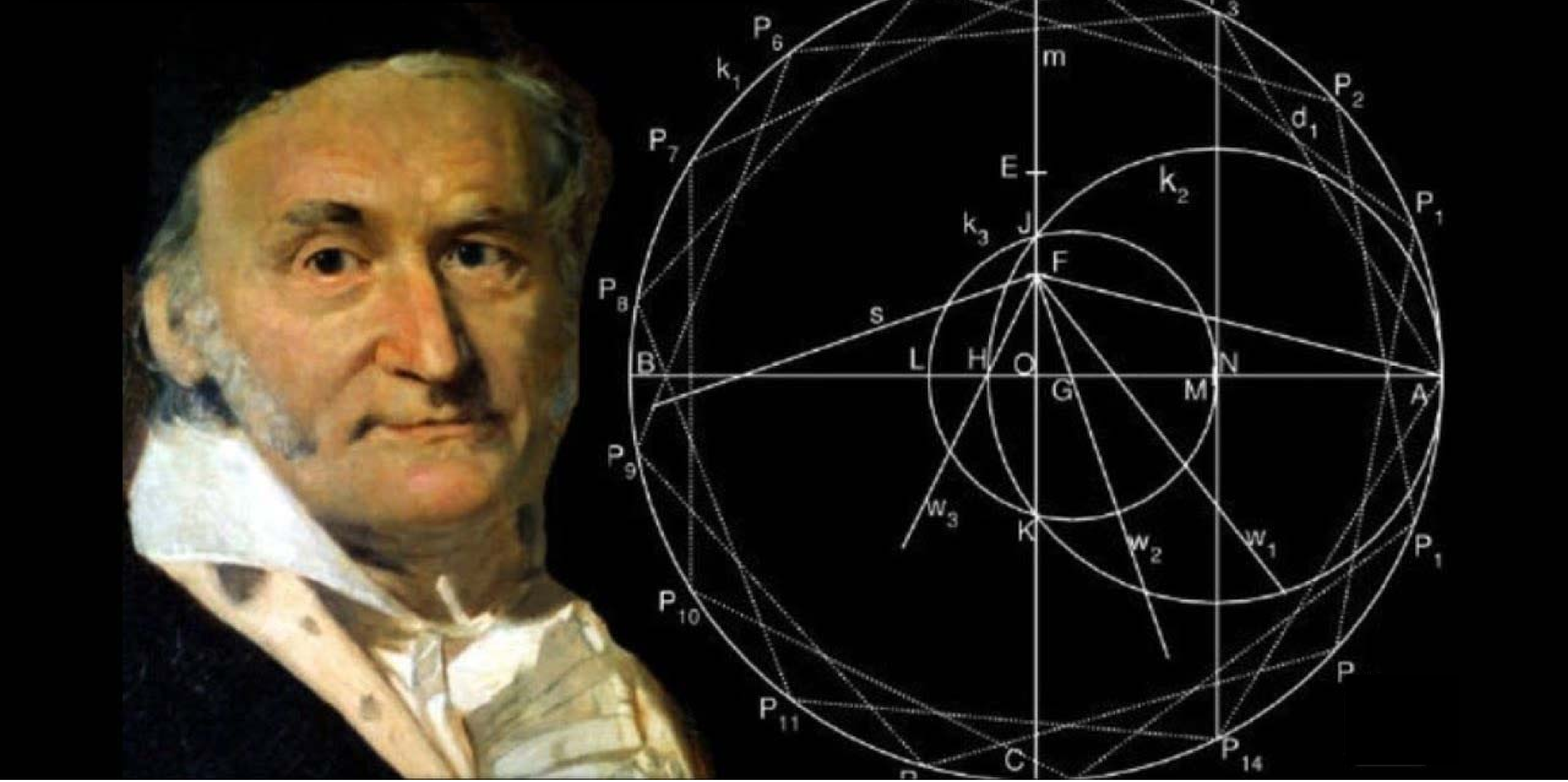
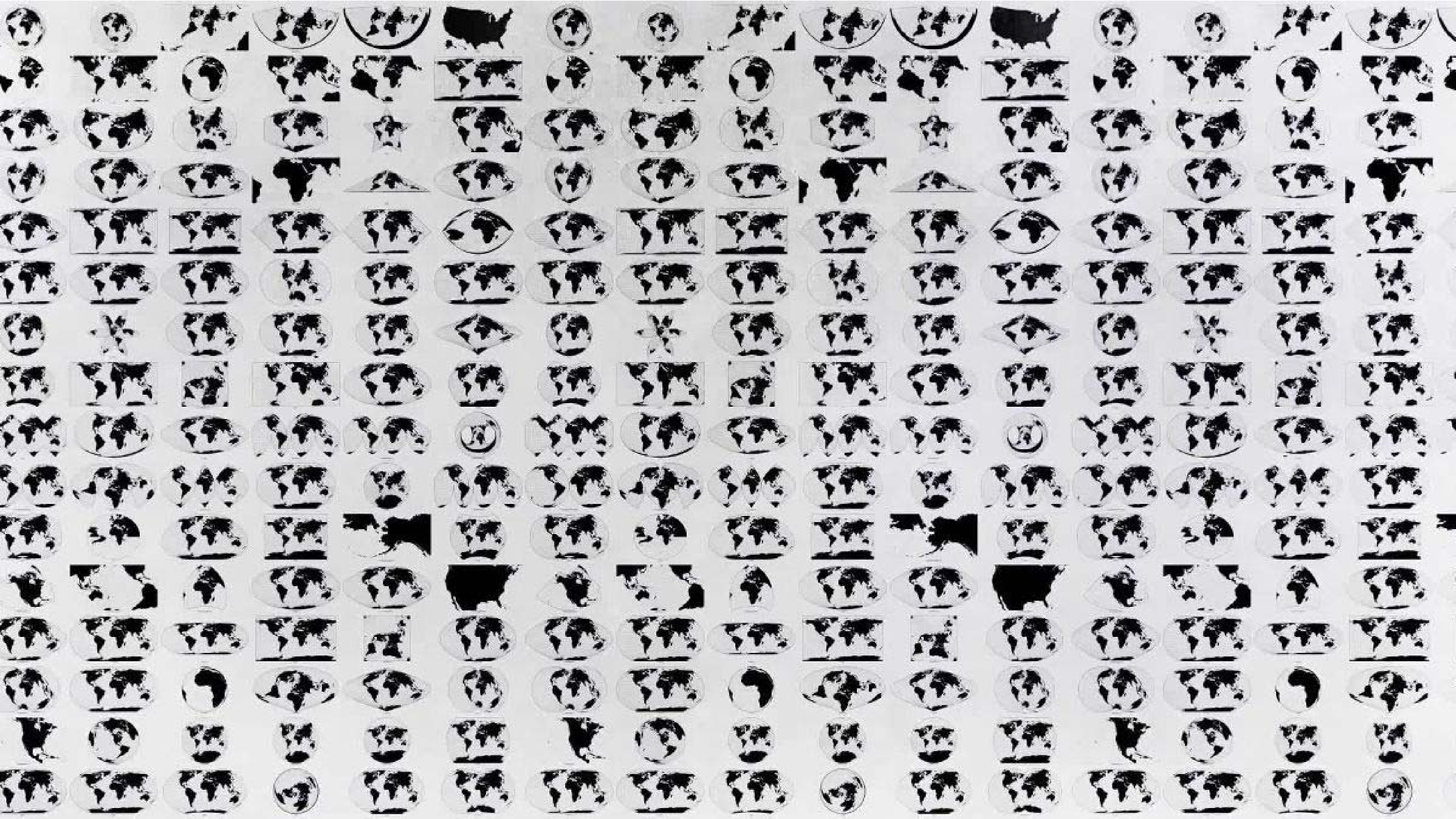
Riemannian Geometry and Machine Learning for Non-Euclidean Data

Frank C. Park and C.J. Jang
Seoul National University

Carl Friedrich Gauss (1777-1855)



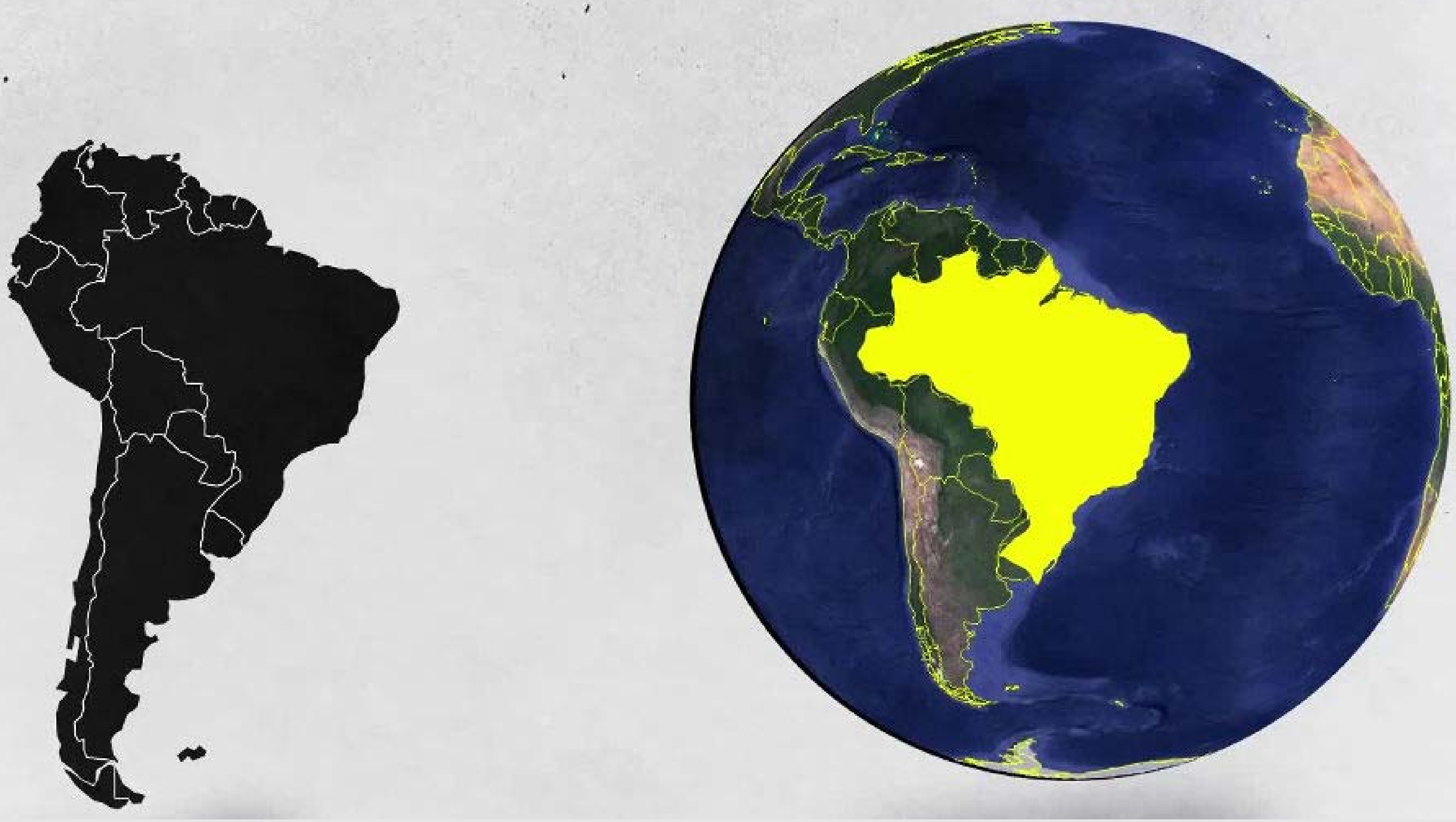
15th Century Mapmaking



It would be nice if straight lines on maps...

