Gaussian process on Riemannian manifolds

Alison Pouplin - GPSS 2025

Gaussian Processes

$$f : X \rightarrow Y$$

Earth	Temperature
Molecule	Energy
EEG	Brain activity

Inputs	Robot mouvement
Time	Motion capture
Age	Shape bone

$$f : \mathscr{M} \to \mathbb{R}^q$$

Manifold-valued inputs

Objective: finding a well defined kernel!

Extrinsic kernels

Lin, Lizhen, Niu Mu, Pokman Cheung, and David Dunson., et al. "Extrinsic Gaussian Processes for Regression and Classification on Manifolds." Bayesian Analysis 14.3 (2019): 887-906.

Naive generalisation

Aasa Feragen, Francois Lauze, and Soren Hauberg. "Geodesic exponential kernels: When curvature and linearity conflict." CVPR 2015.

Intrinsic kernels

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, Marc Deisenroth. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS 2020*.

$$f : \mathbb{R}^d \to \mathscr{M}$$

Manifold-valued outputs

Objective: wrapping everything correctly on the manifold!

Wrapped GPs

Anton Mallasto and Aasa Feragen. "Wrapped Gaussian process regression on Riemannian manifolds." *CVPR* 2018.

Wrapped GPLVMs

Anton Mallasto, Soren Hauberg and Aasa Feragen. "Probabilistic Riemannian submanifold learning with wrapped Gaussian process latent variable models." *AISTATS* 2019.

WGPLVM with the pullback metric

Leonel Rozo, Miguel González-Duque, Noemie Jaquier, Soren Hauberg, "Riemann-2: Learning Riemannian Submanifolds from Riemannian Data." *AISTATS 2025*

Geometric interlude

Smooth manifold: *M*

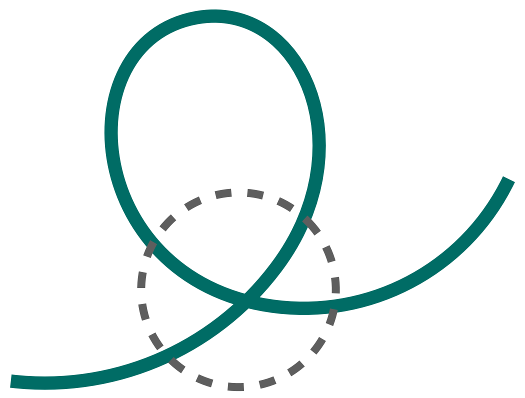
A smooth manifold, also called differentiable manifold, is a topological manifold on which we can perform calculus.

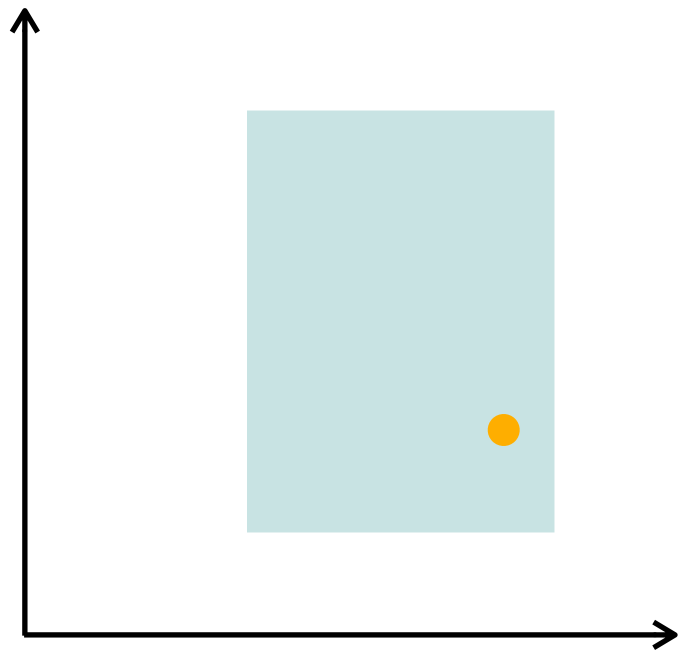
What is a topological manifold?

Second Countable
It has a countable collections of open sets that cover the space.

Hausdorff
Any two distinct points can be separated by neighborhoods that do not overlap.

Locally Euclidean
For every point, there exists a neighborhood that is homeomorphic to an open subset of R^n





$\mathbb{R}^n$

Smooth manifold: *M*

A smooth manifold, also called differentiable manifold, is a topological manifold on which we can perform calculus.

What is a topological manifold?

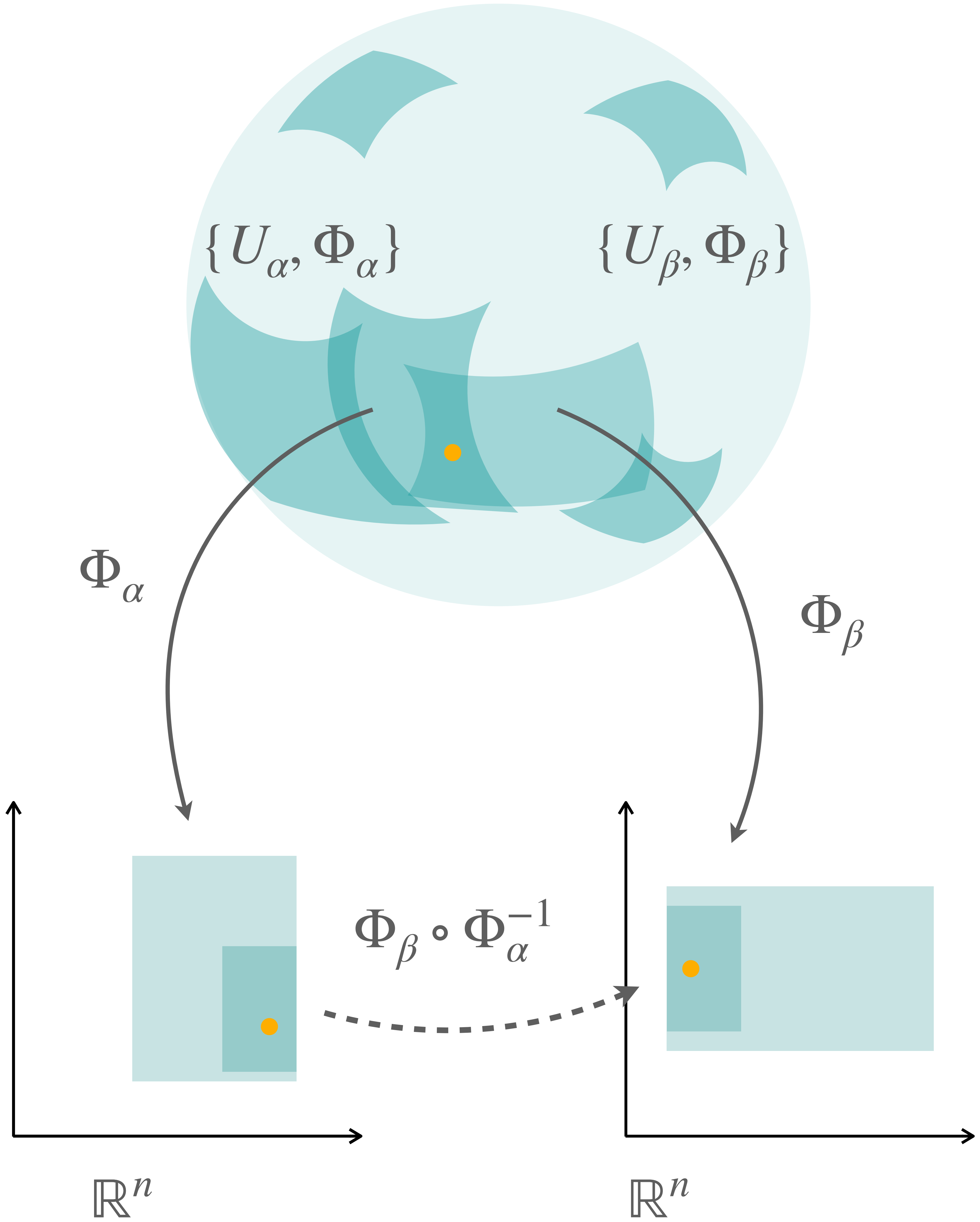
Second Countable
It has a countable collections of open sets that cover the space.

Hausdorff
Any two distinct points can be separated by neighborhoods that do not overlap.

Locally Euclidean
For every point, there exists a neighborhood that is homeomorphic to an open subset of R^n

Performing calculus?

The manifold M can be seen an **atlas of charts** $\{(U_a, \Phi_a)\}$ where $U_a \subset M$ is homeomorphic to an open subset of R^n .
For any overlapping charts $\{(U_a, \Phi_a)\}$ and $\{(U_b, \Phi_b)\}$, the transition map $\Phi_b \circ \Phi_a^{-1}$ between R^n-open sets is infinitely **differentiable**.



Smooth manifold: *M*

A smooth manifold, also called differentiable manifold, is a topological manifold on which we can perform calculus.

Riemannian metric: g

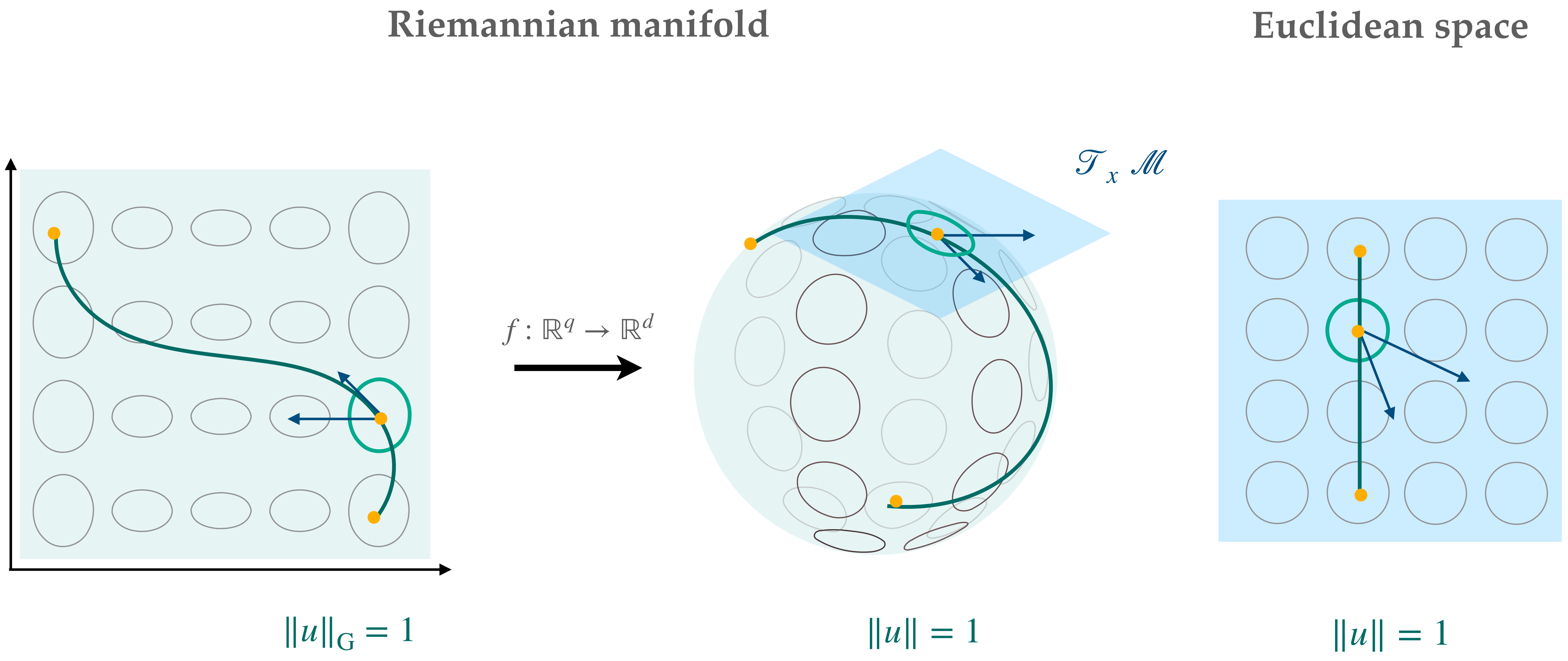
A map assigning at each point p, an inner product: $g_p(u,v)=u^T \mathbf{G_p} v$, with $\mathbf{G_p}$ a symmetric positive definite bilinear map (matrix, called metric tensor).

Riemannian norm

Positive:	$\|x\|_G \geq 0$
Definite:	$\|x\|_G = 0 \implies x = 0$
Symmetric:	$\|x\|_G = \|-x\|_G$
Triangle inequality:	$\|x+y\|_G \leq \|x\|_G + \|y\|_G$

Length and Energy functional

$$\mathscr{L}(\gamma) = \int \|\dot{\gamma}(t)\|_G \, dt \qquad \mathscr{E}(\gamma) = \frac{1}{2} \int \|\dot{\gamma}(t)\|_G^2 \, dt$$



Smooth manifold: *M*

A smooth manifold, also called differentiable manifold, is a topological manifold on which we can perform calculus.

Riemannian metric: g

A map assigning at each point p, an inner product: $g_p(u,v)=u^T \mathbf{G_p} v$, with $\mathbf{G_p}$ a symmetric positive definite bilinear map (matrix, called metric tensor).

