Matérn Gaussian Processes on Riemannian manifolds

Viacheslav Borovitskiy (Slava)

22.01.2021

Mathematics and Computer Science Department, St. Petersburg University
St. Petersburg Department of Steklov Mathematical Institute

Matérn Gaussian processes on Riemannian manifolds

NEURAL INFORMATION PROCESSING SYSTEMS




Viacheslav Borovitskiy*

Alexander Terenin*

Peter Mostowsky*

Marc Deisenroth

St Petersburg University

Imperial College London

UCL

* Equal contribution

NeurIPS 2020

Talk structure

Talk structure

Gaussian processes primer

Matérn kernels

Defining Matérn kernels on compact Riemannian manifolds

Solving for Matérn kernels on compact Riemannian manifolds

Toy examples

Conclusion

Talk structure

Gaussian processes primer

Matérn kernels

Defining Matérn kernels on compact Riemannian manifolds

Solving for Matérn kernels on compact Riemannian manifolds

Toy examples

Conclusion

Talk structure

Gaussian processes primer

Matérn kernels

Defining Matérn kernels on compact Riemannian manifolds

Solving for Matérn kernels on compact Riemannian manifolds

Toy examples

Conclusion

Talk structure

Gaussian processes primer

Matérn kernels

Defining Matérn kernels on compact Riemannian manifolds

Solving for Matérn kernels on compact Riemannian manifolds

Toy examples

Conclusion

Talk structure

Gaussian processes primer

Matérn kernels

Defining Matérn kernels on compact Riemannian manifolds

Solving for Matérn kernels on compact Riemannian manifolds

Toy examples

Conclusion

Gaussian processes primer

Gaussian processes

Gaussian processes (GPs) are distributions over functions.

They are used as non-parametric priors for machine learning tasks.

A Gaussian process is determined by its mean and covariance functions:

$$f \sim GP(m(\cdot), k(\cdot, \cdot)),$$

where

- $m(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is any function,
- $k(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a positive-definite function (kernel).

This means that for any $X = (x_1, \ldots, x_N)^\top$ the joint distribution of the vector $f(X) = (f(x_1), .., f(x_N))^\top$ is

$$f(X) \sim N(m_X, K_{XX}),$$

with

$$(K_{XX})_{ij} := k(x_i, x_j),$$
$$m_X := (m(x_1), \ldots, m(x_N))^\top.$$

Gaussian processes

Gaussian processes (GPs) are distributions over functions.

They are used as non-parametric priors for machine learning tasks.

A Gaussian process is determined by its mean and covariance functions:

$$f \sim GP(m(\cdot), k(\cdot, \cdot)),$$

where

- $m(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is any function,
- $k(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a positive-definite function (kernel).

This means that for any $X = (x_1, \ldots, x_N)^\top$ the joint distribution of the vector $f(X) = (f(x_1), .., f(x_N))^\top$ is

$$f(X) \sim N(m_X, K_{XX}),$$

with

$$(K_{XX})_{ij} := k(x_i, x_j),$$
$$m_X := (m(x_1), \ldots, m(x_N))^\top.$$

Gaussian processes

Gaussian processes (GPs) are distributions over functions.

They are used as non-parametric priors for machine learning tasks.

A Gaussian process is determined by its mean and covariance functions:

$$f \sim GP(m(\cdot), k(\cdot, \cdot)),$$

where

- $m(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is any function,
- $k(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a positive-definite function (kernel).

This means that for any $X = (x_1, \dots, x_N)^\top$ the joint distribution of the vector $f(X) = (f(x_1), .., f(x_N))^\top$ is

$$f(X) \sim N(m_X, K_{XX}),$$

with

$$(K_{XX})_{ij} := k(x_i, x_j),$$
$$m_X := (m(x_1), \dots, m(x_N))^\top.$$

Gaussian processes

Gaussian processes (GPs) are distributions over functions.

They are used as non-parametric priors for machine learning tasks.

A Gaussian process is determined by its mean and covariance functions:

$$f \sim GP(m(\cdot), k(\cdot, \cdot)),$$

where

- $m(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is any function,
- $k(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a positive-definite function (kernel).

This means that for any $X = (x_1, \ldots, x_N)^\top$ the joint distribution of the vector $f(X) = (f(x_1), .., f(x_N))^\top$ is

$$f(X) \sim N(m_X, K_{XX}),$$

with

$$\begin{aligned}(K_{XX})_{ij} &:= k(x_i, x_j),\\ m_X &:= (m(x_1), \ldots, m(x_N))^\top.\end{aligned}$$

Gaussian process regression

Consider a set of observations (x_n, y_n), $x_n \in \mathbb{R}^d$, $y_n \in \mathbb{R}$, $n = 1, \ldots, N$.

