)

Coffee breaks at 10:00 and 14:45 (no afternoon break on Wednesday)

UNIVERSITY OF COPENHAGEN

Welcome to Copenhagen!

Social Programme!

- **Today:** Pizza and walking tour!
  - 17:15 Pizza dinner in lecture hall
  - 18:00 Departure from lecture hall (with Metro – we have tickets)
  - 19:00 Walking tour of old university



- **Wednesday:** Boat tour, Danish beer and dinner
  - 15:20 Bus from KUA to Nyhavn
  - 16:00-17:00 Boat tour
  - 17:20 Bus from Nyhavn to NÃ¸rrebro bryghus (NB, brewery)
  - 18:00 Guided tour of NB
  - 19:00 Dinner at NB

UNIVERSITY OF COPENHAGEN

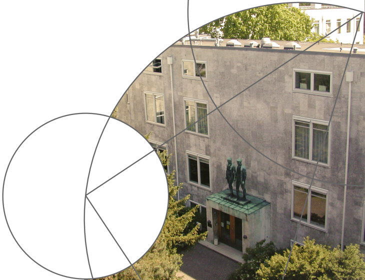
Welcome to Copenhagen!

- Lunch on your own – canteens and coffee on campus
- Internet connection
  - Eduroam
  - Alternative will be set up ASAP
- Emergency? Call Aasa: +4526220498
- Questions?

 UNIVERSITY OF COPENHAGEN

A Very Brief Introduction to Differential and Riemannian Geometry

Aasa Feragen and François Lauze
Department of Computer Science
University of Copenhagen



Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

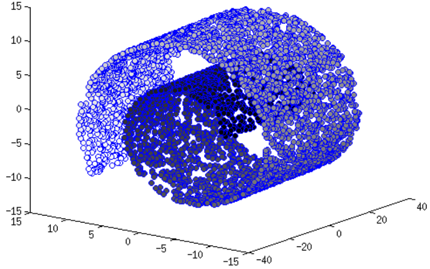
UNIVERSITY OF COPENHAGEN

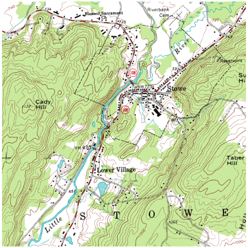
Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

 UNIVERSITY OF COPENHAGEN

Why do we care about nonlinearity?

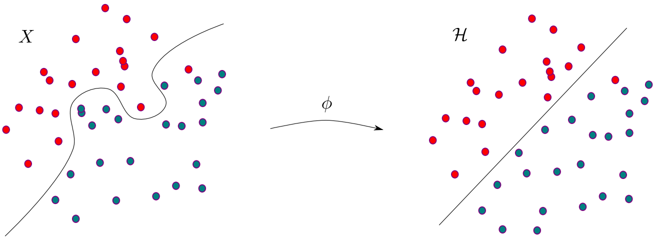




- Nonlinear relations between data objects
- True distances not reflected by linear representation

Mildly nonlinear: Nonlinear transformations between different linear representations

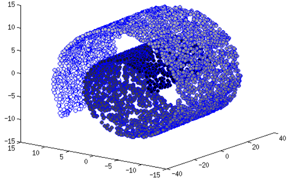
- Kernels!
- Feature map = nonlinear transformation of (linear?) data space X into linear feature space $\mathcal{H}$
- Learning problem is (usually) linear in $\mathcal{H}$, not in X.

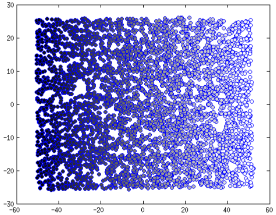


UNIVERSITY OF COPENHAGEN

Mildly nonlinear: Nonlinearly embedded subspaces whose intrinsic metric is linear

- Manifold learning!
  - Find intrinsic dataset distances
  - Find an $\mathbb{R}^d$ embedding that minimally distorts those distances

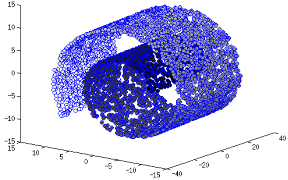


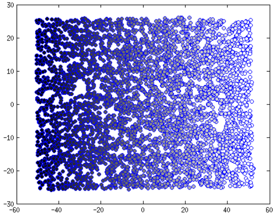


 UNIVERSITY OF COPENHAGEN

Mildly nonlinear: Nonlinearly embedded subspaces whose intrinsic metric is linear

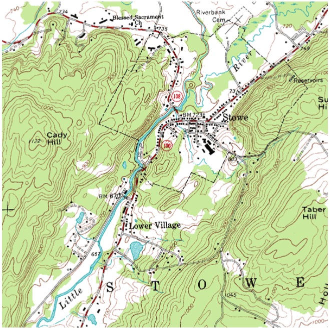
- Manifold learning!
  - Find intrinsic dataset distances
  - Find an $\mathbb{R}^d$ embedding that minimally distorts those distances
- Searches for the folded-up *Euclidean* space that best fits the data
  - the *embedding* of the data in feature space is *nonlinear*
  - the recovered intrinsic distance structure is *linear*

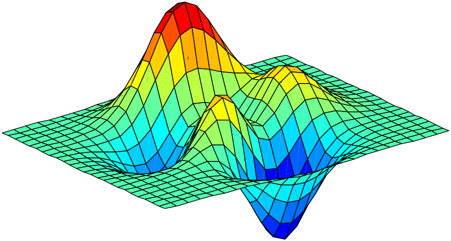




 UNIVERSITY OF COPENHAGEN

More nonlinear: Data spaces which are intrinsically nonlinear





- Distances distorted in nonlinear way, varying spatially
- We shall see: the distances cannot always be linearized

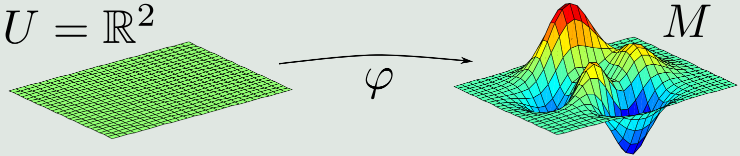
"Topographic map example". Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Topographic_map_example.png#mediaviewer/File:Topographic_map_example.png

 UNIVERSITY OF COPENHAGEN

Intrinsically nonlinear data spaces: Smooth manifolds

Definition

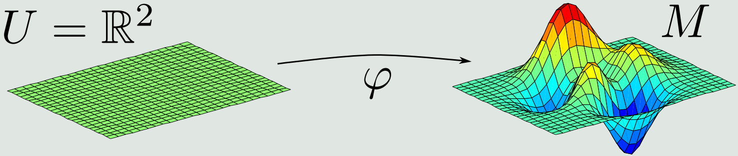
A *manifold* is a set M with an associated one-to-one map $\varphi\colon U \to M$ from an open subset $U \subset \mathbb{R}^m$ called a *global chart* or a *global coordinate system* for M.



Intrinsically nonlinear data spaces: Smooth manifolds

Definition

A *manifold* is a set M with an associated one-to-one map $\varphi\colon U \to M$ from an open subset $U \subset \mathbb{R}^m$ called a *global chart* or a *global coordinate system* for M.

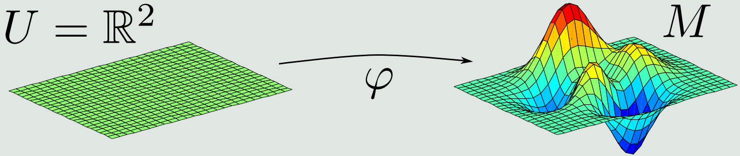


- Open set $U \subset \mathbb{R}^m$ = set that does not contain its boundary
- Manifold M gets its *topology (= definition of open sets)* from U via φ

Intrinsically nonlinear data spaces: Smooth manifolds

Definition

A *manifold* is a set M with an associated one-to-one map $\varphi\colon U \to M$ from an open subset $U \subset \mathbb{R}^m$ called a *global chart* or a *global coordinate system* for M.



- Open set $U \subset \mathbb{R}^m$ = set that does not contain its boundary
- Manifold M gets its *topology (= definition of open sets)* from U via φ
- What are the implications of getting the topology from U?

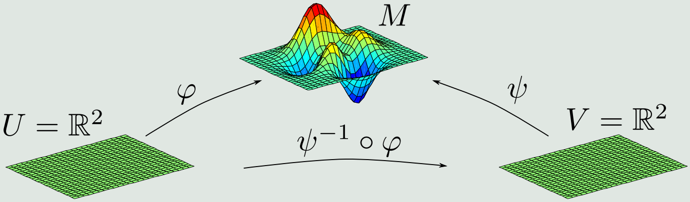
UNIVERSITY OF COPENHAGEN

Intrinsically nonlinear data spaces: Smooth manifolds

Definition

A *smooth* manifold is a pair $(M, \mathcal{A})$ where

- M is a set
- $\mathcal{A}$ is a family of one-to-one global charts $\varphi\colon U \to M$ from some open subset $U = U_\varphi \subset \mathbb{R}^m$ for M,
- for any two charts $\varphi\colon U \to \mathbb{R}^m$ and $\psi\colon V \to \mathbb{R}^m$ in $\mathcal{A}$, their corresponding *change of variables* is a smooth diffeomorphism $\psi^{-1} \circ \varphi\colon U \to V \subset \mathbb{R}^m$.



UNIVERSITY OF COPENHAGEN

Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

UNIVERSITY OF COPENHAGEN

Differentiable and smooth functions

- $f : U$ open $\subset \mathbb{R}^n \to \mathbb{R}^q$ continuous: write

$$(y_1, \dots, y_q) = f(x_1, \dots, x_n)$$

 UNIVERSITY OF COPENHAGEN

Differentiable and smooth functions

- $f : U$ open $\subset \mathbb{R}^n \to \mathbb{R}^q$ continuous: write

$$(y_1, \dots, y_q) = f(x_1, \dots, x_n)$$

- f is of class $\mathcal{C}^r$ if f has continuous partial derivatives

$$\frac{\partial^{r_1+\cdots+r_n} y_k}{\partial x_1^{r_1} \dots \partial x_n^{r_n}}$$

$k = 1 \dots q$, $r_1 + \dots r_n \leq r$.

Differentiable and smooth functions

- $f : U$ open $\subset \mathbb{R}^n \to \mathbb{R}^q$ continuous: write

$$(y_1, \ldots, y_q) = f(x_1, \ldots, x_n)$$

- f is of class $\mathcal{C}^r$ if f has continuous partial derivatives

$$\frac{\partial^{r_1+\cdots+r_n} y_k}{\partial x_1^{r_1} \ldots \partial x_n^{r_n}}$$

  $k = 1 \ldots q$, $r_1 + \ldots r_n \leq r$.
- When $r = \infty$, f is smooth. Our focus.

Differential, Jacobian Matrix

- Differential of f in $\mathbf{x}$: unique linear map (if exists) $d_x f : \mathbb{R}^n \to \mathbb{R}^q$ s.t.

