Gaussian Process Surrogates

Fehmi Cirak
University of Cambridge

UKACM-SEMNI Autumn School
Data-Centric Engineering in Computational Mechanics
16–19 September 2024

Today's Schedule

- 09:00–09:50 Bayesian Inverse Problems
  - 10:00–11:00 Hands-on Session

- 11:15–12:10 Gaussian Process Surrogates
  - 13:30–14:30 Hands-on Session
- 14:45–15:35 Statistical Finite Elements
  - 15:45–16:45 Hands-on Session
- 16:45–17:00 Summary & Discussion

Asking questions by typing in Q&A or raising hand both are fine

Need for Surrogate Modeling

- Advanced computational models are costly to evaluate
  - Problematic for multi-query tasks, like optimisation, uncertainty quantification, inverse problems and model calibration
- Surrogates provide quick-to-evaluate approximation to expensive computational models
  - Trained offline by fitting to data collected from a computational model
- Various surrogate modeling techniques exist
  - Neural networks, polynomial chaos expansions, Gaussian processes, …
- Surrogate modeling can be posed as a Bayesian inference problem
  - Probabilistic surrogates provide a mean and uncertainty estimates

Outline

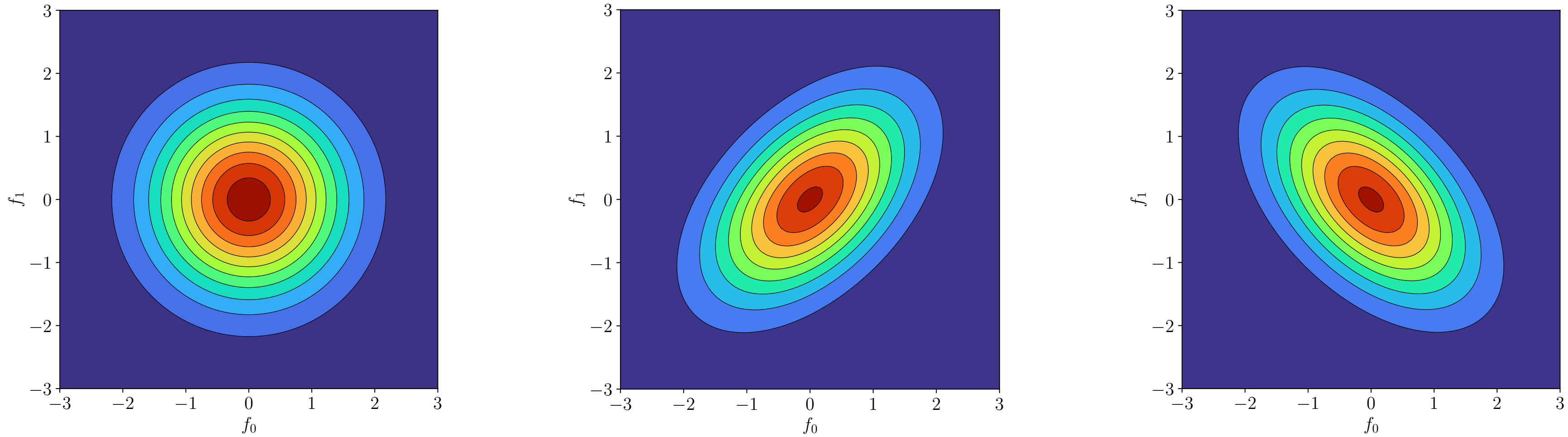
- Introduction to Gaussian Processes (GPs)
  - GPs as generalisation of Gaussian random vectors
  - Covariance functions
  - Conditioning and marginalization of multivariate Gaussians
- GP regression
  - GP priors
  - Observation model
  - Bayes rule and GP posterior
- Advanced topics
  - Stochastic PDE representation of random fields
  - GP regression on non-Euclidean bounded domains
  - Sparse precision formulation of GP regression

Bivariate Gaussians (Recap) –1–

- Probability density of the random vector $\boldsymbol{f} \in \mathbb{R}^2$

$$p(\boldsymbol{f}) = N(\overline{\boldsymbol{f}}, \boldsymbol{C}) = \frac{1}{2\pi\sqrt{\det \boldsymbol{C}}} \exp\left(-\frac{1}{2}(\boldsymbol{f} - \overline{\boldsymbol{f}})^{\mathsf{T}} \boldsymbol{C}^{-1} (\boldsymbol{f} - \overline{\boldsymbol{f}})\right)$$

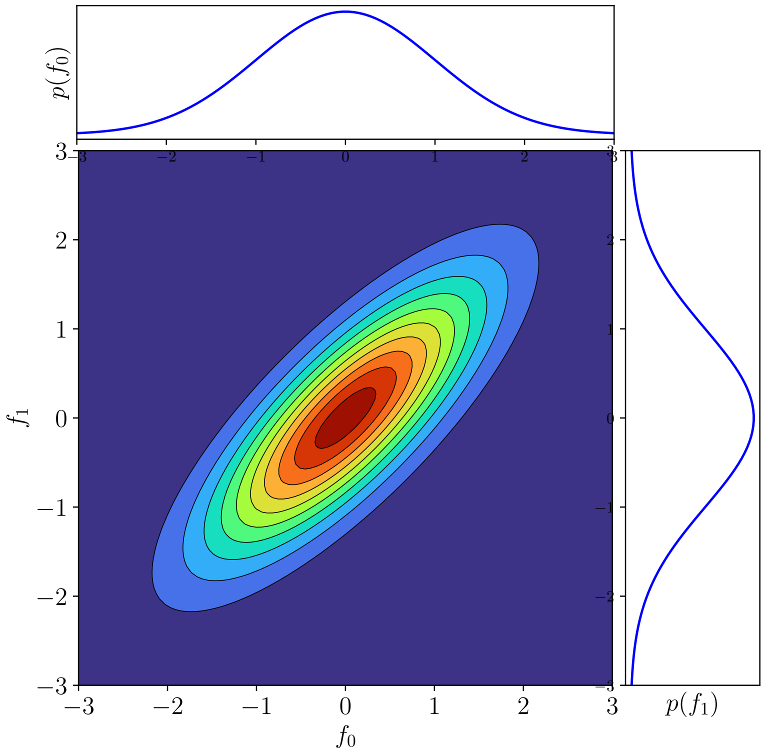
  - With the mean $\overline{\boldsymbol{f}} \in \mathbb{R}^2$ and covariance matrix $\boldsymbol{C} \in \mathbb{R}^{2\times 2}$
  - Properties of the random vector mostly governed by the covariance matrix

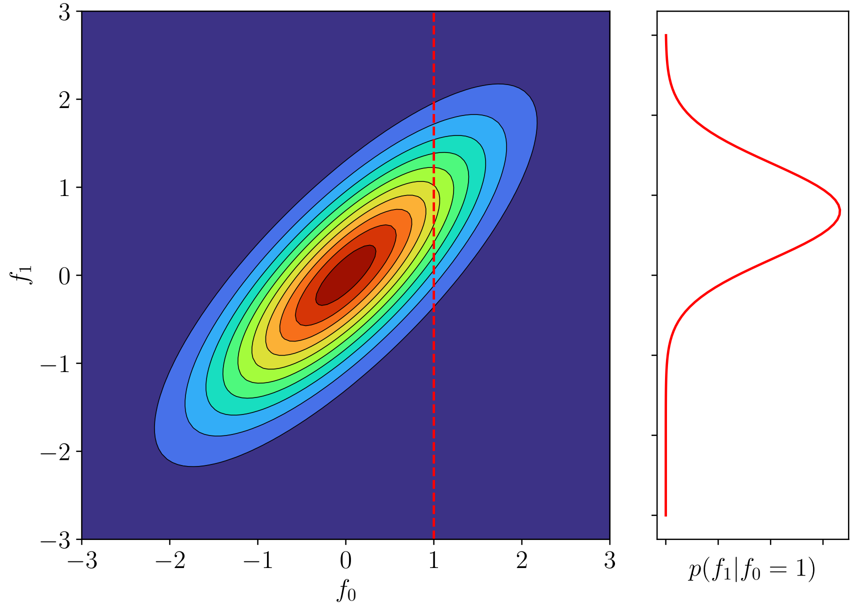


$\boldsymbol{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\boldsymbol{C} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ $\boldsymbol{C} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$

