Meeting Mécanique-Energétique Department

Gaussian Process Regression on Vector Fields and Uncertainty Quantification

Rémy Hosseinkhan Boucher, Stella Douka, Onofrio Semeraro, Lionel Mathelin[1]

Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, CNRS

Doctoral School: *Sciences and Technologies of Information and Communication (STIC)*

Granted by the Agence Nationale de la Recherche (ANR) under projet ANR-21-CE46-0008 Reinforcement Learning as Optimal control for Shear Flows (REASON)

Goal, Notations and Definitions

Dynamics (Discrete Time)

$X_{k+1} = F^*(X_k)$

State Space

$\mathcal{X} \subset \mathbb{R}^{d_\mathcal{X}}$

Vector Field

$F^* : \mathcal{X} \to \mathcal{X}$

Dataset

$\mathcal{D} = ((x_i, F^*(x_i)))_{i=1}^n = \left(x^\mathcal{D}, F^{*,\mathcal{D}}\right)$

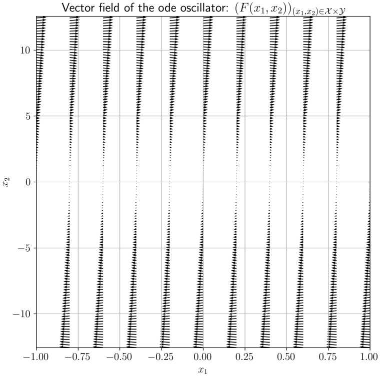
Goal

- Infer F^* from $\mathcal{D}$
- Quantify the uncertainty on F^*
- Sampling strategy to improve the inference

Example (Noisy Oscillator)

$$F^* : ((x_1, x_2)) = \begin{pmatrix} x_2 \\ -x_1^3 + x_1 \end{pmatrix}$$

$d_\mathcal{X} = 2$



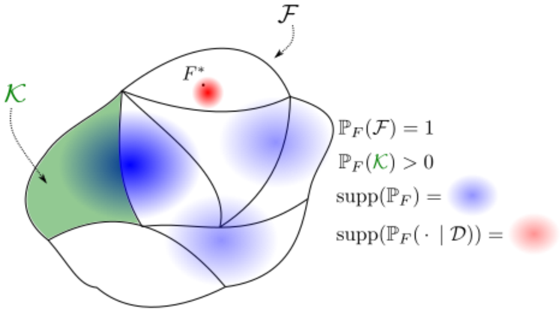
Bayesian Nonparametric Regression

F is unknown

- Hypothesis space (class of function) $\mathcal{F}$
- Suppose $F^* \in \mathcal{F}$

Uncertainty on F

- Define $\mathbb{P}_F : \mathcal{A} \subset \mathcal{P}(\mathcal{F}) \to [0, 1]$ such that $\mathbb{P}_F(\mathcal{F}) = 1$
- $\mathbb{P}_F$ is called a **probability measure**
- Given $\mathcal{K} \subset \mathcal{F}$, $\mathbb{P}_F(\mathcal{K}) \in [0, 1]$ measures the confidence or **belief** that $F \in \mathcal{K}$
- $\mathbb{P}_F$ is arbitrarily chosen a priori such that it it possible to sample from it

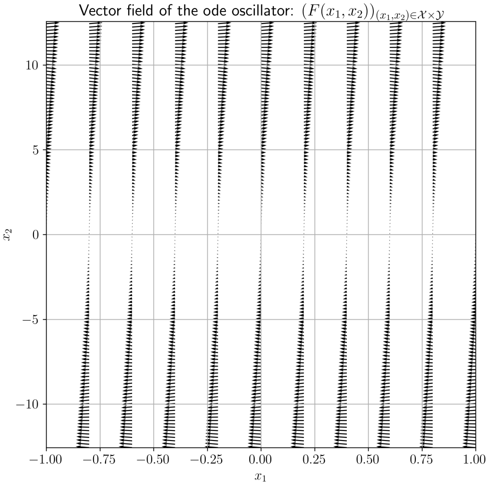


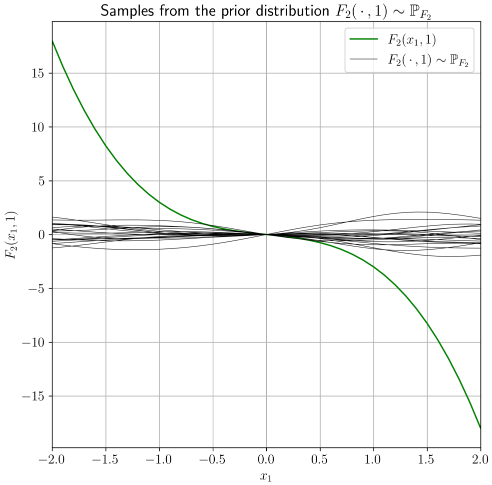
Functional Probability Space

Goal

- Use the dataset $\mathcal{D}$ to infer a posterior measure $\mathbb{P}_F(\cdot \mid \mathcal{D})$