Assume a Gaussian process prior on $f \sim GP(0, k)$ and the Gaussian likelihood parameterized by the noise σ^2

$$p(Y \mid f(X)) = N(f(X), \sigma^2 I)$$

Then we can compute the posterior:

$$f(\cdot) | Y \sim GP(\widehat{m}(\cdot), \widehat{k}(\cdot, \cdot))$$

with

$$\widehat{m}(*) = \mathsf{K}_{*X}(\mathsf{K}_{XX} + \sigma^2 I)^{-1} Y,$$
$$\widehat{k}(*, *') = k(*, *') - \mathsf{K}_{*X} \underbrace{(\mathsf{K}_{XX} + \sigma^2 I)^{-1}}_{N \times N \text{ matrix}} \mathsf{K}_{X*'}.$$

Gaussian process regression

Consider a set of observations (x_n, y_n), $x_n \in \mathbb{R}^d$, $y_n \in \mathbb{R}$, $n = 1, \ldots, N$.

Assume a Gaussian process prior on $f \sim GP(0, k)$ and the Gaussian likelihood parameterized by the noise σ^2

$$p(Y \mid f(X)) = N(f(X), \sigma^2 I)$$

Then we can compute the posterior:

$$f(\cdot)|Y \sim GP(\widehat{m}(\cdot), \widehat{k}(\cdot, \cdot))$$

with

$$\widehat{m}(*) = \mathrm{K}_{*X}(\mathrm{K}_{XX} + \sigma^2 I)^{-1} Y,$$
$$\widehat{k}(*, *') = k(*, *') - \mathrm{K}_{*X} \underbrace{(\mathrm{K}_{XX} + \sigma^2 I)^{-1}}_{N \times N \text{ matrix}} \mathrm{K}_{X*'}.$$

Gaussian process regression

Consider a set of observations (x_n, y_n), $x_n \in \mathbb{R}^d$, $y_n \in \mathbb{R}$, $n = 1, \ldots, N$.

Assume a Gaussian process prior on $f \sim GP(0, k)$ and the Gaussian likelihood parameterized by the noise σ^2

$$p(Y \mid f(X)) = N(f(X), \sigma^2 I)$$

Then we can compute the posterior:

$$f(\cdot)|Y \sim GP(\widehat{m}(\cdot), \widehat{k}(\cdot, \cdot))$$

with

$$\widehat{m}(*) = \mathrm{K}_{*X}(\mathrm{K}_{XX} + \sigma^2 I)^{-1} Y,$$
$$\widehat{k}(*, *') = k(*, *') - \mathrm{K}_{*X} \underbrace{(\mathrm{K}_{XX} + \sigma^2 I)^{-1}}_{N \times N \text{ matrix}} \mathrm{K}_{X*'}.$$

Gaussian process regression

Consider a set of observations (x_n, y_n), $x_n \in \mathbb{R}^d$, $y_n \in \mathbb{R}$, $n = 1, \ldots, N$.

Assume a Gaussian process prior on $f \sim GP(0, k)$ and the Gaussian likelihood parameterized by the noise σ^2

$$p(Y \mid f(X)) = N(f(X), \sigma^2 I)$$

Then we can compute the posterior:

$$f(\cdot)|Y \sim GP(\widehat{m}(\cdot), \widehat{k}(\cdot, \cdot))$$

with

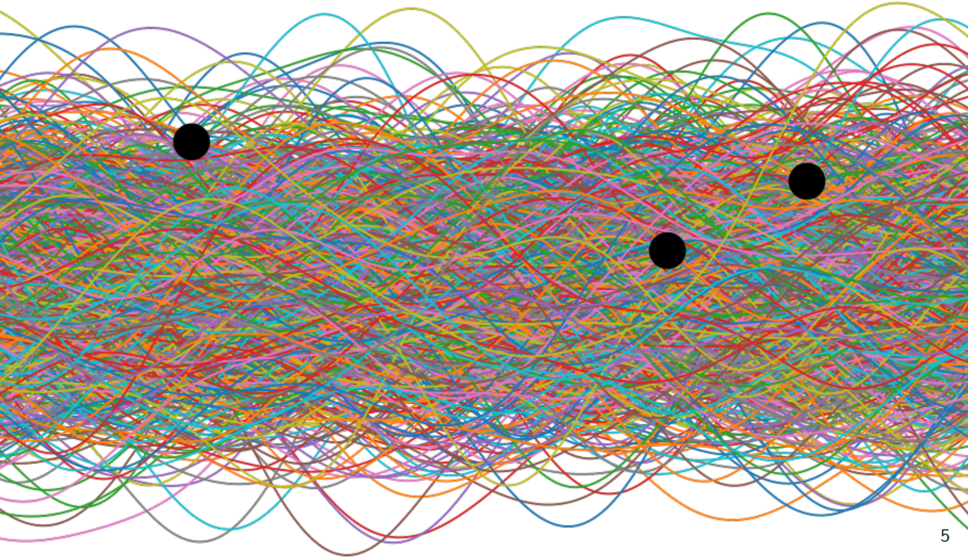
$$\begin{aligned} \widehat{m}(*) &= \mathrm{K}_{*X}(\mathrm{K}_{XX} + \sigma^2 I)^{-1} Y, \\ \widehat{k}(*, *') &= k(*, *') - \mathrm{K}_{*X} \underbrace{(\mathrm{K}_{XX} + \sigma^2 I)^{-1}}_{N \times N \text{ matrix}} \mathrm{K}_{X*'}. \end{aligned}$$