Bivariate Gaussians (Recap) –2–

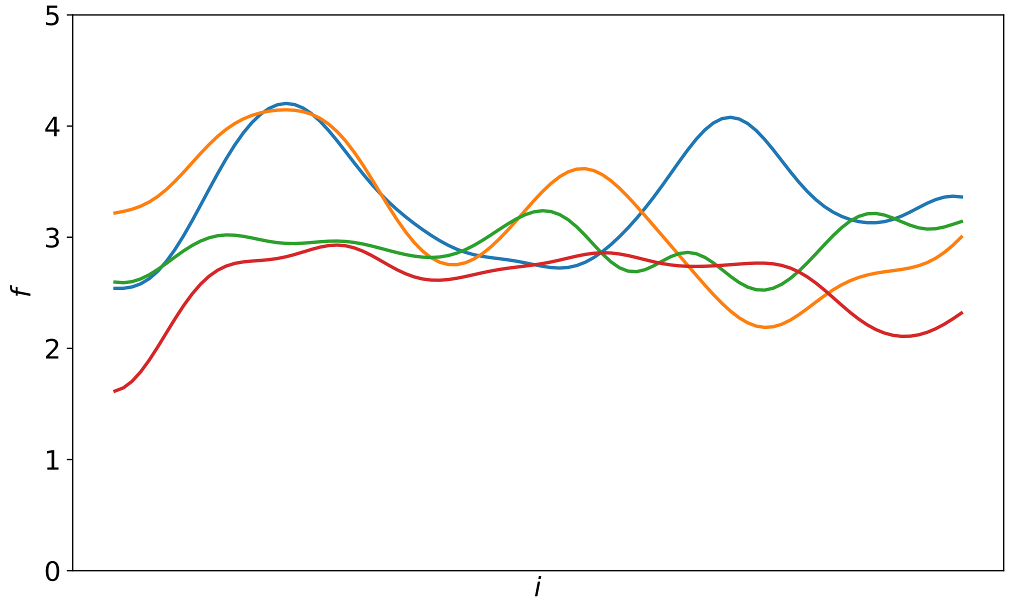
- Marginal probability densities $p(f_0), p(f_1)$ and conditional densities $p(f_0|f_1), p(f_1|f_0)$ are all Gaussians





Multivariate Gaussians

- Random vector $\boldsymbol{f} \in \mathbb{R}^{\gg 2}$
  - Components plotted over the indices



  - Each line is a sample from a multivariate Gaussian, i.e., $\boldsymbol{f} \sim \mathcal{N}(\overline{\boldsymbol{f}}, \boldsymbol{C})$
    - Each component f_i is a Gaussian because of the marginalisation property
- Gaussian processes are the infinite dimensional generalisation of multivariate Gaussians
  - *(Random) processes* are also referred to as a *random fields* or *random functions*

Gaussian Processes (GPs)

- A GP is a set of infinite collection of random variables $f \in \mathbb{R}$ indexed by $x \in \mathbb{R}$ such that

$$f(x) \sim \mathcal{GP}(\overline{f}(x), c(x, x'))$$

  - Mean function: $\mathbb{E}[f(x)] = \overline{f}(x)$
  - Covariance function: $\operatorname{cov}(f(x), f(x')) = c(x, x')$
  - GP fully specified by the mean and covariance functions

- Evaluating the GP at a set of arbitrary points $\boldsymbol{x} \in \mathbb{R}^n$ yields a n −dimensional multivariate Gaussian distribution $\mathcal{N}(\overline{\boldsymbol{f}}, \boldsymbol{C})$

  - Mean vector: $\overline{\boldsymbol{f}} = \left(\overline{f}(x_0), \overline{f}(x_1), \ldots, \overline{f}(x_{n-1})\right)^{\mathsf{T}}$
  - Covariance matrix: $\boldsymbol{C} = \begin{pmatrix} c(x_0, x_0) & c(x_0, x_1) & \cdots & c(x_0, x_{n-1}) \\ c(x_1, x_0) & c(x_1, x_1) & \cdots & c(x_1, x_{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ c(x_{n-1}, x_0) & c(x_{n-1}, x_1) & \cdots & c(x_{n-1}, x_{n-1}) \end{pmatrix}$

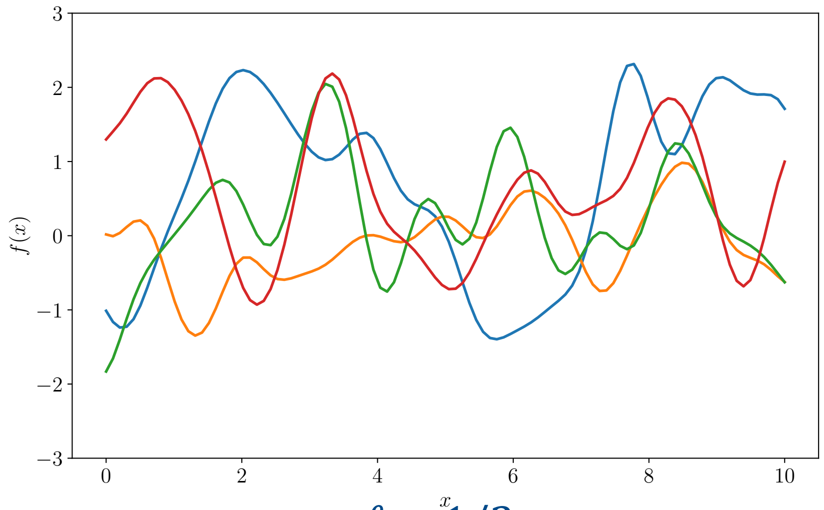
Squared-Exponential Covariance

- Properties of the GP mostly governed by the covariance function
- Consider a GP with $\bar{f}(x) = 0$ and the squared exponential covariance function

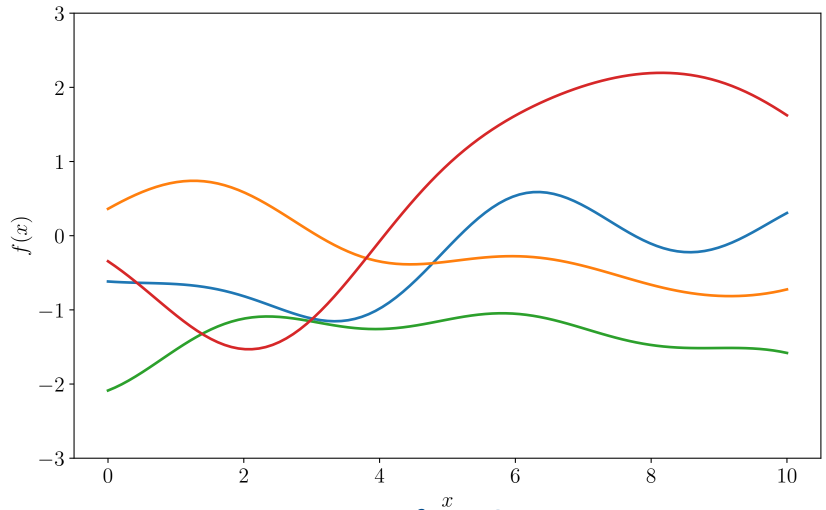
$$c(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\ell^2}\right)$$

  - Parameterised by the length-scale parameter ℓ
  - Depends only on distance between evaluation points x and x'

  - Four random functions sampled from the GP



$\ell = 1/2$



$\ell = 3$

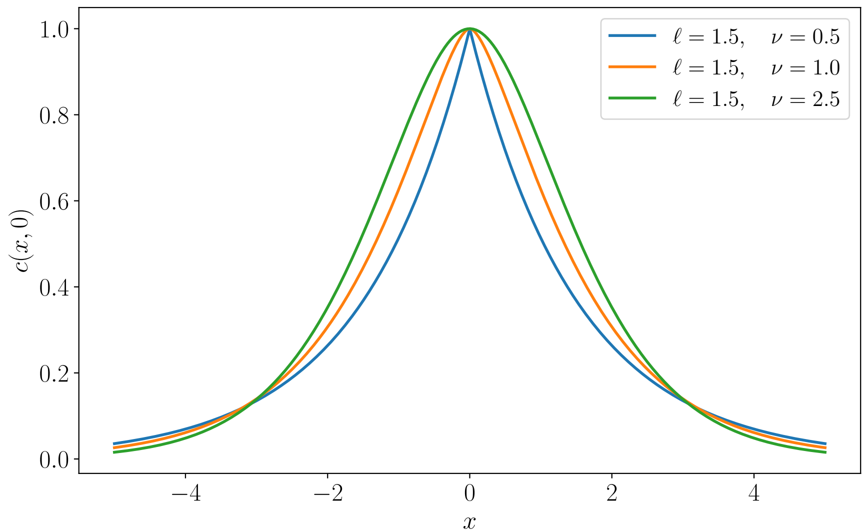
Matérn Covariance

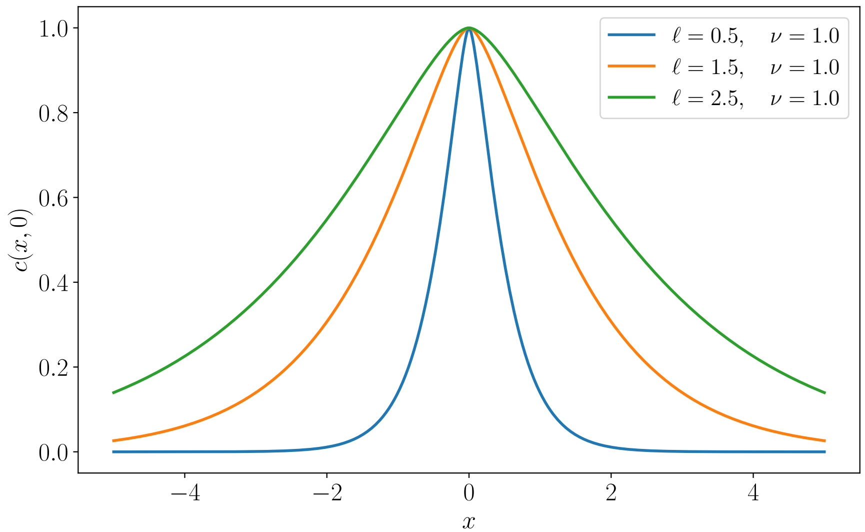
Matérn covariance function

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

- Parameterised by length-scale and smoothness parameters ℓ and ν
- $\Gamma(\nu)$ and K_{ν} are the Gamma and modified Bessel functions of the second kind