Function Sampling





Gaussian Vector vs. Gaussian Process

F can be **sampled** from $\mathcal{F}$ according to the weights given by $\mathbb{P}_F$. Notation: $F \sim \mathbb{P}_F$

Gaussian Vector

Suppose $\mathcal{X} = \{x_1, \ldots, x_m\} \in \left(\mathbb{R}^{d_\mathcal{X}}\right)^m$ is finite

$F := (F_{x_1}, \ldots, F_{x_m}) = (F_{x_i})_{x_i \in \mathcal{X}}$

Mean Vector

$(\mu_{x_i})_{x_i \in \mathcal{X}} := (\mathbb{E}[F_{x_i}])_{x_i \in \mathcal{X}}$

Covariance Matrix

$\left(k_{x_i,x_j}\right)_{x_i,x_j \in \mathcal{X}\times\mathcal{X}} := \left(\mathrm{Cov}\left(F_{x_i}, F_{x_j}\right)\right)_{x_i,x_j \in \mathcal{X}\times\mathcal{X}}$

Gaussian Process

Suppose $\mathcal{X} = \{x \in \mathcal{X}\} \subset \mathbb{R}^{d_\mathcal{X}}$ is infinite

$F := (F_x)_{x \in \mathcal{X}}$

Mean Operator

$(\mu_x)_{x \in \mathcal{X}} := (\mathbb{E}[F_x])_{x \in \mathcal{X}}$

Covariance Operator

$(k_{x,x'})_{x,x' \in \mathcal{X}\times\mathcal{X}} := (\mathrm{Cov}(F_x, F_{x'}))_{x,x' \in \mathcal{X}\times\mathcal{X}}$

In both cases, for any finite collection $(F_{x_1}, \ldots, F_{x_q}) \sim \mathcal{N}\left(\left(\mu_{x_1}, \ldots, \mu_{x_q}\right), \left(k_{x_i,x_j}\right)_{1 \le i,j \le q}\right)$

Particularly, $F_x \sim \mathcal{N}(\mu_x, k_{x,x})$

$\mathbb{P}_{F_{x_1},\ldots,F_{x_q}} = f_{\mathcal{N}((\mu_{x_1},\ldots,\mu_{x_q}),(k_{x_i,x_j})_{i,j})} dx_1 \ldots dx_q$, $\quad f_{\mathcal{N}((\mu_{x_1},\ldots,\mu_{x_q}),(k_{x_i,x_j})_{i,j})}$ is the associated Gaussian density

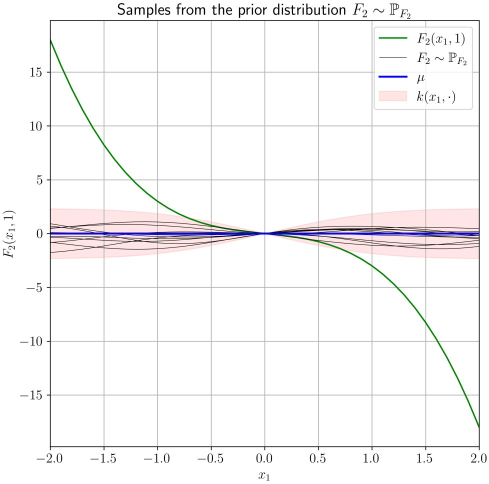
Notation: $F \sim \mathcal{GP}(\mu, k)$ if F is a Gaussian Process (GP)

Gaussian Process Regression

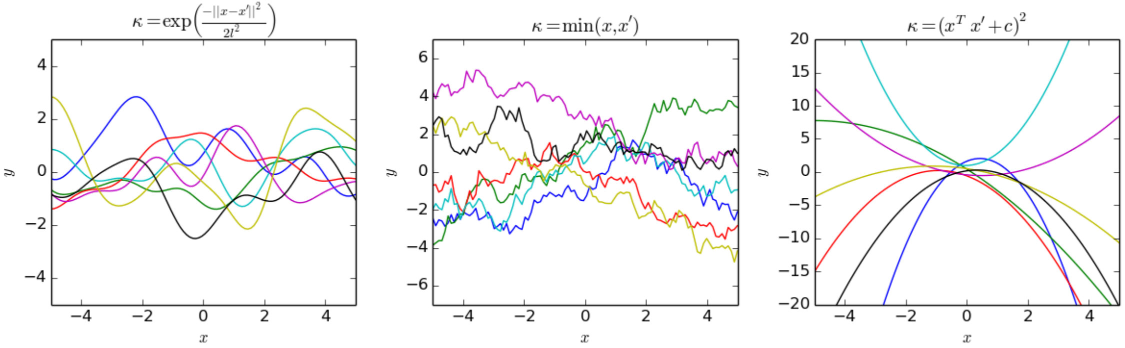
- GP are **characterised** by μ and k.
- Choose $F \sim \mathcal{GP}(\mu, k)$, *i.e.*, $\mathbb{P}_F = \mathcal{GP}(\mu, k)$
- Recall that $\mathcal{X}$ is **infinite**, thus $(k_{x,x'})_{x,x' \in \mathcal{X}\times\mathcal{X}} \coloneqq (\mathrm{Cov}\,(F_x, F_{x'}))_{x,x' \in \mathcal{X}\times\mathcal{X}}$ needs to be specified by the kernel operator k