Levi-Civita connection: ∇

It is the unique covariant derivative (directional derivative) of a tensor field defined on a manifold, that is compatible with the Riemannian metric.

$$\nabla_v u = v^j(\partial_j u^i)e_i + u^i v^j \Gamma^k_{ij} e_k$$

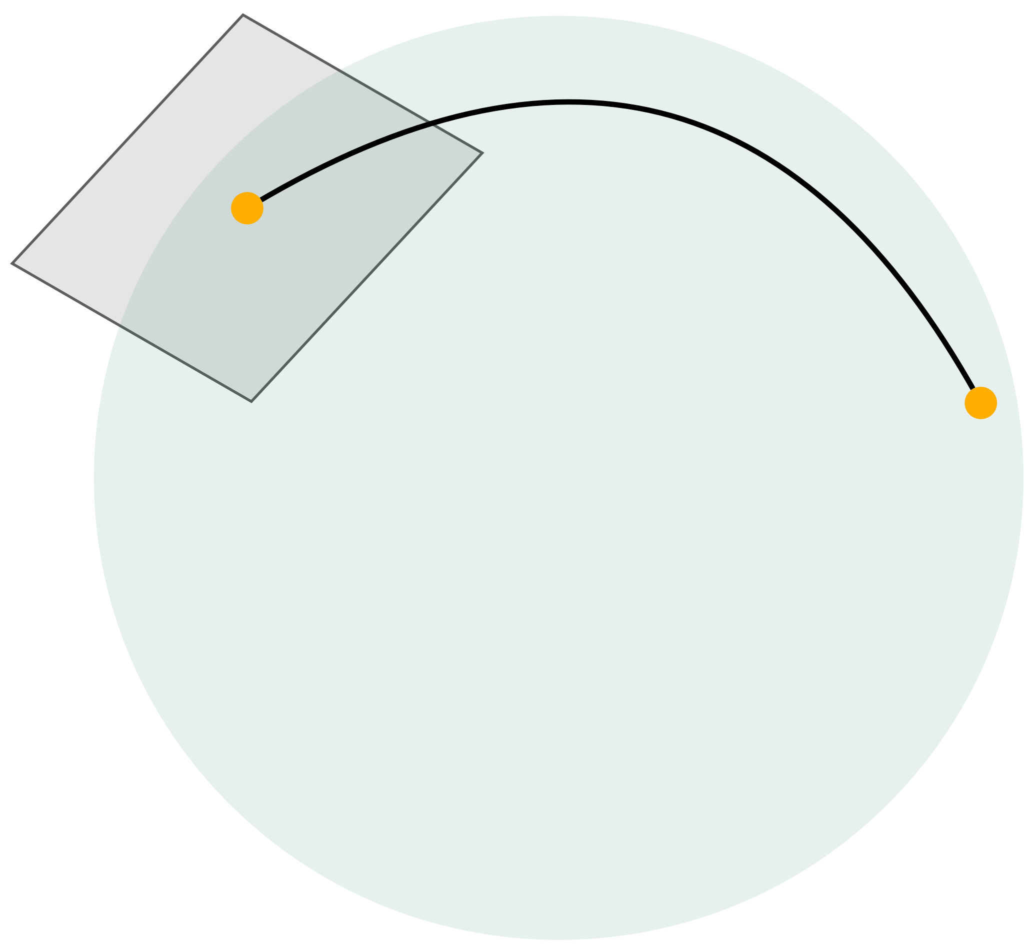
$$[x, y] = \nabla_x y - \nabla_y x$$

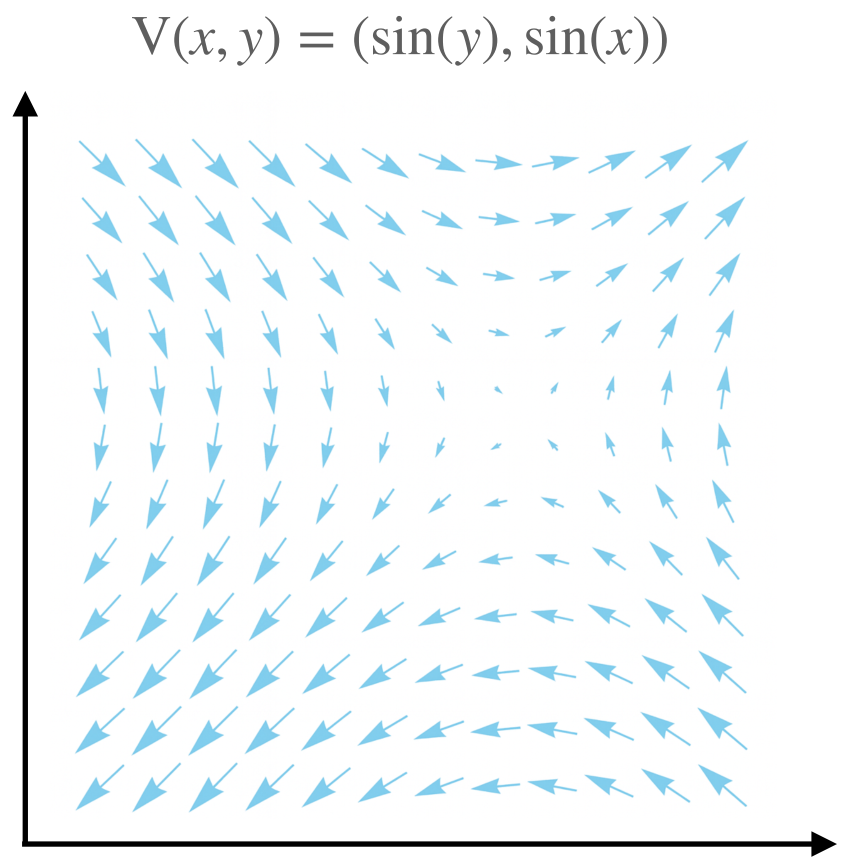
$$\nabla_x g(y, z) = g(\nabla_x y, z) + g(y, \nabla_x z) \qquad (\nabla_k g_{ij} = 0)$$

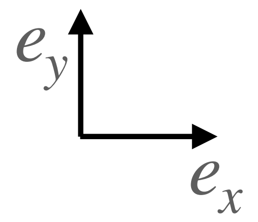
Laplace Beltrami Operator: Δ

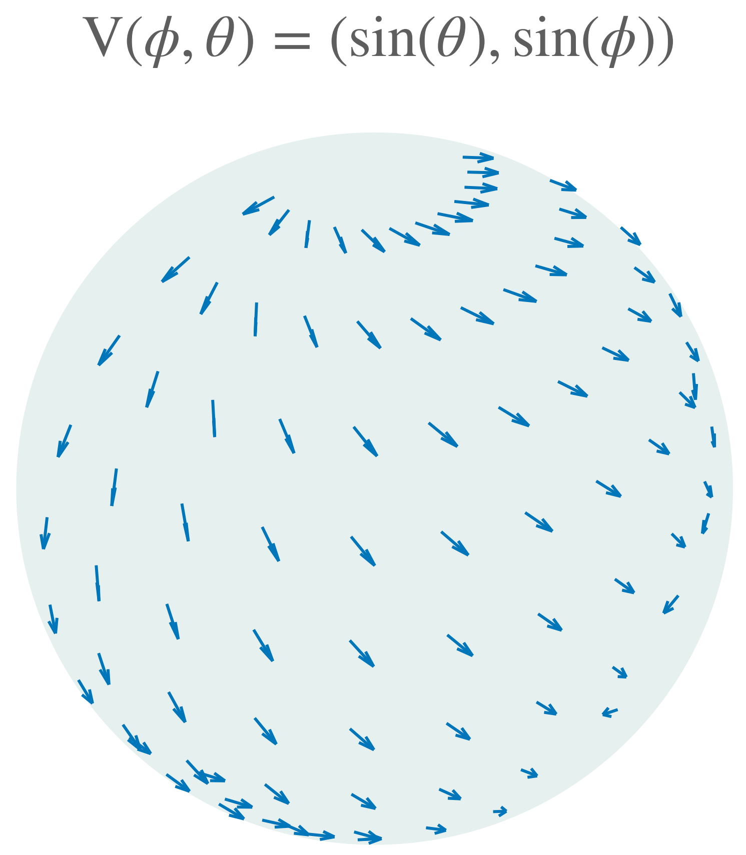
Generalisation of Laplacian on Riemannian manifolds.

$$\Delta f = \nabla \cdot \nabla f \qquad \Delta f = \frac{1}{\sqrt{|G|}}\partial_i\left(\sqrt{G}g^{ij}\partial_j f\right)$$









$$u = u^i e_i \qquad v = v^j e_j \qquad \nabla_v u = v^j \partial_j (u^i) e_i$$

$$\nabla_v u = v^j \partial_j (u^i e_i) = v^j(\partial_j u^i e_i + u^i \partial_j e_i)$$

$$= v^j(\partial_j u^i) e_i + u^i v^j \Gamma^k_{ij} e_k$$

Smooth manifold: $\mathscr{M}$

A smooth manifold, also called differentiable manifold, is a topological manifold on which we can perform calculus.

Riemannian metric: g

A map assigning at each point p, an inner product: $g_p(u,v)=u^T \mathbf{G_p} v$, with $\mathbf{G_p}$ a symmetric positive definite bilinear map (matrix, called metric tensor).

Levi-Civita connection: ∇

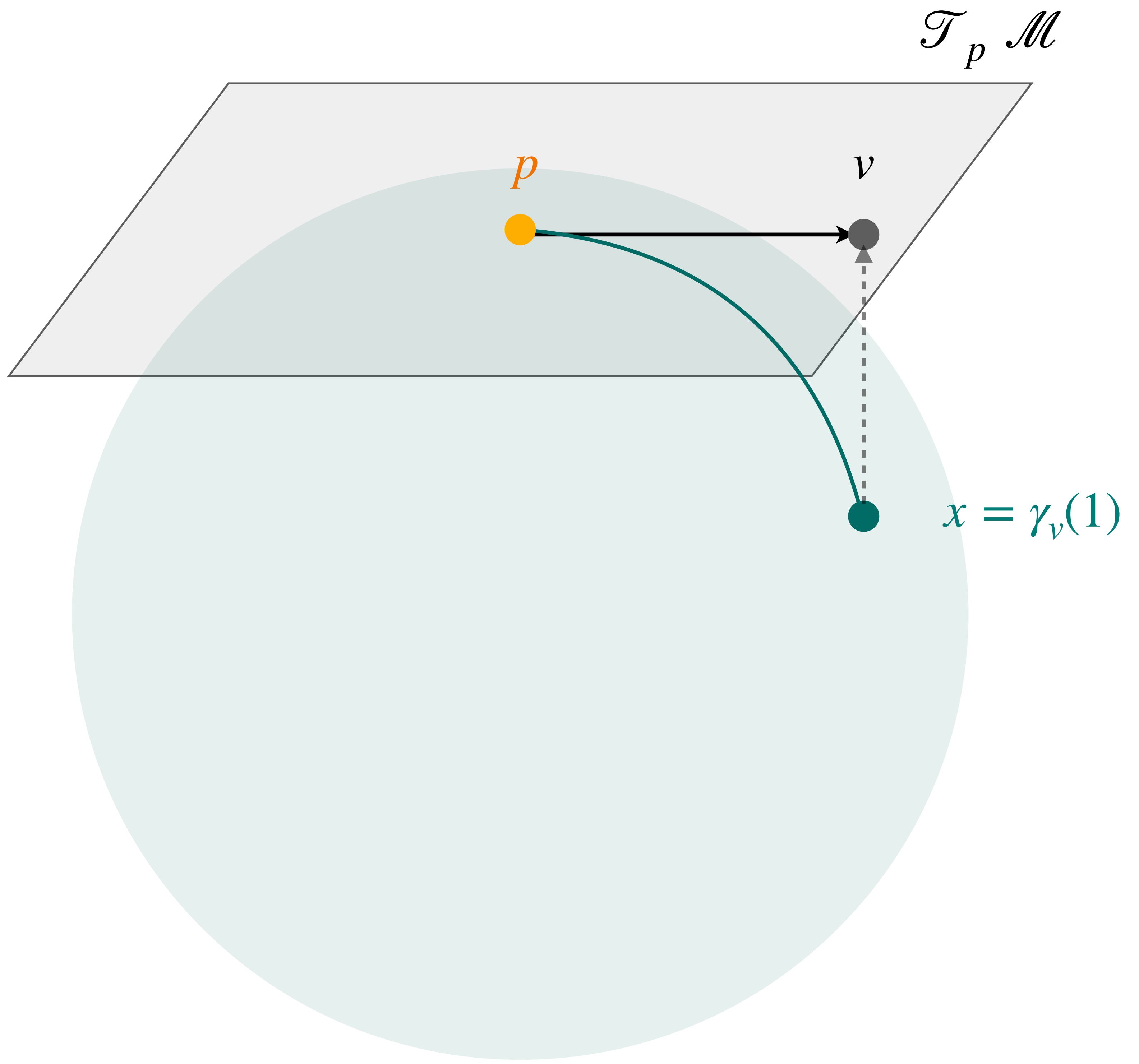
It is the unique covariant derivative (directional derivative) of a tensor field defined on a manifold, that is compatible with the Riemannian metric.

Laplace Beltrami Operator: Δ

Generalisation of Laplacian on Riemannian manifolds.