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + d_x f(\mathbf{h}) + o(\mathbf{h}).$$

Differential, Jacobian Matrix

- Differential of f in $\mathbf{x}$: unique linear map (if exists) $d_x f : \mathbb{R}^n \to \mathbb{R}^q$ s.t.

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + d_x f(\mathbf{h}) + o(\mathbf{h}).$$

- Jacobian matrix of f: matrix $q \times n$ of partial derivatives of f:

$$J_{\mathbf{x}} f = \begin{pmatrix} \frac{\partial y_1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial y_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial y_q}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial y_q}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

Differential, Jacobian Matrix

- Differential of f in $\mathbf{x}$: unique linear map (if exists) $d_x f : \mathbb{R}^n \to \mathbb{R}^q$ s.t.

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + d_x f(\mathbf{h}) + o(\mathbf{h}).$$

- Jacobian matrix of f: matrix $q \times n$ of partial derivatives of f:

$$J_{\mathbf{x}} f = \begin{pmatrix} \frac{\partial y_1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial y_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial y_q}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial y_q}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

- What is the meaning of the Jacobian? The differential? How do they differ?

Diffeomorphism

- **When $n = q$:**
  - If f is 1-1, f and f^{-1} both $\mathcal{C}^r$
  - $\rightsquigarrow$ f is a $\mathcal{C}^r$-diffeomorphism.
  - Smooth diffeomorphisms are simply referred to as a diffeomorphisms.

Diffeomorphism

- **When $n = q$:**
  - If f is 1-1, f and f^{-1} both $\mathcal{C}^r$
  - $\rightsquigarrow$ f is a $\mathcal{C}^r$-diffeomorphism.
  - Smooth diffeomorphisms are simply referred to as a diffeomorphisms.
- Inverse Function Theorem:
  - f diffeomorphism $\Rightarrow \det(J_{\mathbf{x}} f) \neq 0$.
  - $\det(J_{\mathbf{x}} f) \neq 0 \Rightarrow f$ local diffeomorphism in a neighborhood of $\mathbf{x}$.

Diffeomorphism

- **When $n = q$:**
  - If f is 1-1, f and f^{-1} both $\mathcal{C}^r$
  - $\rightsquigarrow$ f is a $\mathcal{C}^r$-diffeomorphism.
  - Smooth diffeomorphisms are simply referred to as a diffeomorphisms.
- Inverse Function Theorem:
  - f diffeomorphism $\Rightarrow \det(J_{\mathbf{x}} f) \neq 0$.
  - $\det(J_{\mathbf{x}} f) \neq 0 \Rightarrow f$ local diffeomorphism in a neighborhood of $\mathbf{x}$.
- What is the meaning of $J_{\mathbf{x}} f$? Of $\det(J_{\mathbf{x}} f) \neq 0$?

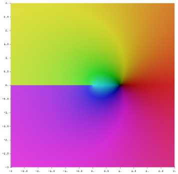
 UNIVERSITY OF COPENHAGEN

Diffeomorphism

- f may be a local diffeomorphism everywhere but fail to be a global diffeomorphism. Examples:
  - Complex exponential:

$$f : \mathbb{R}^2 \backslash 0 \to \mathbb{R}^2, \quad (x, y) \to (e^x \cos(y), e^x \sin(y)).$$

Recall its inverse (the complex log) has infinitely many branches.



"Complex log" by Jan Homann; Color encoding image comment author Hal Lane, September 28, 2009 - Own work. This mathematical image was created with Mathematica. Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Complex_log.jpg#mediaviewer/File:Complex_log.jpg

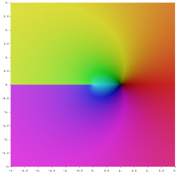
 UNIVERSITY OF COPENHAGEN

Diffeomorphism

- f may be a local diffeomorphism everywhere but fail to be a global diffeomorphism. Examples:
  - Complex exponential:

$$f : \mathbb{R}^2 \backslash 0 \to \mathbb{R}^2, \quad (x, y) \to (e^x \cos(y), e^x \sin(y)).$$

Recall its inverse (the complex log) has infinitely many branches.



- If f is 1-1 and a local diffeomorphism everywhere, it is a global diffeomorphism.

"Complex log" by Jan Homann; Color encoding image comment author Hal Lane, September 28, 2009 - Own work. This mathematical image was created with Mathematica. Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Complex_log.jpg#mediaviewer/File:Complex_log.jpg

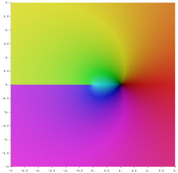
 UNIVERSITY OF COPENHAGEN

Diffeomorphism

- f may be a local diffeomorphism everywhere but fail to be a global diffeomorphism. Examples:
  - Complex exponential:

$$f : \mathbb{R}^2 \backslash 0 \to \mathbb{R}^2, \quad (x, y) \to (e^x \cos(y), e^x \sin(y)).$$

Recall its inverse (the complex log) has infinitely many branches.



- If f is 1-1 and a local diffeomorphism everywhere, it is a global diffeomorphism.
- What is the intuitive meaning of a diffeomorphism?

"Complex log" by Jan Homann; Color encoding image comment author Hal Lane, September 28, 2009 - Own work. This mathematical image was created with Mathematica. Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Complex_log.jpg#mediaviewer/File:Complex_log.jpg

Outline

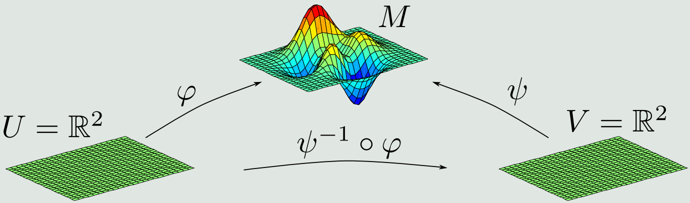
1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

Back to smooth manifolds

Definition

A *smooth* manifold is a pair $(M, \mathcal{A})$ where

- M is a set
- $\mathcal{A}$ is a family of one-to-one global charts $\varphi\colon U \to M$ from some open subset $U = U_\varphi \subset \mathbb{R}^m$ for M,
- for any two charts $\varphi\colon U \to \mathbb{R}^m$ and $\psi\colon V \to \mathbb{R}^m$, their corresponding *change of variables* is a smooth diffeomorphism $\psi^{-1} \circ \varphi\colon U \to V \subset \mathbb{R}^m$.

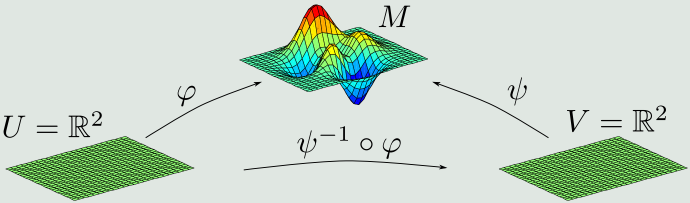


Back to smooth manifolds

Definition

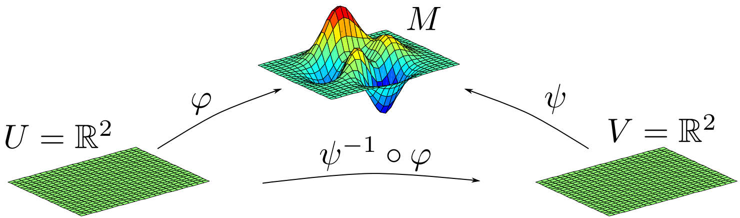
A *smooth* manifold is a pair $(M, \mathcal{A})$ where

- M is a set
- $\mathcal{A}$ is a family of one-to-one global charts $\varphi \colon U \to M$ from some open subset $U = U_\varphi \subset \mathbb{R}^m$ for M,
- for any two charts $\varphi \colon U \to \mathbb{R}^m$ and $\psi \colon V \to \mathbb{R}^m$, their corresponding *change of variables* is a smooth diffeomorphism $\psi^{-1} \circ \varphi \colon U \to V \subset \mathbb{R}^m$.



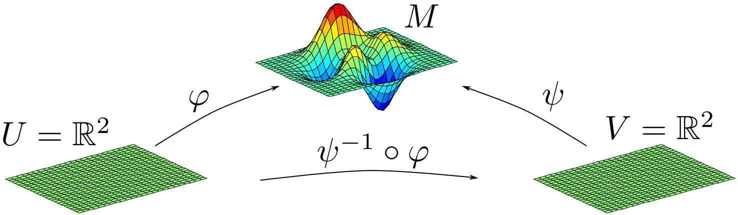
- What are the implications of inheriting structure through $\mathcal{A}$?

Back to smooth manifolds



- φ and ψ are *parametrizations* of *M*

Back to smooth manifolds



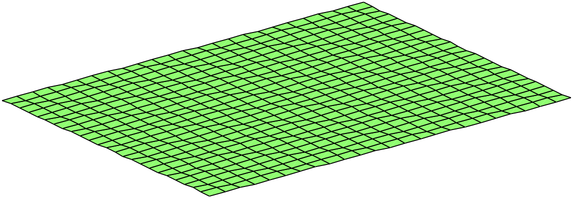
- φ and ψ are *parametrizations* of M
- Set $\varphi_j(P) = (y^1(P), \ldots, y^n(P))$, then

$$\varphi_j \circ \varphi_i^{-1}(x_1, \ldots, x_m) = (y_1, \ldots, y_m)$$

and the $m \times m$ Jacobian matrices $\left(\frac{\partial y^k}{\partial x^h}\right)_{k,h}$ are invertible.

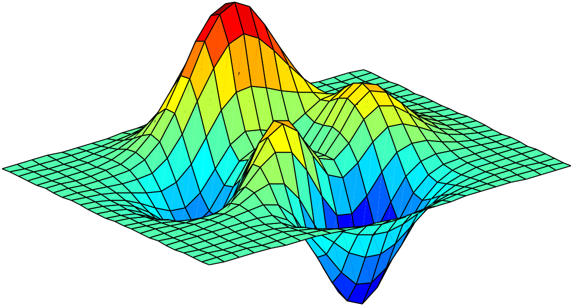
Example: Euclidean space

- The Euclidean space $\mathbb{R}^n$ is a manifold: take $\varphi = \mathit{Id}$ as global coordinate system!



Example: Smooth surfaces

- Smooth surfaces in $\mathbb{R}^n$ that are the image of a smooth map $f\colon \mathbb{R}^2 \to \mathbb{R}^n$.
- A global coordinate system given by f



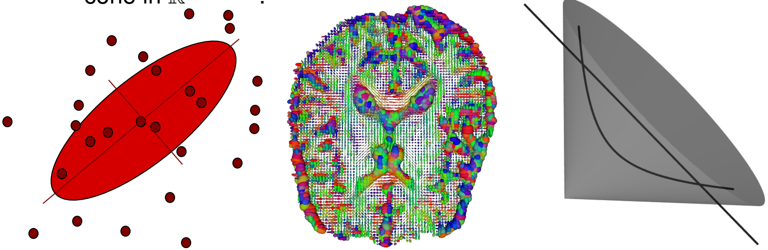
 UNIVERSITY OF COPENHAGEN

Example: Symmetric Positive Definite Matrices

- $\mathcal{P}(n) \subset GL_n$ consists of all symmetric $n \times n$ matrices A that satisfy

$$xAx^T > 0 \text{ for any } x \in \mathbb{R}^n, \qquad \text{(positive definite – PD – matrices)}$$

  - $\mathcal{P}(n) =$ the set of covariance matrices on $\mathbb{R}^n$
  - $\mathcal{P}(3) =$ the set of (diffusion) tensors on $\mathbb{R}^3$
- **Global chart:** $\mathcal{P}(n)$ is an open, convex subset of $\mathbb{R}^{(n^2+n)/2}$
  - $A, B \in \mathcal{P}(n) \rightarrow aA + bB \in \mathcal{P}(n)$ for all $a, b > 0$ so $\mathcal{P}(n)$ is a convex cone in $\mathbb{R}^{(n^2+n)/2}$.