A priori, μ and k are chosen:

- $\mu = 0$
- $k_{x,x'} = \exp\left(-\frac{\|x-x'\|^2}{2}\right)$



Other kernels



Posterior Probability Measure

Bayes' Theorem (Gaussian Density Measure)

- $\mathbb{P}_{F_{x_1}|F_{x_2}} = \frac{\mathbb{P}_{F_{x_1},F_{x_2}}}{\mathbb{P}_{F_{x_2}}} = \frac{\mathbb{P}_{F_{x_2}|F_{x_1}}\mathbb{P}_{F_{x_1}}}{\mathbb{P}_{F_{x_2}}}$
- In terms of densities functions:
  $f_{F_{x_1}|F_{x_2}} = \frac{f_{F_{x_1},F_{x_2}}}{f_{F_{x_2}}} = \frac{f_{F_{x_2}|F_{x_1}}f_{F_{x_1}}}{f_{F_{x_2}}}$

Gaussian Posterior Density

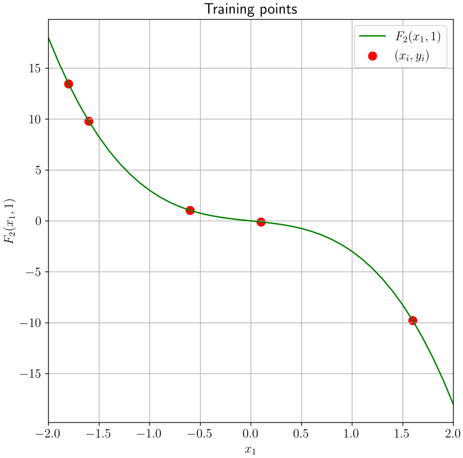
Given: $(F_{x_1}, F_{x_2}) \sim \mathcal{N}\left((\mu_{x_1}, \mu_{x_2}), (k_{x_1,x_1}, k_{x_2,x_2})\right) = \mathbb{P}_{F_{x_1},F_{x_2}}$

$$F_{x_1} \mid F_{x_2} = y_2 \sim \mathcal{N}\left(\mu_{F_{x_1}|F_{x_2}}, k_{F_{x_1}|F_{x_2}}\right)$$

$$\mu_{F_{x_1}|F_{x_2}} = \mu_{x_1} + \frac{k_{x_1,x_2}}{k_{x_2,x_2}} k_{x_2,x_1} (y_2 - \mu_{x_2})$$

$$k_{F_{x_1}|F_{x_2}} = k_{x_1,x_1} - \frac{k_{x_1,x_2}}{k_{x_2,x_2}} k_{x_2,x_1}$$

This generalise by replacing F_{x_2} by a **whole dataset of observations** $F^{*,\mathcal{D}}$ from $\mathcal{D} = ((x_i, F^*(x_i)))_{i=1}^n = \left(x^{\mathcal{D}}, F^{*,\mathcal{D}}\right)$



Posterior Probability Measure

Learning

Let $x' \in \mathcal{X}$ be a test point

Hence, $F_{x'} \mid F^{*,\mathcal{D}} \sim \mathcal{N}\left(\mu_{x'}^{\mathcal{D}}, k_{x',x'}^{\mathcal{D}}\right)$

Define the **posterior covariance matrix** $K^{\mathcal{D}} = \left(k_{x_i,x_j}\right)_{x_i,x_j \in \mathcal{D}}$

Define the **test covariance vector** $k_{x'}^{\mathcal{D}} = \left(k_{x_i,x'}\right)_{x_i \in \mathcal{D}}$

$$\mu_{x'}^{\mathcal{D}} = \langle k_{x'}^{\mathcal{D}},\, K^{\mathcal{D}^{-1}} F^{*,\mathcal{D}} \rangle = \langle k_{x'}^{\mathcal{D}},\, F^{*,\mathcal{D}} \rangle_{K^{\mathcal{D}^{-1}}}$$

$$k_{x',x'}^{\mathcal{D}} = k_{x',x'} - \langle k_{x'}^{\mathcal{D}},\, K^{\mathcal{D}^{-1}} k_{x'}^{\mathcal{D}} \rangle = k_{x',x'} - \|k_{x'}^{\mathcal{D}}\|_{K^{\mathcal{D}^{-1}}}^2$$