...were shortest paths on the sphere (but in most cases they're not)

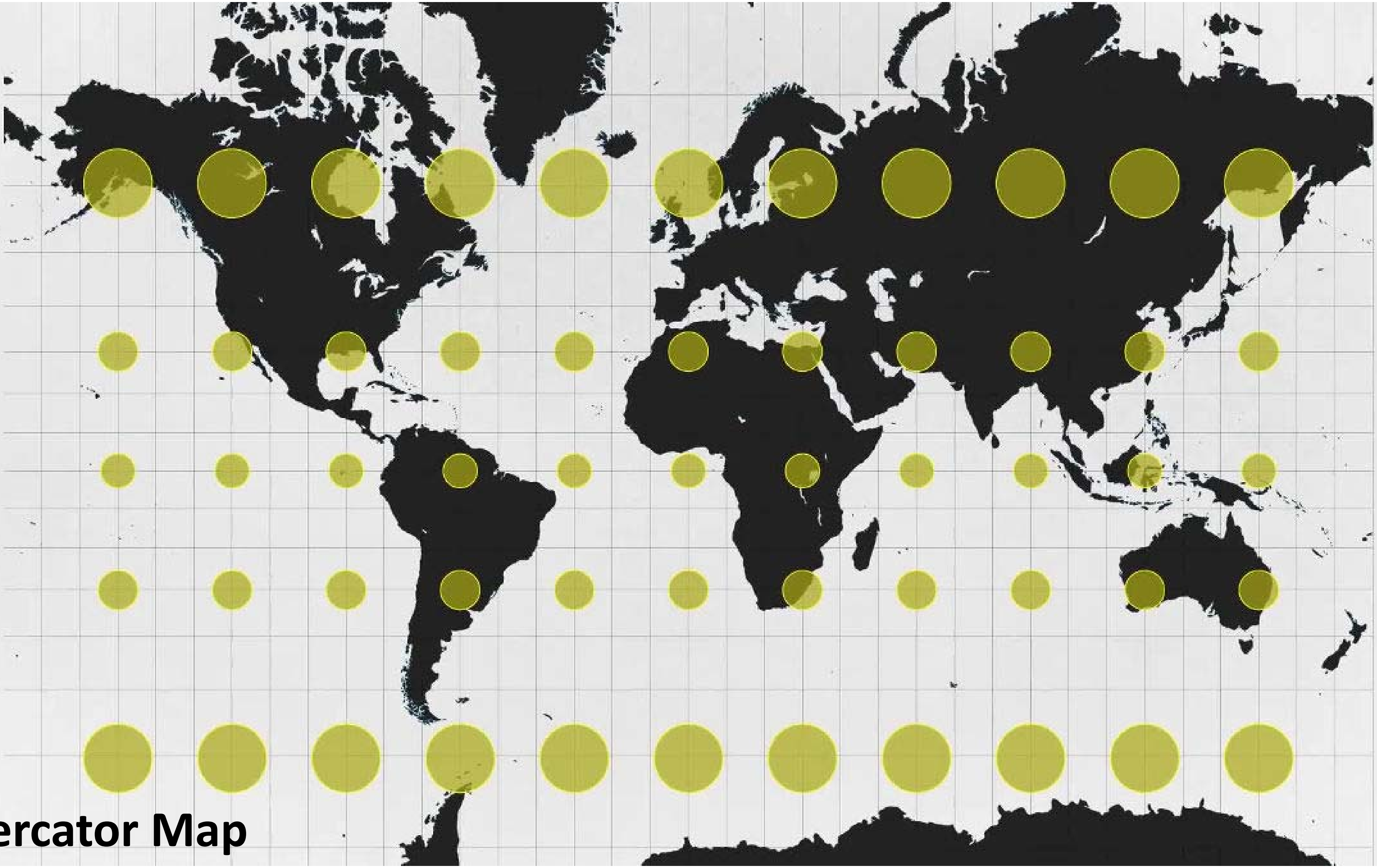
Google Maps (Mercator projection)



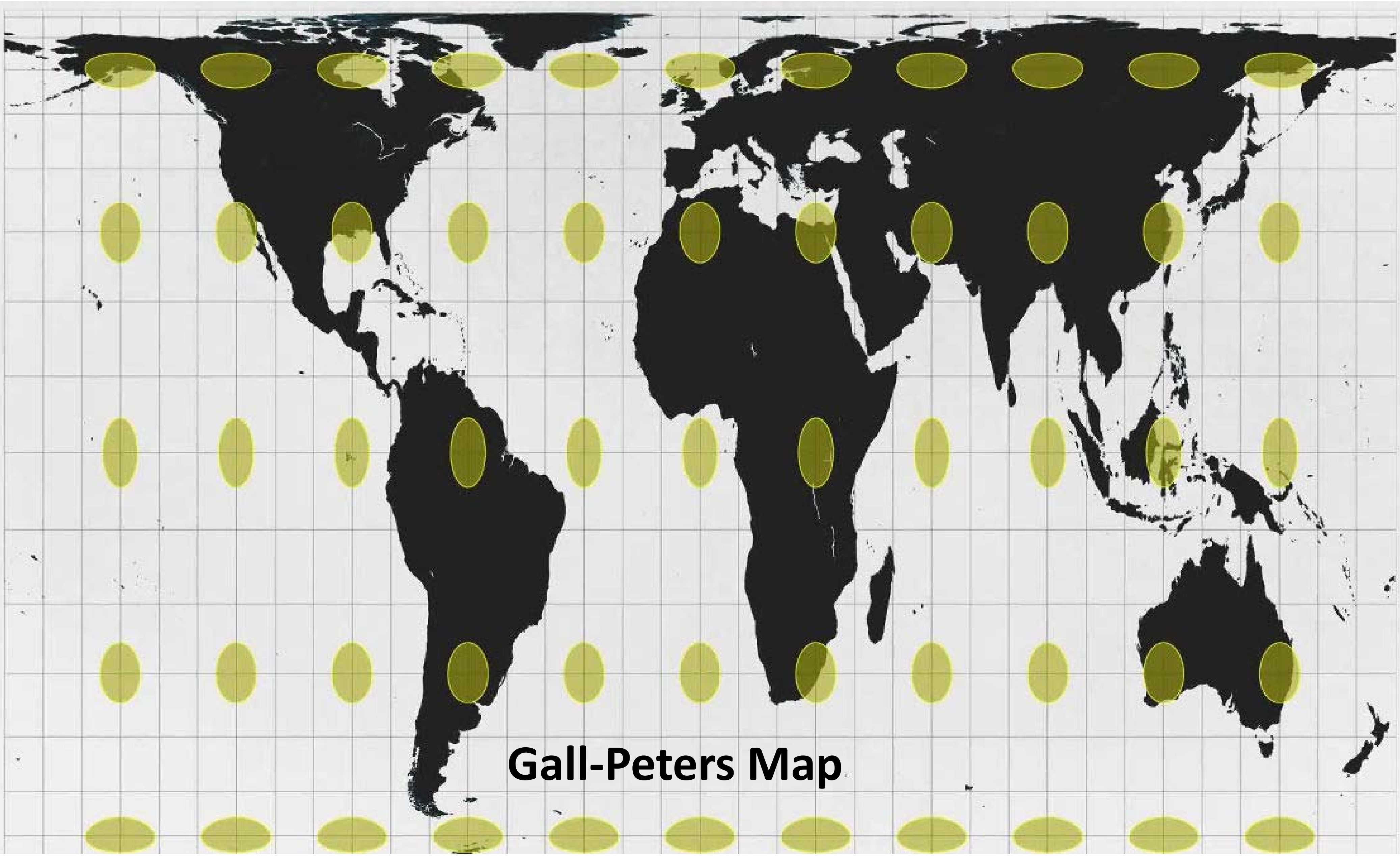
Mercator maps are very accurate for countries near the equator (e.g., Brazil)