Visual guide to Gaussian process regression

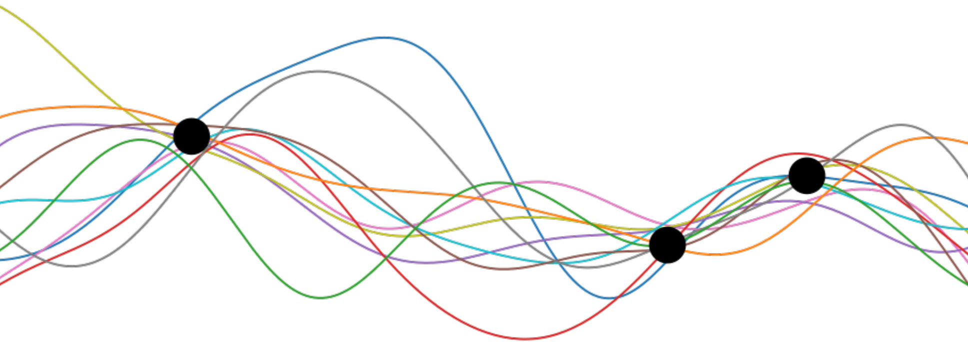
Visual guide to Gaussian process regression

Visual guide to Gaussian process regression

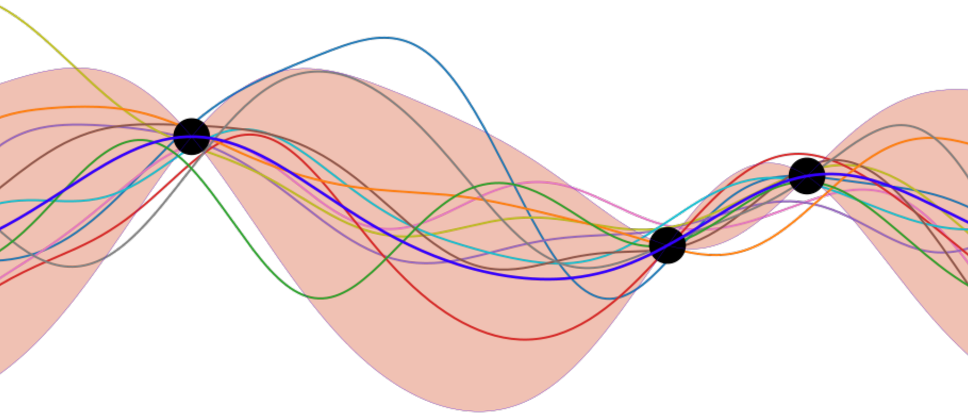
Visual guide to Gaussian process regression



Visual guide to Gaussian process regression



Visual guide to Gaussian process regression



Picking the prior

Pick a parametric family $\{k_\theta(\cdot,\cdot)\}_{\theta\in\Theta}$.

Maximize log likelihood over θ:

$$\log p(Y) = -\frac{1}{2} Y^\top (\mathrm{K}_{XX} + \sigma^2 I)^{-1} Y - \frac{1}{2}\log\left|(\mathrm{K}_{XX} + \sigma^2 I)\right| - \frac{n}{2}\log 2\pi$$

Picking the prior

Pick a parametric family $\{k_\theta(\cdot,\cdot)\}_{\theta\in\Theta}$.

Maximize log likelihood over θ:

$$\log p(Y) = -\frac{1}{2} Y^\top (\mathsf{K}_{XX} + \sigma^2 I)^{-1} Y - \frac{1}{2} \log |(\mathsf{K}_{XX} + \sigma^2 I)| - \frac{n}{2} \log 2\pi$$

Picking the prior

Pick a parametric family $\{k_\theta(\cdot,\cdot)\}_{\theta\in\Theta}$.