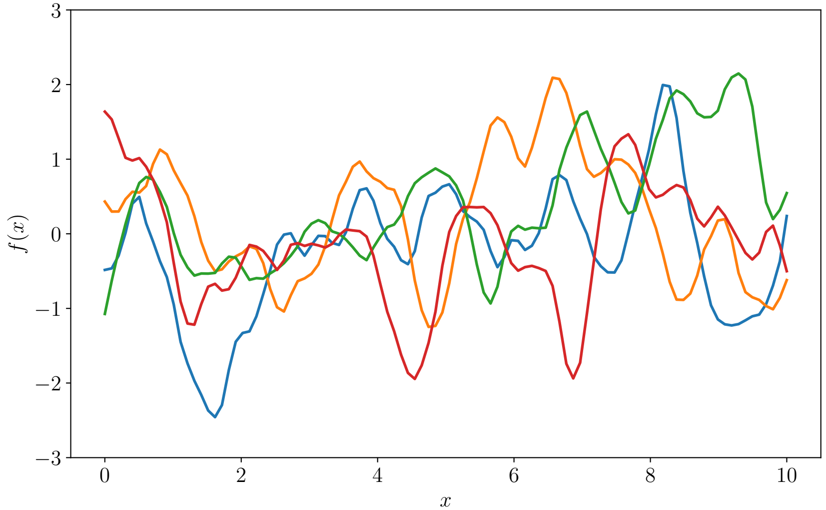
Matérn covariance function $c(x, x' = 0)$ for different ν and ℓ



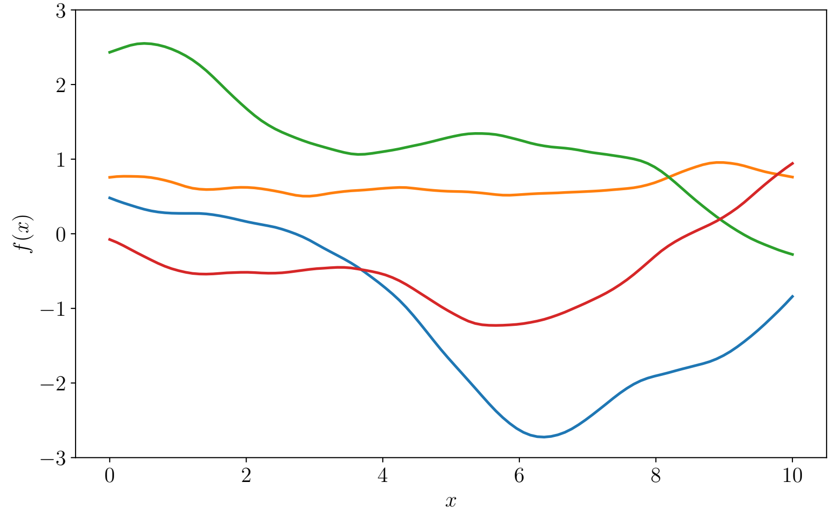


Matérn Covariance – Samples

- Samples from a GP with $\bar{f}(x) = 0$ and Matérn covariance function

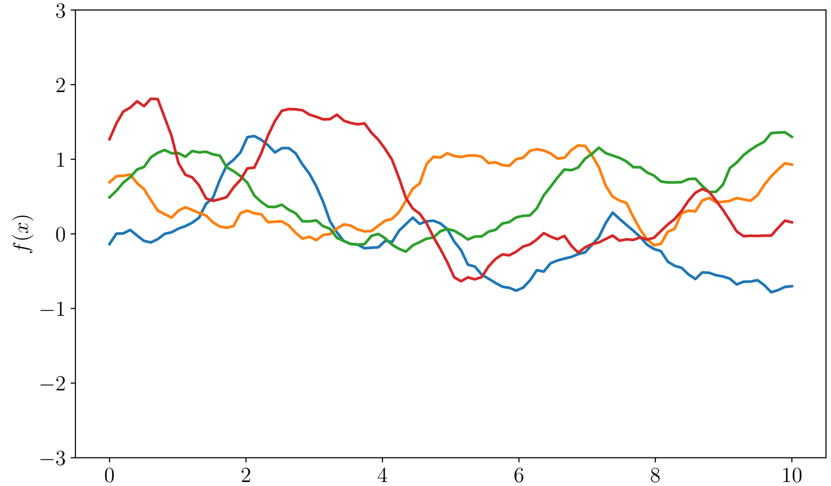


$\ell = 0.5, \nu = 2.5$



$\ell = 3, \nu = 2.5$

$\ell = 0.5, \nu = 1$



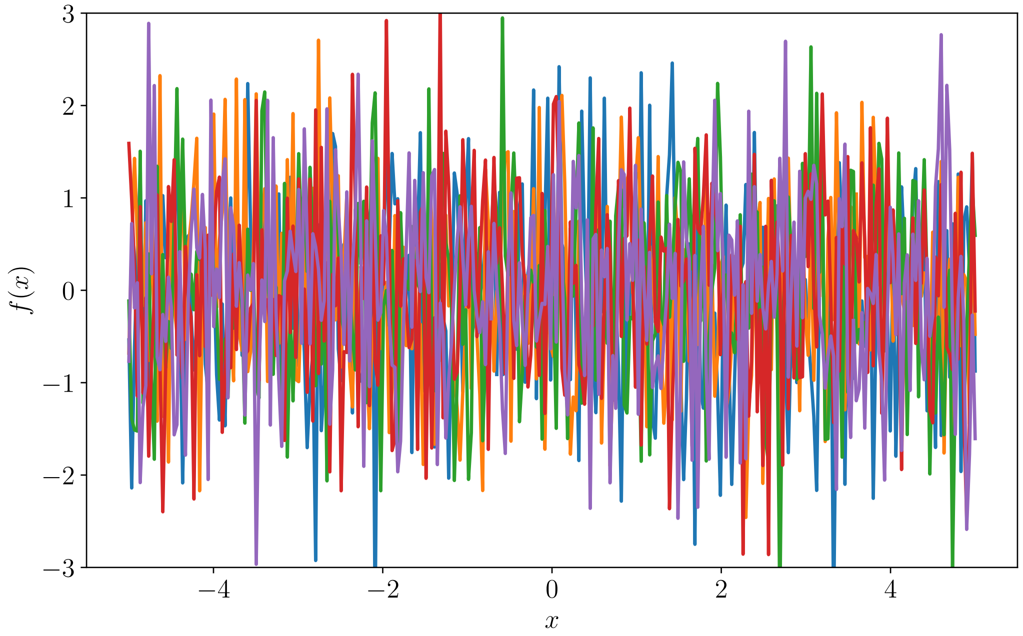
$\ell = 3, \nu = 1$

White Noise Covariance

White Noise

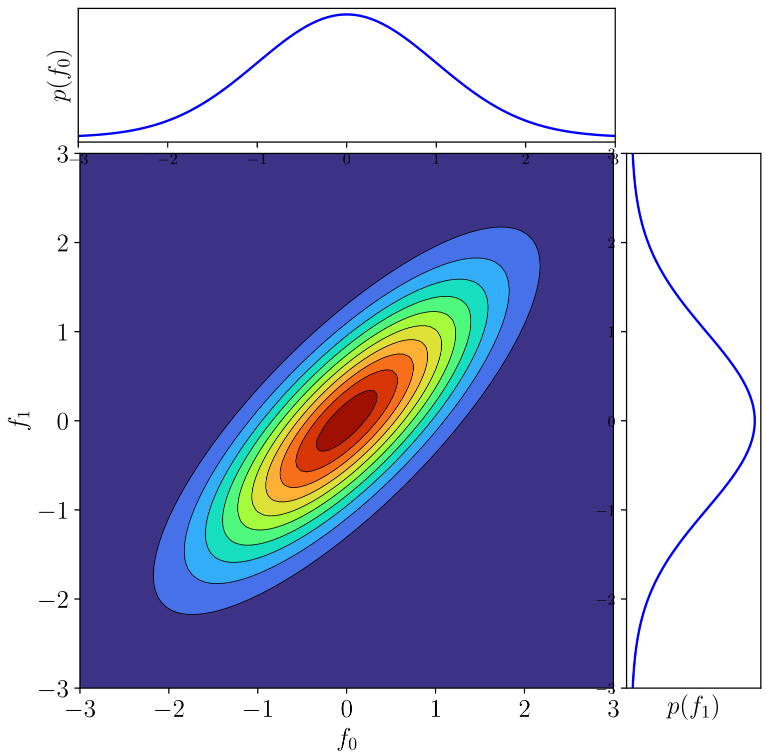
$$c(x, x') = \delta(x - x') = \begin{cases} 1 & \text{if } x = x' \\ 0 & \text{if } x \neq x' \end{cases}$$