Greenland *vs* Africa: Sizes on Mercator Map

Greenland
***vs* Africa:**
Actual Size
Comparison

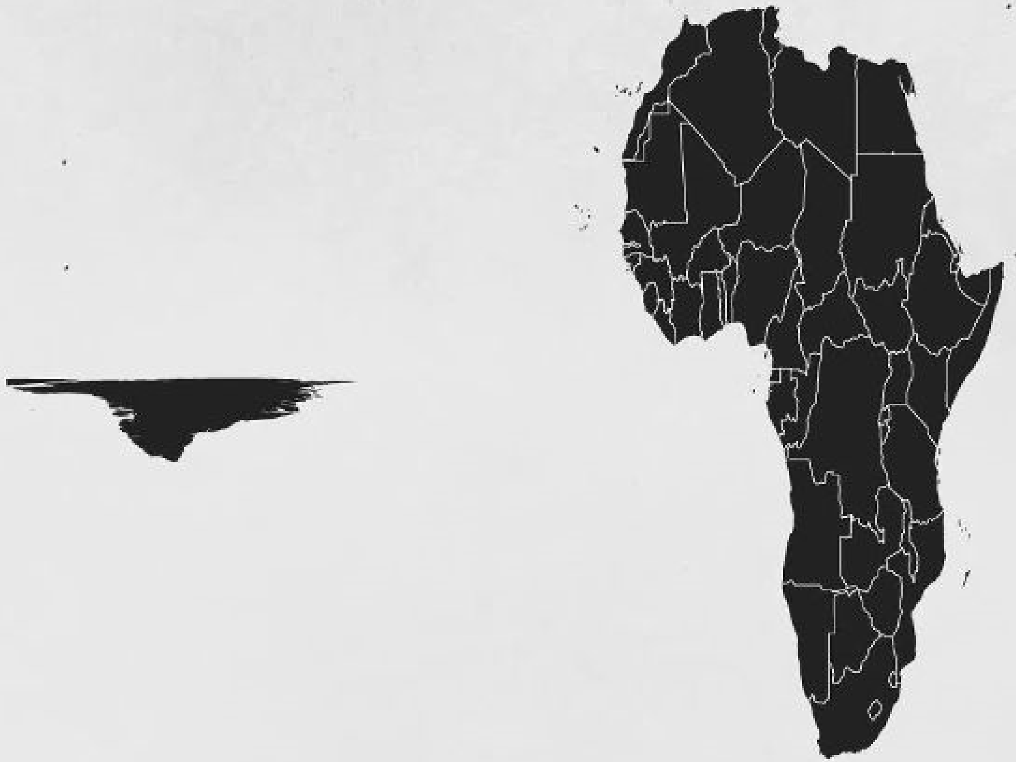


Mercator Map



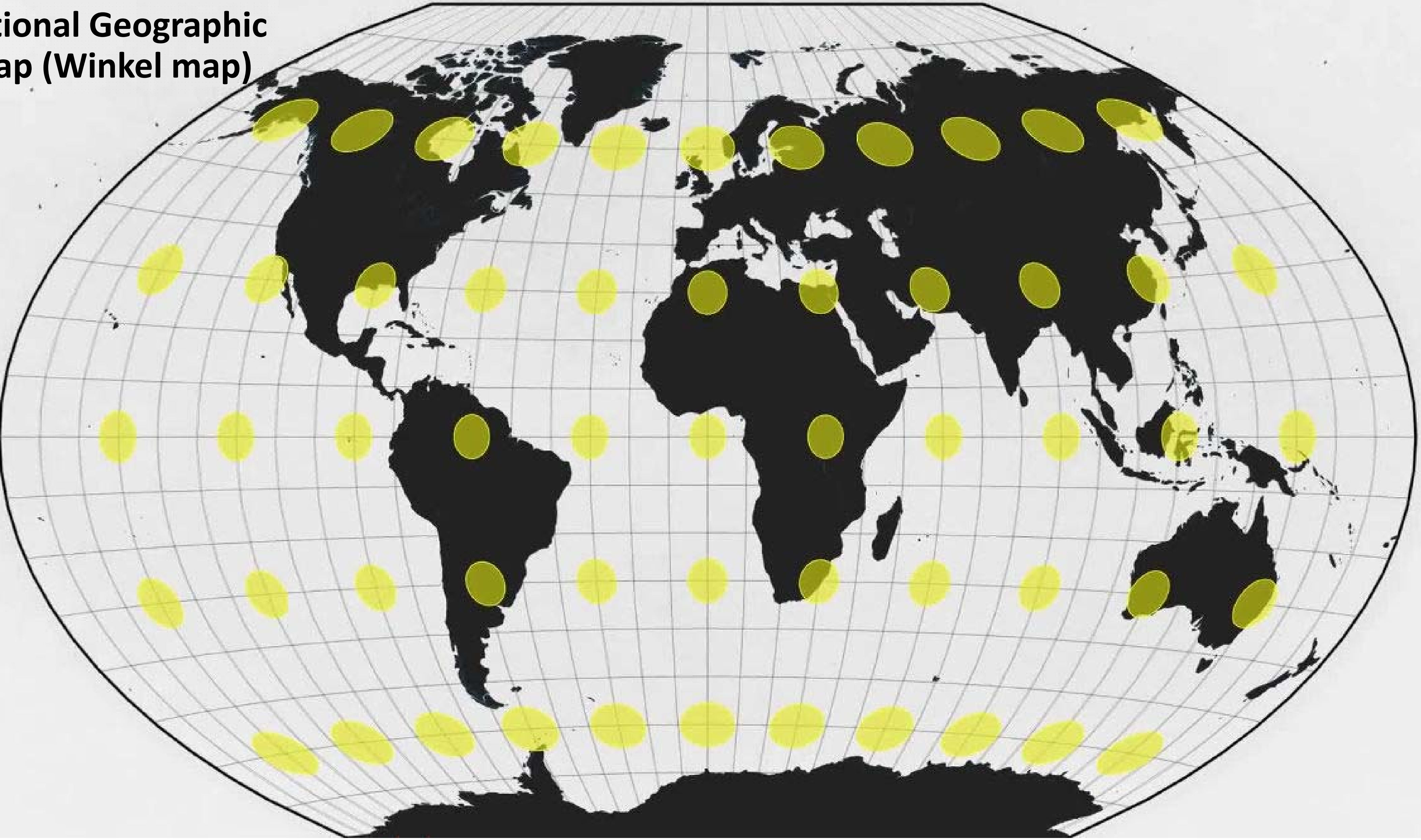
Gall-Peters Map

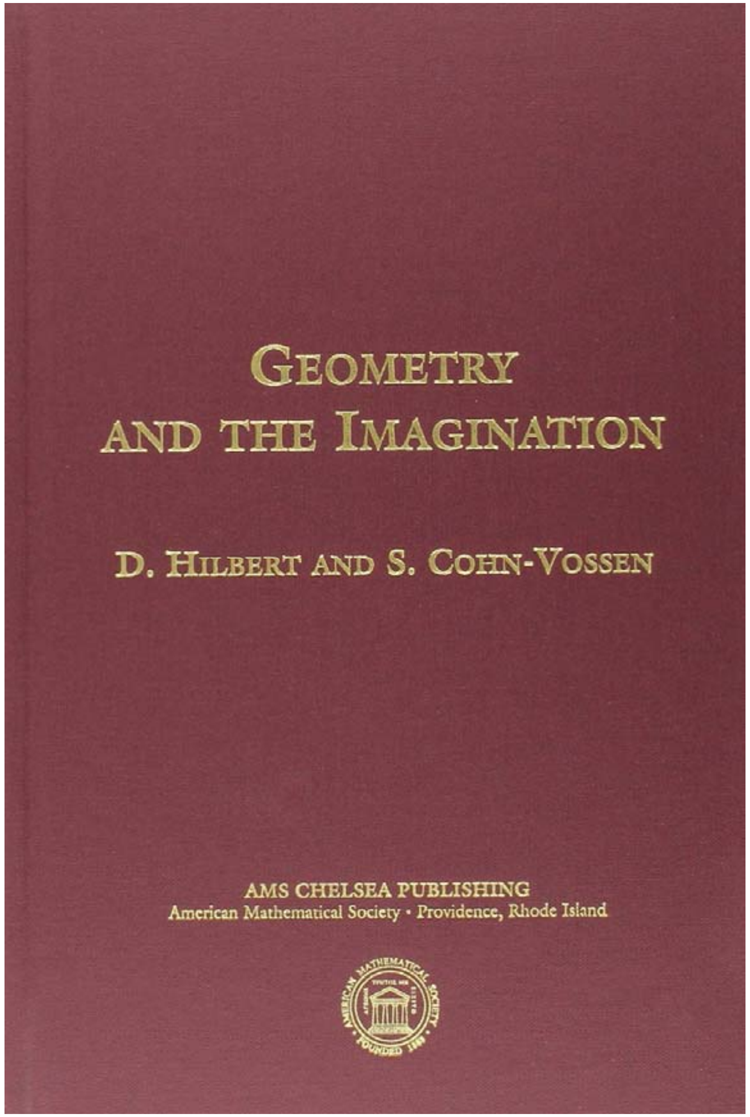
Gall-Peters Map: Greenland *vs* Africa



Relative areas are accurate, but shapes are now distorted

National Geographic Map (Winkel map)





David Hilbert (1862-1943)

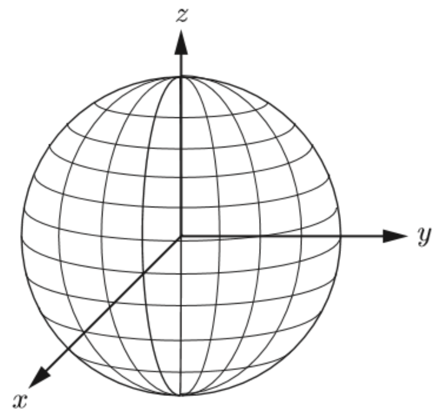
A Hierarchy of Mappings

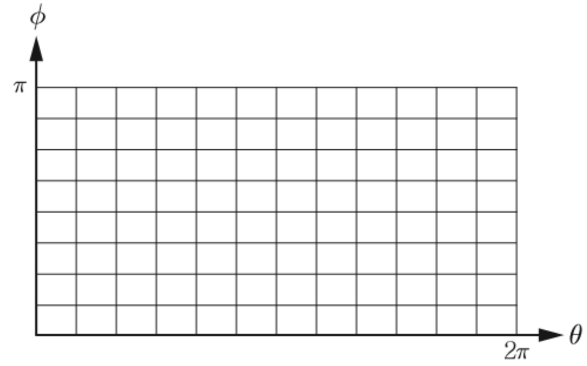
- **Isometry** (distortion-free)
- Area-preserving
- Geodesic-preserving
- Angle-preserving (conformal)
- ....

Calculus on the Sphere

The unit **two-sphere** is parametrized as $x^2 + y^2 + z^2 = 1$. **Spherical coordinates**:

$$
\begin{aligned}
\mathrm{x} &= \cos\theta \sin\phi \\
\mathrm{y} &= \sin\theta \sin\phi \\
z &= \cos\phi
\end{aligned}
$$





Other coordinate parametrizations are possible, e.g., stereographic projection:

$$x = \frac{2u}{1 + u^2 + v^2}, \quad y = \frac{2v}{1 + u^2 + v^2}, \quad z = \frac{-1 + u^2 + v^2}{1 + u^2 + v^2}$$

Calculus on the Sphere

Given a curve $(x(t), y(t), z(t))$ on the sphere, its incremental arclength is

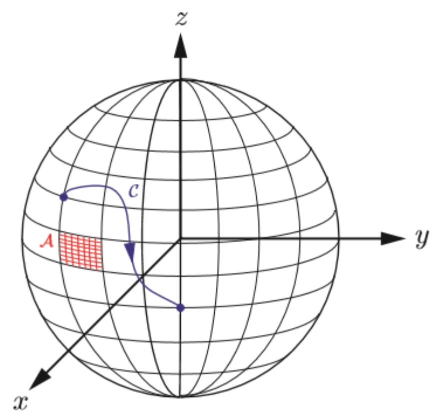
$$ds^2 = dx^2 + dy^2 + dz^2 = d\phi^2 + \sin^2\phi \, d\theta^2$$

$$= [d\theta \; d\phi] \begin{bmatrix} \sin^2\phi & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d\theta \\ d\phi \end{bmatrix}$$

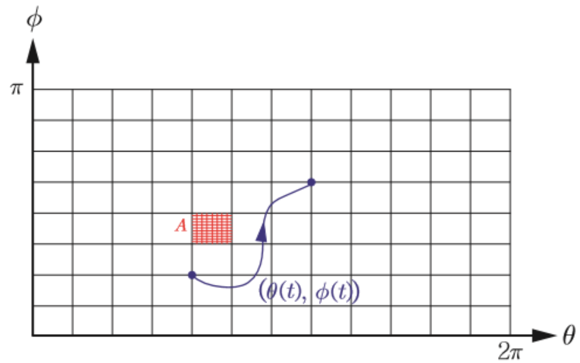
The matrix $G = \begin{bmatrix} \sin^2\phi & 0 \\ 0 & 1 \end{bmatrix}$ is called the **first fundamental form** in classical differential geometry (we'll call it the **Riemannian metric**).

Calculus on the Sphere

Calculating lengths and areas on the sphere using spherical coordinates:



- Length of $\mathcal{C} = \int ds$
$$= \int_0^T \sqrt{\dot{\phi}^2 + \dot{\theta}^2 \sin\phi}\ dt$$
- Area of $\mathcal{A} = \iint_A dA = \iint_A |\sin\phi|\, d\phi\, d\theta$



Note that the **area element** $dA = |\sin\phi|\, d\phi\, d\theta$ is $\sqrt{\det G}\ d\phi\, d\theta$

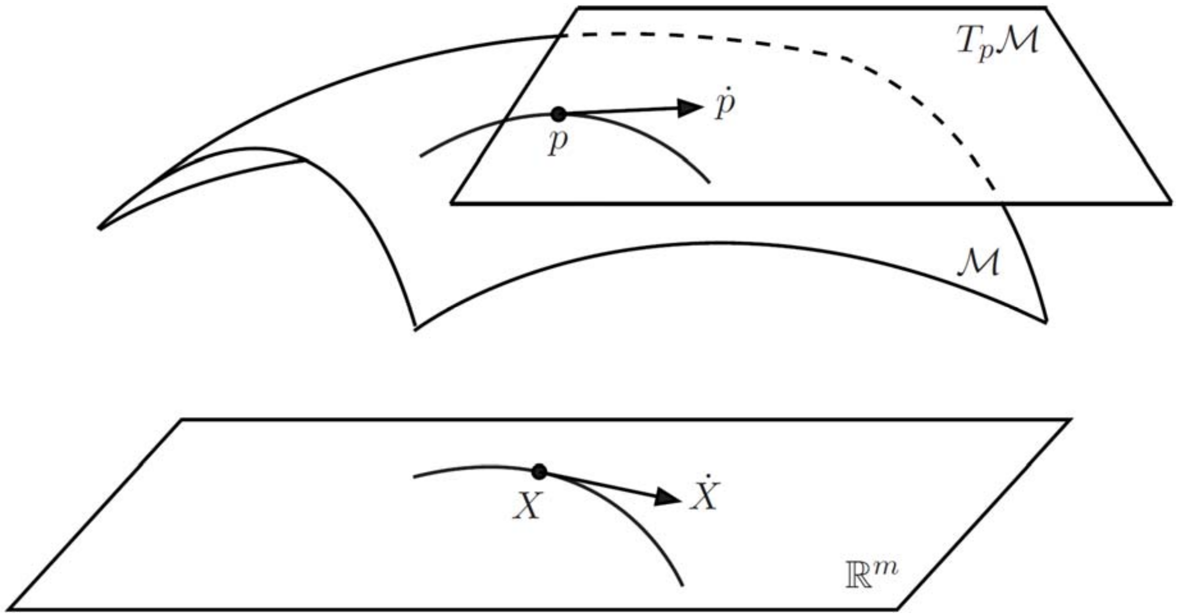
Calculus on the Sphere: The Setup So Far

- **Local coordinates**: (θ, ϕ)
- The **Riemannian metric**: $G(\theta, \phi) = \begin{bmatrix} \sin^2 \phi & 0 \\ 0 & 1 \end{bmatrix}$
- *Note 1:* Other local coordinates are possible.
- *Note 2:* Other choices of Riemannian metric are also possible by defining ds^2 differently, e.g., choose any symmetric positive-definite 3x3 matrix $\left(a_{ij}(x, y, z)\right)$ and set

$$ds^2 = [dx\ dy\ dz] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}$$

Calculus on Riemannian Manifolds

A **differentiable manifold** is a space that is locally diffeomorphic* to Euclidean space (e.g., a multidimensional surface)



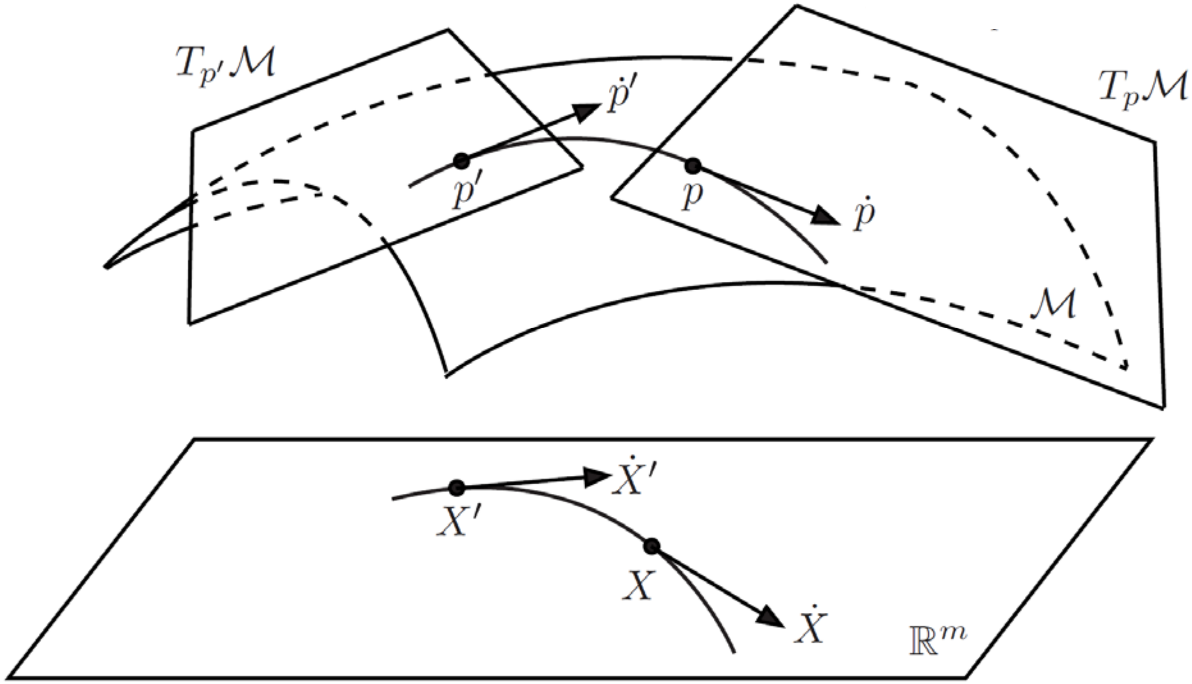
Manifold $\mathcal{M}$

local coordinates x

*Invertible with a differentiable inverse. Essentially, one can be smoothly deformed into the other.

Calculus on Riemannian Manifolds

A **Riemannian metric** is an inner product defined on each tangent space that varies smoothly over $\mathcal{M}$.