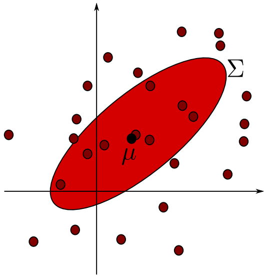


Middle figure from Fillard et al., A Riemannian Framework for the Processing of Tensor-Valued Images, LNCS 3753, 2005, pp 112-123. Rightmost figure from Fletcher, Joshi, Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors, CVAMIA04

UNIVERSITY OF COPENHAGEN

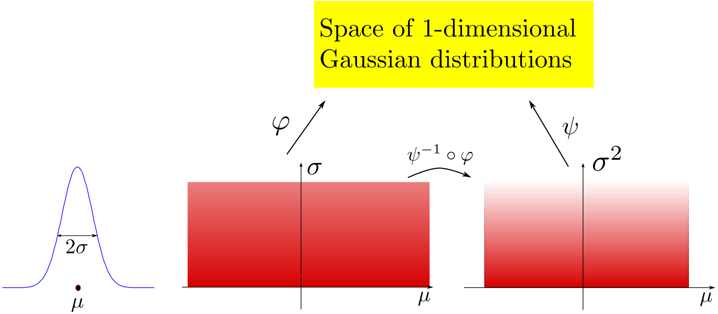
Example: Space of Gaussian distributions

- The space of n-dimensional Gaussian distributions is a smooth manifold
- **Global chart:** $(\mu, \Sigma) \in \mathbb{R}^n \times \mathcal{P}(n)$.



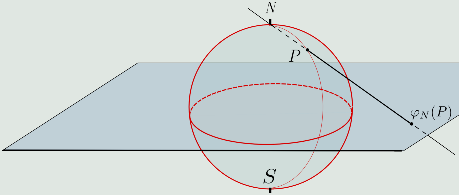
Example: Space of 1-dimensional Gaussian distributions

- The space of 1-dimensional Gaussian distributions is parametrized by $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$, mean μ, standard deviation σ
- Also parametrized by $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$, mean μ, variance σ^2
- Smooth reparametrization $\psi^{-1} \circ \varphi$



In general: Manifolds requiring multiple charts

The sphere $\mathcal{S}^2 = \{(x, y, z), x^2 + y^2 + z^2 = 1\}$



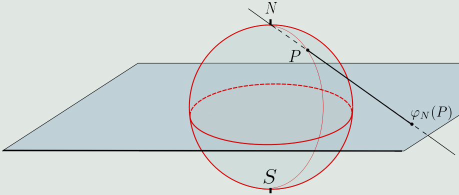
For instance the projection from North Pole, given, for a point $P = (x, y, z) \neq N$ of the sphere, by

$$\varphi_N(P) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$$

is a (large) local coordinate system (around the south pole).

In general: Manifolds requiring multiple charts

The sphere $\mathcal{S}^2 = \{(x, y, z), x^2 + y^2 + z^2 = 1\}$



For instance the projection from North Pole, given, for a point $P = (x, y, z) \neq N$ of the sphere, by

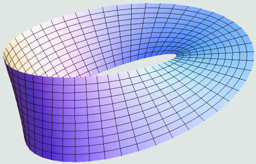
$$\varphi_N(P) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$$

is a (large) local coordinate system (around the south pole).

In these cases, we also require the charts to overlap ”nicely”

In general: Manifolds requiring multiple charts

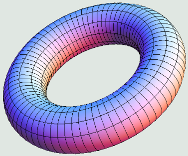
The Moebius strip



$$u \in [0, 2\pi], v \in [\frac{1}{2}, \frac{1}{2}]$$

$$\begin{pmatrix} \cos(u)\left(1 + \frac{1}{2} v \cos(\frac{u}{2})\right) \\ \sin(u)\left(1 + \frac{1}{2} v \cos(\frac{u}{2})\right) \\ \frac{1}{2} v \sin(\frac{u}{2}) \end{pmatrix}$$

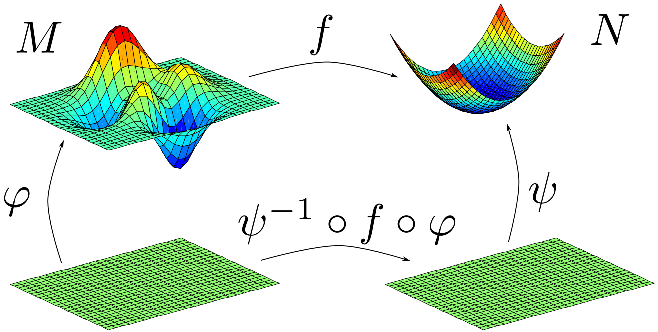
The 2D-torus



$$(u, v) \in [0, 2\pi]^2, R \gg r > 0$$

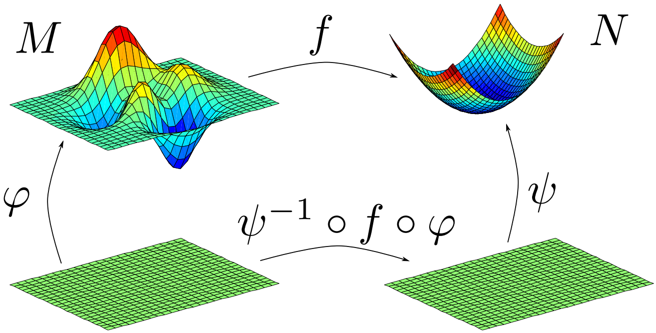
$$\begin{pmatrix} \cos(u)\left(R + r\cos(v)\right) \\ \sin(u)\left(R + r\cos(v)\right) \\ r\sin(v) \end{pmatrix}$$

Smooth maps between manifolds



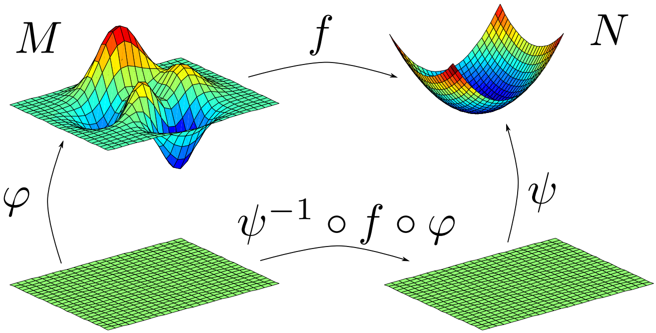
- $f : M \to N$ is smooth if its expression in any global coordinates for M and N is.

Smooth maps between manifolds



- $f\colon M \to N$ is smooth if its expression in any global coordinates for M and N is.
- φ global coordinates for M, ψ global coordinates for N

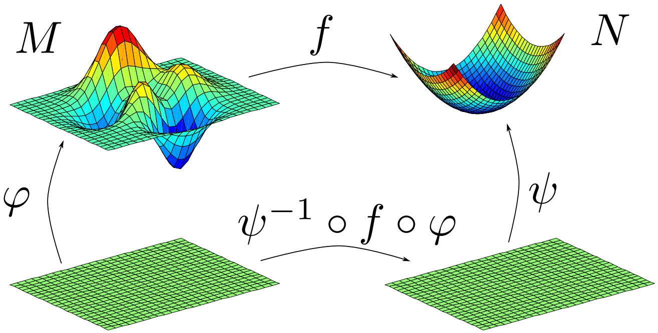
Smooth maps between manifolds



- $f\colon M \to N$ is smooth if its expression in any global coordinates for M and N is.
- φ global coordinates for M, ψ global coordinates for N

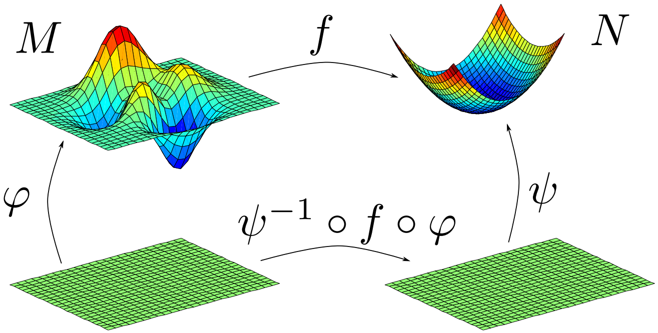
$$\varphi^{-1} \circ f \circ \psi \text{ smooth.}$$

Smooth diffeomorphism between manifolds



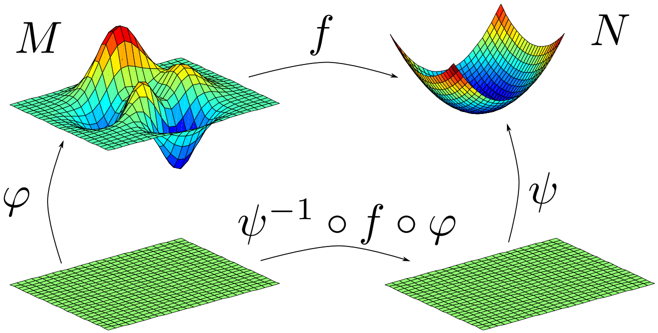
- $f\colon M \to N$ is a smooth diffeomorphism if its expression in any global coordinates for M and N is.

Smooth diffeomorphism between manifolds



- $f\colon M \to N$ is a smooth diffeomorphism if its expression in any global coordinates for M and N is.
- φ global coordinates for M, ψ global coordinates for N

Smooth diffeomorphism between manifolds



- $f\colon M \to N$ is a smooth diffeomorphism if its expression in any global coordinates for M and N is.
- φ global coordinates for M, ψ global coordinates for N

$$\varphi^{-1} \circ f \circ \psi \text{ smooth diffeomorphism .}$$

Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

UNIVERSITY OF COPENHAGEN

Submanifolds of $\mathbb{R}^N$

- Take $f : U \in \mathbb{R}^m \to \mathbb{R}^n$, $n \leq m$ smooth.

UNIVERSITY OF COPENHAGEN

Submanifolds of $\mathbb{R}^N$

- Take $f : U \in \mathbb{R}^m \to \mathbb{R}^n$, $n \leq m$ smooth.
- Set $M = f^{-1}(0)$.