Maximize log likelihood over θ:

$$\log p(Y) = -\frac{1}{2} Y^\top (K_{XX} + \sigma^2 I)^{-1} Y - \frac{1}{2}\log|(K_{XX} + \sigma^2 I)| - \frac{n}{2}\log 2\pi$$

Matérn kernels

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

σ^2: variance κ: length scale ν: smoothness

$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

σ^2: variance $\quad \kappa$: length scale $\quad \nu$: smoothness

$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

σ^2: variance $\quad \kappa$: length scale $\quad \nu$: smoothness

$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

σ^2: variance κ: length scale ν: smoothness

$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

σ^2: variance $\quad$ κ: length scale $\quad$ ν: smoothness

$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

σ^2: variance κ: length scale ν: smoothness

$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x-x'\|}{\kappa} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{\|x-x'\|}{\kappa} \right)$$

σ^2: variance κ: length scale ν: smoothness

$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Matérn kernels

This is the most frequently used parametric family of kernels for GPs.

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

σ^2: variance $\quad \kappa$: length scale $\quad \nu$: smoothness

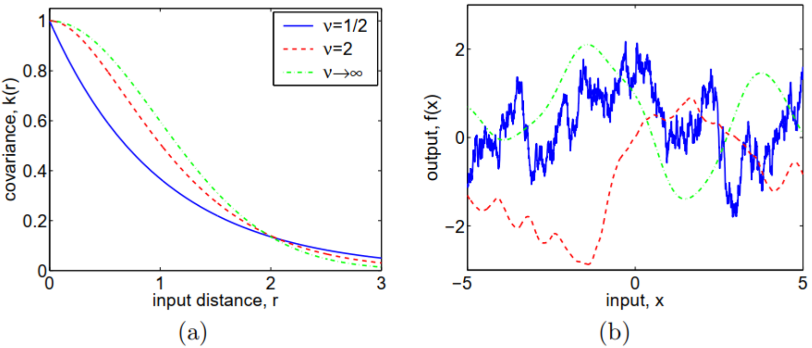
$\nu \to \infty$: recovers square exponential kernel

We have $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$.

Visual guide to Matérn kernels

(a) Matérn kernels as functions of $\|x - x'\|$; (b) GP sample paths

Visual guide to Matérn kernels



(a) Matérn kernels as functions of $\|x - x'\|$; (b) GP sample paths

Defining Matérn kernels on compact Riemannian manifolds

How to define analogous parametric families on a manifold?

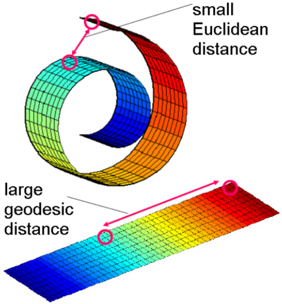
First try: embed a manifold into Euclidean space $\mathbb{R}^d$ and take Matérn kernels from this ambient space.

Bad because:

How to define analogous parametric families on a manifold?

First try: embed a manifold into Euclidean space $\mathbb{R}^d$ and take Matérn kernels from this ambient space.

Bad because:



How to define analogous parametric families on a manifold?

Second try: substitute geodesic distance $d_M(x, x')$ instead of $\|x - x'\|$ into the formula for Matérn kernels.

Surprisingly, this doesn't work, the kernel generally fails to be positive-definite (Feragen et al.):

Geodesic Exponential Kernels: When Curvature and Linearity Conflict

Aasa Feragen
DIKU, University of Copenhagen
Denmark
aasa@diku.dk

François Lauze
DIKU, University of Copenhagen
Denmark
francois@diku.dk

Søren Hauberg
DTU Compute
Denmark
sohau@dtu.dk

How to define analogous parametric families on a manifold?

Second try: substitute geodesic distance $d_M(x, x')$ instead of $\|x - x'\|$ into the formula for Matérn kernels.

Surprisingly, this doesn't work, the kernel generally fails to be positive-definite (Feragen et al.):

Geodesic Exponential Kernels: When Curvature and Linearity Conflict

Aasa Feragen
DIKU, University of Copenhagen
Denmark
aasa@diku.dk

François Lauze
DIKU, University of Copenhagen
Denmark
francois@diku.dk

Søren Hauberg
DTU Compute
Denmark
sohau@dtu.dk

How to define analogous parametric families on a manifold?

Second try: substitute geodesic distance $d_M(x, x')$ instead of $\|x - x'\|$ into the formula for Matérn kernels.

Surprisingly, this doesn't work, the kernel generally fails to be positive-definite (Feragen et al.):

This CVPR2015 paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

Geodesic Exponential Kernels: When Curvature and Linearity Conflict

Aasa Feragen
DIKU, University of Copenhagen
Denmark
aasa@diku.dk

François Lauze
DIKU, University of Copenhagen
Denmark
francois@diku.dk

Søren Hauberg
DTU Compute
Denmark
sohau@dtu.dk

Matérn SPDE

In 60's Whittle have shown that a GP with Matérn kernel over $\mathbb{R}^d$ satisfies this particular SPDE:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{d}{4}+\frac{\nu}{2}} f = \mathcal{W}.$$

Here $\mathcal{W}$ is the Gaussian white noise.