Exponential and Logarithm maps: Exp, Log

$\text{Exp}_p : T_pM \to M$ $\text{Log}_p : M \to T_pM$



$\text{Exp}_p : v \to \gamma_v(1)$

$\text{Log}_p : x \to v$

Manifold-valued inputs

$f : \mathscr{M} \to \mathbb{R}^q$

Extrinsic kernels

Lin, Lizhen, Niu Mu, Pokman Cheung, and David Dunson., et al. "Extrinsic Gaussian Processes for Regression and Classification on Manifolds." Bayesian Analysis 14.3 (2019): 887-906.

Definition: **Extrinsic kernels**

M is a smooth manifold.
i: R^d→M is a smooth map from R to M.

$$k_{ext}(x, x') = k_{\mathbb{R}^d}(i(z), i(z'))$$

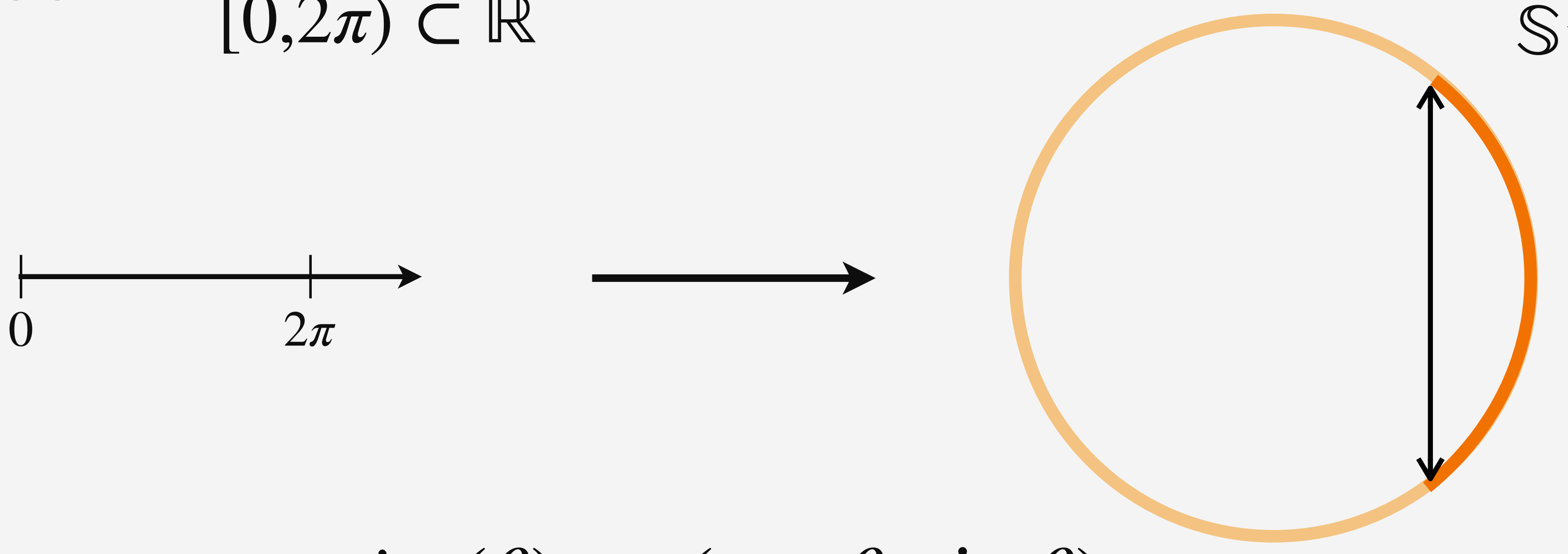
The kernel is properly restricted to the manifold M.

The construction is always PSD because it restricts a Euclidean Gram matrix to the embedded points.

But it does not respect the geometry of the data.

Example 1: a circle

$[0,2\pi) \subset \mathbb{R}$ $\quad$ $\mathbb{S}^1 \subset \mathbb{R}^2$

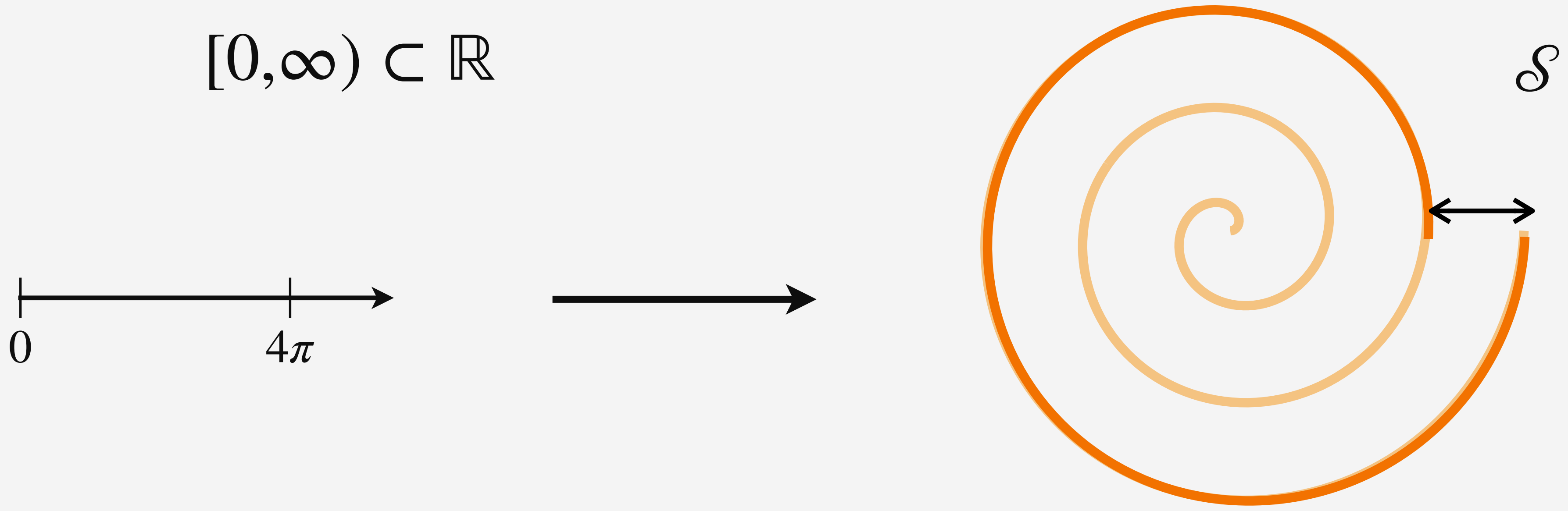


$$i : (\theta) \to (\cos\theta, \sin\theta)$$

$$k_{\mathbb{R}^d}(i(\theta), i(\theta')) = \exp\left(-\frac{\|i(\theta) - i(\theta')\|^2}{2\ell^2}\right) = \cdots \quad = \exp\left(\frac{2}{\ell^2}\sin^2(\frac{\theta-\theta'}{2})\right)$$

Example 2: a spiral

$[0,\infty) \subset \mathbb{R}$ $\quad$ $\mathscr{S} \subset \mathbb{R}^2$



$$i : (\theta) \to (\theta\cos\theta, \theta\sin\theta)$$

$$\|i(\theta) - i(\theta')\|^2 = \theta^2 + \theta'^2 - 2\theta\theta'\cos(\theta - \theta')$$

$$\theta = 2\pi + \theta' \Rightarrow \|i(\theta) - i(\theta')\|^2 = (\theta - \theta')^2$$

$f : \mathcal{M} \to \mathbb{R}^q$ Naive generalisation

Aasa Feragen, Francois Lauze, and Soren Hauberg. "Geodesic exponential kernels: When curvature and linearity conflict." CVPR 2015.

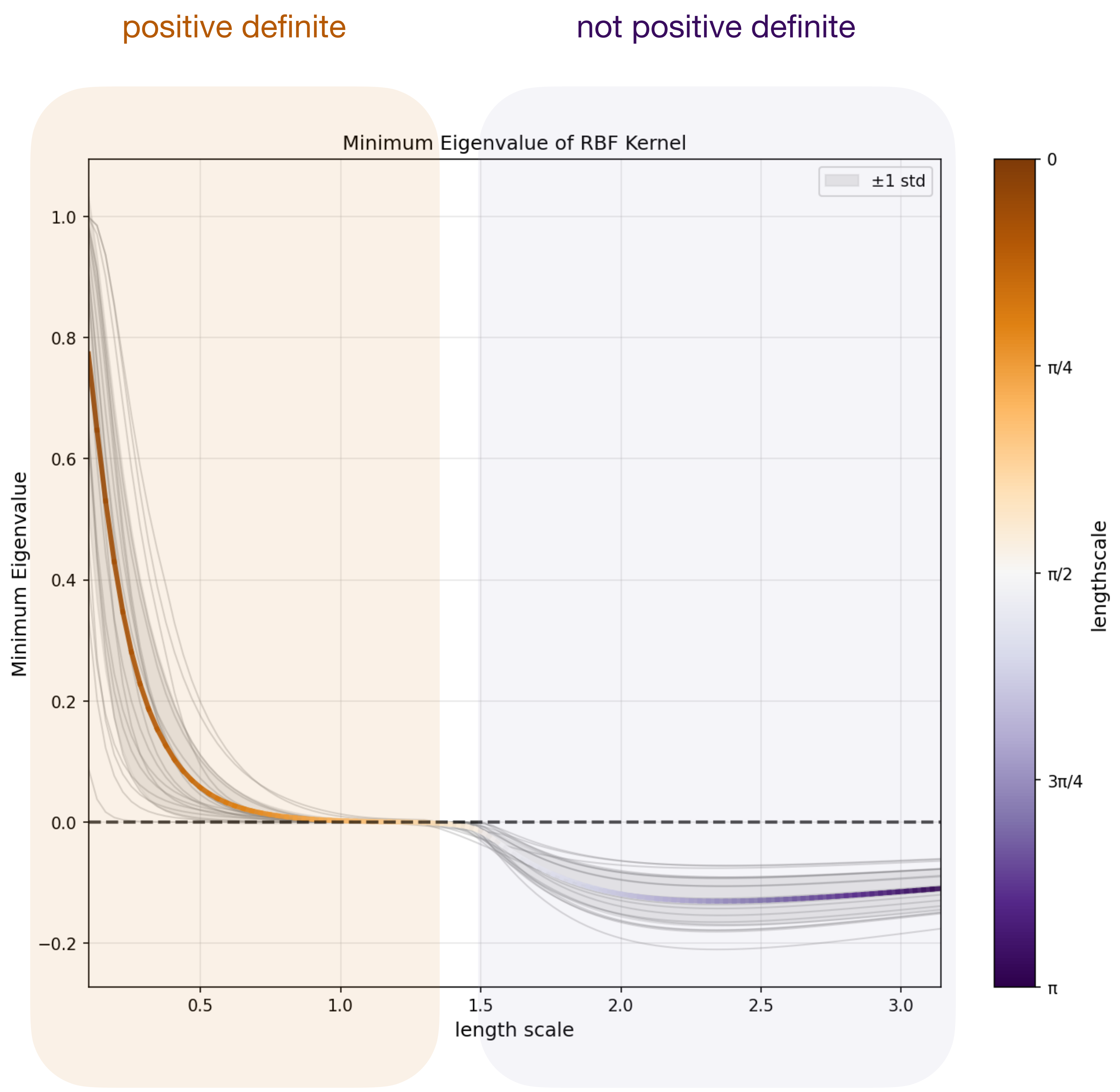
Definition: **Geodesic kernels**

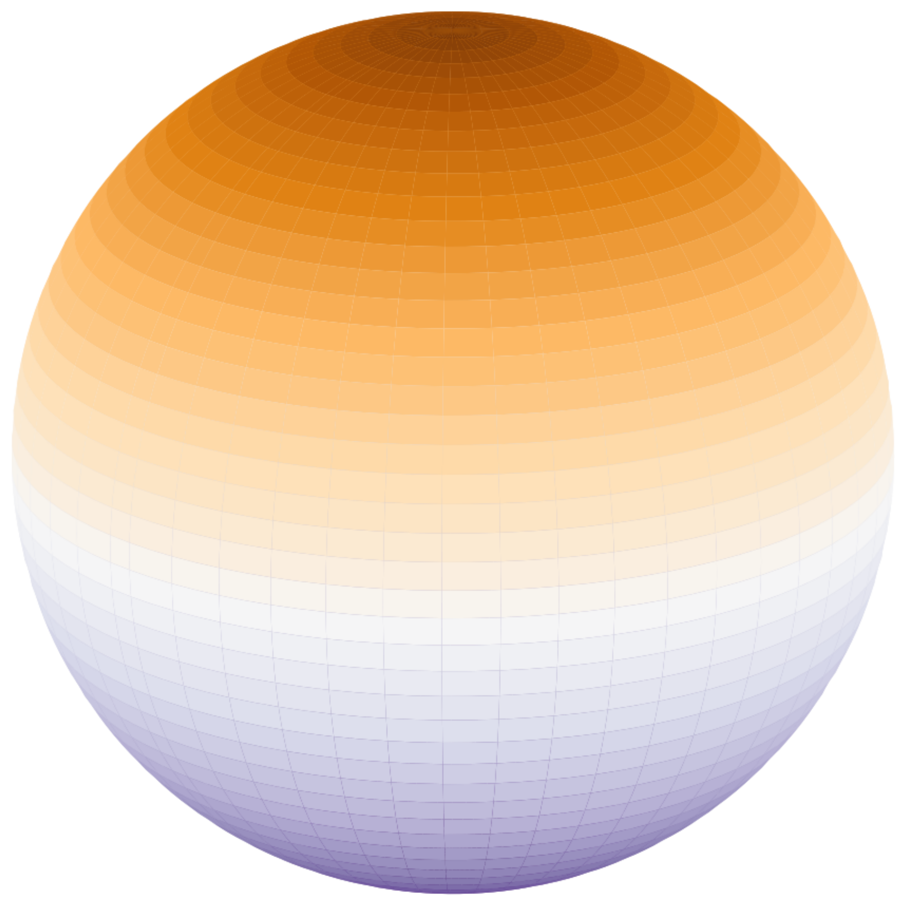
The idea is to extend the RBF definition to a manifold, using the geodesic distance as a way to measure two points.

$$k_{geo}(x, x') = \exp\left(-\frac{\text{dist}_g(x, x')^2}{2\ell^2}\right)$$

But, depending on the lenghtscale chosen, the kernel is **not** positive definite!