UNIVERSITY OF COPENHAGEN

Submanifolds of $\mathbb{R}^N$

- Take $f : U \in \mathbb{R}^m \to \mathbb{R}^n$, $n \leq m$ smooth.
- Set $M = f^{-1}(0)$.
- If for all $\mathbf{x} \in M$, f is a submersion at $\mathbf{x}$ ($d_{\mathbf{x}} f$ has full rank), M is a manifold of dimension $m - n$.

 UNIVERSITY OF COPENHAGEN

Submanifolds of $\mathbb{R}^N$

- Take $f : U \in \mathbb{R}^m \to \mathbb{R}^n$, $n \leq m$ smooth.
- Set $M = f^{-1}(0)$.
- If for all $\mathbf{x} \in M$, f is a submersion at $\mathbf{x}$ ($d_{\mathbf{x}} f$ has full rank), M is a manifold of dimension $m - n$.
- Example:

$$f(x_1, \ldots, x_m) = 1 - \sum_{i=1}^{m} x_i^2 :$$

$f^{-1}(0)$ is the $(m-1)$-dimensional unit sphere $\mathcal{S}^{m-1}$.

 UNIVERSITY OF COPENHAGEN

Submanifolds of $\mathbb{R}^N$

- Take $f : U \in \mathbb{R}^m \to \mathbb{R}^n$, $n \leq m$ smooth.
- Set $M = f^{-1}(0)$.
- If for all $\mathbf{x} \in M$, f is a submersion at $\mathbf{x}$ ($d_\mathbf{x} f$ has full rank), M is a manifold of dimension $m - n$.
- Example:

$$f(x_1, \ldots, x_m) = 1 - \sum_{i=1}^{m} x_i^2 :$$

  $f^{-1}(0)$ is the $(m-1)$-dimensional unit sphere $\mathcal{S}^{m-1}$.
- The graph $\Gamma = (x, f(x)) \in \mathbb{R}^m \times \mathbb{R}^n$ is smooth for any smooth map $f \colon \mathbb{R}^m \to \mathbb{R}^n$.

UNIVERSITY OF COPENHAGEN

Submanifolds of $\mathbb{R}^N$

- Take $f : U \in \mathbb{R}^m \to \mathbb{R}^n$, $n \leq m$ smooth.
- Set $M = f^{-1}(0)$.
- If for all $\mathbf{x} \in M$, f is a submersion at $\mathbf{x}$ ($d_\mathbf{x} f$ has full rank), M is a manifold of dimension $m - n$.
- Example:

$$f(x_1, \ldots, x_m) = 1 - \sum_{i=1}^{m} x_i^2 :$$

  $f^{-1}(0)$ is the $(m-1)$-dimensional unit sphere $\mathcal{S}^{m-1}$.
- The graph $\Gamma = (x, f(x)) \in \mathbb{R}^m \times \mathbb{R}^n$ is smooth for any smooth map $f\colon \mathbb{R}^m \to \mathbb{R}^n$.
  - $\Gamma = F^(0)$ for $F\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $F(x, y) = y - f(x)$.

 UNIVERSITY OF COPENHAGEN

Submanifolds of $\mathbb{R}^N$

- Take $f : U \in \mathbb{R}^m \to \mathbb{R}^n$, $n \leq m$ smooth.
- Set $M = f^{-1}(0)$.
- If for all $\mathbf{x} \in M$, f is a submersion at $\mathbf{x}$ ($d_{\mathbf{x}} f$ has full rank), M is a manifold of dimension $m - n$.
- Example:

$$f(x_1, \ldots, x_m) = 1 - \sum_{i=1}^{m} x_i^2 :$$

  $f^{-1}(0)$ is the $(m-1)$-dimensional unit sphere $\mathcal{S}^{m-1}$.
- The graph $\Gamma = (x, f(x)) \in \mathbb{R}^m \times \mathbb{R}^n$ is smooth for any smooth map $f \colon \mathbb{R}^m \to \mathbb{R}^n$.
  - $\Gamma = F^(0)$ for $F \colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $F(x, y) = y - f(x)$.
- Many common examples of manifolds in practice are of that type.

UNIVERSITY OF COPENHAGEN

Product Manifolds

- M and N manifolds, so is $M \times N$.

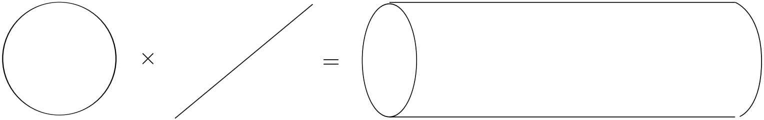
Product Manifolds

- M and N manifolds, so is $M \times N$.
- Just consider the products of charts of M and N!

UNIVERSITY OF COPENHAGEN

Product Manifolds

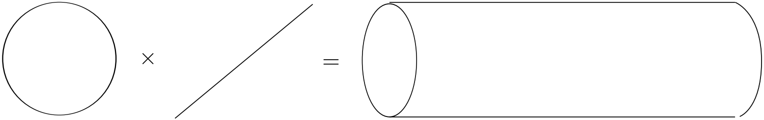
- *M* and *N* manifolds, so is $M \times N$.
- Just consider the products of charts of *M* and *N*!
- Example: $M = \mathcal{S}^1$, $N = \mathbb{R}$: the cylinder



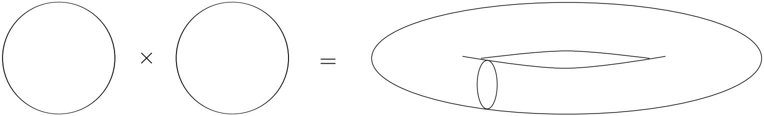
 UNIVERSITY OF COPENHAGEN

Product Manifolds

- M and N manifolds, so is $M \times N$.
- Just consider the products of charts of M and N!
- Example: $M = \mathcal{S}^1$, $N = \mathbb{R}$: the cylinder



- Example: $M = N = \mathcal{S}^1$: the torus



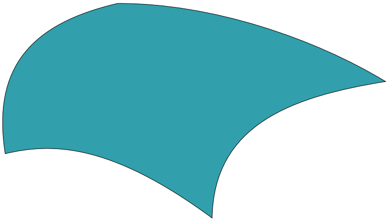
UNIVERSITY OF COPENHAGEN

Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

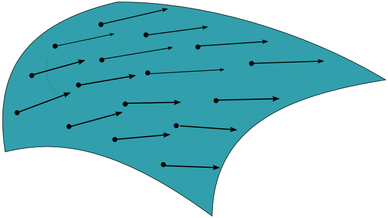
UNIVERSITY OF COPENHAGEN

Tangent vectors informally



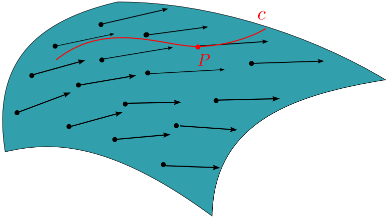
- How can we quantify tangent vectors to a manifold?

Tangent vectors informally



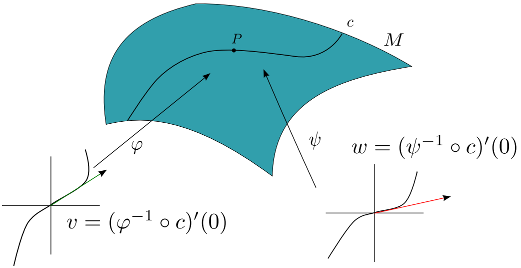
- How can we quantify tangent vectors to a manifold?

Tangent vectors informally

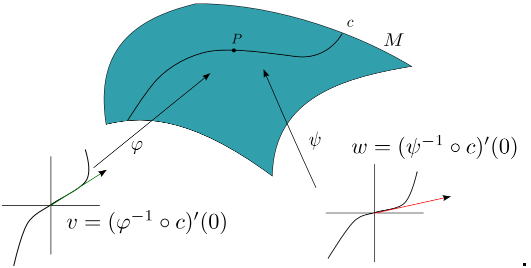


- How can we quantify tangent vectors to a manifold?
- Informally: a tangent vector at $P \in M$: draw a curve $c : (-\varepsilon, \varepsilon) \to M$, $c(0) = P$, then $\dot{c}(0)$ is a tangent vector.

A bit more formally

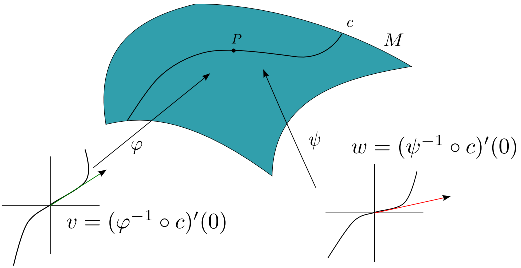


A bit more formally



- $c\colon (-\varepsilon, \varepsilon) \to M$, $c(0) = P$. In chart φ, the map $t \mapsto \varphi \circ c(t)$ is a curve in Euclidean space, and so is $t \mapsto \psi \circ c(t)$.

A bit more formally

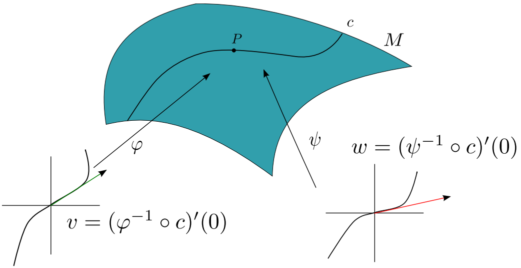


- $c\colon (-\varepsilon, \varepsilon) \to M$, $c(0) = P$. In chart φ, the map $t \mapsto \varphi \circ c(t)$ is a curve in Euclidean space, and so is $t \mapsto \psi \circ c(t)$.
- set $v = \frac{d}{dt}(\varphi \circ c)|_0$, $w = \frac{d}{dt}(\psi \circ c)|_0$ then

$$w = J_0\left(\varphi^{-1} \circ \psi\right) v.$$

($J_0 f$ = Jacobian of f at 0)

A bit more formally



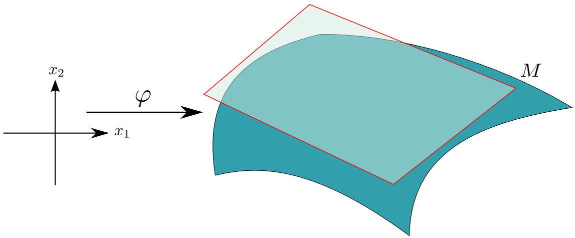
- $c\colon (-\varepsilon,\varepsilon) \to M$, $c(0) = P$. In chart φ, the map $t \mapsto \varphi \circ c(t)$ is a curve in Euclidean space, and so is $t \mapsto \psi \circ c(t)$.
- set $v = \frac{d}{dt}(\varphi \circ c)|_0$, $w = \frac{d}{dt}(\psi \circ c)|_0$ then

$$w = J_0\left(\varphi^{-1} \circ \psi\right) v.$$

  ($J_0 f$ = Jacobian of f at 0)
- Use this relation to identify vectors in different coordinate systems!