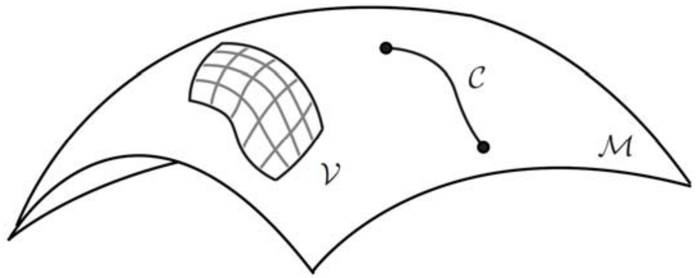
$$ds^2 = \sum_i \sum_j g_{ij}(x)\, dx^i dx^j$$
$$= dx^T G(x)\, dx$$

$G(x) \in \mathbb{R}^{m \times m}$
symmetric positive-definite

Calculus on Riemannian Manifolds

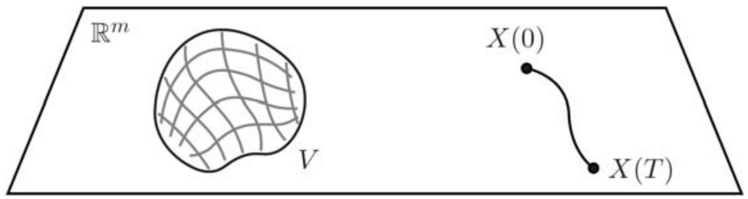
- Length of a curve C on $\mathcal{M}$ (local coordinates $(x_1, \dots, x_m)$):

$$\text{Length} = \int_C ds$$
$$= \int_0^T \sqrt{\dot{x}(t)^T G\big(x(t)\big)\dot{x}(t)}\, dt$$

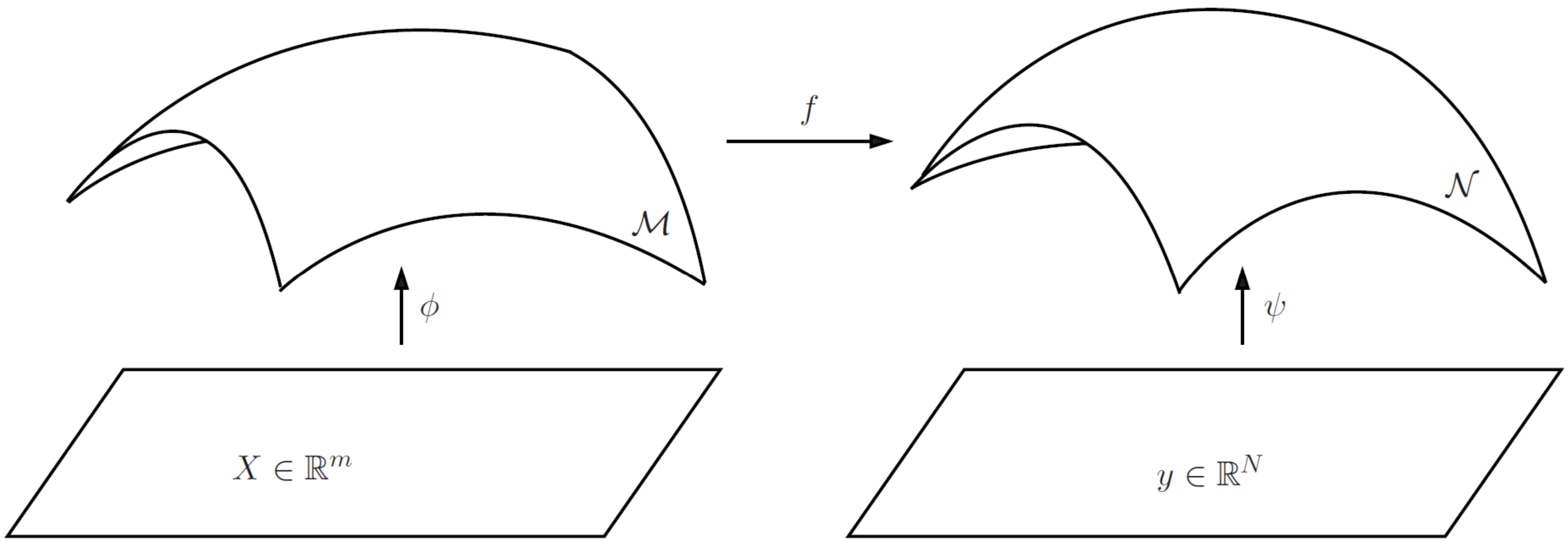


- Volume of a subset $\mathcal{V}$ of $\mathcal{M}$:

$$\text{Volume} = \int_{\mathcal{V}} dV$$
$$= \int \cdots \int_{\mathcal{V}} \sqrt{\det G(x)}\, dx_1 \cdots dx_m$$



Mappings Between Riemannian Manifolds



$$ds^2 = \sum_i \sum_j g_{ij}(x) dx^i dx^j$$

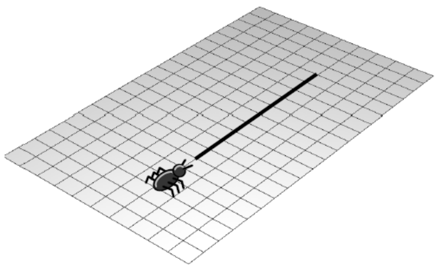
$$ds^2 = \sum_\alpha \sum_\beta h_{\alpha\beta}(y) dy^\alpha dy^\beta$$

Isometry

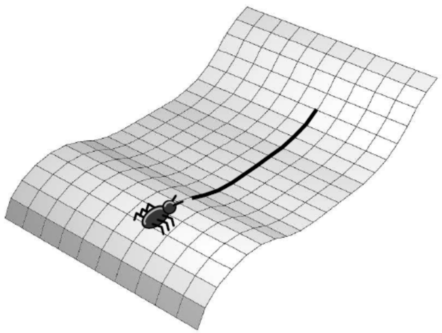
Given two manifolds $\mathcal{M}$ and N, the mapping $f: \mathcal{M} \to N$ is an **isometry** if it preserves distances and angles everywhere:

$$dist_{\boldsymbol{\mathcal{M}}}(x_i, x_j) = dist_{\boldsymbol{N}}(f(x_i), f(x_j)),\ \text{for all } x_i, x_j \text{ in } \mathcal{M}$$

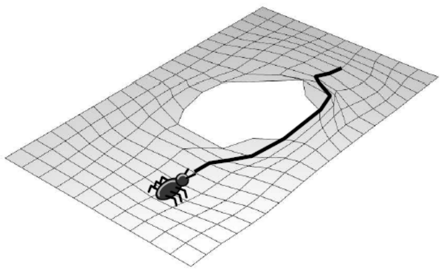
$\mathcal{M}$ and N are then said to be **isometric** to each other; $\mathcal{M}$ can be transformed into N without any stretching or tearing.



Original $\mathcal{M}$



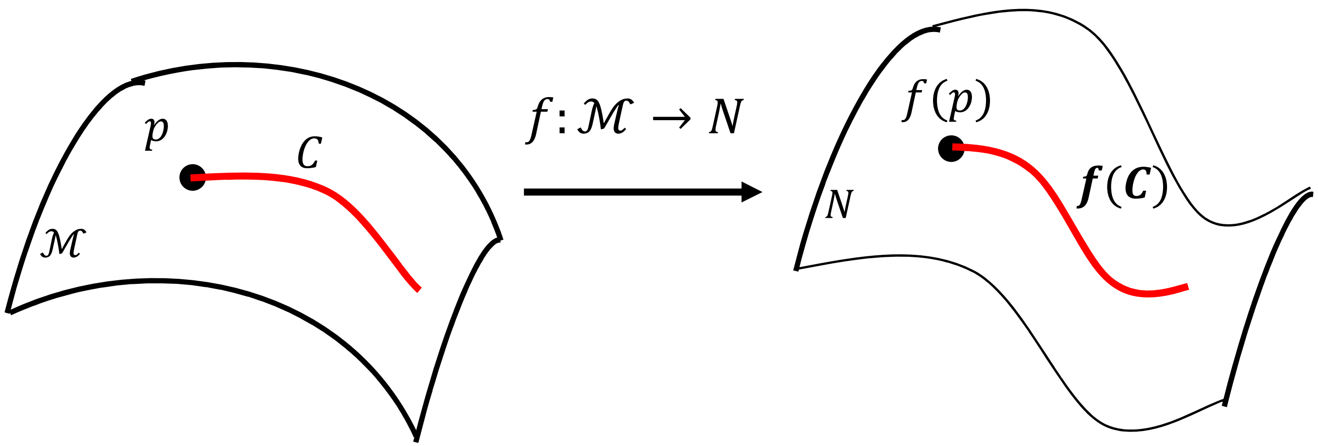
N isometric to $\mathcal{M}$



N not isometric to $\mathcal{M}$

Isometry: Mathematical Formulation

Coordinates x
metric $G(x)$



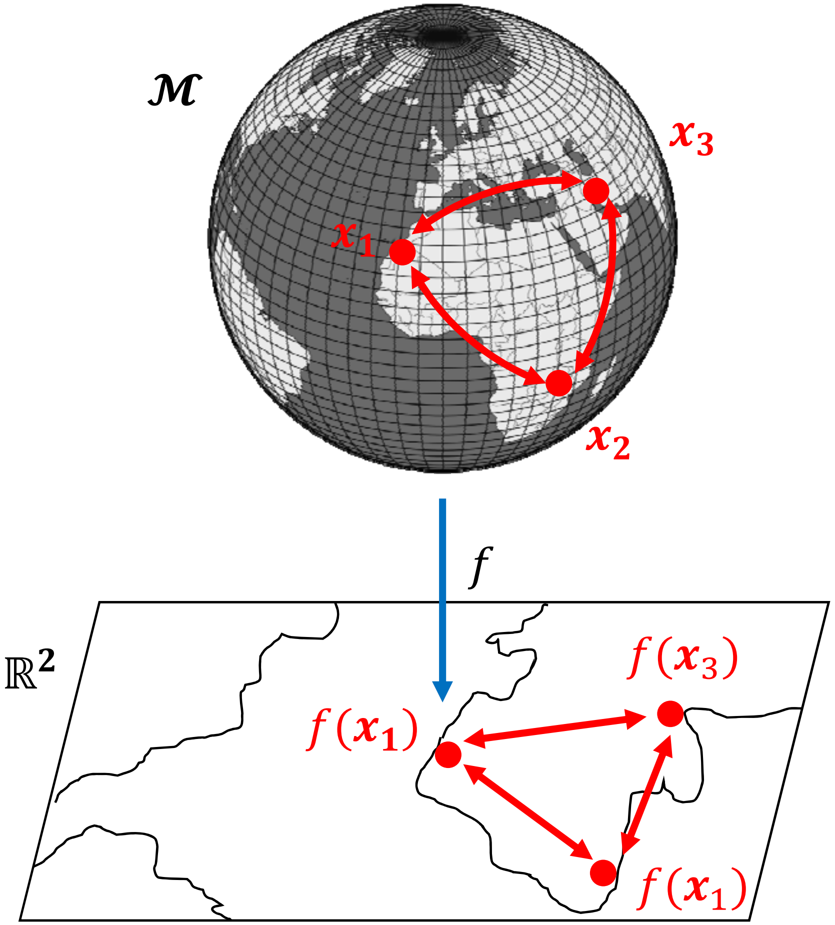
Coordinates f
metric $H(f)$

$$l_C = \int_a^b \sqrt{\dot{x}^T \underline{G(x)} \dot{x}}\, dt \qquad l_{f(C)} = \int_a^b \sqrt{\dot{x}^T \underline{J(x)^T H\big(f(x)\big) J(x)} \dot{x}\, dt}$$

Isometry $\Leftrightarrow \quad l_C = l_{f(C)}$

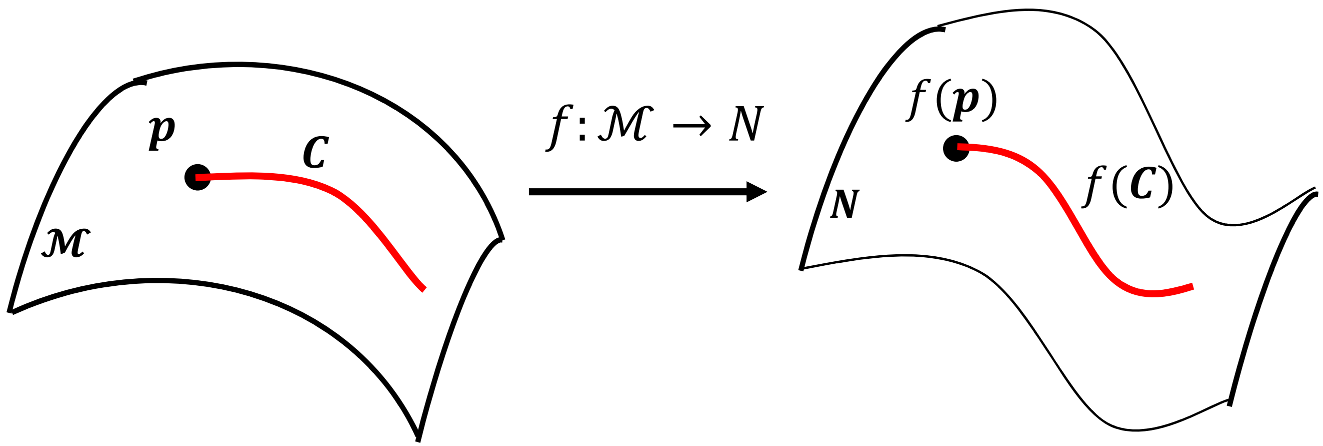
$$G(x) = J(x)^T H\big(f(x)\big) J(x),\ J = \left(\frac{\partial f}{\partial x}\right) \in \mathbb{R}^{n \times m}$$

Isometries and Gaussian Curvature



There is no isometry between manifolds of different Gaussian curvatures. **What's the best one can do in this case?**

Finding Nearly Isometric Maps



Local coordinates x, metric $G(x)$

Local coordinates f, metric $H(f)$

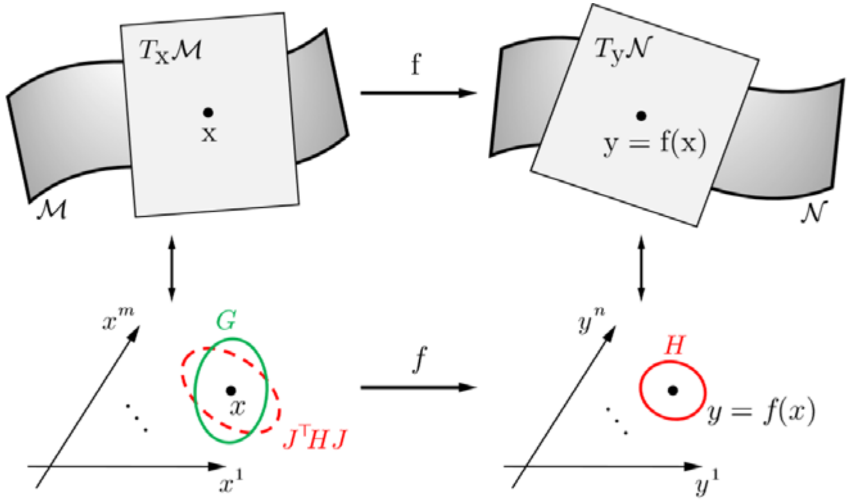
$$\min_f \int_{\mathcal{M}} \mathrm{dist}(G(x), J(x)^T H\big(f(x)\big) J(\mathrm{x}))\, dV$$

Note: The "distance" must be **coordinate-invariant**.