I have zero clue why this is happening 😬





$f : \mathscr{M} \to \mathbb{R}^q$

Naive generalisation

Aasa Feragen, Francois Lauze, and Soren Hauberg. "Geodesic exponential kernels: When curvature and linearity conflict." CVPR 2015.

Definition: **Geodesic kernels**

The idea is to extend the RBF definition to a manifold, using the geodesic distance as a way to measure two points.

$$k_{geo}(x, x') = \exp\left(-\frac{\text{dist}_g(x, x')^2}{2\ell^2}\right)$$

Theorem: **Positive Definiteness of Geodesic kernels**

The geodesic Gaussian kernel $k(x,y)=\exp(-\kappa d^2(x,y))$ is **positive definite for all κ>0** if and only if the Riemannian manifold is (isometric to) Euclidean space.

There is some κ that could make your kernel PD. It is still an **open problem**

And it gets **worse**! … 😵

For an **infinite number of data points**, none of the geodesic gaussian kernels are PD on Riemannian manifold with isometric embeddings to the circle.

For an **infinite number of data points**, non simply connected Riemannian manifold are never PD.

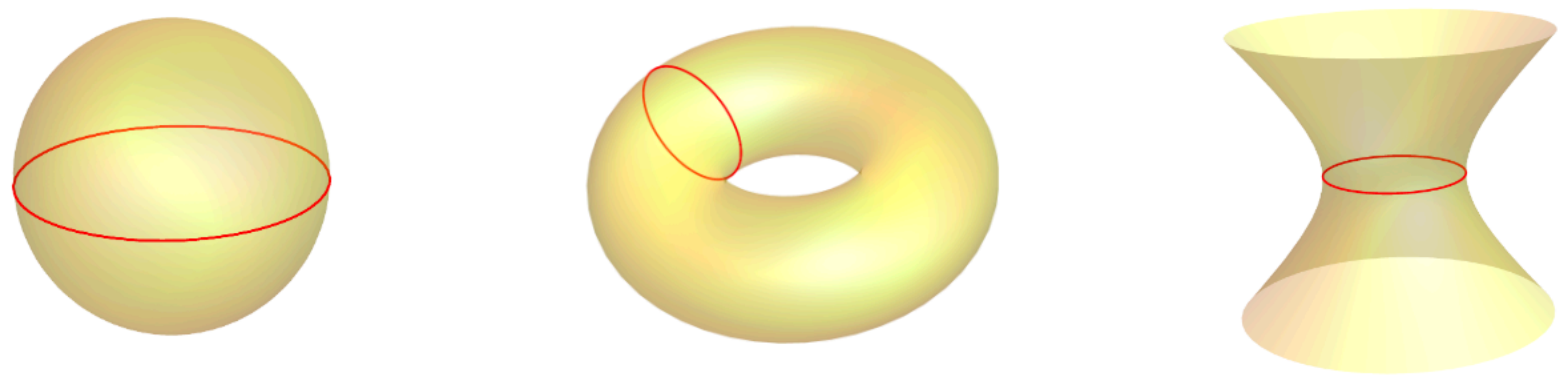
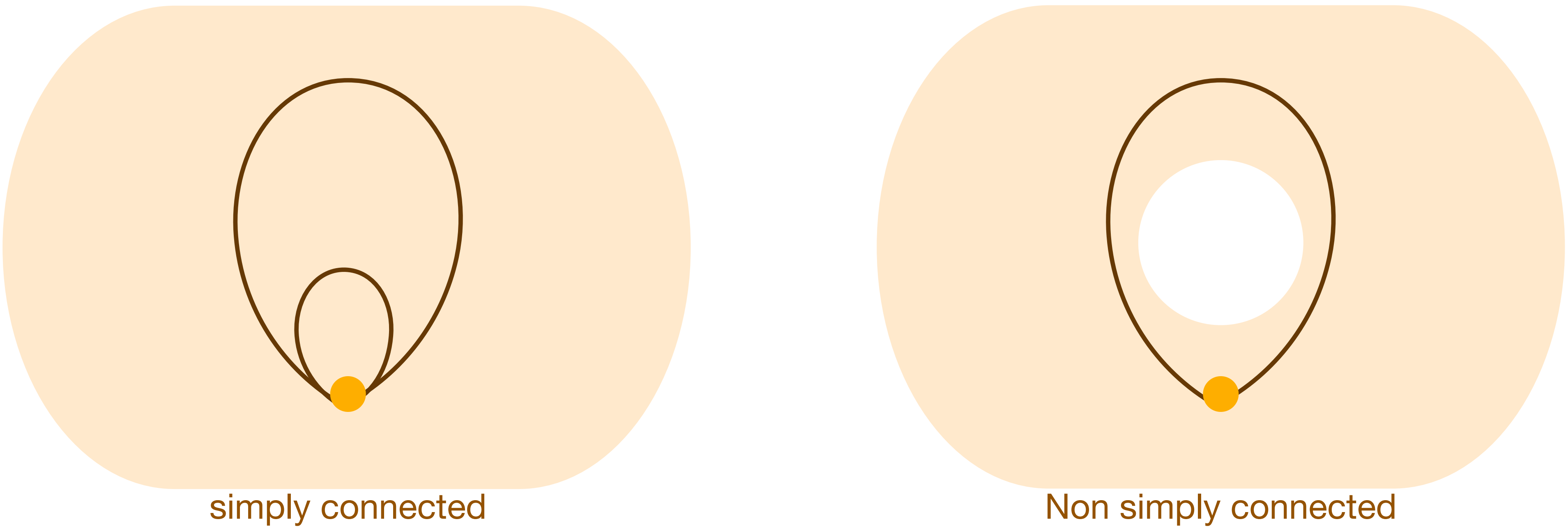


Fig. 2. Examples of Riemannian manifolds that admit isometric embeddings of S^1. From left to right: a sphere, a torus, a hyperbolic hyperboloid.

N. Da Costa et al, "The gaussian kernel on the circle and spaces that admit isometric embeddings of the circle." Geometric Science of Information. GSI 2023.



Li, Siran. "Gaussian kernels on nonsimply connected closed Riemannian manifolds are never positive definite." Bulletin of the London Mathematical Society 56.1 (2024): 263-273.

$f : \mathcal{M} \to \mathbb{R}^q$

Intrinsic kernels

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, Marc Deisenroth. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS 2020*.

Bochner Theorem

A complex-valued function *k* on R[d] is the covariance function of a weakly stationary mean square continuous complex-valued random process on R[d] **if and only if** it can be represented as:

$$k(\tau) = \int_{\mathbb{R}^d} e^{2\pi i s \cdot \tau} \, d\mu(s)$$

with μ a positive finite measure

Sturm Liouville Theorem

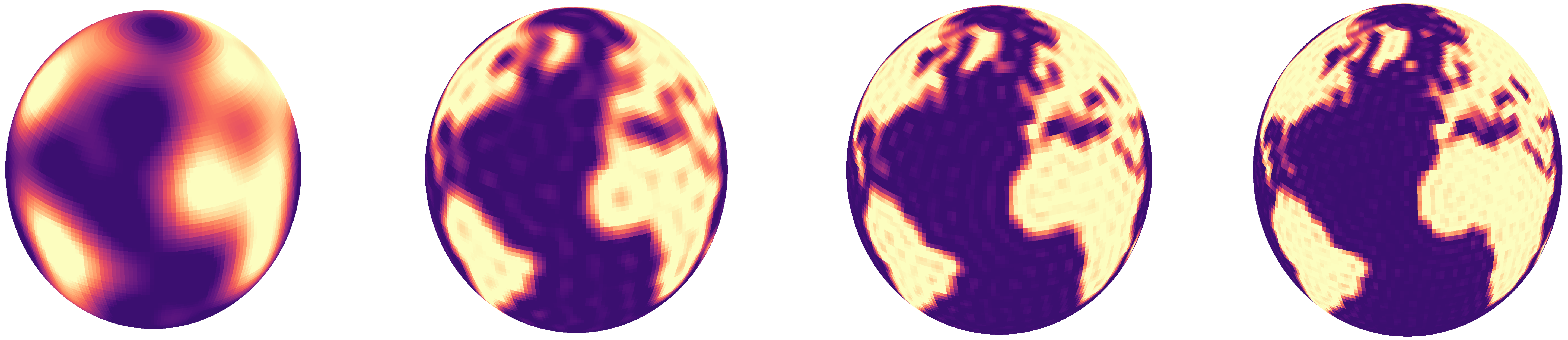
Consider a **compact** Riemannian manifold (M,d) with Δ the Laplace Beltrami operator.

There exists an orthonormal basis $\{u_1, u_2, \ldots, u_n\}$ of the space of square integrable functions, and a sequence of positive numbers $(\lambda_0 < \lambda_1 < \lambda_2 < \ldots)$ such that:

$$-\Delta u_n = \lambda_n u_n \quad \text{and} \quad -\Delta u = \sum_{n \geq 0} \lambda_n \langle u, u_n \rangle u_n .$$

Defining stationary kernels …on compact manifolds

	$(\mathbb{R}^d, \Delta)$	(M, Δ_g)
$K = U \Lambda U^\top$	$k(x,y) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} S(\omega) e^{i\omega(x-y)} \, d\omega$	$k(x,y) = \sum_{n=0}^{\infty} g(\lambda) u_n(x) \overline{u_n}(y)$
$\Lambda \succ 0$	$\forall \omega, \quad S(\omega) \geq 0$	$\forall \lambda, \quad g(\lambda) \geq 0$
	$u_\omega(x) = (2\pi)^{-\frac{d}{2}} e^{i\omega x}$	$u_n(x)$ depends on M
	$-\Delta e^{i\omega} = \lVert\omega\rVert^2 e^{i\omega}$	$-\Delta_g u = \lambda_n u_n$
$\mathcal{F}[-\Delta u] = \lVert\omega\rVert^2 \, \mathcal{F}[u]$	$-\Delta u = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \lVert\omega\rVert^2 \langle u, e^{i\omega x} \rangle e^{i\omega x} d\omega$	$-\Delta_g u = \sum_{n \geq 0} \lambda_n \langle u, u_n \rangle u_n .$



u_{10} u_{50} u_{70} u_∞

$f : \mathcal{M} \to \mathbb{R}^q$

Intrinsic kernels

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, Marc Deisenroth. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS 2020*.

Bochner Theorem

A complex-valued function *k* on R^d is the covariance function of a weakly stationary mean square continuous complex-valued random process on R^d **if and only if** it can be represented as:

$$k(\tau) = \int_{\mathbb{R}^d} e^{2\pi i s \cdot \tau}\, d\mu(s)$$

with μ a positive finite measure

Sturm Liouville Theorem

Consider a **compact** Riemannian manifold (M,d) with Δ the Laplace Beltrami operator.

There exists an orthonormal basis $\{u_1, u_2, \ldots, u_n\}$ of the space of square integrable functions, and a sequence of positive numbers $(\lambda_0 < \lambda_1 < \lambda_2 < \ldots)$ such that:

$$-\Delta u_n = \lambda_n u_n \qquad \text{and} \qquad -\Delta u = \sum_{n \geq 0} \lambda_n \langle u, u_n \rangle u_n .$$

Defining stationary kernels …on compact manifolds

	$(\mathbb{R}^d, \Delta)$	(M, Δ_g)
$K = U \Lambda U^\top$	$k(x,y) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} S(\omega) e^{i\omega(x-y)}\, d\omega$	$k(x,y) = \sum_{n=0}^{\infty} g(\lambda) u_n(x) \overline{u_n}(y)$
$\Lambda \succ 0$	$\forall \omega, \quad S(\omega) \geq 0$	$\forall \lambda, \quad g(\lambda) \geq 0$
	$u_\omega(x) = (2\pi)^{-\frac{d}{2}} e^{i\omega x}$	$u_n(x) = ?$
	$-\Delta e^{i\omega} = \Vert\omega\Vert^2 e^{i\omega}$	$-\Delta_g u = \lambda_n u_n$
$\mathcal{F}[-\Delta u] = \Vert\omega\Vert^2\, \mathcal{F}[u]$	$-\Delta u = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \Vert\omega\Vert^2 \langle u, e^{i\omega x} \rangle e^{i\omega x} d\omega$	$-\Delta_g u = \sum_{n \geq 0} \lambda_n \langle u, u_n \rangle u_n .$
RBF	$S(\omega) = e^{-t\Vert\omega\Vert^2}$	$g(\lambda) = e^{-t\lambda_n}$ $\quad e^{-t\Delta_g} u = W_g$
Matern	$S(\omega) = (\kappa^2 + \Vert\omega\Vert^2)^{-p}$	$g(\lambda) = \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{-p}$ $\quad \left(\frac{2\nu}{\kappa^2} - \Delta_g\right)^{p} u = W_g$
RBF	$k(r) = (4\pi t)^{-\frac{d}{2}} e^{-\frac{r^2}{4t}}$	$k(x,y) = \sum_{n=0}^{\infty} e^{-t\lambda_n} u_n(x)\, u_n(y)$
Matern	$k(r) = \frac{\kappa^{-\alpha}(\kappa r)^\alpha K_\alpha(\kappa r)}{(2\pi)^{d/2}\, 2^{p-1} \Gamma(p)}, \quad \alpha = p - \frac{d}{2}$	$k(x,y) = \sum_{n=0}^{\infty} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{-p} u_n(x)\, u_n(y)$

$f : \mathcal{M} \to \mathbb{R}^q$

Intrinsic kernels

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, Marc Deisenroth. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS 2020*.