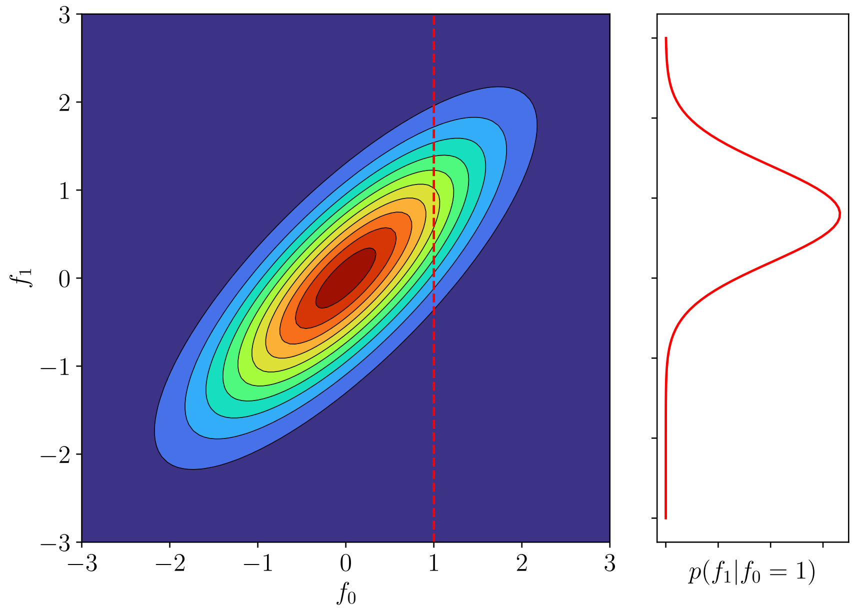
- Note that the respective covariance matrix is a diagonal unit matrix
- Four random functions sampled from a GP with $\bar{f}(x) = 0$ and white noise covariance



Multivariate Gaussians –1–

- Gaussian probability densities are very easy to work with
- Marginal probability densities $p(f_0), p(f_1)$ and conditional densities $p(f_0|f_1), p(f_1|f_0)$ are all Gaussians





Multivariate Gaussians –2–

- Joint probability density of the random vectors $\boldsymbol{f} \in \mathbb{R}^n$ and $\boldsymbol{g} \in \mathbb{R}^m$

$$\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \overline{\boldsymbol{f}} \\ \overline{\boldsymbol{g}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{C}_{ff} & \boldsymbol{C}_{fg} \\ \boldsymbol{C}_{gf} & \boldsymbol{C}_{gg} \end{pmatrix} \right)$$

  - $\overline{\boldsymbol{f}} \in \mathbb{R}^n$, $\overline{\boldsymbol{g}} \in \mathbb{R}^m$, $\boldsymbol{C}_{ff} \in \mathbb{R}^{n\times n}$, $\boldsymbol{C}_{fg} \in \mathbb{R}^{n\times m}$, $\boldsymbol{C}_{gf} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{C}_{gg} \in \mathbb{R}^{m\times m}$

- Conditional probability density of $\boldsymbol{f}$ given $\boldsymbol{g}$

$$p(\boldsymbol{f}|\boldsymbol{g}) = \mathcal{N}\left(\overline{\boldsymbol{f}} + \boldsymbol{C}_{fg}\boldsymbol{C}_{gg}^{-1}(\boldsymbol{g} - \overline{\boldsymbol{g}}), \boldsymbol{C}_{ff} - \boldsymbol{C}_{fg}\boldsymbol{C}_{gg}^{-1}\boldsymbol{C}_{gf}\right)$$

  - Easy to derive using "completion of square" and noting $p(\boldsymbol{f}, \boldsymbol{g}) = p(\boldsymbol{f}|\boldsymbol{g})p(\boldsymbol{g})$

- Marginal probability density of $\boldsymbol{f}$

$$\boldsymbol{f} \sim \mathcal{N}\left(\overline{\boldsymbol{f}}, \boldsymbol{C}_{ff}\right)$$

  - Easy to derive by integrating $p(\boldsymbol{f}) = \int p(\boldsymbol{f}, \boldsymbol{g})\, d\boldsymbol{g}$

Gaussian Process Regression –1–

Data set (independent and identically distributed)

$$\mathcal{D} = \{(x_0, y_0),\ (x_1, y_1),\ \dots,\ (x_{n-1}, y_{n-1})\}$$

Observation model

$$y = f(x) + \epsilon$$

- GP prior $f(x) \sim \mathcal{GP}(\overline{f}(x), c(x, x'))$
- Noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$

Bayesian inference of $f(x)$ in light of data

- Data likelihood $p(\boldsymbol{y}|\boldsymbol{f}) = \prod_{i=0}^{n-1} \mathcal{N}(f(x_i), \sigma^2) = \mathcal{N}(\boldsymbol{f}, \sigma^2 \boldsymbol{I})$
- Prior $p(\boldsymbol{f}) = \mathcal{N}(\overline{\boldsymbol{f}}, \boldsymbol{C}_f)$
- Bayes rule $p(\boldsymbol{f}|\boldsymbol{y}) = \dfrac{p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{y})}$
- Predictive posterior for f_* at x_* not observed $p(f_*|\boldsymbol{y}) = \int p(f_*|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{y})\, d\boldsymbol{f}$

Gaussian Process Regression –2–

- Alternatively, posterior can be derived from the joint probability density

$$\begin{pmatrix} f_* \\ \boldsymbol{y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \overline{f} \\ \overline{\boldsymbol{f}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{C}_{x_* x_*} & \boldsymbol{C}_{x_* \boldsymbol{x}} \\ \boldsymbol{C}_{\boldsymbol{x} x_*} & \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}} + \sigma^2 \boldsymbol{I} \end{pmatrix} \right)$$

  - Covariance components

$$\boldsymbol{C}_{x_* x_*} = \operatorname{cov}(f_*, f_*) = c(x_*, x_*) \,,$$

$$\boldsymbol{C}_{x_* x} = \operatorname{cov}(f_*, \boldsymbol{f}) = \begin{pmatrix} c(x_*, x_0) & c(x_*, x_1) \cdots c(x_*, x_{n-1}) \end{pmatrix}$$

$$\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}} = \operatorname{cov}(\boldsymbol{f}, \boldsymbol{f}) + \sigma^2 \boldsymbol{I}$$