Tangent space

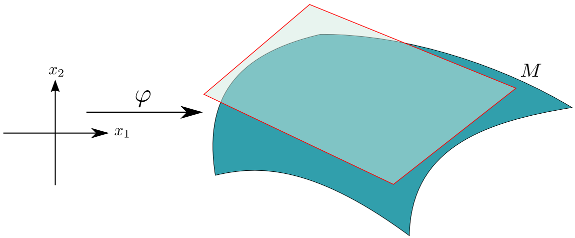
- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.



UNIVERSITY OF COPENHAGEN

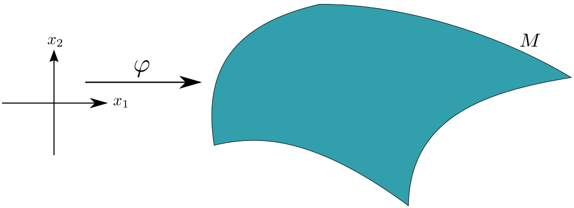
Tangent space

- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.
- It is a vector space of dimension m:



Tangent space

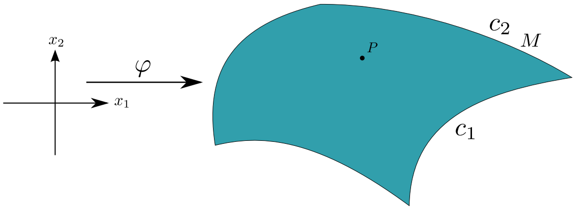
- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.
- It is a vector space of dimension m:
  - Let φ be a global chart for M



UNIVERSITY OF COPENHAGEN

Tangent space

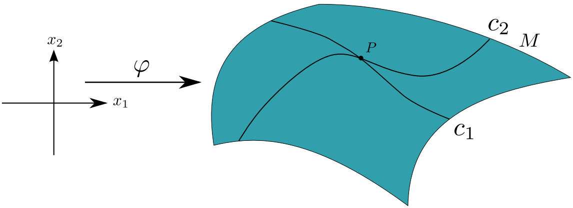
- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.
- It is a vector space of dimension m:
  - Let φ be a global chart for M
  - $\varphi(x_1, \ldots, x_m) \in M$ with $\varphi(0) = P$.



Tangent space

- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.
- It is a vector space of dimension m:
  - Let φ be a global chart for M
  - $\varphi(x_1, \ldots, x_m) \in M$ with $\varphi(0) = P$.
  - Define curves

$$c_i \colon t \mapsto \varphi(0, \ldots, 0, t, 0, 0).$$

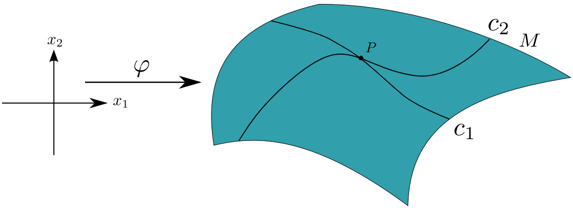


Tangent space

- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.
- It is a vector space of dimension m:
  - Let φ be a global chart for M
  - $\varphi(x_1, \dots, x_m) \in M$ with $\varphi(0) = P$.
  - Define curves

$$c_i \colon t \mapsto \varphi(0, \dots, 0, t, 0, 0).$$

  - They go through P when $t = 0$ and follow the axes.

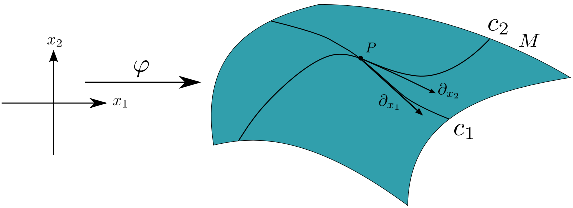


Tangent space

- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.
- It is a vector space of dimension m:
  - Let φ be a global chart for M
  - $\varphi(x_1, \dots, x_m) \in M$ with $\varphi(0) = P$.
  - Define curves

$$c_i \colon t \mapsto \varphi(0, \dots, 0, t, 0, 0).$$

  - They go through P when $t = 0$ and follow the axes.
  - Their derivative at 0 are denoted ∂_{x_i}, or sometimes $\frac{\partial}{\partial x_i}$.

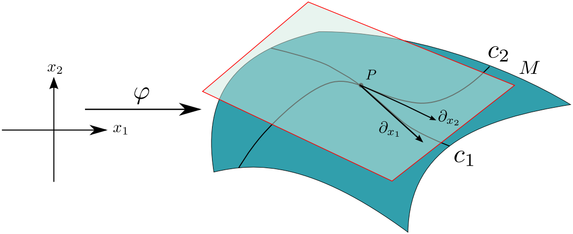


Tangent space

- The set of tangent vectors to the m-dimensional manifold M at point P is the tangent space of M at P denoted $T_P M$.
- It is a vector space of dimension m:
  - Let φ be a global chart for M
  - $\varphi(x_1, \dots, x_m) \in M$ with $\varphi(0) = P$.
  - Define curves

$$c_i \colon t \mapsto \varphi(0, \dots, 0, t, 0, 0).$$

  - They go through P when $t = 0$ and follow the axes.
  - Their derivative at 0 are denoted ∂_{x_i}, or sometimes $\frac{\partial}{\partial x_i}$.
  - The ∂_{x_i} form a basis of $T_P M$.



UNIVERSITY OF COPENHAGEN

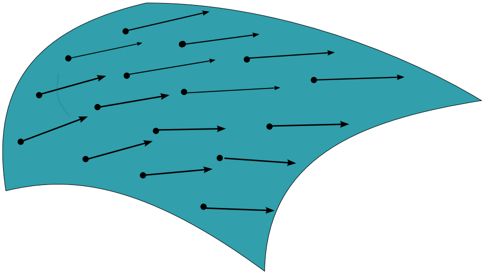
Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

UNIVERSITY OF COPENHAGEN

Vector fields

- A vector field is a smooth map that sends $P \in M$ to a vector $v(P) \in T_P M$.



Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

Differential of a smooth map

- $f : M \to N$ smooth, $P \in M$, $f(P) \in N$

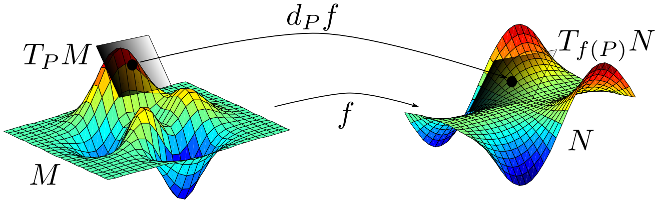
UNIVERSITY OF COPENHAGEN

Differential of a smooth map

- $f : M \to N$ smooth, $P \in M$, $f(P) \in N$
- $d_P f : T_P M \to T_{f(P)} N$ linear map corresponding to the Jacobian matrix of f in local coordinates.

Differential of a smooth map

- $f : M \to N$ smooth, $P \in M$, $f(P) \in N$
- $d_P f : T_P M \to T_{f(P)} N$ linear map corresponding to the Jacobian matrix of f in local coordinates.
- When $N = \mathbb{R}$, $d_P f$ is a linear form $T_P M \to \mathbb{R}$.



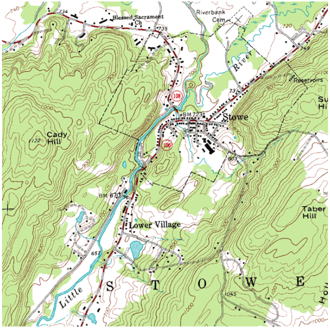
UNIVERSITY OF COPENHAGEN

Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

 UNIVERSITY OF COPENHAGEN

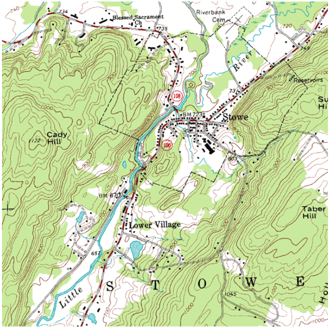
Tools needed in intrinsically nonlinear spaces?



- Comparison of objects in a nonlinear space?

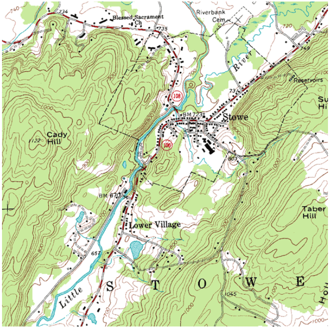
"Topographic map example". Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Topographic_map_example.png#mediaviewer/File:Topographic_map_example.png

Tools needed in intrinsically nonlinear spaces?



- Comparison of objects in a nonlinear space?
  - Distance metric? Kernel?
  - Varying local inner product = Riemannian metric!

Tools needed in intrinsically nonlinear spaces?

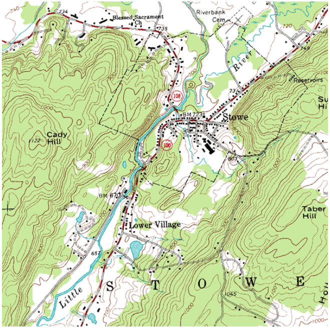


- Comparison of objects in a nonlinear space?
  - Distance metric? Kernel?
  - Varying local inner product = Riemannian metric!
- Optimization over such spaces?

"Topographic map example". Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Topographic_map_example.png#mediaviewer/File:Topographic_map_example.png

 UNIVERSITY OF COPENHAGEN

Tools needed in intrinsically nonlinear spaces?

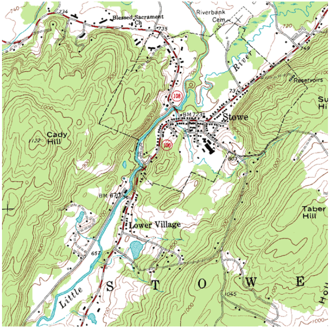


- Comparison of objects in a nonlinear space?
  - Distance metric? Kernel?
  - Varying local inner product = Riemannian metric!
- Optimization over such spaces?
  - Gradients!

"Topographic map example". Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Topographic_map_example.png#mediaviewer/File:Topographic_map_example.png

 UNIVERSITY OF COPENHAGEN

Tools needed in intrinsically nonlinear spaces?



- Comparison of objects in a nonlinear space?
  - Distance metric? Kernel?
  - Varying local inner product = Riemannian metric!
- Optimization over such spaces?
  - Gradients!
- $\rightsquigarrow$ ***Riemannian* geometry**

Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

UNIVERSITY OF COPENHAGEN

Recall: Inner Products

- *Euclidean/Hilbertian Inner Product* on vector space E: bilinear, symmetric, positive definite mapping $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{R}$.

Recall: Inner Products

- *Euclidean/Hilbertian Inner Product* on vector space E: bilinear, symmetric, positive definite mapping $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{R}$.
- Simplest example: usual dot-product on $\mathbb{R}^n$: $\mathbf{x} = (x_1, \dots, x_n)^t$, $\mathbf{y} = (y_1, \dots, y_n)^t$,

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = \mathbf{x}^t \, Id \, \mathbf{y}$$

Id $n \times n$ identity matrix.