Coordinate-Invariance

This is Spinal Tap (1984)

Coordinate-Invariant Functionals



$(\mathcal{M}, g)$, local coord. $x = (x^1, \cdots, x^m)$
Riemannian metric $G = (g_{ij})$

$(\mathcal{N}, h)$, local coord. $y = (y^1, \cdots, y^n)$
Riemannian metric $H = (h_{\alpha\beta})$

A **coordinate-invariant** functional of $f: \mathcal{M} \to \mathcal{N}$ has the general form

$$D(f) = \int_{\mathcal{M}} \sigma(\lambda_1, \cdots, \lambda_m)\sqrt{\det G}\, dx^1 \cdots dx^m$$

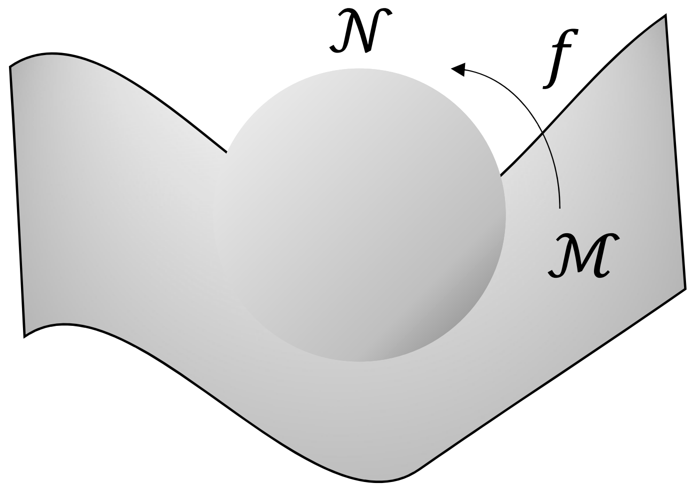
where $\sigma(\cdot)$ is any symmetric function, and $\lambda_1, \cdots, \lambda_m$ are the roots of

$$\det(J^\top HJ - \lambda G) = 0\,, \; J = \left(\frac{\partial f^\alpha}{\partial x^i}\right) \in \mathbb{R}^{n\times m}.$$

Harmonic Maps

- **Intuition**: Take $\mathcal{M}$ to be made of elastic (e.g., rubber) and N to be rigid (e.g., made of steel).

- Wrap the elastic $\mathcal{M}$ so that it covers all of N, and and let $\mathcal{M}$ settle to its elastic equilibrium state. This is the **harmonic map** solution [Eells and Sampson 1964].



Harmonic Maps: Formulation

- $\sigma(\lambda_1, \cdots, \lambda_m) = \sum_{i=1}^{m} \lambda_i$, with boundary conditions $\partial\mathcal{N} = f(\partial\mathcal{M})$
- The harmonic mapping functional is

$$D(f) = \int_{\mathcal{M}} Tr(J(x)^\top H(f(x))J(x)G(x)^{-1})\sqrt{\det G(x)}\, dx^1 \cdots dx^m$$

- Variational equations:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{1}{\sqrt{\det G}} \frac{\partial}{\partial x^i}\left(\frac{\partial f^\alpha}{\partial x^j} g^{ij}\sqrt{\det G}\right) + \sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} g^{ij}\Gamma^\alpha_{\beta\gamma}\frac{\partial f^\beta}{\partial x^i}\frac{\partial f^\gamma}{\partial x^j} = 0$$

where g^{ij} is (i, j) entry of G^{-1}, $\Gamma^\alpha_{\beta\gamma}$ are the Christoffel symbols of the second kind

Examples of Harmonic Maps

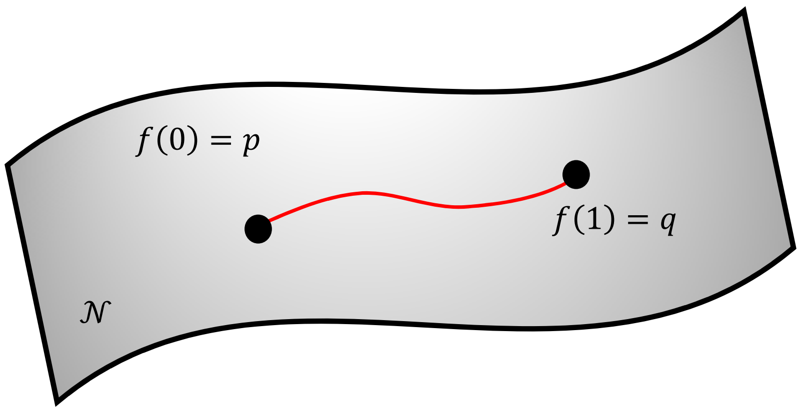
Finding the minimum distortion map from the unit interval [0,1] to itself:

- Find the mapping $f\colon [0,1] \to [0,1]$ that maps the interval [0,1] onto [0,1] so as to minimize $D(f) = \int_0^1 \dot{f}^2 dt$
- Variational equations are $\ddot{f} = 0$, which correspond to the equations for the line $f = t$.

Examples of Harmonic Maps

Geodesics: Given two points on the Riemannian manifold $\mathcal{N}$, find the path of shortest distance connecting these two points:

- Find the mapping $f\colon [0,1] \to \mathcal{N}$ with endpoints specified that minimizes $D(f) = \int_0^1 \dot{f}^\top H(f(t))\dot{f}\,dt$
- Variational equations: $\frac{d^2 f^\alpha}{dt^2} + \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} \Gamma_{\beta\gamma}^{\alpha} \frac{df^\beta}{dt}\frac{df^\gamma}{dt} = 0$

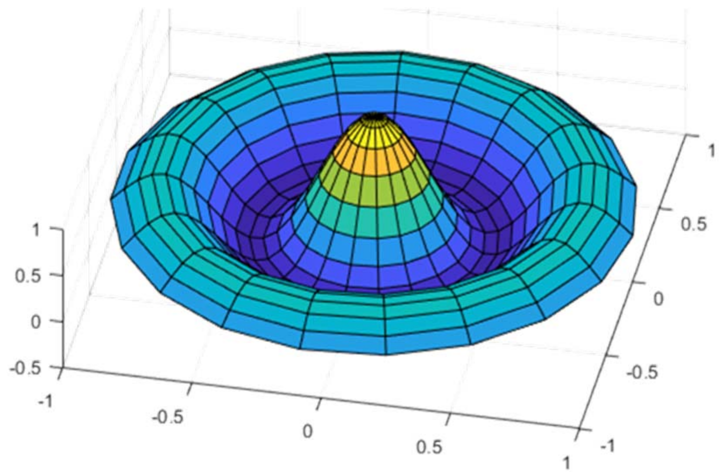


Examples of Harmonic Maps

Harmonic Functions: Find the equilibrium temperature distribution over a planar region with the boundary temperatures specified:

- Find the mapping $f\colon\ \mathbb{R}^2 \to\ \mathbb{R}$ with values for f specified on the boundary of the region.
- Variational equations: $\nabla^2 f = 0$

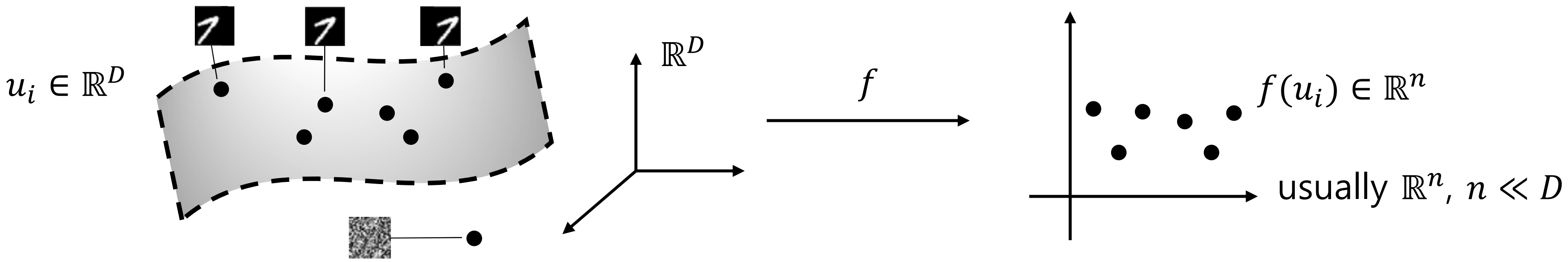
  (Laplace's equation)



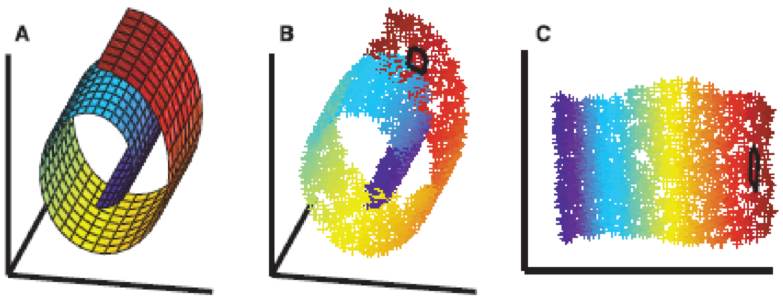
Manifold Learning Revisited

Manifold Learning

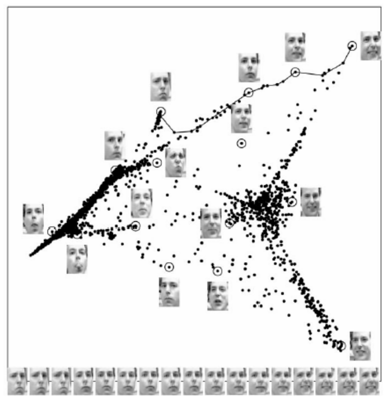
- Find a lower-dimensional, **minimum distortion,** Euclidean representation of high-dimensional data:



- Examples from locally linear embedding (LLE) (Roweis et al. 2000)