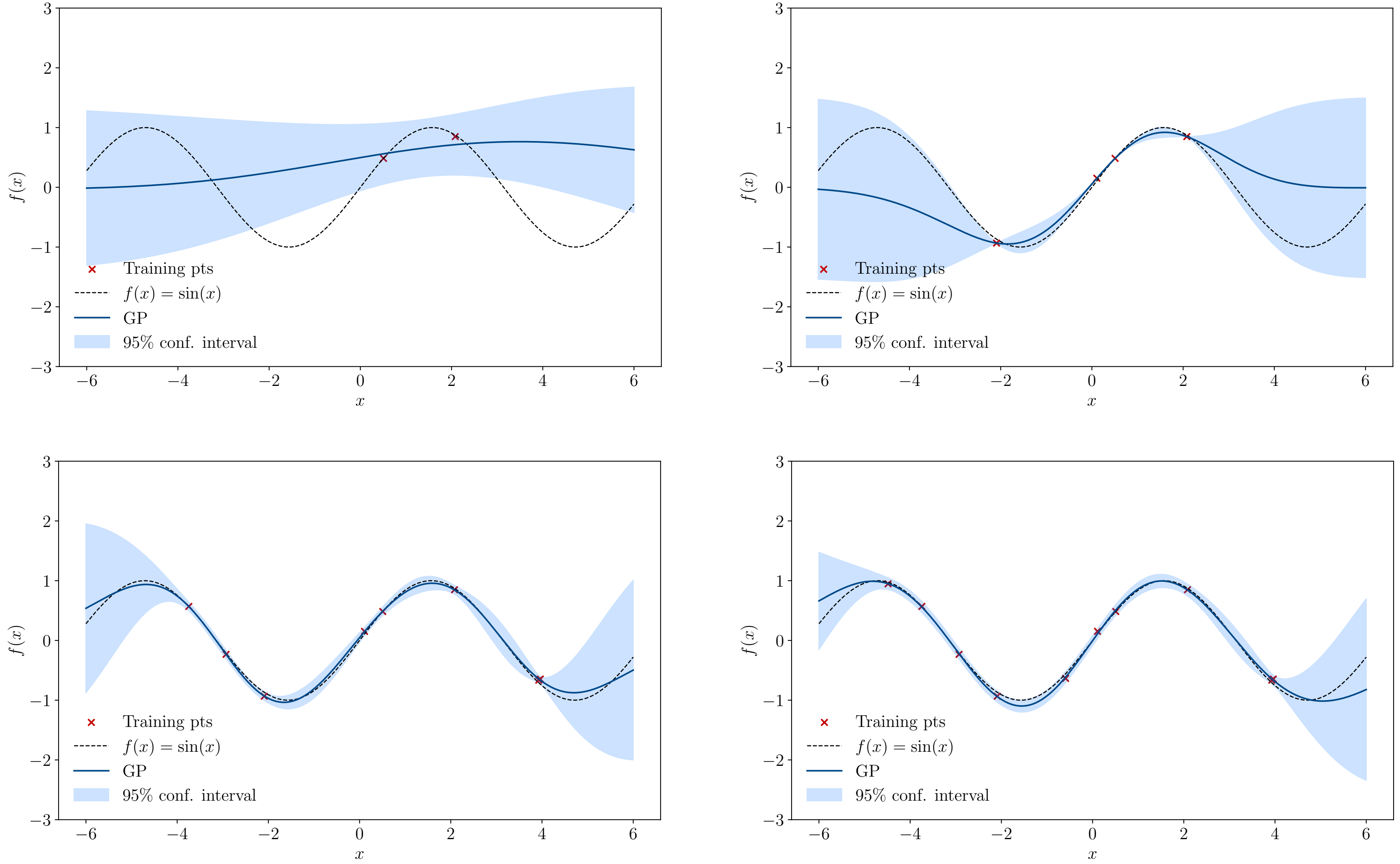
$$= \begin{pmatrix} c(x_0, x_0) + \sigma^2 & c(x_0, x_1) & \cdots & c(x_0, x_{n-1}) \\ c(x_1, x_0) & c(x_1, x_1) + \sigma^2 & \cdots & c(x_1, x_{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ c(x_{n-1}, x_0) & c(x_{n-1}, x_1) & \cdots & c(x_{n-1}, x_{n-1}) + \sigma^2 \end{pmatrix}$$

- Conditional density of f_* given $\boldsymbol{y}$

$$f_{*|\boldsymbol{y}} \sim \mathcal{N}\left(\overline{f}_{x_*} + \boldsymbol{C}_{x_* \boldsymbol{x}} \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}}^{-1} (\boldsymbol{y} - \overline{\boldsymbol{f}}), \, \boldsymbol{C}_{x_* x_*} - \boldsymbol{C}_{x_* \boldsymbol{x}} \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}}^{-1} \boldsymbol{C}_{\boldsymbol{x} x_*} \right)$$

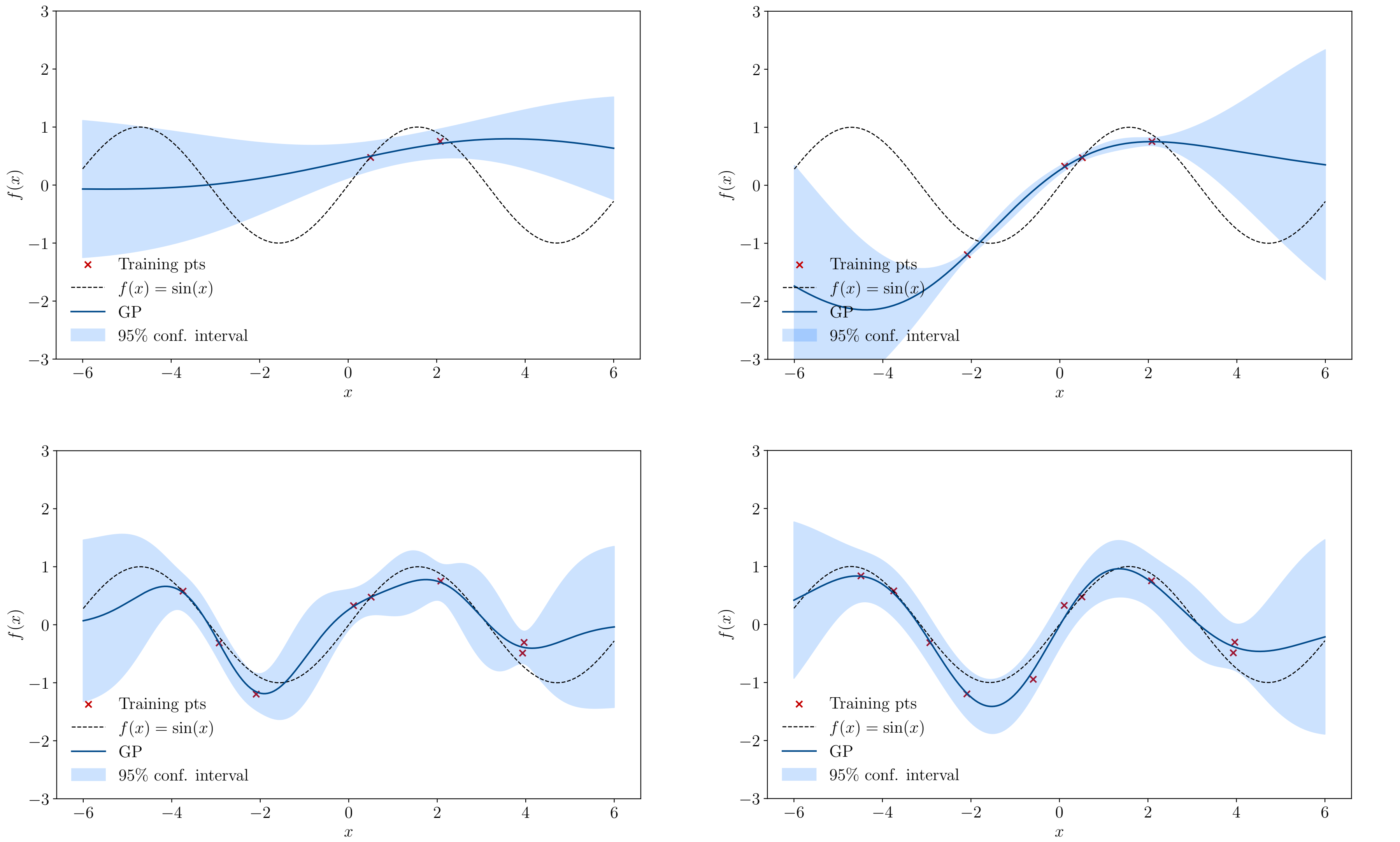
Gaussian Process Regression – Example

- Noise standard deviation $\sigma = 0.05$



Gaussian Process Regression – Example

- Noise standard deviation $\sigma = 0.25$



Hyperparameters

- Covariance matrix of the GP prior depends on hyperparameters controlling length-scale, smoothness, etc
- Hyperparameters can be determined by maximising the marginal likelihood, recall

$$p(\boldsymbol{f}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{y})}$$

  - Marginal likelihood

$$p(\boldsymbol{y}) = \mathcal{N}(\overline{\boldsymbol{f}}, \boldsymbol{C}_{\boldsymbol{xx}} + \sigma^2 \boldsymbol{I})$$