Recall: Inner Products

- *Euclidean/Hilbertian Inner Product* on vector space E: bilinear, symmetric, positive definite mapping $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{R}$.
- Simplest example: usual dot-product on $\mathbb{R}^n$: $\mathbf{x} = (x_1, \dots, x_n)^t$, $\mathbf{y} = (y_1, \dots, y_n)^t$,

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = \mathbf{x}^t \, Id \, \mathbf{y}$$

  Id $n \times n$ identity matrix.
- $\mathbf{x}^T A \mathbf{y}$, A symmetric, positive definite: inner product, $\langle \mathbf{x}, \mathbf{y} \rangle_A$.

UNIVERSITY OF COPENHAGEN

Recall: Inner Products

- *Euclidean/Hilbertian Inner Product* on vector space E: bilinear, symmetric, positive definite mapping $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{R}$.
- Simplest example: usual dot-product on $\mathbb{R}^n$: $\mathbf{x} = (x_1, \dots, x_n)^t$, $\mathbf{y} = (y_1, \dots, y_n)^t$,

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = \mathbf{x}^t \, Id \, \mathbf{y}$$

  Id $n \times n$ identity matrix.
- $\mathbf{x}^T A \mathbf{y}$, A symmetric, positive definite: inner product, $\langle \mathbf{x}, \mathbf{y} \rangle_A$.
- Without subscript $\langle -, - \rangle$ will denote standard Euclidean dot-product (i.e. $A = Id$).

UNIVERSITY OF COPENHAGEN

Orthogonality – Norm – Distance

- Orthogonality, vector norm, distance from inner products.

$$\mathbf{x}\perp_A\mathbf{y} \iff \langle\mathbf{x},\mathbf{y}\rangle_A = 0, \quad \|\mathbf{x}\|_A^2 = \langle\mathbf{x},\mathbf{x}\rangle_A, \quad d_A(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_A.$$

 UNIVERSITY OF COPENHAGEN

Orthogonality – Norm – Distance

- Orthogonality, vector norm, distance from inner products.

$$\mathbf{x} \perp_A \mathbf{y} \iff \langle \mathbf{x}, \mathbf{y} \rangle_A = 0, \quad \|\mathbf{x}\|_A^2 = \langle \mathbf{x}, \mathbf{x} \rangle_A, \quad d_A(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_A.$$

- There are norms and distances not from an inner product.

Inner Products and Duality

Linear form $h = (h_1, \dots, h_n) : \mathbb{R}^n \to \mathbb{R}$: $h(\mathbf{x}) = \sum_{i=1}^n h_i x_i$.

- inner product $\langle -, - \rangle_A$ on $\mathbb{R}^n$: h represented by a unique vector $\mathbf{h}_A$ s.t

$$h(\mathbf{x}) = \langle \mathbf{h}_A, \mathbf{x} \rangle_A$$

  $\mathbf{h}_A$ is the dual of h (w.r.t $\langle -, - \rangle_A$).

Inner Products and Duality

Linear form $h = (h_1, \dots, h_n) : \mathbb{R}^n \to \mathbb{R}$: $h(\mathbf{x}) = \sum_{i=1}^n h_i x_i$.

- inner product $\langle -, - \rangle_A$ on $\mathbb{R}^n$: h represented by a unique vector $\mathbf{h}_A$ s.t

$$h(\mathbf{x}) = \langle \mathbf{h}_A, \mathbf{x} \rangle_A$$

  $\mathbf{h}_A$ is the dual of h (w.r.t $\langle -, - \rangle_A$).

- for standard dot product:

$$\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = h^T!$$

Inner Products and Duality

Linear form $h = (h_1, \ldots, h_n) : \mathbb{R}^n \to \mathbb{R}$: $h(\mathbf{x}) = \sum_{i=1}^n h_i x_i$.

- inner product $\langle -, - \rangle_A$ on $\mathbb{R}^n$: h represented by a unique vector $\mathbf{h}_A$ s.t

$$h(\mathbf{x}) = \langle \mathbf{h}_A, \mathbf{x} \rangle_A$$

  $\mathbf{h}_A$ is the dual of h (w.r.t $\langle -, - \rangle_A$).

- for standard dot product:

$$\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = h^T!$$

- for general inner product $\langle -, - \rangle_A$

$$\mathbf{h}_A = A^{-1}\mathbf{h} = A^{-1}h^T.$$

UNIVERSITY OF COPENHAGEN

Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

UNIVERSITY OF COPENHAGEN

Riemannian Metric

- Riemannian metric on an $m-$dimensional manifold $=$ smooth family g_P of inner products on the tangent spaces T_PM of M
  - $u, v \in T_PM \mapsto g_p(u, v) := \langle u, v \rangle_P \in \mathbb{R}$
  - With it, one can compute length of vectors in tangent spaces, check orthogonality, etc...

Riemannian Metric

- Riemannian metric on an $m-$dimensional manifold $=$ smooth family g_P of inner products on the tangent spaces $T_P M$ of M
  - $u, v \in T_P M \mapsto g_p(u, v) := \langle u, v \rangle_P \in \mathbb{R}$
  - With it, one can compute length of vectors in tangent spaces, check orthogonality, etc...
- Given a global parametrization $\varphi \colon (\mathbf{x}) = (x_1, \ldots, x_n) \mapsto \varphi(\mathbf{x}) \in M$, it corresponds to a smooth family of symmetric positive definite matrices:

$$g_{\mathbf{x}} = \begin{pmatrix} g_{\mathbf{x}11} & \cdots & g_{\mathbf{x}1n} \\ \vdots & & \vdots \\ g_{\mathbf{x}n1} & \cdots & g_{\mathbf{x}nn} \end{pmatrix}$$

Riemannian Metric

- Riemannian metric on an $m-$dimensional manifold $=$ smooth family g_P of inner products on the tangent spaces T_PM of M
  - $u, v \in T_PM \mapsto g_p(u, v) := \langle u, v \rangle_P \in \mathbb{R}$
  - With it, one can compute length of vectors in tangent spaces, check orthogonality, etc...
- Given a global parametrization $\varphi \colon (\mathbf{x}) = (x_1, \dots, x_n) \mapsto \varphi(\mathbf{x}) \in M$, it corresponds to a smooth family of symmetric positive definite matrices:

$$g_{\mathbf{x}} = \begin{pmatrix} g_{\mathbf{x}11} & \cdots & g_{\mathbf{x}1n} \\ \vdots & & \vdots \\ g_{\mathbf{x}n1} & \cdots & g_{\mathbf{x}nn} \end{pmatrix}$$

- $u = \sum_{i=1}^n u_i \partial_{x_i}$, $v = \sum_{i=1}^n v_i \partial_{x_i}$
  $\langle u, v \rangle_{\mathbf{x}} = (u_1, \dots, u_n) g_{\mathbf{x}} (v_1, \dots, v_n)^t$

Riemannian Metric

- Riemannian metric on an $m-$dimensional manifold $=$ smooth family g_P of inner products on the tangent spaces T_PM of M
  - $u, v \in T_PM \mapsto g_p(u, v) := \langle u, v \rangle_P \in \mathbb{R}$
  - With it, one can compute length of vectors in tangent spaces, check orthogonality, etc...
- Given a global parametrization $\varphi \colon (\mathbf{x}) = (x_1, \dots, x_n) \mapsto \varphi(\mathbf{x}) \in M$, it corresponds to a smooth family of symmetric positive definite matrices:

$$g_{\mathbf{x}} = \begin{pmatrix} g_{\mathbf{x}11} & \cdots & g_{\mathbf{x}1n} \\ \vdots & & \vdots \\ g_{\mathbf{x}n1} & \cdots & g_{\mathbf{x}nn} \end{pmatrix}$$

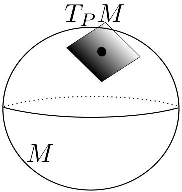
- $u = \sum_{i=1}^n u_i \partial_{x_i}$, $v = \sum_{i=1}^n v_i \partial_{x_i}$
  $\langle u, v \rangle_{\mathbf{x}} = (u_1, \dots, u_n) g_{\mathbf{x}} (v_1, \dots, v_n)^t$
- A smooth manifold with a Riemannian metric is a Riemannian manifold.

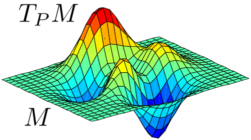
Riemannian Metric

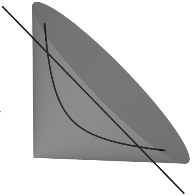
 UNIVERSITY OF COPENHAGEN

Example: Induced Riemannian metric on submanifolds of $\mathbb{R}^n$

- Inner product from $\mathbb{R}^n$ restricts to inner product on $M \subset \mathbb{R}^n$







- Frobenius metric on $\mathcal{P}(n)$
  - $\mathcal{P}(n)$ is a convex subset of $\mathbb{R}^{(n^2+n)/2}$
  - The Euclidean inner product defines a Riemannian metric on $\mathcal{P}(n)$

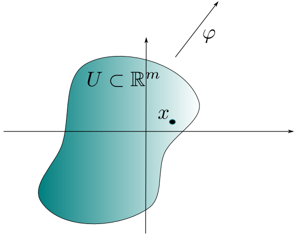
Rightmost figure from Fletcher, Joshi, Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors, CVAMIA04

 UNIVERSITY OF COPENHAGEN

Example: Fisher information metric

- Smooth manifold $M = \varphi(U)$ represents a family of probability distributions (M is a *statistical model*), $U \subset \mathbb{R}^m$
- Each point $P = \varphi(x) \in M$ is a probability distribution $P\colon \mathcal{Z} \to \mathbb{R}_{>0}$

M = probability distributions P on set $\mathcal{Z}$
$P\colon \mathcal{Z} \to \mathbb{R}_{>0}$ $\qquad \int_{\mathcal{Z}} P(z)dz = 1$



- The *Fisher information metric* of M at $P_x = \varphi(x)$ in coordinates φ defined by:

$$g_{ij}(x) = \int_{\mathcal{Z}} \frac{\partial \log P_x(z)}{\partial x_i} \frac{\partial \log P_x(z)}{\partial x_j} P_x(z)dz$$

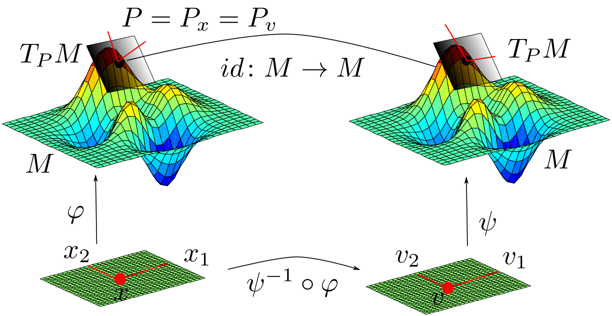
Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