Final result on compact manifolds

Consider a **compact** Riemannian manifold (M,d) with Δ the Laplace Beltrami operator.
Then the Matern kernel and RBF kernel are defined as:

$$k_\nu(x,x') = \frac{\sigma^2}{C_\nu}\sum_{n=0}^{\infty}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-\nu-\frac{d}{2}} f_n(x)f_n(x')$$

$$k_\infty(x,x') = \frac{\sigma^2}{C_\infty}\sum_{n=0}^{\infty} e^{-\frac{\kappa^2}{2}\lambda_n} f_n(x)f_n(x')$$

Torus, hypersphere and meshes

Close-form expression exist, and in the case of the hypersphere, they use the addition theorem. For meshes, the laplacian can be computed numerically and then the series is truncated

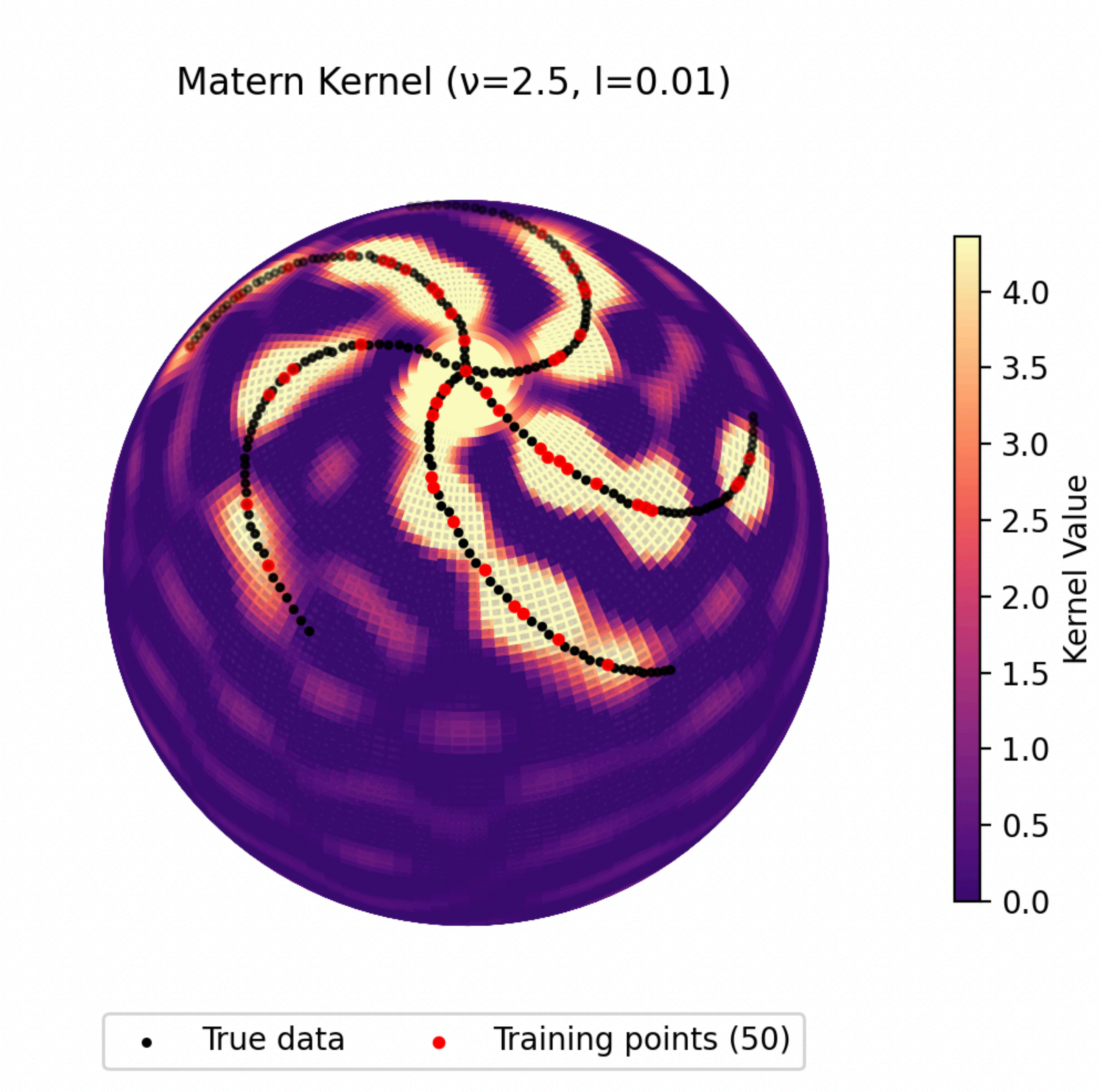
Viacheslav Borovitskiy, Iskander Azangulov, Alexander Terenin, Peter Mostowsky, Marc Deisenroth, and Nicolas Durrande. "Matérn Gaussian Processes on Graphs." AISTATS 21

And graphs!
The Laplace Beltrami operator is replaced with the Hodge Laplacian for graphs, and the kernels are defined in a similar fashion.

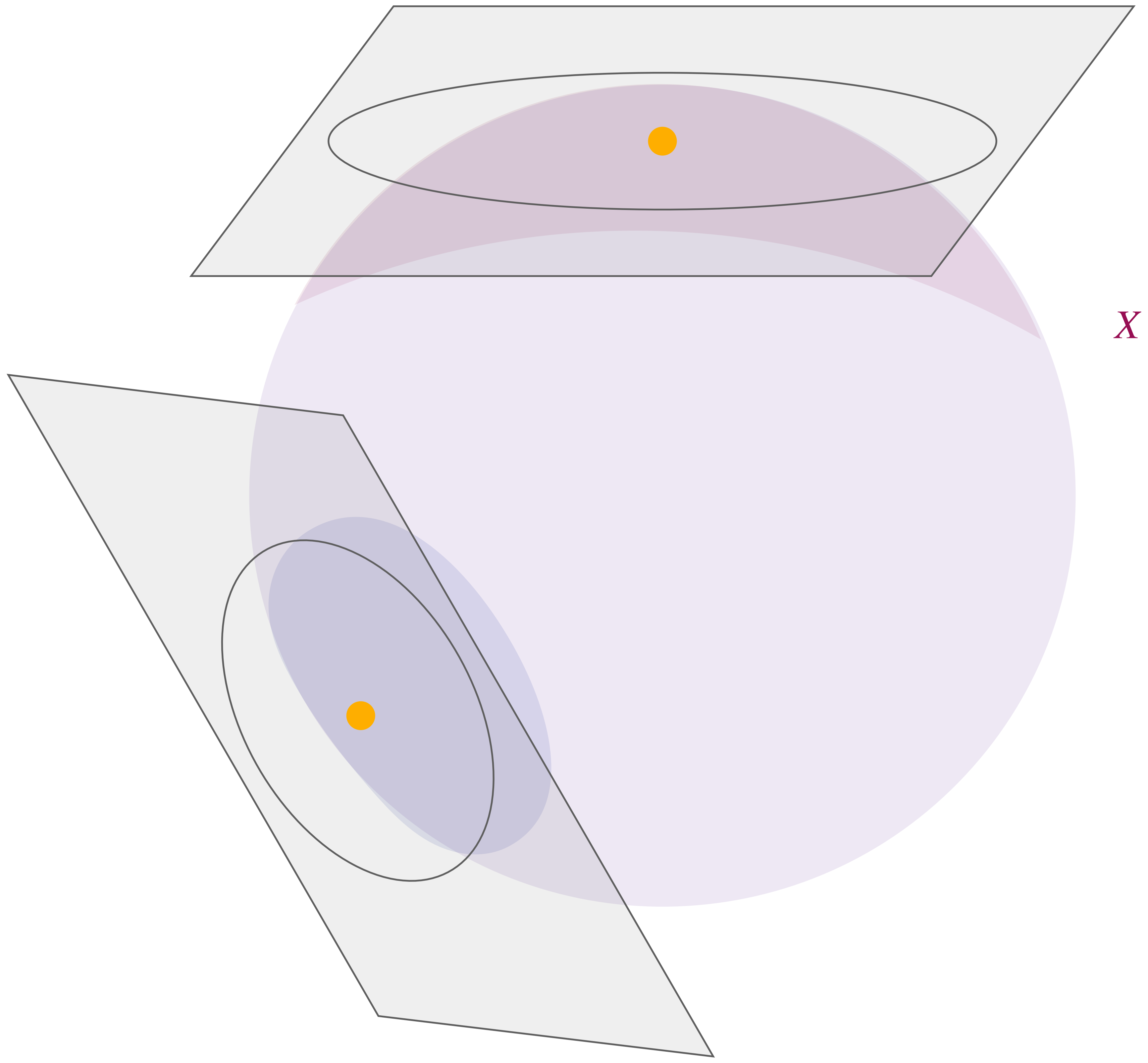
Iskander Azangulov, Andrei Smolensky, Alexander Terenin, and Viacheslav Borovitskiy. "Stationary Kernels and Gaussian Processes on Lie Groups and their Homogeneous Spaces II: non-compact symmetric spaces." JMLR 2024

And non compact spaces?
The theory is different, but the kernels can be approximated!

```
pip install geometric_kernels

from geometric_kernels.spaces import Hypersphere
from geometric_kernels.kernels import MaternGeometricKernel

sphere = Hypersphere(dim=2)
kernel = MaternGeometricKernel(sphere)
params = kernel.init_params()
params['nu'], params['lengthscale'] = 2.5, 0.01

X_obs = [[0.0, 1.0, 0.0], [0.0, -1.0, 0.0]]
X_pred = sphere_grid(num_lats, num_longs)
kernel_values = kernel.K(params, X_obs, X_pred)

plot_sphere_surface(kernel_values)
```



Manifold-valued outputs



$f : \mathbb{R}^d \to \mathcal{M}$

Wrapped Gaussian distributions

Anton Mallasto and Aasa Feragen. "Wrapped Gaussian process regression on Riemannian manifolds." *CVPR* 2018.

Definition: wrapped Gaussian distribution

Let (M,g) be a Riemannian manifold. A random point X on M follows a **wrapped Gaussian distribution**, if for some μ∈M and well defined kernel K:

$$X \sim \left(\text{Exp}_\mu\right)_\# \left(\mathcal{N}(0,K)\right) \quad \text{and we note:} \quad X \sim \mathcal{N}_M(\mu, K)$$

Theorem: Conditionally wrapped Gaussian distribution

Let X1 and X2 be jointly wrapped Gaussian distributed random variables on the manifold M.

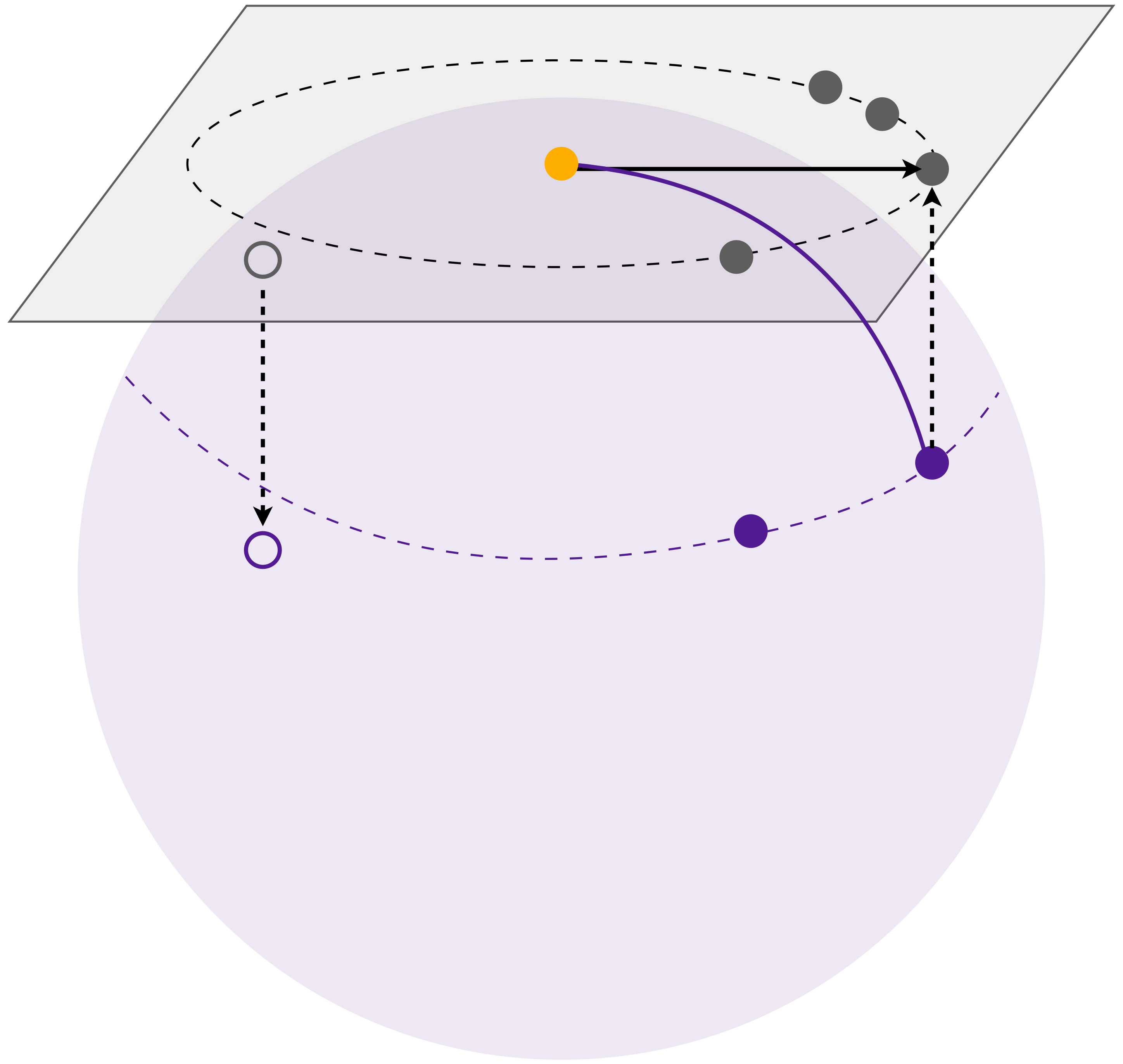
$$(X_1, X_2) \sim \mathcal{N}_{M_1 \times M_2}\left(\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \begin{pmatrix}K_1 & K_{12} \\ K_{12}^\top & K_2\end{pmatrix}\right)$$