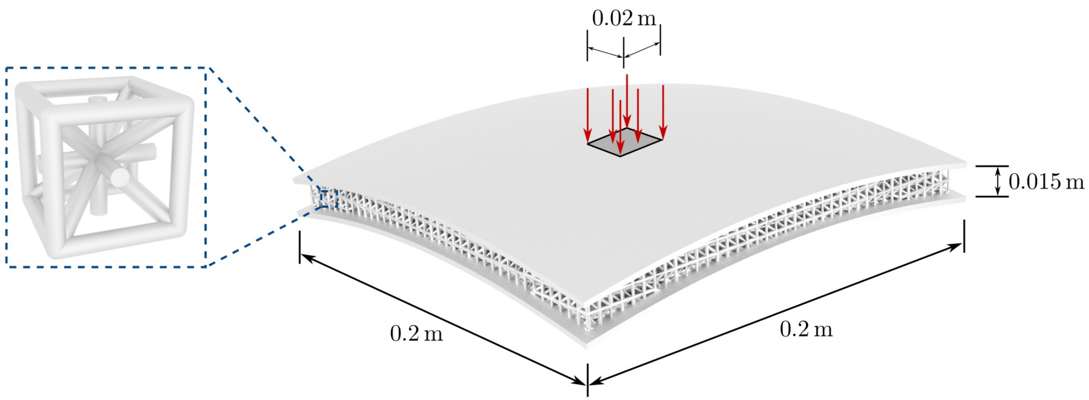
    - Probability of observing $\boldsymbol{y}$ with the considered model
    - Expedient to choose a model for which marginal likelihood is maximum
    - Numerically more stable to maximise the log marginal likelihood

$$\log p(\boldsymbol{y}) = -\frac{1}{2}(\boldsymbol{y} - \overline{\boldsymbol{f}})^\mathsf{T}(\boldsymbol{C}_{\boldsymbol{xx}} + \sigma^2 \boldsymbol{I})^{-1}(\boldsymbol{y} - \overline{\boldsymbol{f}}) - \frac{1}{2}\log\det(\boldsymbol{C}_{\boldsymbol{xx}} + \sigma^2 \boldsymbol{I}) - \frac{n}{2}\log 2\pi$$

      - Not convex certain care must be taken when maximising

Limitations

- GP regression has time and storage complexity of $\mathcal{O}(n^3)$ and $\mathcal{O}(n^2)$
  - Posterior and marginal likelihood computation require the inversion of the dense covariance matrix
  - Impossible to have $\gtrsim 1000$ observation points
  - Sparse GPs can reduce time complexity to $\mathcal{O}(nm^2)$ using m inducing points
- Furthermore, covariance functions introduced depend on Euclidean distance and are isotropic and stationary
  - Not suitable for non-Euclidean geometries like shells or mechanical assemblies



  - Correlation in material parameters may depend non-trivially on Euclidean distance

Solution: Stochastic PDEs

- Matérn random fields can be expressed as the solution of a partial differential equation (PDE) with random forcing
  - Such PDEs are referred to as stochastic partial differential equations (SPDEs)
- SPDEs can be discretised using standard finite elements
  - Inverse of the SPDE stiffness matrix is the covariance matrix of the Matérn random field
    - Inverse of the covariance matrix is called the precision matrix
  - SPDE stiffness matrix (i.e. precision matrix) is sparse

Matérn Fields via Fractional SPDEs –1–

- Stochastic partial differential equation with fractional exponent β

$$\mathcal{L}^{\beta}\hat{u}(x) = (\kappa^2 - \Delta)^{\beta}\hat{u}(x) = \frac{1}{\tau}\hat{g}(x)$$

  - β, κ and τ are parameters and Δ is the Laplacian
  - Gaussian white noise forcing

$$\hat{g}(x) \sim \mathcal{GP}(0, \delta(x, x'))$$

- Solution field $\hat{u}(x)$ is a Matérn random field
  - Choosing the Dirac delta $\delta(\boldsymbol{x}')$ as right-hand side yields Matérn covariance function

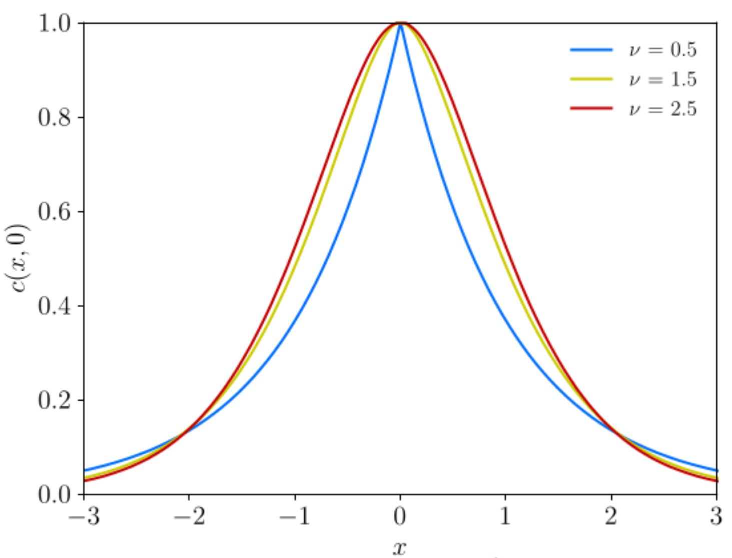
$$\hat{u}(x) = c(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}}{\ell}\|x - x'\|\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}}{\ell}\|x - x'\|\right)$$

  - Correspondence between the parameters

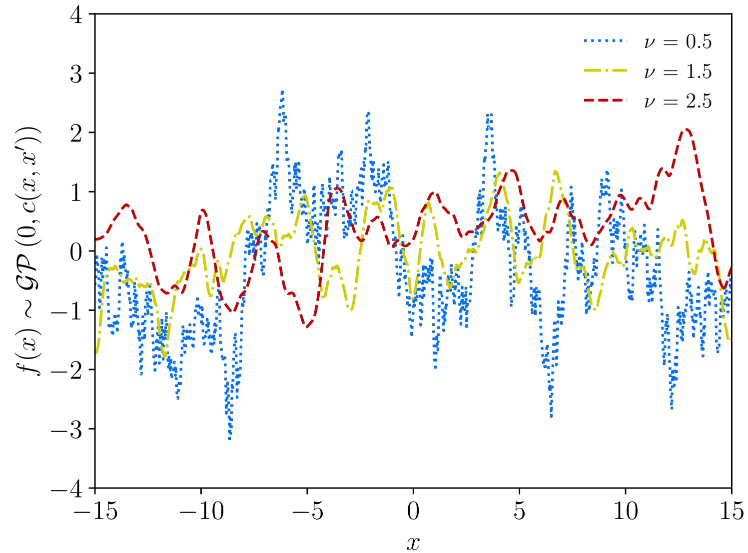
$$\kappa = \frac{\sqrt{2\nu}}{\ell}, \quad \beta = \frac{\nu}{2} + \frac{d}{4}, \quad \tau^2 = \frac{\Gamma(\nu)}{\sigma^2\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}}$$

Matérn Fields via Fractional SPDEs –2–

- Covariance and function and samples



Matérn covariance function



Matérn fields

SPDEs with an Integer Exponent β

- Recursive solution, e.g., for β=2

$$\hat{u}(x) = \mathcal{L}^{-2}\left(\frac{1}{\tau}\hat{g}(x)\right) = \mathcal{L}^{-1}\mathcal{L}^{-1}\left(\frac{1}{\tau}\hat{g}(x)\right)$$

  - Can be reduced to two PDEs

$$\hat{u}_2(x) = \mathcal{L}^{-1}\left(\frac{1}{\tau}\hat{g}(x)\right), \qquad \hat{u}(x) = \mathcal{L}^{-1}\hat{u}_2(x)$$

- FE discretisation of the first PDE

$$\boldsymbol{L}\hat{\boldsymbol{u}}_2 = \left(\kappa^2\boldsymbol{M} + \boldsymbol{A}\right)\hat{\boldsymbol{u}}_2 = \frac{1}{\tau}\hat{\boldsymbol{g}}$$

  - With $\boldsymbol{M}$ the (lumped) mass matrix, $\boldsymbol{A}$ the stiffness matrix and Gaussian forcing

$$\hat{\boldsymbol{g}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{M})$$

  - Probability density

$$\hat{\boldsymbol{u}}_2 \sim \mathcal{N}\left(\boldsymbol{0}, \left(\tau^2\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{L}\right)^{-1}\right) = \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{Q}_\beta^{-1}\right)$$

  - Probability density of the original SPDE

$$\hat{\boldsymbol{u}} \sim \mathcal{N}\left(\boldsymbol{0}, \left(\tau^2\boldsymbol{L}^2\boldsymbol{M}^{-1}\boldsymbol{L}^2\right)^{-1}\right) = \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{Q}^{-1}\right)$$

Generalisation of Matérn Fields

- Anisotropic and non-stationary random fields

$$(\kappa^2(x) - \nabla \cdot (\boldsymbol{H}\nabla))^\beta \hat{u}(x) = \frac{1}{\tau}\hat{g}(x)$$

  - $\boldsymbol{H}$ is an anisotropy matrix

- For random fields on manifolds Laplacian is replaced with Laplace-Beltrami operator