The meaning of the left hand side:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{p} f = \mathcal{F}^{-1}\left(\frac{2\nu}{\kappa^2} + |\zeta|^2\right)^{p} \mathcal{F}f(\zeta)$$

where $\mathcal{F}$ is the Fourier transform.

This generalizes well to the Riemannian setting:

1. Δ becomes the Laplace–Beltrami operator,
2. $\mathcal{W}$ becomes the white noise with respect to the Riemannian measure.

The problem is the implicit nature of this definition.

Matérn SPDE

In 60's Whittle have shown that a GP with Matérn kernel over $\mathbb{R}^d$ satisfies this particular SPDE:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{d}{4}+\frac{\nu}{2}} f = \mathcal{W}.$$

Here $\mathcal{W}$ is the Gaussian white noise.

The meaning of the left hand side:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{p} f = \mathcal{F}^{-1}\left(\frac{2\nu}{\kappa^2} + |\zeta|^2\right)^{p} \mathcal{F}f(\zeta)$$

where $\mathcal{F}$ is the Fourier transform.

This generalizes well to the Riemannian setting:

1. Δ becomes the Laplace–Beltrami operator,
2. $\mathcal{W}$ becomes the white noise with respect to the Riemannian measure.

The problem is the implicit nature of this definition.

Matérn SPDE

In 60's Whittle have shown that a GP with Matérn kernel over $\mathbb{R}^d$ satisfies this particular SPDE:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{d}{4}+\frac{\nu}{2}} f = \mathcal{W}.$$

Here $\mathcal{W}$ is the Gaussian white noise.

The meaning of the left hand side:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{p} f = \mathcal{F}^{-1}\left(\frac{2\nu}{\kappa^2} + |\zeta|^2\right)^{p} \mathcal{F} f(\zeta)$$

where $\mathcal{F}$ is the Fourier transform.

This generalizes well to the Riemannian setting:

1. Δ becomes the Laplace–Beltrami operator,
2. $\mathcal{W}$ becomes the white noise with respect to the Riemannian measure.

The problem is the implicit nature of this definition.

Matérn SPDE

In 60's Whittle have shown that a GP with Matérn kernel over $\mathbb{R}^d$ satisfies this particular SPDE:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{d}{4}+\frac{\nu}{2}} f = \mathcal{W}.$$

Here $\mathcal{W}$ is the Gaussian white noise.

The meaning of the left hand side:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{p} f = \mathcal{F}^{-1}\left(\frac{2\nu}{\kappa^2} + |\zeta|^2\right)^{p} \mathcal{F}f(\zeta)$$

where $\mathcal{F}$ is the Fourier transform.

This generalizes well to the Riemannian setting:

1. Δ becomes the Laplace–Beltrami operator,
2. $\mathcal{W}$ becomes the white noise with respect to the Riemannian measure.

The problem is the implicit nature of this definition.

Matérn SPDE

In 60's Whittle have shown that a GP with Matérn kernel over $\mathbb{R}^d$ satisfies this particular SPDE:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{d}{4}+\frac{\nu}{2}} f = \mathcal{W}.$$

Here $\mathcal{W}$ is the Gaussian white noise.

The meaning of the left hand side:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{p} f = \mathcal{F}^{-1}\left(\frac{2\nu}{\kappa^2} + |\zeta|^2\right)^{p} \mathcal{F}f(\zeta)$$

where $\mathcal{F}$ is the Fourier transform.

This generalizes well to the Riemannian setting:

1. Δ becomes the Laplace–Beltrami operator,
2. $\mathcal{W}$ becomes the white noise with respect to the Riemannian measure.

The problem is the implicit nature of this definition.

Solving for Matérn kernels on compact Riemannian manifolds

Laplace–Beltrami operator

Consider a *compact* Riemannian manifold (M, d) and denote the Laplace–Beltrami operator on (M, d) by $\Delta : C^\infty(M) \mapsto L^2(M)$.

Theorem (Sturm–Liouville decomposition)
There exists an orthonormal basis $(f_n)_{n \in \mathbb{Z}_+}$ *of the space* $L^2(M)$*, and a sequence of non-negative numbers* $0 = \lambda_0 < \lambda_1 \leq \lambda_n \leq \ldots$ *such that*

$$-\Delta f_n = \lambda_n f_n$$

and

$$-\Delta f = \sum_{n \geq 0} \lambda_n \langle f, f_n \rangle f_n$$

Imagine f_n as a substitute for sines and cosines in Fourier series.