Mapping 3-dim data to 2-dim space

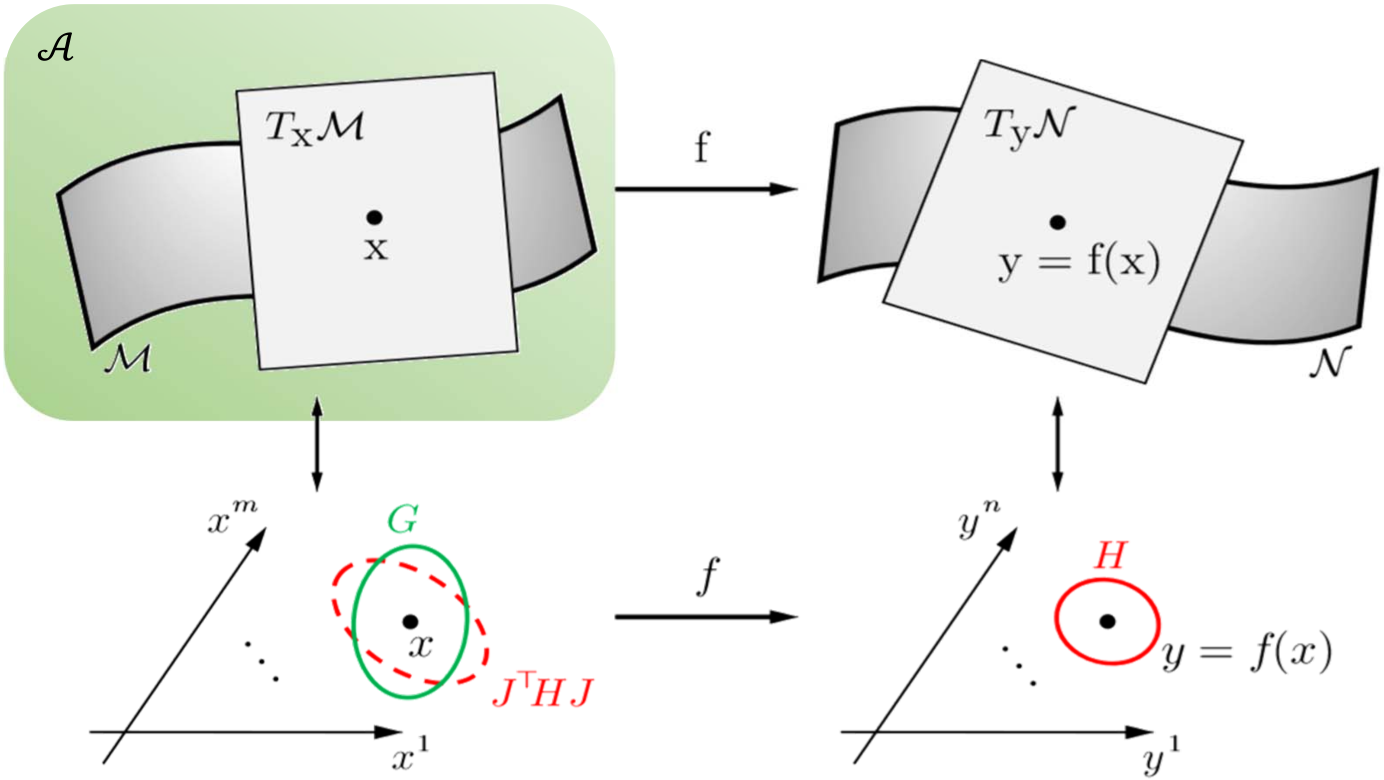


Face images mapped into 2-dim space

Riemannian Manifold Learning

- Recall the general setup of our global distortion measure:

$$\int_{\mathcal{M}} \sigma(\lambda_1, \cdots, \lambda_m)\sqrt{\det G}\, dx^1 \cdots dx^m$$



Riemannian Manifold Learning

Choices need to be made:

- Manifolds $\mathcal{M}$ and N
- Metric G in $\mathcal{M}$
- Metric H in N
- Integrand function $\sigma(\lambda_1, \cdots, \lambda_m)$
- Constraints, boundary conditions
- Discretization method

*$\boldsymbol{J}\boldsymbol{G}^{-1}\boldsymbol{J}^{\top}$ **can be estimated** using $(y_1, \ldots, y_N), y_i = f(x_i)$ from Laplace-Beltrami operator based method

A classification scheme for existing manifold learning algorithms

A roadmap for finding new manifold learning methods and algorithms (for example, the harmonic mapping distortion)

Example: Harmonic Mapping Distortion Details

- Discretized objective function for $\sigma(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \lambda_i$:

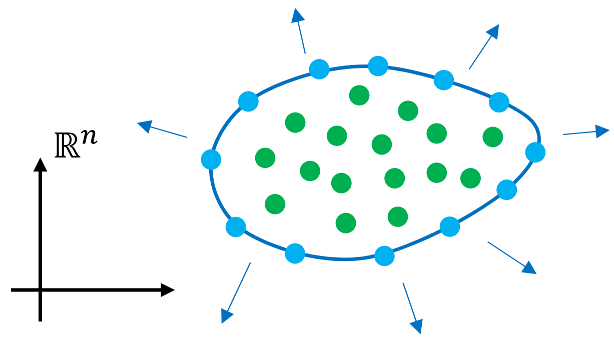
$$\mathcal{D}(Y) = \frac{1}{2}\mathrm{Tr}\big(Y(\widetilde{D} - \widetilde{K})Y^\top\big)$$

$$= \frac{1}{2}\mathrm{Tr}\big(Y_b(\widetilde{D}_{bb} - \widetilde{K}_{bb})Y_b^\top - 2Y_b\widetilde{K}_{br}Y_r^\top + Y_r(\widetilde{D}_{rr} - \widetilde{K}_{rr})Y_r^\top\big)$$

where $Y = [Y_b \ \ Y_r] \in \mathbb{R}^{n\times N}$: embedding points in $\mathbb{R}^n$

$Y_b \in \mathbb{R}^{n\times N_b}$: embedding of boundary points

$$\widetilde{D} = \begin{bmatrix} \widetilde{D}_{bb} & 0 \\ 0 & \widetilde{D}_{rr} \end{bmatrix}, \widetilde{K} = \begin{bmatrix} \widetilde{K}_{bb} & \widetilde{K}_{br} \\ \widetilde{K}_{br}^\top & \widetilde{K}_{rr} \end{bmatrix}$$



- Given Y_b, $Y_r = Y_b W$ for $W = \widetilde{K}_{br}\big(\widetilde{D}_{rr} - \widetilde{K}_{rr}\big)^{-1}$
- If Y_b is unspecified, Y_b can be optimized with respect to other global distortion measures

A Taxonomy of Manifold Learning Algorithms (1)

	G^{-1} (inverse pseudo-metric)	$\sigma(\lambda)$	Volume element	Constraint
LLE (Locally Linear Embedding) **(Roweis et al. 2000)**	Rank-one matrix $\Delta x \Delta x^\top$	$\sum_{i=1}^{m} \lambda_i$	$\rho(x)dx$	$\int_{\mathcal{M}} f(x)f(x)^\top \cdot \rho(x)dx = I$
LE (Laplacian Eigenmap) **(Belkin et al. 2003)**	Kernel-weighted covariance matrix $\int_{\mathcal{M}} k(x,z)(x-z)(x-z)^\top \rho(z)dz$	$\sum_{i=1}^{m} \lambda_i$	$\rho(x)dx$	$\int_{\mathcal{M}} f(x)f(x)^\top \cdot \frac{\rho^2}{\sqrt{\det A}} dx = I$
DM (Diffusion Map) **(Coifman et al. 2006)**	Projected metric from $\mathbb{R}^D$	$\sum_{i=1}^{m} \lambda_i$	$\sqrt{\det G}\, dx$	$\int_{\mathcal{M}} f(x)f(x)^\top \cdot \sqrt{\det G}\, dx = I$

Manifold learning algorithms such as LLE, LE, DM share the similar objective to **harmonic maps** while having equality constraint to avoid trivial solution $f = const. \in \mathbb{R}^d$

- Δx in LLE is local reconstruction error obtained when running the algorithm
- A in LE method represents the projected metric from $\mathbb{R}^D$

A Taxonomy of Manifold Learning Algorithms (2)

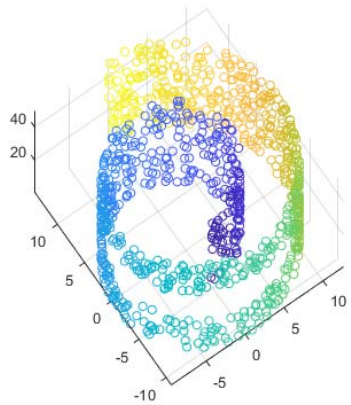
	G^{-1} (inverse pseudo-metric)	$\sigma(\lambda)$	Volume element	Constraint
RR (Riemannian Relaxation) **(McQueen et al. 2016)**	Projected metric from the ambient manifold ($JG^{-1}J$ is estimated from Laplace-Beltrami operator based method)	$\max_i(\lambda_i - 1)^2$	$\sqrt{\det G}\, dx$	
LS (Least-squares spectral distortion)	Same as above	$\sum_{i=1}^{m}(\lambda_i - 1)^2$	$\sqrt{\det G}\, dx$	
PD (P(n) distance metric distortion)	Same as above	$\sum_{i=1}^{m}(\log(\lambda_i))^2$	$\sqrt{\det G}\, dx$	
HM (Harmonic mapping distortion)	Same as above	$\sum_{i=1}^{m}\lambda_i$	$\sqrt{\det G}\, dx$	$\mathrm{f}(\partial\mathcal{M}) = \partial\mathcal{N}$

- LS and PD can be thought of as variants of RR with different $\sigma(\lambda)$
- For HM, further optimization is possible when boundary $\partial\mathcal{N}$ is not specified

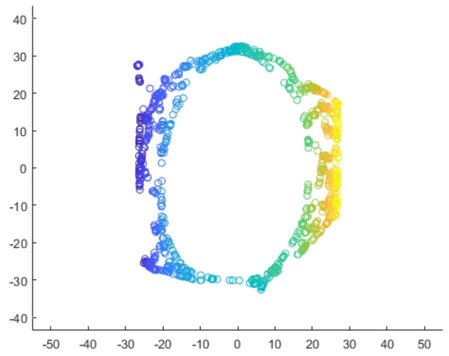
Example: Swiss Roll

Swiss roll data
(2-dim manifold in 3-dim space)

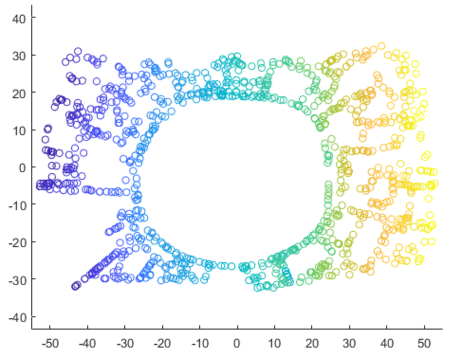
○ : data points



Diffusion map embedding

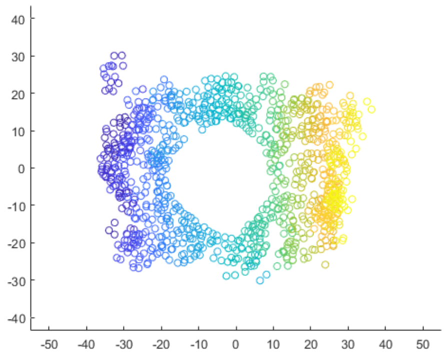


Isomap embedding

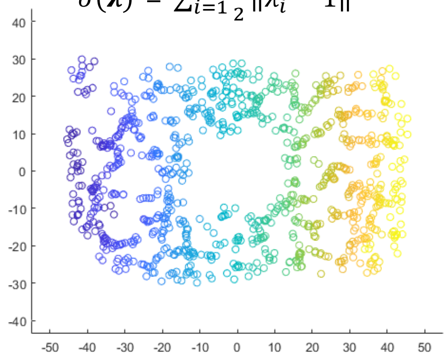


Riemannian distortion results

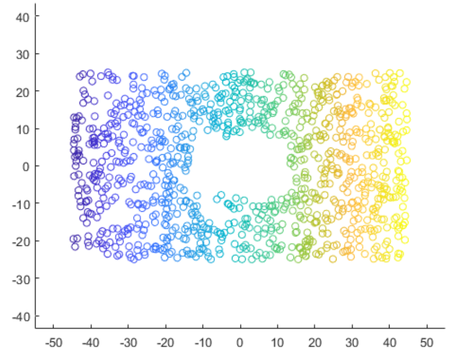
$\sigma(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \frac{1}{2} \|\lambda_i - 1\|^2$



Harmonic mapping with boundary ($\partial\mathcal{N}$) to minimize
$\sigma(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \frac{1}{2} \|\lambda_i - 1\|^2$



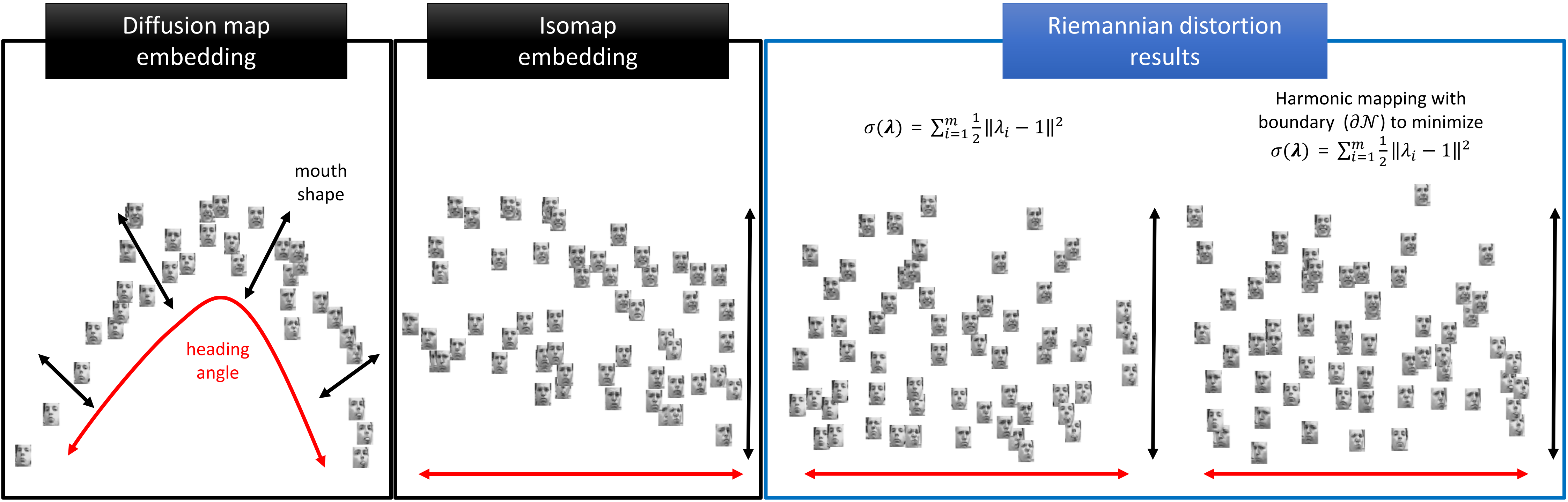
Flattened Swiss roll



Minimum distortion results are closer to flattened swiss roll

Example: Faces

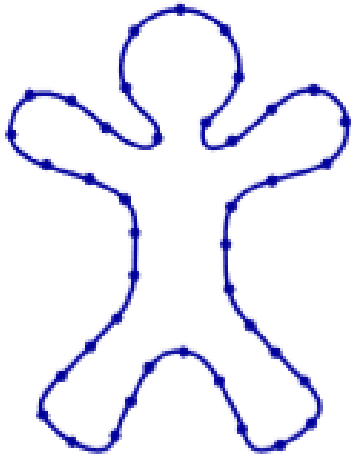
- Face images for the corresponding two-dim. embeddings