- Finite element discretisation of the respective problems proceeds as usual

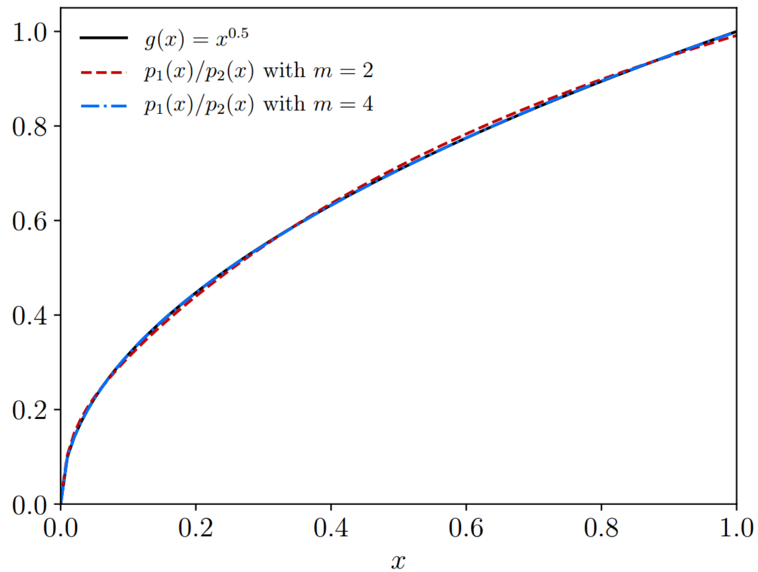
SPDEs with a Fractional Exponent β

- Rational approximation of a power function

$$x^\beta \approx \frac{\sum_{i=0}^{m} a_i x^i}{\sum_{j=0}^{m} b_j x^j} = \frac{a_m \prod_{i=0}^{m-1}(x - c_i)}{b_m \prod_{j=0}^{m-1}(x - d_j)}$$

  - Coefficients $a_i,\ b_j, c_i, d_j$ obtained from an external software library
  - Rational approximation is more accurate than other expansions, like Taylor

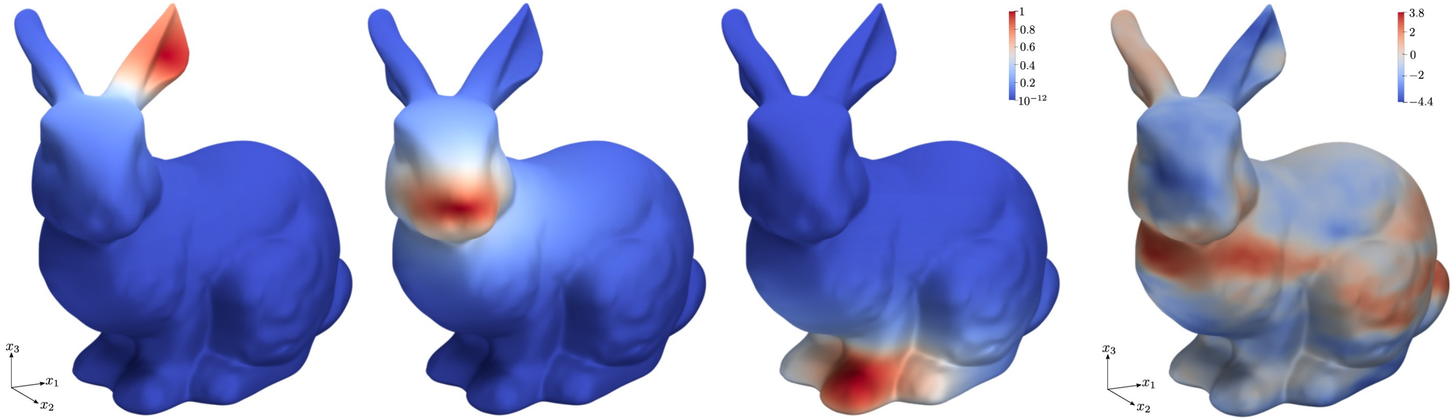
- Key Idea for solving the SPDE is to replace x with the operator $\boldsymbol{L}$



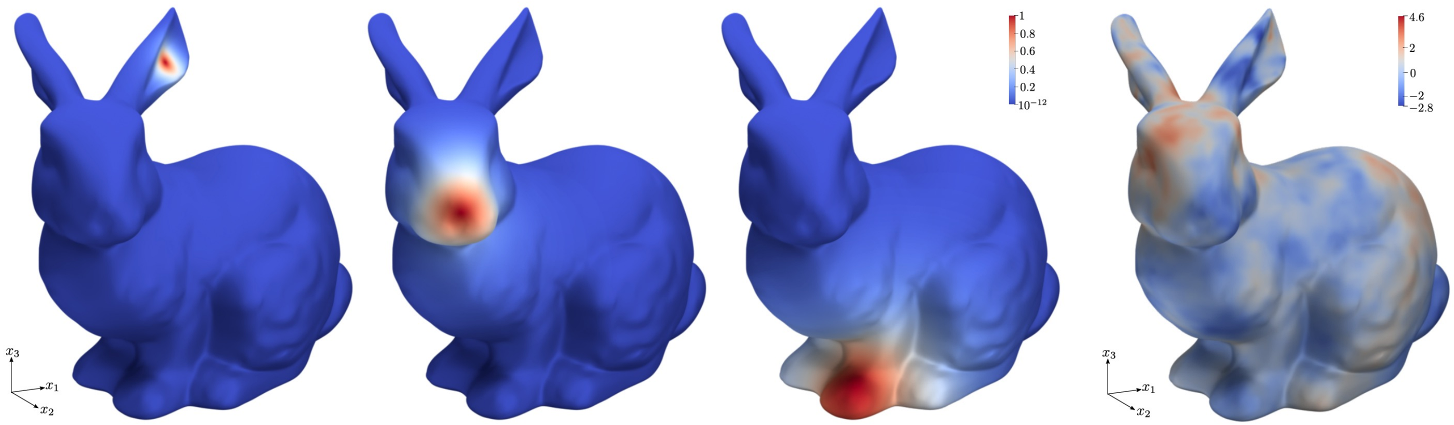
Rational approximation of $\sqrt{x}$

Example

- Anisotropic Matérn random field



- Non-stationary Matérn random field



Sparse GP Regression –1–

- Data set (independent and identically distributed

$$\mathcal{D} = \{(\boldsymbol{x}_0, y_0),\ (\boldsymbol{x}_1, y_1),\ \dots,\ (\boldsymbol{x}_{n-1}, y_{n-1})\}$$

- Observation model

$$\boldsymbol{y} = \boldsymbol{P}\hat{\boldsymbol{u}} + \boldsymbol{e}$$

  - GP prior from SPDE $p(\hat{\boldsymbol{u}}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{C}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$
  - Noise $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I})$
  - Data likelihood $p(\boldsymbol{y}|\hat{\boldsymbol{u}}) = \mathcal{N}(\boldsymbol{P}\hat{\boldsymbol{u}}, \sigma_e^2 \boldsymbol{I})$

- Bayes rule

$$p(\hat{\boldsymbol{u}}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\hat{\boldsymbol{u}})p(\hat{\boldsymbol{u}})}{p(\boldsymbol{y})}$$

Sparse GP Regression –2–

- Posterior can also be derived from the joint probability density
  - Covariance matrix formulation

$$p\left(\hat{\boldsymbol{u}}, \boldsymbol{y}\right) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{C} & \boldsymbol{C}\boldsymbol{P}^{\mathsf{T}} \\ \boldsymbol{P}\boldsymbol{C} & \boldsymbol{P}\boldsymbol{C}\boldsymbol{P}^{\mathsf{T}} + \sigma_e^2 \boldsymbol{I} \end{pmatrix}\right)$$

  - Precision matrix formulation (by inverting the 2×2 covariance)

$$p\left(\hat{\boldsymbol{u}}, \boldsymbol{y}\right) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \frac{1}{\sigma_e^2}\begin{pmatrix} \sigma_e^2 \boldsymbol{Q} + \boldsymbol{P}^{\mathsf{T}}\boldsymbol{P} & -\boldsymbol{P}^{\mathsf{T}} \\ -\boldsymbol{P} & \boldsymbol{I} \end{pmatrix}^{-1}\right)$$

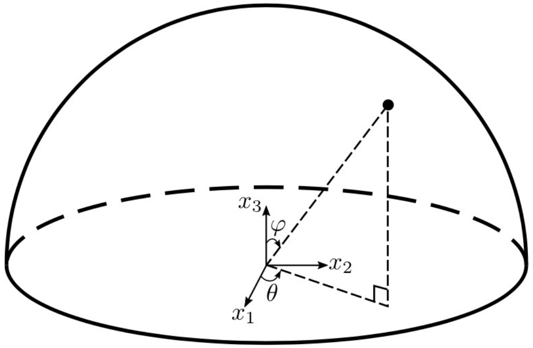
- Posterior (according to well-known results for Gaussians)

$$p(\hat{\boldsymbol{u}}|\boldsymbol{y}) = \mathcal{N}\left(\overline{\hat{\boldsymbol{u}}}_{|y}, \boldsymbol{Q}_{|y}^{-1}\right)$$