Laplace–Beltrami operator

Consider a *compact* Riemannian manifold (M, d) and denote the Laplace–Beltrami operator on (M, d) by $\Delta : C^\infty(M) \mapsto L^2(M)$.

Theorem (Sturm–Liouville decomposition)
There exists an orthonormal basis $(f_n)_{n\in\mathbb{Z}_+}$ of the space $L^2(M)$, and a sequence of non-negative numbers $0 = \lambda_0 < \lambda_1 \leq \lambda_n \leq \ldots$ such that

$$-\Delta f_n = \lambda_n f_n$$

and

$$-\Delta f = \sum_{n\geq 0} \lambda_n \langle f, f_n \rangle f_n$$

Imagine f_n as a substitute for sines and cosines in Fourier series.

Laplace–Beltrami operator

Consider a *compact* Riemannian manifold (M, d) and denote the Laplace–Beltrami operator on (M, d) by $\Delta : C^\infty(M) \mapsto L^2(M)$.

Theorem (Sturm–Liouville decomposition)
There exists an orthonormal basis $(f_n)_{n\in\mathbb{Z}_+}$ *of the space* $L^2(M)$*, and a sequence of non-negative numbers* $0 = \lambda_0 < \lambda_1 \leq \lambda_n \leq \dots$ *such that*

$$-\Delta f_n = \lambda_n f_n$$

and

$$-\Delta f = \sum_{n\geq 0} \lambda_n \langle f, f_n\rangle f_n$$

Imagine f_n as a substitute for sines and cosines in Fourier series.

Laplace–Beltrami operator

Consider a *compact* Riemannian manifold (M, d) and denote the Laplace–Beltrami operator on (M, d) by $\Delta : C^\infty(M) \mapsto L^2(M)$.

Theorem (Sturm–Liouville decomposition)
There exists an orthonormal basis $(f_n)_{n \in \mathbb{Z}_+}$ *of the space* $L^2(M)$*, and a sequence of non-negative numbers* $0 = \lambda_0 < \lambda_1 \leq \lambda_n \leq \dots$ *such that*

$$-\Delta f_n = \lambda_n f_n$$

and

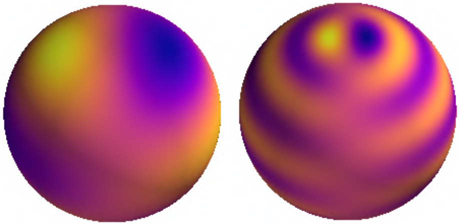
$$-\Delta f = \sum_{n \geq 0} \lambda_n \langle f, f_n \rangle f_n$$

Imagine f_n as a substitute for sines and cosines in Fourier series.

Visual guide to Laplace–Beltrami eigenfunctions

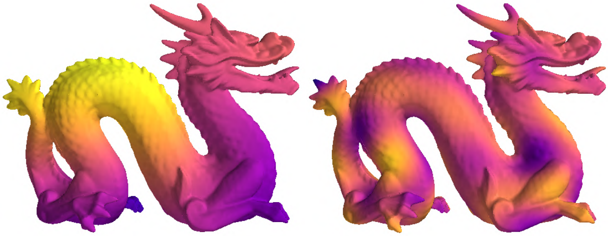
Sample eigenfunction on the sphere

Visual guide to Laplace–Beltrami eigenfunctions



Sample eigenfunction on the sphere

Visual guide to Laplace–Beltrami eigenfunctions



Sample eigenfunction on the dragon

Solving Mátern SPDE on a compact Riemannian manifold

With

$$-\Delta f = \sum_{n \geq 0} \lambda_n \langle f, f_n \rangle f_n$$

it is natural to define

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^p f = \sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n \rangle f_n$$

Somewhat informally, we can represent the Gaussian white noise by

$$\mathcal{W} = \sum_{n \geq 0} w_n f_n, \qquad w_n \sim N(0, 1) \text{ (i.i.d.)}$$

Then the equation can be represented in form

$$\sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n \rangle f_n = \sum_{n \geq 0} w_n f_n.$$

Solving Mátern SPDE on a compact Riemannian manifold

With

$$-\Delta f = \sum_{n \geq 0} \lambda_n \langle f, f_n \rangle f_n$$

it is natural to define

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^p f = \sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n \rangle f_n$$

Somewhat informally, we can represent the Gaussian white noise by

$$\mathcal{W} = \sum_{n \geq 0} w_n f_n, \qquad w_n \sim N(0,1) \text{ (i.i.d.)}$$

Then the equation can be represented in form

$$\sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n \rangle f_n = \sum_{n \geq 0} w_n f_n.$$