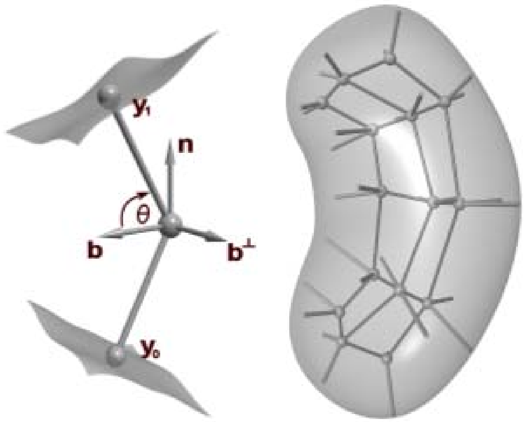
Variations in the face heading angle and mouth shape can be observed along the horizontal and vertical axes respectively

Machine Learning for Non-Euclidean Data

Examples of Non-Euclidean Data



Kendall's shape space
$\mathbb{CP}^{k-2}$

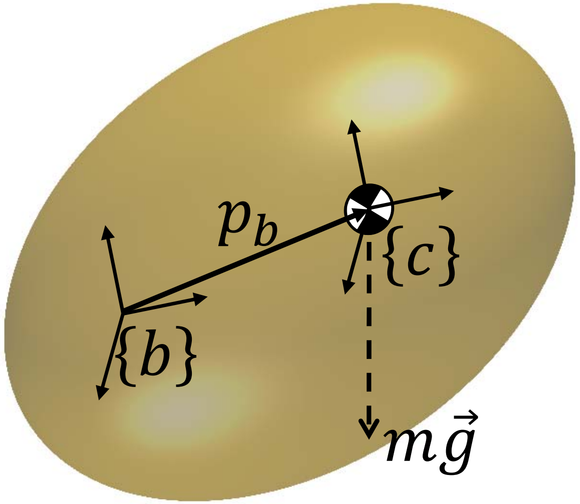


M-Rep $\mathcal{M}^n$
$(\mathcal{M} = \mathbb{R}^3 \times \mathbb{R}^+ \times \mathrm{SO}(3) \times \mathrm{SO}(2))$

Lie Shape G_T^n
$(G_T = \mathrm{SO}(3) \times G_A \times \mathbb{R}^+)$

Rotations SO(3), rigid body motions SE(3), general linear transformations GL(n) and their various subgroups, etc: geometry and distance metrics are now well-established (but still not widely known or used by the community).

Examples of Non-Euclidean Data



- Inertial parameters of a rigid body:

$$\phi = \left[m, h_b^T, I_b^{xx}, I_b^{yy}, I_b^{zz}, I_b^{xy}, I_b^{yz}, I_b^{zx}\right]^T \in \mathbb{R}^{10}$$

(m : mass, $h_b \in \mathbb{R}^3$: first moment, $I_b \in \mathbb{R}^{3\times3}$: moments of inertia)

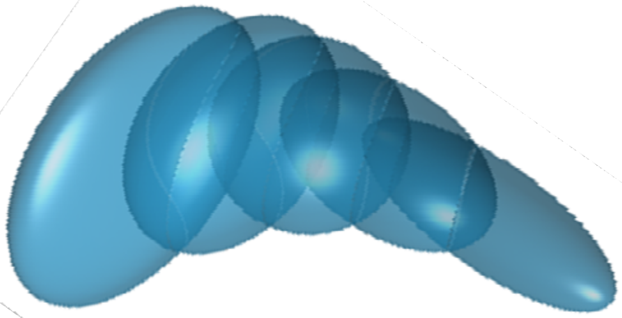
- 4x4 symmetric matrix representation of ϕ:

$$\boldsymbol{\phi \mapsto P(\phi) = \begin{bmatrix} \frac{1}{2} tr(I_b) \cdot \mathbb{I} - I_b & h_b \\ h_b^T & m \end{bmatrix}} \in \mathbb{R}^{4\times4},$$

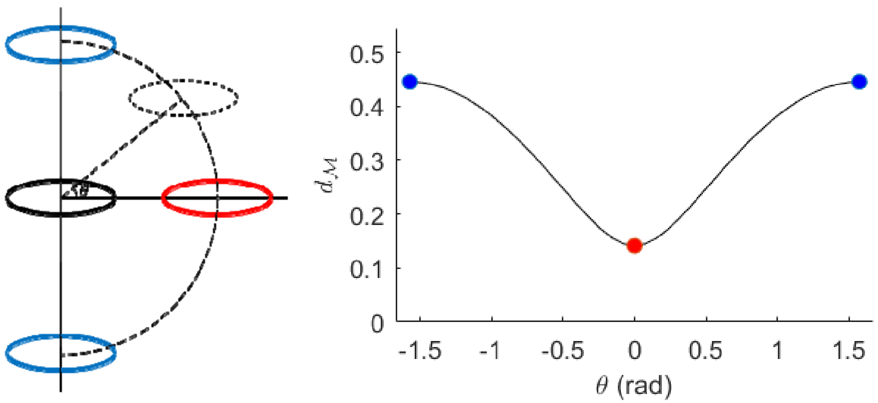
should be **positive definite, i.e., $\boldsymbol{P(\phi) > 0}$.**

P(n): The space of $n \times n$ symmetric positive-definite matrices

Natural Distance on P(n)



Geodesic path on P(4)



Geodesic Distances between Pairs of Inertial Parameters

- **Affine-invariant** metric on $\phi \in \mathbb{R}^{10}$:

$$ds^2 = \frac{1}{2} tr\left((P^{-1}dPP^{-1}dP)^2\right),$$
$$(P = P(\phi) > 0)$$

- Geodesic distance on P(n):

$$d_{\mathcal{M}}(\phi_1, \phi_2)^2 = d_{\mathcal{P}(4)}(P_1, P_2)^2$$
$$= \sum_{i=1}^{4} (\log \lambda_i (P_1^{-1} P_2))^2$$

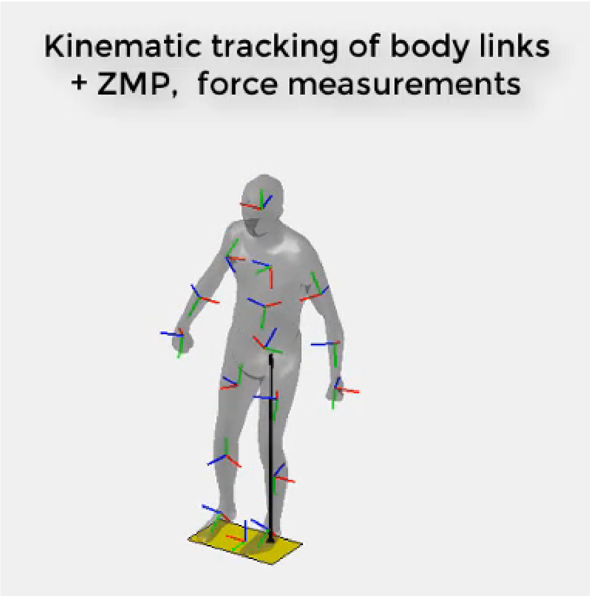
✓ **Well-defined** on positive definite matrix manifold $P(\phi) \in \mathcal{P}(4)$

✓ **Invariant** to reference frames, physical units

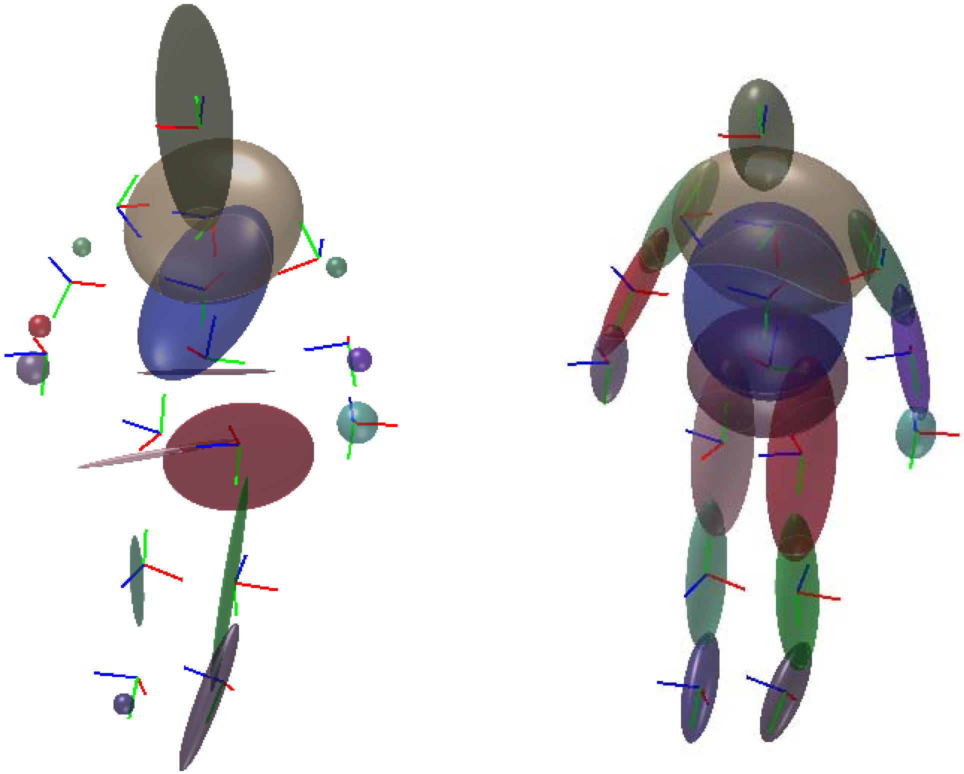
✓ **Dimensionless**

✓ Better encodes **natural distance between positive mass distributions**

Example: Human Dynamic Modeling



- **High dimensional system**
- **Insufficient, noisy measurements**

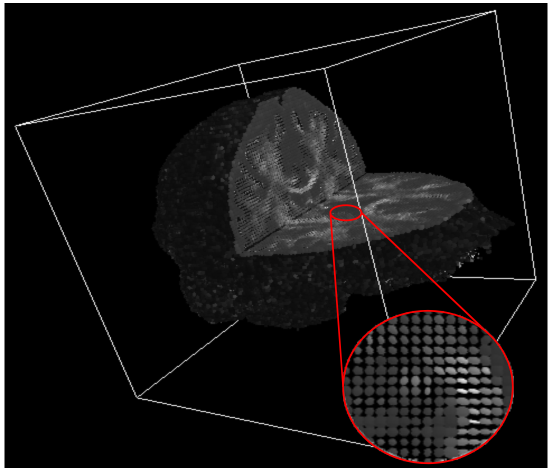


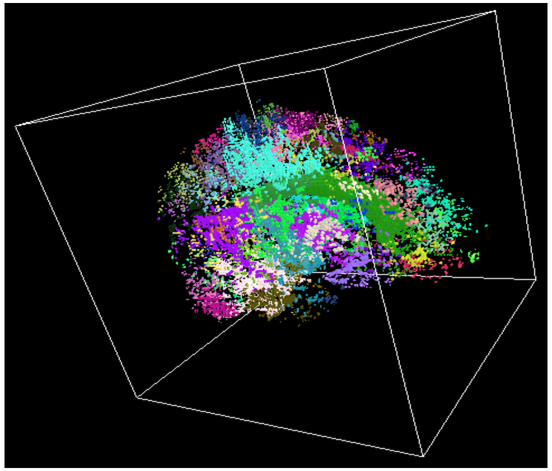
Existing Vector Space Methods **Geometric Method**

T. Lee, F. C. Park, "A Geometric Algorithm for Robust Multibody Inertial Parameter Identification," RA-Letters, 2018

T. Lee, P. M. Wensing, F. C. Park, "Geometric Robot Dynamic Identification: A Convex Programming Approach," submitted to TRO, 2018