Invariance of the Fisher information metric

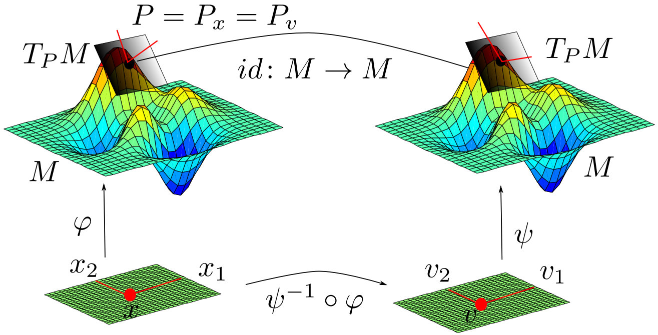
- Obtaining a Riemannian metric g on the left chart by pulling $\tilde{g}$ back from the right chart:

$$g_{ij} = g\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_i}\right) = \tilde{g}\left(d(\psi^{-1}\circ\varphi)\left(\frac{\partial}{\partial x_i}\right), d(\psi^{-1}\circ\varphi)\left(\frac{\partial}{\partial x_j}\right)\right).$$



- **Claim:** $g_{ij} = \sum_{k=1}^{m}\sum_{l=1}^{m} \tilde{g}_{kl}\frac{\partial v_k}{\partial x_i}\frac{\partial v_l}{\partial x_j}$

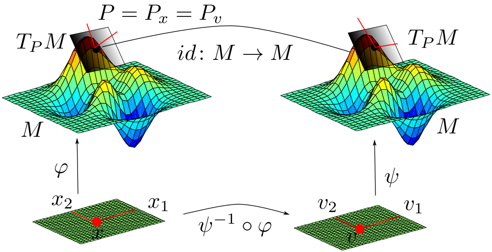
Invariance of the Fisher information metric



- **Claim:** $g_{ij} = \sum_{k=1}^{m} \sum_{l=1}^{m} \tilde{g}_{kl} \frac{\partial v_k}{\partial x_i} \frac{\partial v_l}{\partial x_j}$
- **Proof:**

$$
\begin{aligned}
g_{ij} &= g\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right) \\
&= \tilde{g}\left(\sum_{k=1}^{m} \frac{\partial v_k}{\partial x_i} \frac{\partial}{\partial v_k}, \sum_{l=1}^{m} \frac{\partial v_l}{\partial x_j} \frac{\partial}{\partial v_l}\right) \\
&= \sum_{k=1}^{m} \sum_{l=1}^{m} \frac{\partial v_k}{\partial x_i} \frac{\partial v_l}{\partial x_j} \tilde{g}\left(\frac{\partial}{\partial v_k}, \frac{\partial}{\partial v_l}\right) \\
&= \sum_{k=1}^{m} \sum_{l=1}^{m} \tilde{g}_{kl} \frac{\partial v_k}{\partial x_i} \frac{\partial v_l}{\partial x_j}
\end{aligned}
$$

Invariance of the Fisher information metric



- **Fact:** $g_{ij} = \sum_{k=1}^m \sum_{l=1}^m \tilde{g}_{kl} \frac{\partial v_k}{\partial x_i} \frac{\partial v_l}{\partial x_j}$
- Pulling the Fisher information metric $\tilde{g}$ from right to left:

$$
\begin{aligned}
g_{ij} &= \sum_{k=1}^m \sum_{l=1}^m \tilde{g}_{kl} \frac{\partial v_k}{\partial x_i} \frac{\partial v_l}{\partial x_j} \\
&= \sum_{k=1}^m \sum_{l=1}^m \int \frac{\partial \log P_v(z)}{\partial v_k} \frac{\partial \log P_v(z)}{\partial v_l} P_v(z) dz \cdot \frac{\partial v_k}{\partial x_i} \frac{\partial v_l}{\partial x_j} \\
&= \int \left(\sum_{k=1}^m \frac{\partial \log P_v(z)}{\partial v_k} \frac{\partial v_k}{\partial x_i} \right) \left(\sum_{l=1}^m \frac{\partial \log P_v(z)}{\partial v_l} \frac{\partial v_l}{\partial x_j} \right) P_v(z) dz \\
&= \int \frac{\partial \log P_x(z)}{\partial x_i} \frac{\partial \log P_x(z)}{\partial x_j} P_x(z) dz
\end{aligned}
$$

- **Result:** Formula invariant of parametrization

UNIVERSITY OF COPENHAGEN

Outline

1. Motivation
   - Nonlinearity
   - Recall: Calculus in $\mathbb{R}^n$
2. Differential Geometry
   - Smooth manifolds
   - Building Manifolds
   - Tangent Space
   - Vector fields
   - Differential of smooth map
3. Riemannian metrics
   - Introduction to Riemannian metrics
   - Recall: Inner Products
   - Riemannian metrics
   - Invariance of the Fisher information metric
   - A first take on the geodesic distance metric
   - A first take on curvature

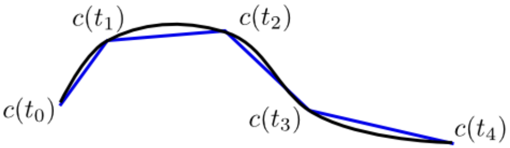
 UNIVERSITY OF COPENHAGEN

Riemannian metrics and distances

Path length in metric spaces:

- Let (X, d) be a metric space. The length of a curve $c\colon [a, b] \to X$ is

$$l(c) = sup_{a=t_0 \leq t_1 \leq \ldots \leq t_n = b} \sum_{i=0}^{n-1} d(c(t_i, t_{i+1})). \qquad (3.1)$$

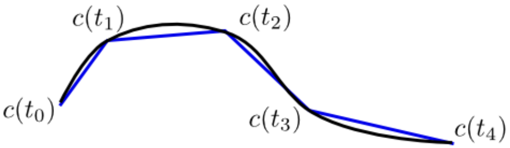


Riemannian metrics and distances

Path length in metric spaces:

- Let (X, d) be a metric space. The length of a curve $c\colon [a, b] \to X$ is

$$l(c) = sup_{a=t_0 \le t_1 \le \ldots \le t_n = b} \sum_{i=0}^{n-1} d(c(t_i, t_{i+1})). \tag{3.1}$$



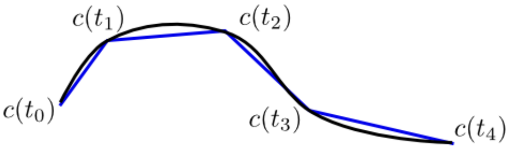
- Approach supremum through segments $c(t_i, t_{i+1})$ of length $\to 0$

Riemannian metrics and distances

Path length in metric spaces:

- Let (X, d) be a metric space. The length of a curve $c\colon [a, b] \to X$ is

$$l(c) = sup_{a=t_0 \leq t_1 \leq \ldots \leq t_n = b} \sum_{i=0}^{n-1} d(c(t_i, t_{i+1})). \quad (3.1)$$



- Approach supremum through segments $c(t_i, t_{i+1})$ of length $\to 0$

Intuitive **path length on Riemannian manifolds:**

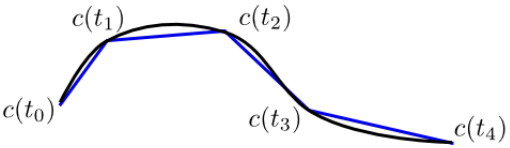
- Riemannian metric g on M defines norm in $T_P M$

Riemannian metrics and distances

Path length in metric spaces:

- Let (X, d) be a metric space. The length of a curve $c\colon [a, b] \to X$ is

$$l(c) = sup_{a=t_0 \leq t_1 \leq \ldots \leq t_n = b} \sum_{i=0}^{n-1} d(c(t_i, t_{i+1})). \qquad (3.1)$$



- Approach supremum through segments $c(t_i, t_{i+1})$ of length $\to 0$

Intuitive **path length on Riemannian manifolds:**

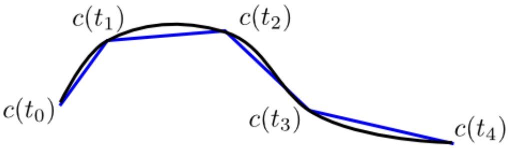
- Riemannian metric g on M defines norm in $T_P M$
- Locally a good approximation for use with (3.1)

Riemannian metrics and distances

Path length in metric spaces:

- Let (X, d) be a metric space. The length of a curve $c\colon [a, b] \to X$ is

$$l(c) = sup_{a=t_0 \leq t_1 \leq \ldots \leq t_n = b} \sum_{i=0}^{n-1} d(c(t_i, t_{i+1})). \tag{3.1}$$



- Approach supremum through segments $c(t_i, t_{i+1})$ of length $\to 0$

***Intuitive* path length on Riemannian manifolds:**

- Riemannian metric g on M defines norm in $T_P M$
- Locally a good approximation for use with (3.1)
- *This will be made precise in Francois' lecture!*

UNIVERSITY OF COPENHAGEN

Geodesics as length-minimizing curves

- We have a concept of path length $l(c)$ for paths $c\colon [a,b] \to M$

UNIVERSITY OF COPENHAGEN

Geodesics as length-minimizing curves

- We have a concept of path length $l(c)$ for paths $c\colon [a,b] \to M$
- A *geodesic* from P to Q in M is a path $c\colon [a,b] \to X$ such that $c(a) = P, c(b) = Q$ and $l(c) = \inf_{c_{P\to Q}} l(c_{P\to Q})$.

Geodesics as length-minimizing curves

- We have a concept of path length $l(c)$ for paths $c\colon [a,b] \to M$
- A *geodesic* from P to Q in M is a path $c\colon [a,b] \to X$ such that $c(a) = P, c(b) = Q$ and $l(c) = \inf_{c_{P\to Q}} l(c_{P\to Q})$.
- The distance function $d(P,Q) = \inf_{c_{P\to Q}} l(c_{P\to Q})$ is a *distance metric* on the Riemannian manifold (M,g).

Geodesics as length-minimizing curves

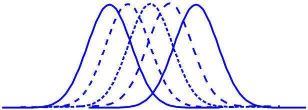
- We have a concept of path length $l(c)$ for paths $c\colon [a, b] \to M$
- A *geodesic* from P to Q in M is a path $c\colon [a, b] \to X$ such that $c(a) = P, c(b) = Q$ and $l(c) = \inf_{c_{P \to Q}} l(c_{P \to Q})$.
- The distance function $d(P, Q) = \inf_{c_{P \to Q}} l(c_{P \to Q})$ is a *distance metric* on the Riemannian manifold (M, g). Can you see why?

Geodesics as length-minimizing curves

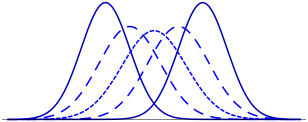
- We have a concept of path length $l(c)$ for paths $c \colon [a, b] \to M$
- A *geodesic* from P to Q in M is a path $c \colon [a, b] \to X$ such that $c(a) = P, c(b) = Q$ and $l(c) = \inf_{c_{P\to Q}} l(c_{P\to Q})$.
- The distance function $d(P, Q) = \inf_{c_{P\to Q}} l(c_{P\to Q})$ is a *distance metric* on the Riemannian manifold (M, g). Can you see why?
- Do geodesics always exist?