Solving Mátern SPDE on a compact Riemannian manifold

With

$$-\Delta f = \sum_{n \geq 0} \lambda_n \langle f, f_n \rangle f_n$$

it is natural to define

$$\left(\frac{2\nu}{\kappa^2} - \Delta \right)^p f = \sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n \right)^p \langle f, f_n \rangle f_n$$

Somewhat informally, we can represent the Gaussian white noise by

$$\mathcal{W} = \sum_{n \geq 0} w_n f_n, \qquad w_n \sim N(0, 1) \text{ (i.i.d.)}$$

Then the equation can be represented in form

$$\sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n \right)^p \langle f, f_n \rangle f_n = \sum_{n \geq 0} w_n f_n.$$

Solving Mátern SPDE on a compact Riemannian manifold

With

$$-\Delta f = \sum_{n\geq 0} \lambda_n \langle f, f_n\rangle f_n$$

it is natural to define

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^p f = \sum_{n\geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n\rangle f_n$$

Somewhat informally, we can represent the Gaussian white noise by

$$\mathcal{W} = \sum_{n\geq 0} w_n f_n, \qquad w_n \sim N(0,1) \text{ (i.i.d.)}$$

Then the equation can be represented in form

$$\sum_{n\geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n\rangle f_n = \sum_{n\geq 0} w_n f_n.$$

Solving Mátern SPDE on a compact Riemannian manifold

The equation from the end of the previous slide:

$$\sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle f_n = \sum_{n\geq 0} w_n f_n.$$

Continuing our informal computation, we get

$$\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle = w_n \qquad \Longrightarrow \qquad \langle f, f_n\rangle = \left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n$$

Hence

$$f = \sum_{n\geq 0}\langle f, f_n\rangle f_n = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n f_n.$$

Finally

$$k(x, x') = \mathrm{Cov}(f(x), f(x')) = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-2p} f_n(x) f_n(x').$$

This is the formula for the Matérn kernel!

Solving Mátern SPDE on a compact Riemannian manifold

The equation from the end of the previous slide:

$$\sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle f_n = \sum_{n\geq 0} w_n f_n.$$

Continuing our informal computation, we get

$$\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle = w_n \quad \Longrightarrow \quad \langle f, f_n\rangle = \left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n$$

Hence

$$f = \sum_{n\geq 0}\langle f, f_n\rangle f_n = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n f_n.$$

Finally

$$k(x,x') = \mathrm{Cov}(f(x), f(x')) = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-2p} f_n(x) f_n(x').$$

This is the formula for the Matérn kernel!

Solving Mátern SPDE on a compact Riemannian manifold

The equation from the end of the previous slide:

$$\sum_{n\geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n \rangle f_n = \sum_{n\geq 0} w_n f_n.$$

Continuing our informal computation, we get

$$\left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^p \langle f, f_n \rangle = w_n \qquad \Longrightarrow \qquad \langle f, f_n \rangle = \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{-p} w_n$$

Hence

$$f = \sum_{n\geq 0} \langle f, f_n \rangle f_n = \sum_{n\geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{-p} w_n f_n.$$

Finally

$$k(x, x') = \mathrm{Cov}(f(x), f(x')) = \sum_{n\geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{-2p} f_n(x) f_n(x').$$

This is the formula for the Matérn kernel!

Solving Mátern SPDE on a compact Riemannian manifold

The equation from the end of the previous slide:

$$\sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle f_n = \sum_{n\geq 0} w_n f_n.$$

Continuing our informal computation, we get

$$\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle = w_n \qquad \Longrightarrow \qquad \langle f, f_n\rangle = \left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n$$

Hence

$$f = \sum_{n\geq 0}\langle f, f_n\rangle f_n = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n f_n.$$

Finally

$$k(x, x') = \mathrm{Cov}(f(x), f(x')) = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-2p} f_n(x) f_n(x').$$

This is the formula for the Matérn kernel!

Solving Mátern SPDE on a compact Riemannian manifold

The equation from the end of the previous slide:

$$\sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle f_n = \sum_{n\geq 0} w_n f_n.$$

Continuing our informal computation, we get

$$\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^p \langle f, f_n\rangle = w_n \qquad \Longrightarrow \qquad \langle f, f_n\rangle = \left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n$$

Hence

$$f = \sum_{n\geq 0}\langle f, f_n\rangle f_n = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-p} w_n f_n.$$

Finally

$$k(x,x') = \mathrm{Cov}(f(x), f(x')) = \sum_{n\geq 0}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-2p} f_n(x) f_n(x').$$

This is the formula for the Matérn kernel!

Riemannian Matérn kernels on compact spaces

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

λ_n, f_n are Laplace–Beltrami eigenpairs (known analytically or approximated numerically).

This is the Karhunen–Loéve type expansion: $f_n(\cdot)$ are analogous to Fourier features.