Examples of Non-Euclidean Data

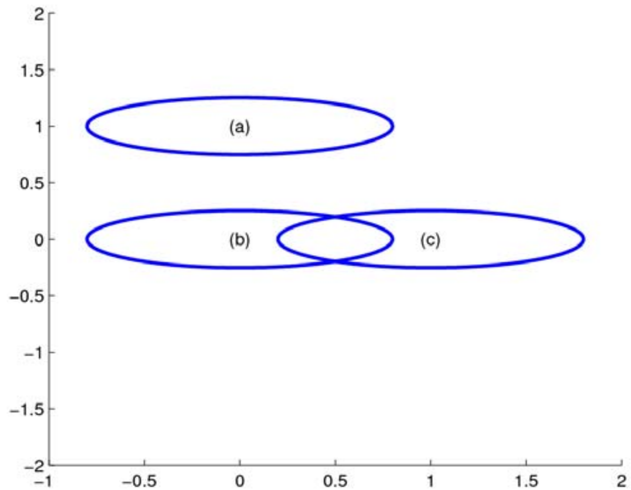




Diffusion tensor images (DTI)

Each voxel is a 3D multivariate normal distribution: the mean indicates the position, while the covariance indicates the direction of diffusion of water molecules. Segmentation of a DTI image requires a metric on the manifold of **multivariate Gaussian distributions.**

Geometry of DTI Segmentation

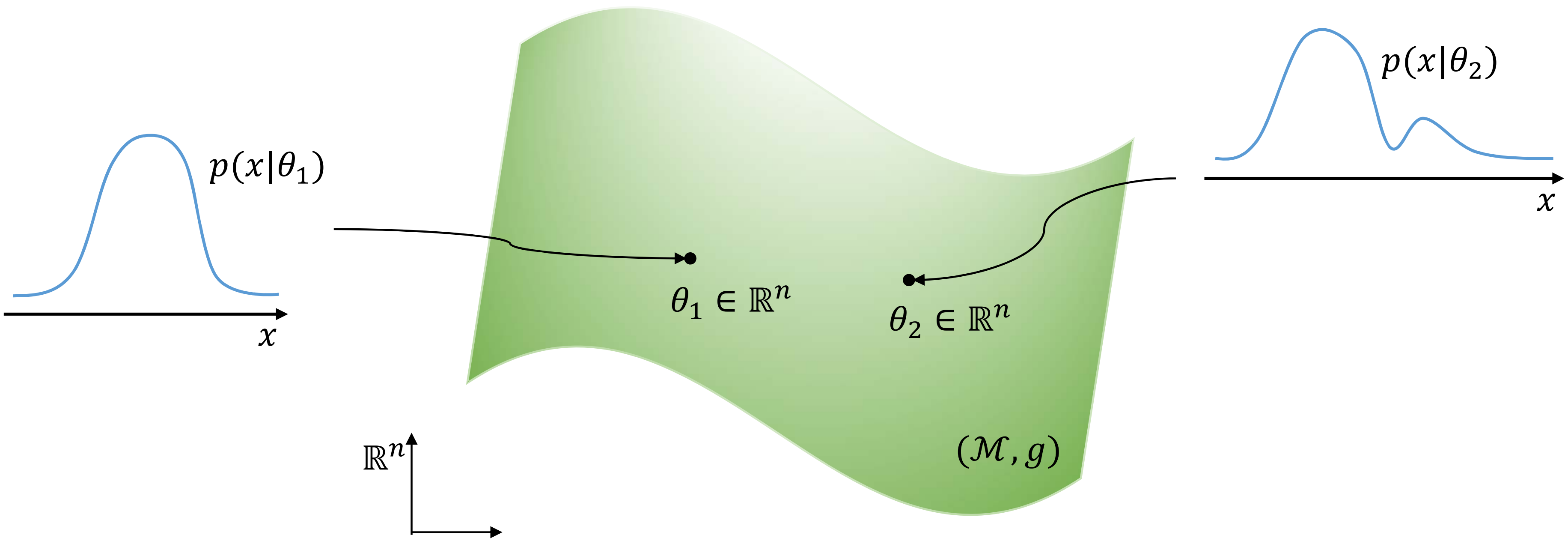


In this example, water molecules are able to move more easily in the *x*-axis direction. Therefore, diffusion tensors (b) and (c) are closer than (a) and (b)

Using the standard approach of calculating distances on the means and covariances separately, and summing the two for the total distance, results in dist(a,b) = dist(b,c), which is unsatisfactory.

Geometry of Statistical Manifolds

An n-dimensional statistical manifold $\mathcal{M}$ is a set of probability distributions parametrized by some smooth, continuously-varying parameter $\theta \in \mathbb{R}^n$.



Geometry of Statistical Manifolds

- The Fisher information defines a Riemannian metric g on a statistical manifold $\mathcal{M}$:

$$g_{ij}(\theta) = \mathbb{E}_{x\sim p(.|\theta)}\left[\frac{\partial \log p(x|\theta)}{\partial \theta_i}\frac{\partial \log p(x|\theta)}{\partial \theta_j}\right]$$

- Connection to KL divergence:

$$D_{KL}\big(p(.\,|\theta)||p(.\,|\theta + d\theta)\big) = \frac{1}{2}d\theta^T g(\theta)d\theta + o(\|d\theta\|^2)$$

Geometry of Gaussian Distributions

- The manifold of Gaussian distributions $\mathcal{N}(n)$

$$\mathcal{N}(n) = \{\theta = (\mu, \Sigma) \mid \mu \in \mathbb{R}^n, \Sigma \in \mathcal{P}(n)\},$$

where $\mathcal{P}(n) = \{P \in \mathbb{R}^{n\times n} | P = P^T, P \succ 0\}$

- Fisher information metric on $\mathcal{N}(n)$

$$ds^2 = d\theta^T g(\theta) d\theta = d\mu^T \Sigma^{-1} d\mu + \frac{1}{2} tr\left((\Sigma^{-1} d\Sigma)^2\right)$$

- Euler-Lagrange equations for geodesics on $\mathcal{N}(n)$

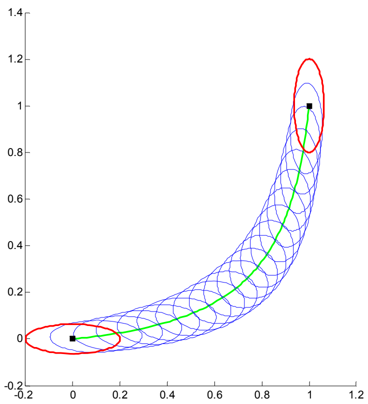
$$\frac{d^2\mu}{dt^2} - \frac{d\Sigma}{dt}\Sigma^{-1}\frac{d\mu}{dt} = 0$$

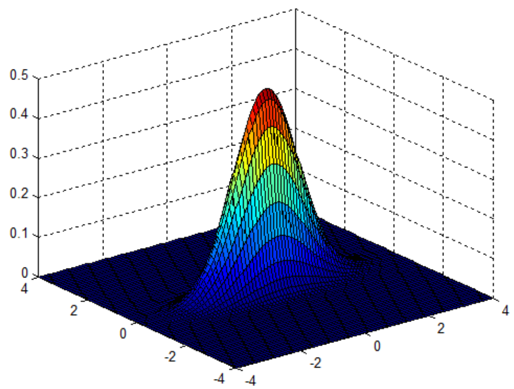
$$\frac{d^2\Sigma}{dt^2} + \frac{d\mu}{dt}\frac{d\mu^T}{dt} - \frac{d\Sigma}{dt}\Sigma^{-1}\frac{d\Sigma}{dt} = 0$$

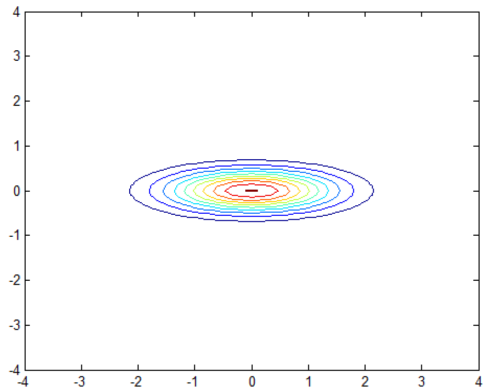
Geometry of Gaussian Distributions

- Geodesic Path on $\mathcal{N}(2)$

$$\mu_0 = \begin{bmatrix}0\\0\end{bmatrix}, \Sigma_0 = \begin{bmatrix}1 & 0\\0 & 0.1\end{bmatrix}, \qquad \mu_1 = \begin{bmatrix}1\\1\end{bmatrix}, \Sigma_1 = \begin{bmatrix}0.1 & 0\\0 & 1\end{bmatrix}$$







Restriction to Covariances

- Fisher information metric on $\mathcal{N}(n)$ with fixed mean $\bar{\mu}$

$$ds^2 = \frac{1}{2} tr\left((\Sigma^{-1} d\Sigma)^2\right)$$

Affine-invariant metric on $\mathcal{P}(n)$

- Invariant under general linear group $GL(n)$ action

$$\Sigma \to S^T \Sigma S,\ S \in GL(n)$$

which implies **coordinate invariance**.

- **Closed-form** geodesic distance

$$d_{\mathcal{P}(n)}(\Sigma_1, \Sigma_2) = \left[\sum_{i=1}^{n} (\log \lambda_i(\Sigma_1^{-1}\Sigma_2))^2\right]^{1/2}$$

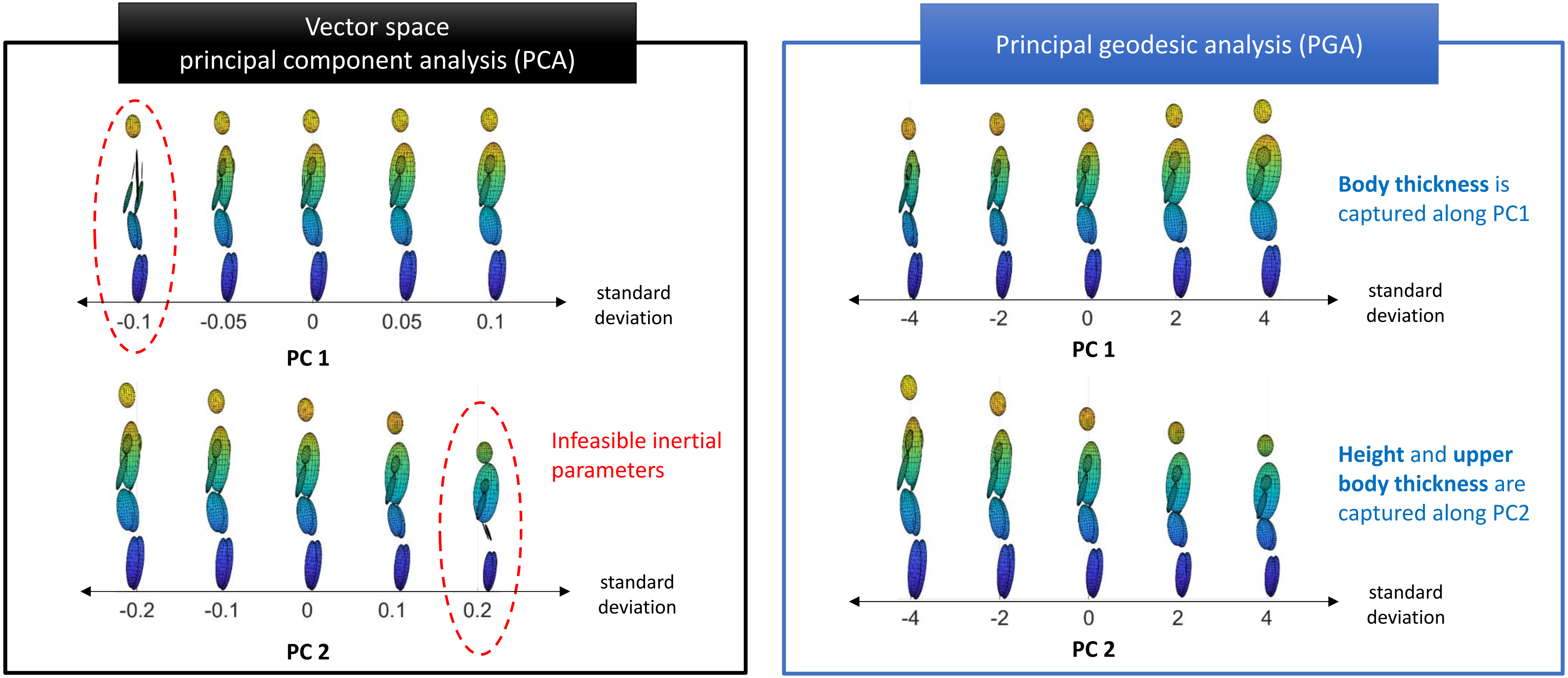
Results of Segmentation for Brain DTI

Using covarianceand Euclidean distance

Using MND distance

Example: Human Mass-Inertia Data

- Manifold learning for human mass-inertia data:



Concluding Remarks

Concluding Remarks

- ML for non-Euclidean data is receiving greater attention from the ML research community:
    - Application to autoencoders;
    - CNNs for geometric data;
- Many problems in engineering are analogous to trying to **fit a square peg into a round hole.**
- Often the things we work with are not vectors, but elements of a manifold.
- The geometric methods and distortion measures described in this talk can be helpful in addressing such problems.