Then the conditional X1|X2=p2 is a mixture of wrapped Gaussians on M.

$$X_1 \,\Big|\, (X_2 = p_2) \sim (\text{Exp}_{\mu_1})_\# \, \mathcal{N}(\mu_{1|2}, K_{1|2})$$

$$\mu_{1|2} = K_{12} K_2^{-1} \text{Log}_{\mu_2}(p_2)$$

$$K_{1|2} = K_1 - K_{12}^\top K_2^{-1} K_{12}^\top$$



$f : \mathbb{R}^d \to \mathscr{M}$ Wrapped Gaussian processes

Anton Mallasto and Aasa Feragen. "Wrapped Gaussian process regression on Riemannian manifolds." *CVPR* 2018.

Definition: From wrapped distribution to wrapped GPs

A **Wrapped Gaussian Process** is a collection of manifold-valued random points whose finite marginals are jointly WGD; informally

$$f \sim \left(\mathrm{Exp}_m\right)_\#\left(GP(0,k)\right)$$

The pair (m,k) are the basepoint function and tangent-space covariance function.

Two routes:

(i) Naïve tangent-space GP: pick a single base point m, log-map data, do Euclidean GP, then Exp back.

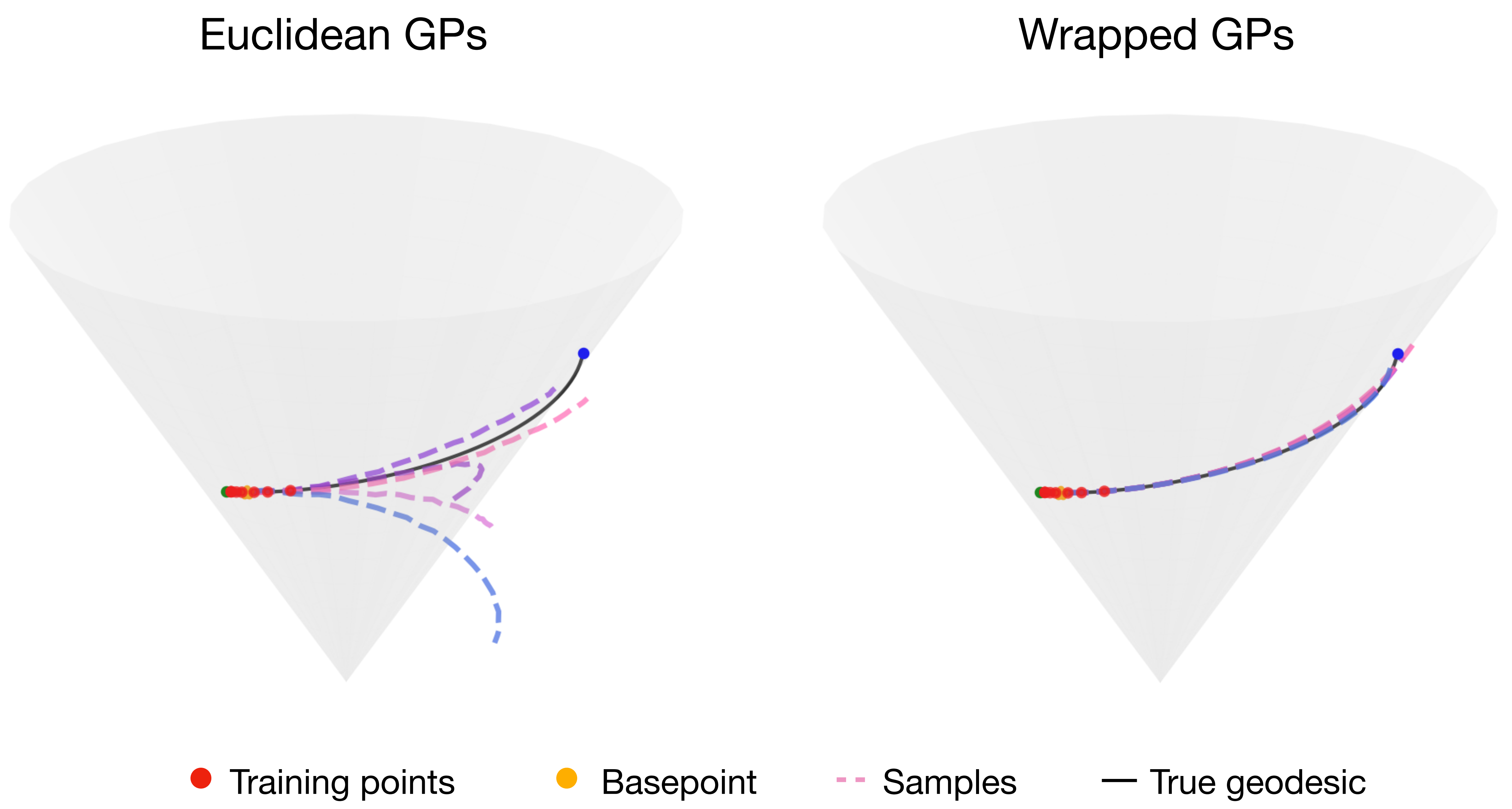
(ii) WGP regression: place a WGP prior with (m,k), condition analytically, then wrap back with Exp.

Algorithm:

1. Pick a prior basepoint function m
2. Log-map the training inputs
3. Train your standard GPs
4. Predict your new inputs
5. Project back to the manifold

$f: \mathbb{R}^d \to \mathscr{M}$ Wrapped Gaussian processes

Anton Mallasto and Aasa Feragen. "Wrapped Gaussian process regression on Riemannian manifolds." *CVPR* 2018.



Definition: From wrapped distribution to wrapped GPs

A **Wrapped Gaussian Process** is a collection of manifold-valued random points whose finite marginals are jointly WGD; informally

$$f \sim \left(\text{Exp}_m\right)_\# \left(GP(0,k)\right)$$

The pair (m,k) are the basepoint function and tangent-space covariance function.

Two routes:

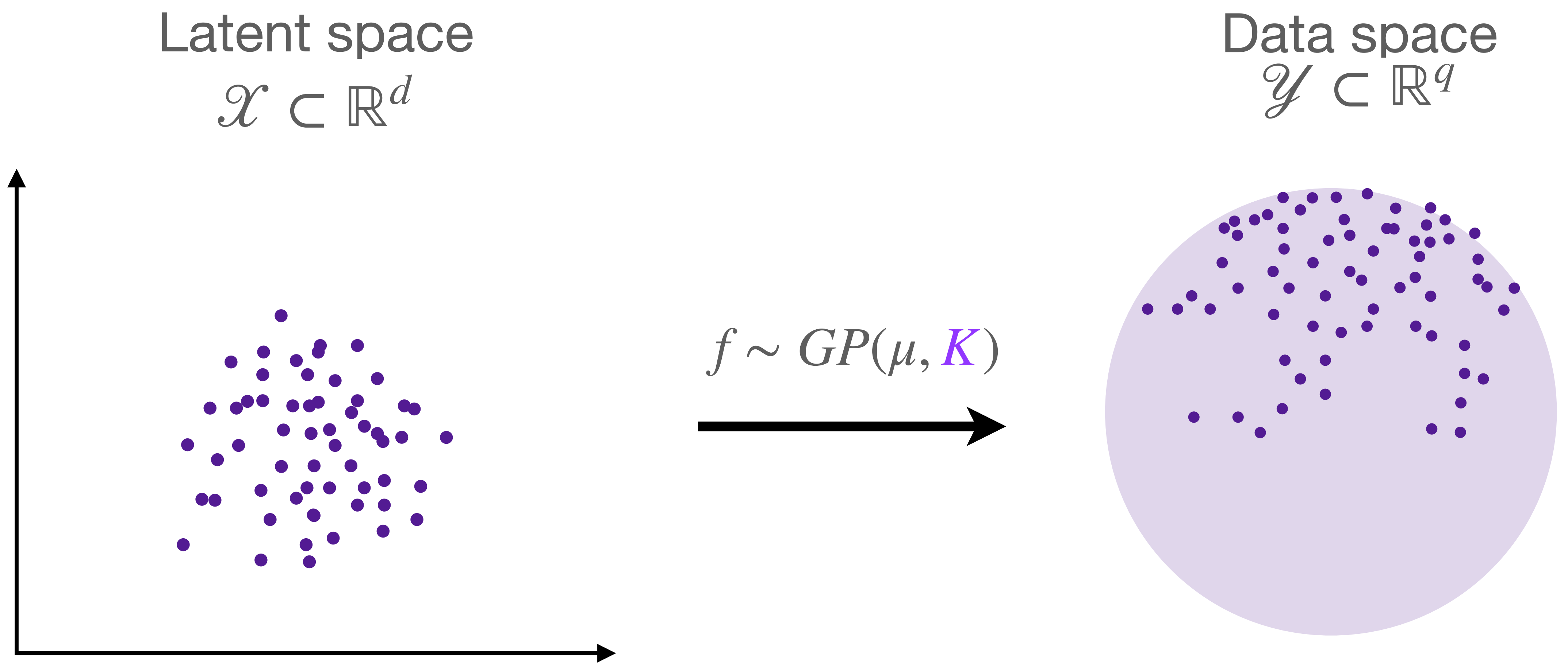
(i) Naïve tangent-space GP: pick a single base point m, log-map data, do Euclidean GP, then Exp back.

(ii) WGP regression: place a WGP prior with (m,k), condition analytically, then wrap back with Exp.

Algorithm:

1. Pick a prior basepoint function m
2. Log-map the training inputs
3. Train your standard GPs
4. Predict your new inputs
5. Project back to the manifold

GPs interlude: Gaussian Latent Variable Model



$f : \mathbb{R}^d \rightarrow \mathbb{R}^q$

GPLVM

Neil Lawrence. "Gaussian process latent variable models for visualisation of high dimensional data." *NeurIPS 2003.*

Algorithm: Gaussian Process

Inputs: $X \in \mathbb{R}^{nd}, \quad Y \in \mathbb{R}^{nq}, \quad K_\theta = k_\theta(X, X) + \sigma^2 I$

Objective: $\log p(y \,|\, \theta) = -\frac{1}{2}\text{Tr}\left(Y^\top K_\theta^{-1} Y\right) - \frac{q}{2}\log|K_\theta| - \frac{nq}{2}log(2\pi)$