Figure: values of Matérn kernel $k_{1/2}(x, \cdot)$. x is marked with a red dot.

Riemannian Matérn kernels on compact spaces

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

λ_n, f_n are Laplace–Beltrami eigenpairs (known analytically or approximated numerically).

This is the Karhunen–Loéve type expansion: $f_n(\cdot)$ are analogous to Fourier features.

Figure: values of Matérn kernel $k_{1/2}(x,\cdot)$. x is marked with a red dot.

Riemannian Matérn kernels on compact spaces

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

λ_n, f_n are Laplace–Beltrami eigenpairs (known analytically or approximated numerically).

This is the Karhunen–Loéve type expansion: $f_n(\cdot)$ are analogous to Fourier features.

Figure: values of Matérn kernel $k_{1/2}(x, \cdot)$. x is marked with a red dot.

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

λ_n, f_n are Laplace–Beltrami eigenpairs (known analytically or approximated numerically).

This is the Karhunen–Loéve type expansion: $f_n(\cdot)$ are analogous to Fourier features.

Figure: values of Matérn kernel $k_{1/2}(x, \cdot)$. x is marked with a red dot.

Riemannian Matérn kernels on compact spaces

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

λ_n, f_n are Laplace–Beltrami eigenpairs (known analytically or approximated numerically).

This is the Karhunen–Loéve type expansion: $f_n(\cdot)$ are analogous to Fourier features.

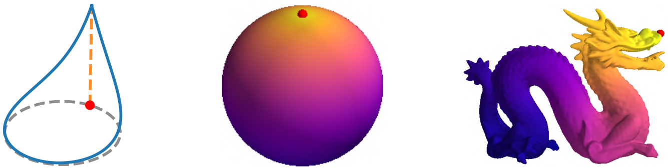


Figure: values of Matérn kernel $k_{1/2}(x, \cdot)$. x is marked with a red dot.

Toy examples

A Gaussian process regression problem on the dragon

(a) Ground truth

(b) Posterior mean

(c) Standard deviation

(d) A sample path

A Gaussian process regression problem on the dragon

(a) Ground truth

(b) Posterior mean

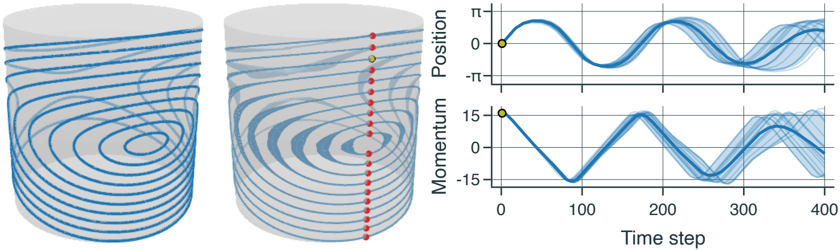
(c) Standard deviation

(d) A sample path

Pendulum dynamics

(a) Ground truth **(b)** 95%-confidence **(c)** Trajectories

Pendulum dynamics



(a) Ground truth **(b)** 95%-confidence **(c)** Trajectories

Conclusion

Summary

Why useful:

- if you know Laplace–Beltrami eigenpairs, you have explicit formulas for Matérn kernels,
- these kernels can be used in the usual GP framework including sparse regression, Fourier feature sampling etc.

What's next:

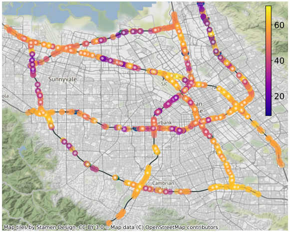
- experiments with Bayesian optimization,
- vector fields,
- also...

Matérn Gaussian Processes on Graphs

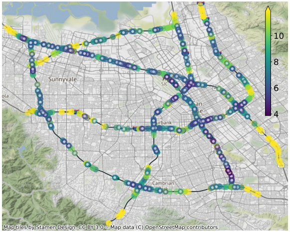
Viacheslav Borovitskiy*[1,5] Iskander Azangulov*[1] Alexander Terenin*[2]
Peter Mostowsky[1] Marc Peter Deisenroth[3] Nicolas Durrande[4]

[1]St. Petersburg State University [2]Imperial College London [3]University College London [4]Secondmind
[5]St. Petersburg Department of Steklov Mathematical Institute of Russian Academy of Sciences

In review for AISTATS 2021



(a) Mean



(b) Standard deviation

Concluding remarks

Thank you for your attention!

Special thanks to Peter Mostowsky for his help with these slides.

Blog post:

https://sml-group.cc/blog/2020-gp-sampling/

GitHub:

https://github.com/spbu-math-cs/Riemannian-Gaussian-Processes

Feel free to email me via viacheslav.borovitskiy@gmail.com

St Petersburg University

Imperial College London

UCL