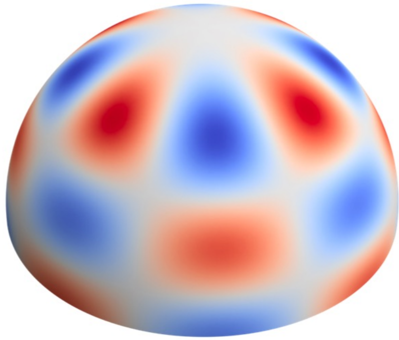
$$\overline{\hat{\boldsymbol{u}}}_{|y} = \frac{1}{\sigma_e^2}\boldsymbol{Q}_{|y}^{-1}\boldsymbol{P}^{\mathsf{T}}\boldsymbol{y} \qquad \boldsymbol{Q}_{|y} = \boldsymbol{Q} + \sigma_e^2 \boldsymbol{P}^{\mathsf{T}}\boldsymbol{P}$$

  - Mean and components of the covariance computed using only sparse matrix operations

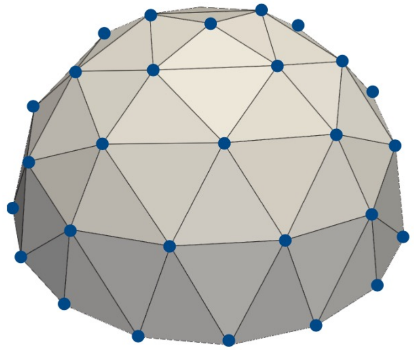
GP Regression on a Hemisphere –1–



Two-manifold



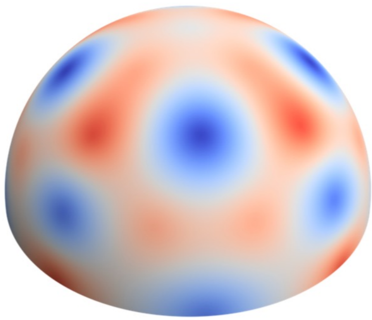
True function



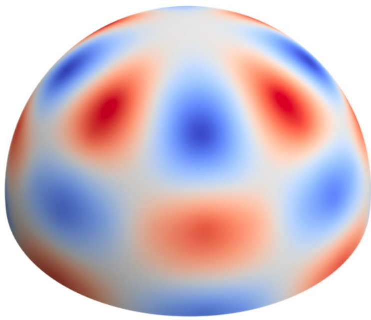
Coarse mesh

- SPDE discretization with about one million linear elements

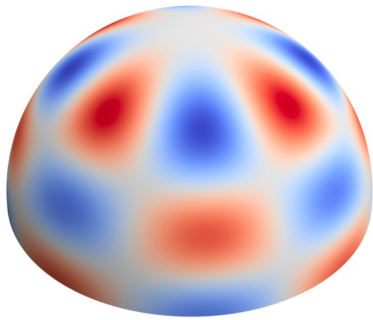
- Inferred function for different number of observation points



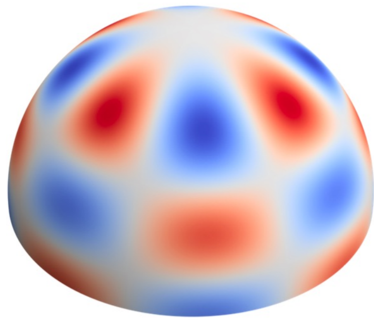
41 observ.



145 observ.



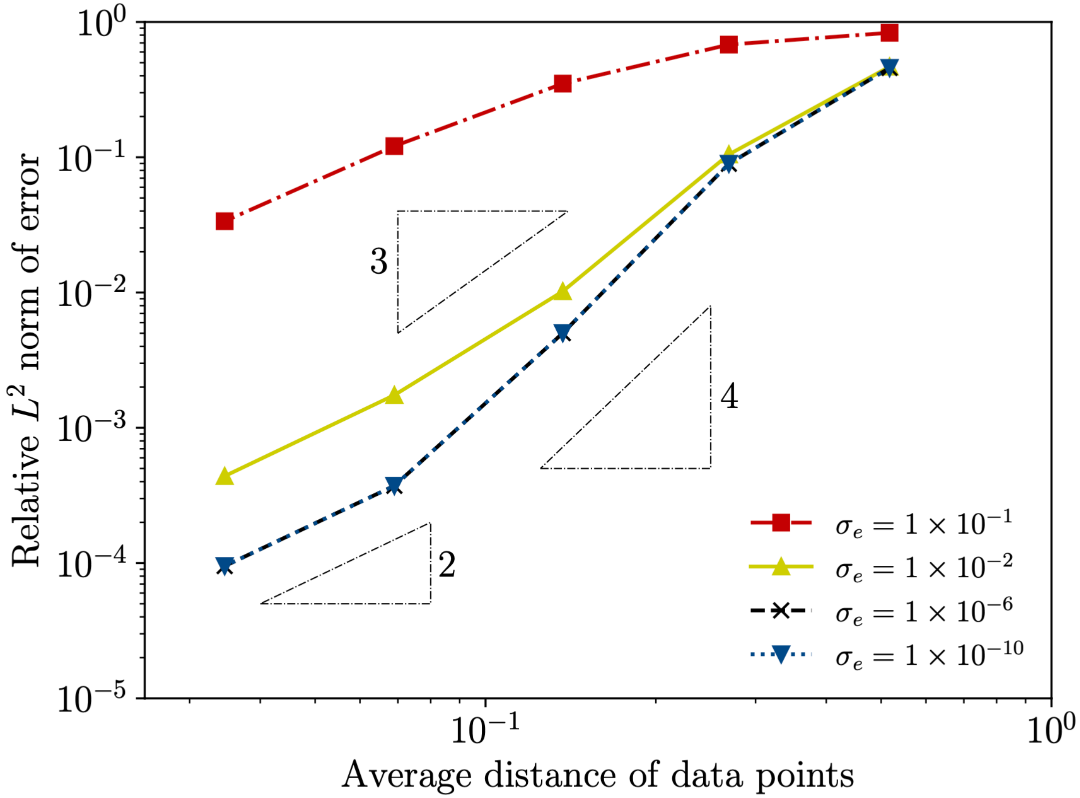
545 observ.



8321 observ.

GP Regression on a Hemisphere –2–

- Convergence with increasing number of observation points



Summary

- GP regression is a powerful probabilistic approximation technique for noisy observation data
  - Properties of the prior are governed by the selected covariance function
  - Posterior is a Gaussian, with closed-form expressions for the mean and covariance
  - Prediction uncertainties are governed by the posterior covariance

- Stochastic representation of the prior is crucial for:
  - Improved scalability through the use of sparse matrices
  - Generalisation to non-Euclidean domains and non-standard random fields

- References

  - Ben-Yelun, Yuksel, Cirak Robust topology optimisation of lattice structures with spatially correlated uncertainties, SMO, 2024
  - Koh, Cirak, Stochastic PDE representation of random fields for large-scale Gaussian process regression and statistical finite element analysis, CMAME, 2023
  - Girolami, Febrianto, Yin, Cirak The statistical finite element method (statFEM) for coherent synthesis of observation data and model predictions, CMAME, 2021

Properties of Gaussians

- Gaussian probability densities are very easy to work with
- Consider, e.g., Gaussian random variables

$$x \sim \mathcal{N}\left(\overline{x}, \sigma_x^2\right), \quad y \sim \mathcal{N}\left(\overline{y}, \sigma_y^2\right)$$

  - Sum of two Gaussian variables is a Gaussian variable

$$x + y \sim \mathcal{N}\left(\overline{x} + \overline{y}, \sigma_x^2 + \sigma_y^2\right)$$

  - Product of a Gaussian variable with a constant is a Gaussian variable

$$cx \sim \mathcal{N}\left(c\overline{x}, c^2\sigma_x^2\right)$$

- Same relationships hold for Gaussian random vectors

$$\boldsymbol{x} \sim \mathcal{N}\left(\overline{\boldsymbol{x}}, \boldsymbol{C}_x\right), \quad \boldsymbol{y} \sim \mathcal{N}\left(\overline{\boldsymbol{y}}, \boldsymbol{C}_y\right)$$

$$\boldsymbol{x} + \boldsymbol{y} \sim \mathcal{N}\left(\overline{\boldsymbol{x}} + \overline{\boldsymbol{y}}, \boldsymbol{C}_x + \boldsymbol{C}_y\right)$$

$$\boldsymbol{A}\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{A}\overline{\boldsymbol{x}}, \boldsymbol{A}\boldsymbol{C}_x\boldsymbol{A}^{\mathsf{T}}\right),$$

Today’s Schedule

- 09:00–09:50 Bayesian Inverse Problems
  - 10:00–11:00 Hands-on Session
- 11:15–12:10 Gaussian Process Surrogates
  - 13:30–14:30 Hands-on Session
- 14:45–15:35 Statistical Finite Elements
  - 15:45–16:45 Hands-on Session
- 16:45–17:00 Summary & Discussion

Asking questions by typing in Q&A or raising hand both are fine