Update: $\theta \longleftarrow \theta + \eta \nabla_\theta L$

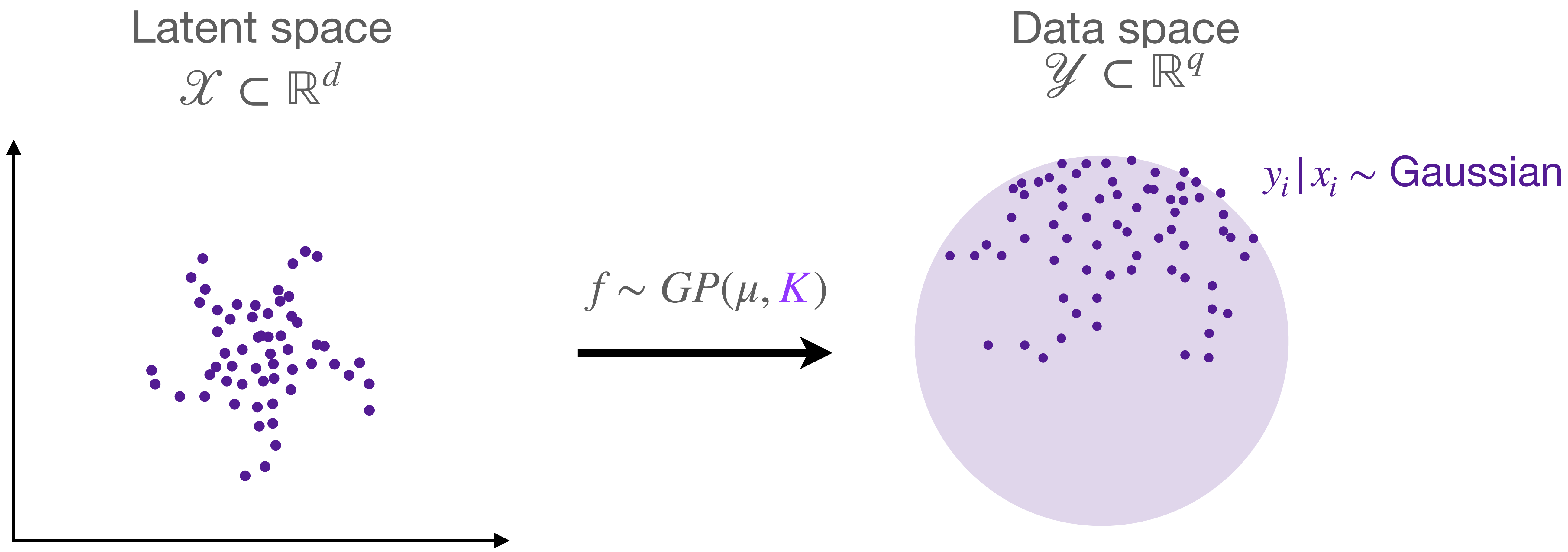
Algorithm: Gaussian Process Latent Variable Model

Inputs: $X \in \mathbb{R}^{nd}, \quad Y \in \mathbb{R}^{nq}, \quad K_{\theta,X} = k_\theta(X, X) + \sigma^2 I$

Objective: $\log p(y \,|\, \theta, X) = -\frac{1}{2}\text{Tr}\left(Y^\top K_{\theta,X}^{-1} Y\right) - \frac{q}{2}\log|K_{\theta,X}| - \frac{nq}{2}log(2\pi)$

Update: $\theta \longleftarrow \theta + \eta \nabla_\theta L \qquad X \longleftarrow X + \eta \nabla_X L$

GPs interlude: Gaussian Latent Variable Model



$f : \mathbb{R}^d \to \mathbb{R}^q$

GPLVM

Neil Lawrence. "Gaussian process latent variable models for visualisation of high dimensional data." *NeurIPS 2003.*

Algorithm: Gaussian Process

Inputs: $X \in \mathbb{R}^{nd}, \quad Y \in \mathbb{R}^{nq}, \quad K_\theta = k_\theta(X, X) + \sigma^2 I$

Objective: $\log p(y | \theta) = -\frac{1}{2}\text{Tr}\left(Y^\top K_\theta^{-1} Y\right) - \frac{q}{2}\log|K_\theta| - \frac{nq}{2} log(2\pi)$

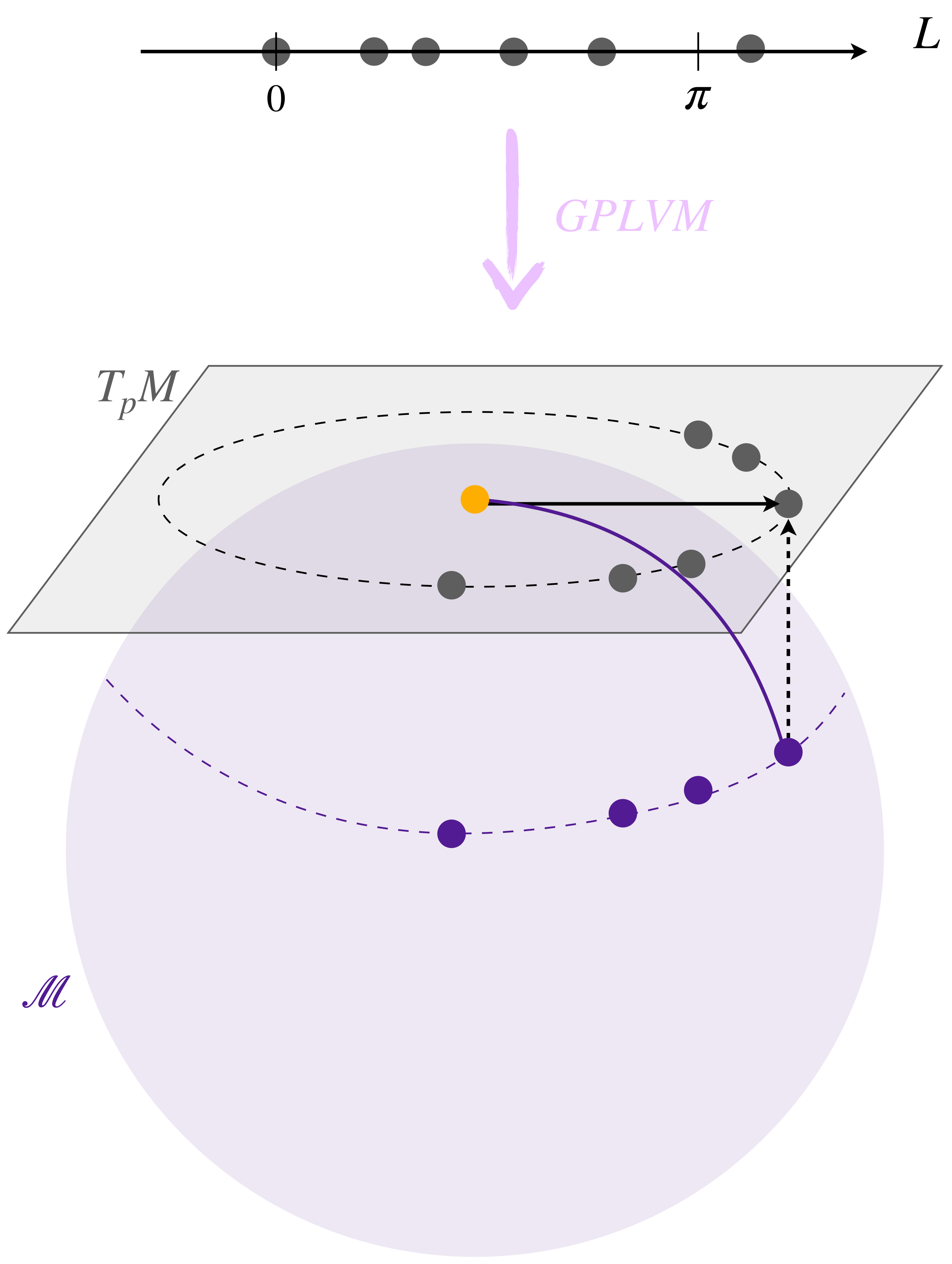
Update: $\theta \longleftarrow \theta + \eta \nabla_\theta L$

Algorithm: Gaussian Process Latent Variable Model

Inputs: $X \in \mathbb{R}^{nd}, \quad Y \in \mathbb{R}^{nq}, \quad K_{\theta,X} = k_\theta(X, X) + \sigma^2 I$

Objective: $\log p(y | \theta, X) = -\frac{1}{2}\text{Tr}\left(Y^\top K_{\theta,X}^{-1} Y\right) - \frac{q}{2}\log|K_{\theta,X}| - \frac{nq}{2} log(2\pi)$

Update: $\theta \longleftarrow \theta + \eta \nabla_\theta L \qquad X \longleftarrow X + \eta \nabla_X L$



$f : \mathbb{R}^d \to \mathscr{M}$

Wrapped GPLVMs

Anton Mallasto, Soren Hauberg and Aasa Feragen. "Probabilistic Riemannian submanifold learning with wrapped Gaussian process latent variable models." *AISTATS* 2019.

Result: Generalising the log likelihood

We can approximate the likelihood of a point y on the manifold by the likelihood of its projection to the tangent space T_mM

$$\mathbb{P}(y \mid x, \theta) \approx \mathcal{N}(\mathrm{Log}_m(y) \,|\, 0, K_{X,\theta})$$

And so, we can get the objective function that we we use to train the wrapped GPLVM:

$$\log p(y \mid x, \theta) \approx -\frac{d}{2} \log|K_{x,\theta}| - \frac{1}{2} \mathrm{Log}_m(y)^{\top} K_{x,\theta}^{-1} \mathrm{Log}_m(y) + \text{const.}$$

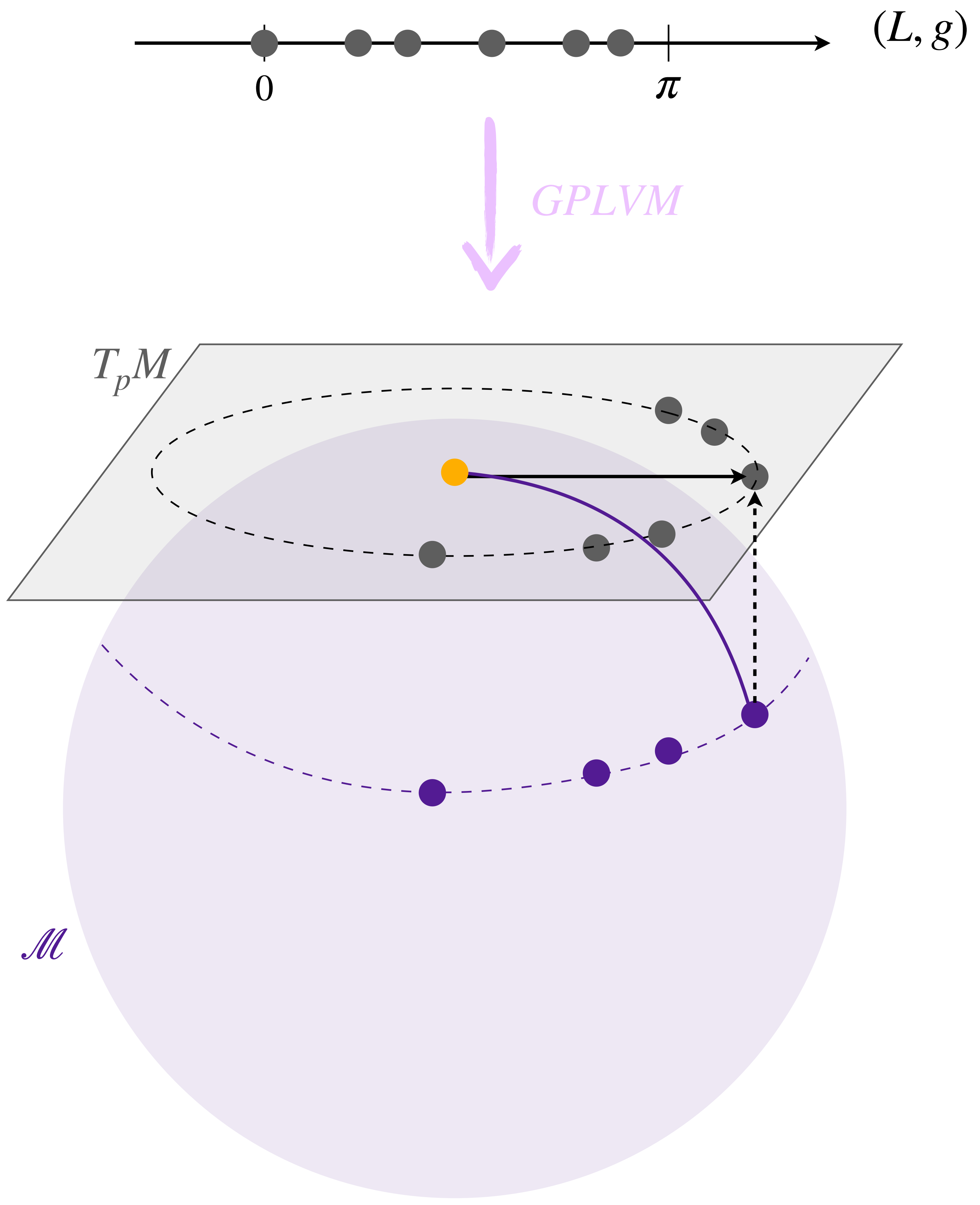
Algorithm:

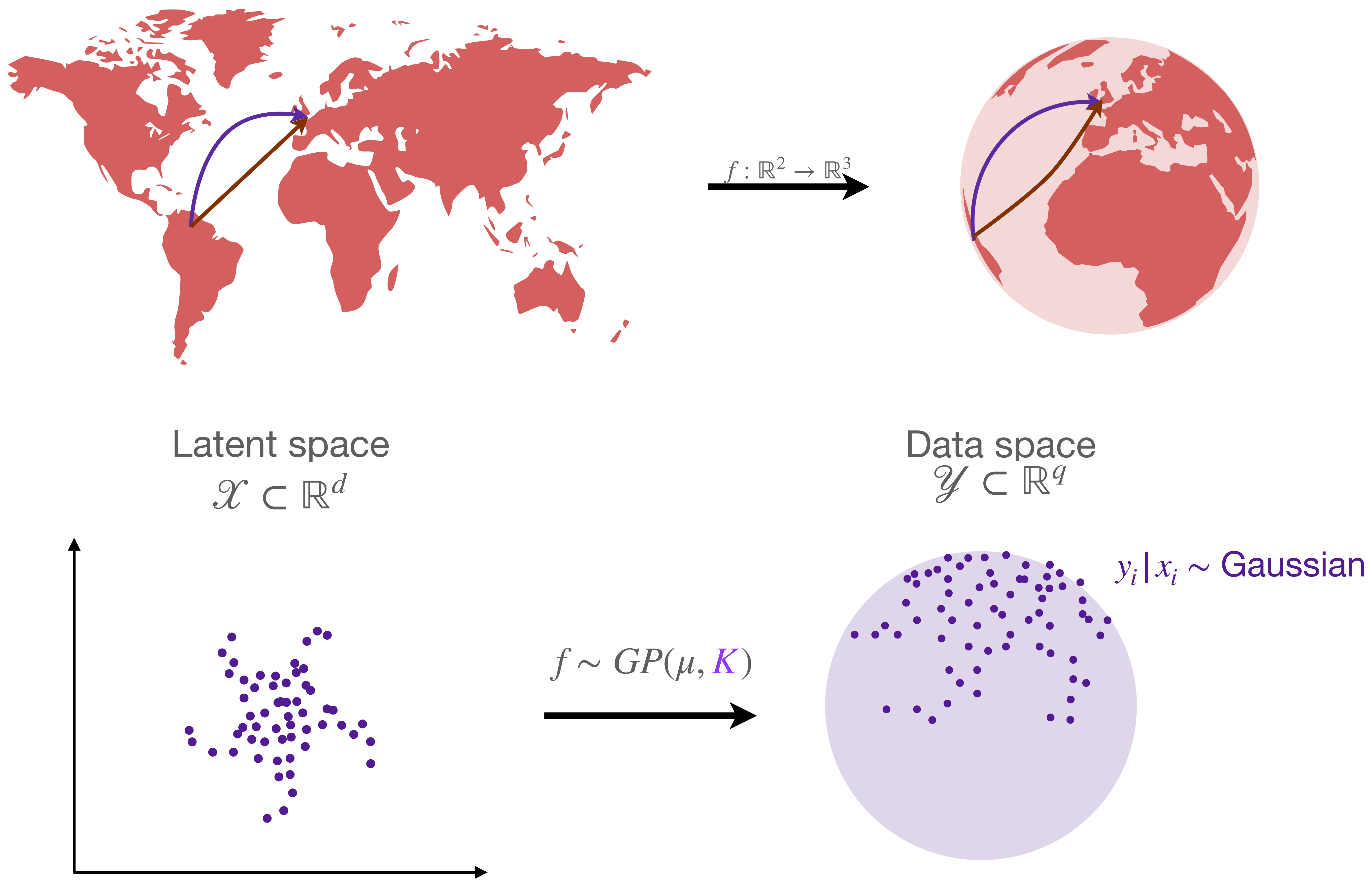
1. Pick a prior basepoint function m
2. Log-map the training inputs
3. Train your standard GPLVMs
4. Predict your new inputs
5. Project back to the manifold

$f : \mathbb{R}^d \rightarrow \mathcal{M}$

Pullback WGPLVM

Leonel Rozo, Miguel González-Duque, Noemie Jaquier, Soren Hauberg, “Riemann-2: Learning Riemannian Submanifolds from Riemannian Data." *AISTATS 2025*





$$G = J_f^\top J_f \implies \mathbb{E}[G] = \mathbb{E}[J]^\top \mathbb{E}[J] + q\Sigma$$

$$f \sim GP(m, K) \implies \mathrm{J}_f \sim \mathcal{N}(\mu, \Sigma) \implies \mathrm{G} = \mathrm{J}_f^\top \mathrm{J}_f \sim \mathcal{W}_d(q, \Sigma, \Sigma^{-1}\mu^\top \mu)$$

Alessandra Tosi, Soren Hauberg, Alfredo Vellido, Neil Lawrence, “Metrics for probabilistic Geometries." *UAI 2014*

$f : \mathbb{R}^d \to \mathcal{M}$ Pullback WGPLVM

Leonel Rozo, Miguel González-Duque, Noemie Jaquier, Soren Hauberg, “Riemann-2: Learning Riemannian Submanifolds from Riemannian Data." *AISTATS 2025*

One step back: Pulling back the metric through a GPLVM

We want to navigate the latent space. We need a Riemannian metric G, obtained by **pulling back** the data to the latent space through immersion f:

$$G = J_f^\top J_f$$

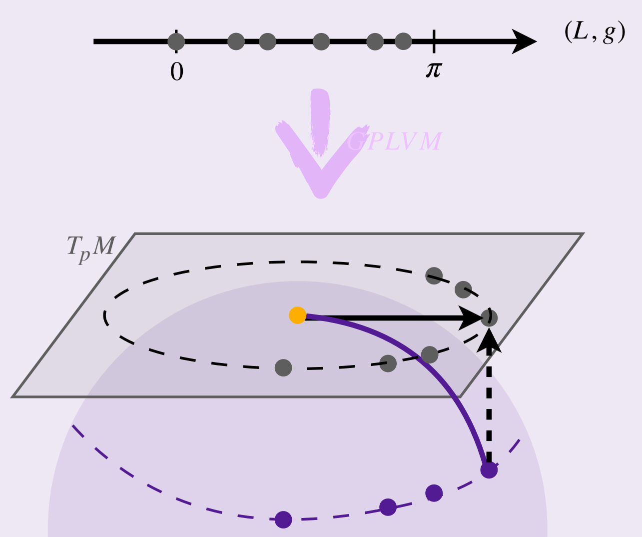
For all points, f is a normal distribution, and its derivatives too. G is a Wishart distribution, and we know its mean.

$$\mathbb{E}[G] = \mathbb{E}[J]^\top \mathbb{E}[J] + q\Sigma$$

But the GP still outputs to the Euclidean space!

Rozo et al (2025) are *extending* Tosi et al (2014) framework, by wrapping the posterior y|x to the manifold with the WGPLVM.

…. and the key is to find the *right* Riemannian metric! Which involves a lot of **chain rules**.



$f : \mathbb{R}^d \rightarrow \mathcal{M}$

Pullback WGPLVM

Pullback Wrapped GPLVMs (Riemann2) explained

Leonel Rozo, Miguel González-Duque, Noemie Jaquier, Soren Hauberg, "Riemann-2: Learning Riemannian Submanifolds from Riemannian Data." *AISTATS 2025*

$J_{fE}^\top J_{Exp}^\top G J_{Exp} J_{fE}$ $\quad$ $J_{Exp}^\top G J_{Exp}$ $\quad$ G

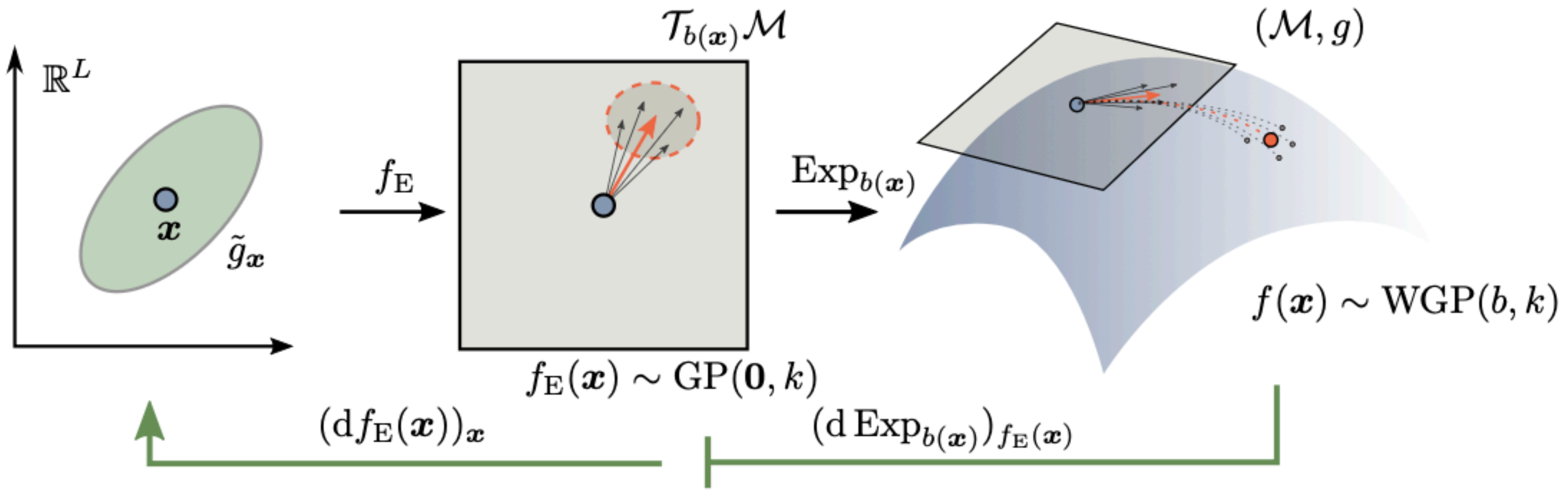


Figure 1: RIEMANN^2 : To learn a Riemannian submanifold from Riemannian data, our method pulls back a Riemannian metric $\tilde{g}_{\boldsymbol{x}}$ to a latent space via a Wrapped GPLVM. In this model, each latent code $\boldsymbol{x} \in \mathbb{R}^L$ defines a distribution of tangent vectors $f_{\mathrm{E}}(\boldsymbol{x}) \sim \mathrm{GP}(\mathbf{0}, k)$, which is then pushed forward onto the manifold $\mathcal{M}$ via the exponential map $\mathrm{Exp}_{b(\cdot)}$. Our framework enables geodesics that, when decoded, comply with the data manifold and are guaranteed to lie on $\mathcal{M}$.

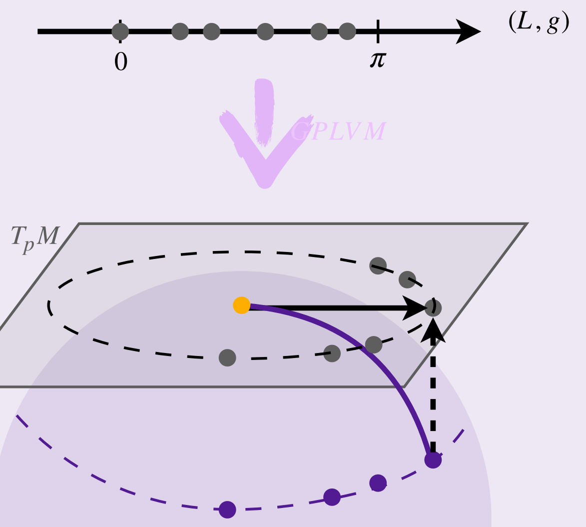
Theorem: Expected pullback metric of a wrapped GPLVM

The pullback metric of (M,g) through the composition of an multi-task Euclidean GPs wrapped to the manifold via the exponential map, is obtained via the chain rule:

$$\tilde{G} = J_{fE}^\top \check{G} J_{fE} = J_{fE}^\top J_{Exp}^\top G J_{Exp} J_{fE}$$

$$\mathbb{E}[\tilde{G}] = \mathbb{E}[J_{fE}]^\top \check{G} \mathbb{E}[J_{fE}] + \mathrm{Tr}[\check{G}^\top K^f] \Sigma_r(J_{fE}^\top)$$

When G=I, $\quad \mathbb{E}[\tilde{G}] = \mathbb{E}[J_{fE}]^\top \mathbb{E}[J_{fE}] + q\Sigma_r(J_{fE})$



$$f : \mathcal{M} \to \mathbb{R}^q$$

Manifold-valued inputs

Defining proper kernels

Extrinsic kernels

Lin, Lizhen, Niu Mu, Pokman Cheung, and David Dunson., et al. "Extrinsic Gaussian Processes for Regression and Classification on Manifolds." Bayesian Analysis 14.3 (2019): 887-906.

Naive generalisation

Aasa Feragen, Francois Lauze, and Soren Hauberg. "Geodesic exponential kernels: When curvature and linearity conflict." CVPR 2015.

Intrinsic kernels

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, Marc Deisenroth. "Matérn Gaussian processes on Riemannian manifolds." *NeurIPS 2020*.

$$f : \mathbb{R}^d \to \mathcal{M}$$

Manifold-valued outputs

Wrapping everything

Wrapped GPs

Anton Mallasto and Aasa Feragen. "Wrapped Gaussian process regression on Riemannian manifolds." *CVPR* 2018.

Wrapped GPLVMs

Anton Mallasto, Soren Hauberg and Aasa Feragen. "Probabilistic Riemannian submanifold learning with wrapped Gaussian process latent variable models." *AISTATS* 2019.

WGPLVM with the pullback metric

Leonel Rozo, Miguel González-Duque, Noemie Jaquier, Soren Hauberg, “Riemann-2: Learning Riemannian Submanifolds from Riemannian Data." *AISTATS 2025*

Thank you!

Special thanks to the GPSS organisers 🧡

Mauricio Alvarez, Carl Henrik Ek, Richard Wilkinson, Henry Moss, Neil Lawrence

And to those inspiring researchers that led to this talk 💜

Aasa Feragen, Alexander Terenin, Andrei Smolensky, Anton Mallasto, David Hilbert, François Lauze, Carl Friedrich Gauss, Iskander Azangulov, Joseph-Louis Lagrange, Pierre-Simon Laplace, Leonel Rozo, Marc Peter Deisenroth, Michael John Hutchinson, Miguel Gonzalez-Duque, Nathael Da Costa, Noemie Jaquier, Peter Mostowsky, Bernhard Riemann, Salomon Bochner, Jacques Charles François Sturm, Joseph Liouville, Siran Li, Søren Hauberg, Viacheslav Borovitskiy, Vincent Dutordoir and many more…