Example: Riemannian geodesics between 1-dimensional Gaussian distributions

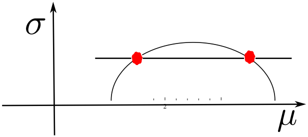
- Space parametrized by $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$
- Metric 1: Euclidean inner product $\Rightarrow$ Euclidean geodesics



- Metric 2: Fisher information metric



- View in plane:



Middle figure from Costa et al, Fisher information distance: a geometrical reading

Outline

1 Motivation
Nonlinearity
Recall: Calculus in $\mathbb{R}^n$

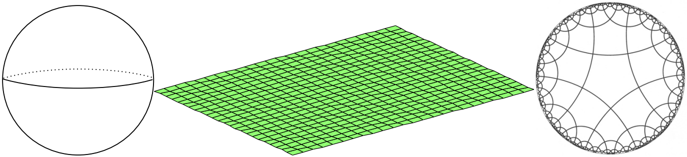
2 Differential Geometry
Smooth manifolds
Building Manifolds
Tangent Space
Vector fields
Differential of smooth map

3 Riemannian metrics
Introduction to Riemannian metrics
Recall: Inner Products
Riemannian metrics
Invariance of the Fisher information metric
A first take on the geodesic distance metric
A first take on curvature

 UNIVERSITY OF COPENHAGEN

A first take on curvature

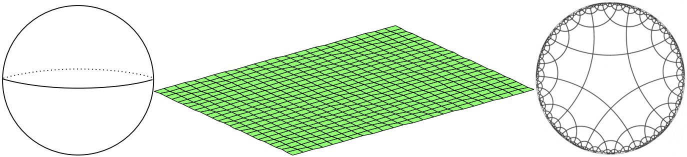
- Curvature in metric spaces defined by comparison with *model spaces* of known curvature.



 UNIVERSITY OF COPENHAGEN

A first take on curvature

- Curvature in metric spaces defined by comparison with *model spaces* of known curvature.
  - Positive curvature model spaces. Spheres of curvature $\kappa > 0$:
  - Flat model space: Euclidean plane
  - Negatively curved model spaces: Hyperbolic space of curvature $\kappa > 0$



 UNIVERSITY OF COPENHAGEN

A first take on curvature

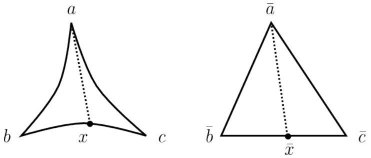
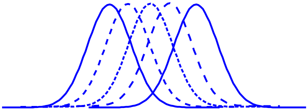


Figure : **Left:** Geodesic triangle in a negatively curved space. **Right:** Comparison triangle in the plane.

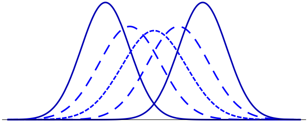
- A $CAT(\kappa)$ space is a metric space in which geodesic triangles are "thinner" than for their comparison triangles in the model space M_κ; that is, $d(x,a) \leq d(\bar{x},\bar{a})$.
- A locally $CAT(\kappa)$ space has curvature bounded from above by κ.
- Geodesic triangles are useful for intuition and proofs!

Example: The two metrics on 1-dimensional Gaussian distributions

- Metric 1: Euclidean inner product: **FLAT**

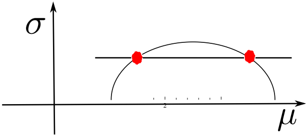


- Metric 2: Fisher information metric: **HYPERBOLIC**



(**OBS: Not hyperbolic for any family of distributions**)

- View in plane:

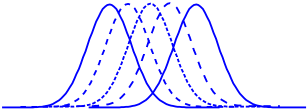


Middle figure from Costa et al. Fisher information distance: a geometrical reading

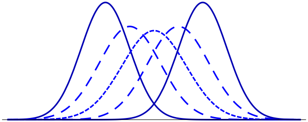
 UNIVERSITY OF COPENHAGEN

Example: The two metrics on 1-dimensional Gaussian distributions

- Metric 1: Euclidean inner product: **FLAT**

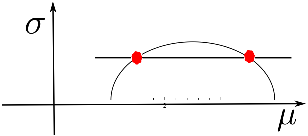


- Metric 2: Fisher information metric: **HYPERBOLIC**



(**OBS: Not hyperbolic for any family of distributions**)

- View in plane:

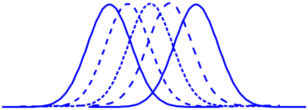


Middle figure from Costa et al. Fisher information distance: a geometrical reading

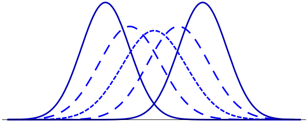
 UNIVERSITY OF COPENHAGEN

Example: The two metrics on 1-dimensional Gaussian distributions

- Metric 1: Euclidean inner product: **FLAT**

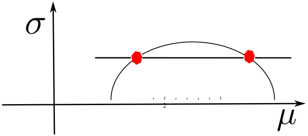


- Metric 2: Fisher information metric: **HYPERBOLIC**



(**OBS: Not hyperbolic for any family of distributions**)

- View in plane:



- You will see these again with Stefan!

Middle figure from Costa et al. Fisher information distance: a geometrical reading

Relation to sectional curvature

- *CAT*(κ) is a weak notion of curvature
- Stronger notion of sectional curvature (requires a little more Riemannian geometry)

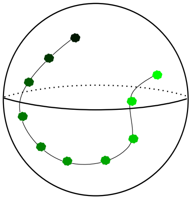
Theorem

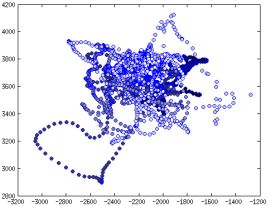
A smooth Riemannian manifold M is (locally) CAT(κ) *if and only if the sectional curvature of M is* $\leq \kappa$. □

UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you

- Given a distance matrix $D_{ij} = d(x_i, x_j)$ for a dataset $X = \{x_1, \ldots, x_n\}$ residing on a manifold M, where d is a geodesic metric.

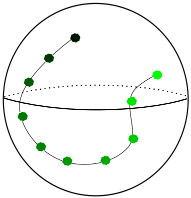


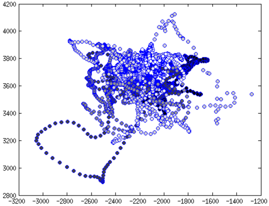


UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you

- Given a distance matrix $D_{ij} = d(x_i, x_j)$ for a dataset $X = \{x_1, \ldots, x_n\}$ residing on a manifold M, where d is a geodesic metric.
- Assume that $Z = \{z_1, \ldots, z_n\} \subset \mathbb{R}^d$ is an embedding of X obtained through MDS or manifold learning.

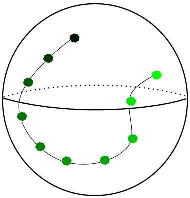


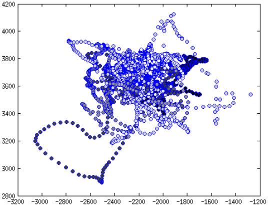


UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you

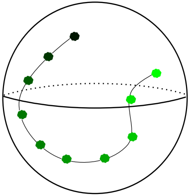
- Given a distance matrix $D_{ij} = d(x_i, x_j)$ for a dataset $X = \{x_1, \dots, x_n\}$ residing on a manifold M, where d is a geodesic metric.
- Assume that $Z = \{z_1, \dots, z_n\} \subset \mathbb{R}^d$ is an embedding of X obtained through MDS or manifold learning.
- **Common belief:** If d large, then Z is a good (perfect?) representation of X.

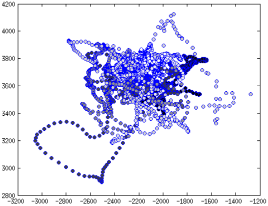




 UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you

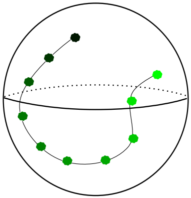


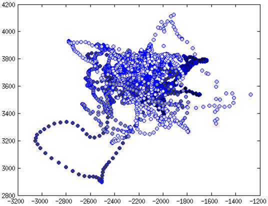


- **Truth:** If there exists a map $f\colon M \to \mathbb{R}^d$ such that $\|f(a) - f(b)\| = d(a,b)$ for all $a, b \in M$, then

 UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you

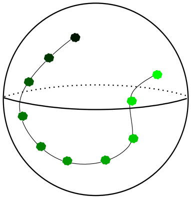


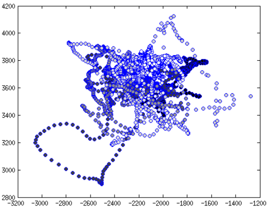


- **Truth:** If there exists a map $f\colon M \to \mathbb{R}^d$ such that $\|f(a) - f(b)\| = d(a, b)$ for all $a, b \in M$, then
  - f maps geodesics to straight lines
  - M is $CAT(0)$
  - M is not $CAT(\kappa)$ for any $\kappa < 0$

 UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you

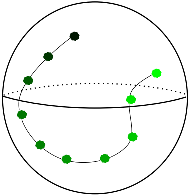


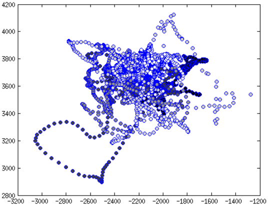


- **Truth:** If there exists a map $f\colon M \to \mathbb{R}^d$ such that $\|f(a) - f(b)\| = d(a, b)$ for all $a, b \in M$, then
    - f maps geodesics to straight lines
    - M is $CAT(0)$
    - M is not $CAT(\kappa)$ for any $\kappa < 0$
    - $\Rightarrow M$ is flat

 UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you

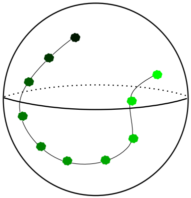


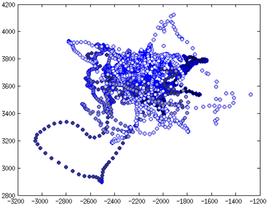


- **Truth:** If there exists a map $f\colon M \to \mathbb{R}^d$ such that $\|f(a) - f(b)\| = d(a, b)$ for all $a, b \in M$, then
  - f maps geodesics to straight lines
  - M is $CAT(0)$
  - M is not $CAT(\kappa)$ for any $\kappa < 0$
  - $\Rightarrow$ M is flat
- That is, if M is not flat, MDS and manifold learning lie to you

 UNIVERSITY OF COPENHAGEN

Example of insight with $CAT(\kappa)$: MDS and manifold learning lie to you





- **Truth:** If there exists a map $f\colon M \to \mathbb{R}^d$ such that $\|f(a) - f(b)\| = d(a, b)$ for all $a, b \in M$, then
  - f maps geodesics to straight lines
  - M is $CAT(0)$
  - M is not $CAT(\kappa)$ for any $\kappa < 0$
  - $\Rightarrow$ M is flat
- That is, if M is not flat, MDS and manifold learning lie to you
- (but sometimes lies are useful)

UNIVERSITY OF COPENHAGEN

A message from Stefan for tomorrow’s practical

Check out course webpage for installation instructions!
`http://image.diku.dk/MLLab/IG4.php`