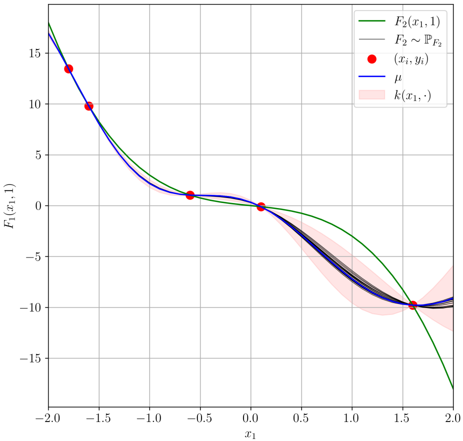
Connection with Functional Analysis

Links with *Reproducing Kernel Hilbert Space* (RKHS), in Kernel Regression, $\mathcal{H}_k := \overline{\operatorname{span}\{k_{\cdot,x} \mid x \in \mathcal{X}\}}$

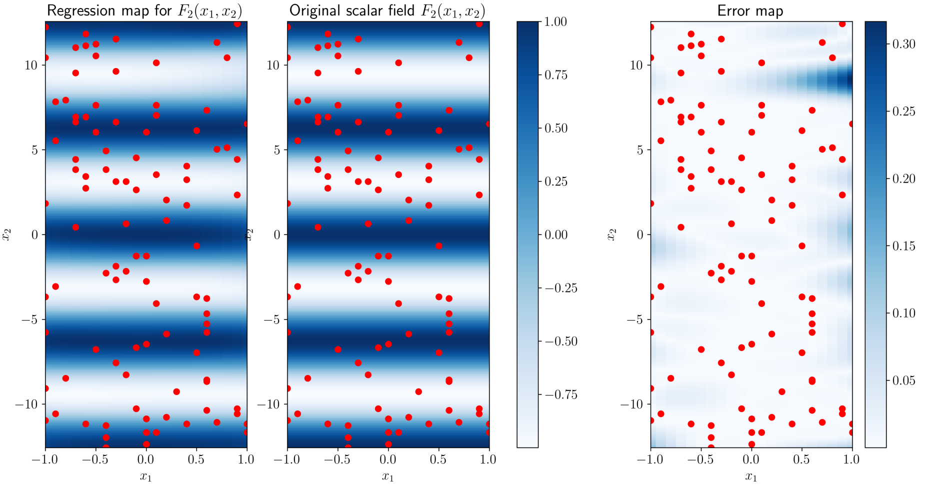
$$\mu_{x'}^{\mathcal{D}} := \arg\min_{\mu \in \mathcal{H}_k} \|\mu\|_{\mathcal{H}_k}^2 \;\; s.t. \;\; \mu_{x_i} = F_{x_i}$$

Observe the link with *Proper Orthogonal Decomposition (POD)* through the Mercer's theorem

$$k_{x,x'} = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x')$$



Scalar field regression on the whole state space $\mathcal{X}$



Entropy Map (Boltzmann 1870, Shannon 1948, Kolmogorov 1956)

Pick a point $x' \in \mathcal{X}$

How to quantify the uncertainty on $F_{x'}$?

Let dx an infinitesimal volume element of $\mathcal{X}$

Uncertainty on dx

$I(dx) = \log(\frac{1}{\mathbb{P}_{F_{x'}}(dx)}) > 0$

Entropy (average uncertainty)

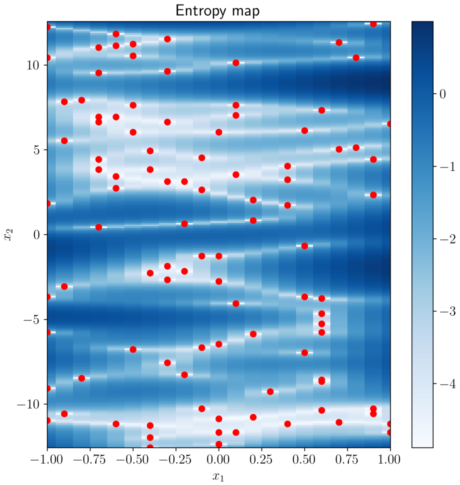
$\mathcal{H}(F_{x'}) = \int_{\mathbb{R}} \log \frac{1}{\mathbb{P}_{F_{x'}}(dx)} \mathbb{P}_{F_{x'}}(dx)$

Gaussian Entropy

$\mathbb{P}_{F_{x'}}(dx) = f_{\mathcal{N}(\mu_{x'}, k_{x',x'})}(x)dx$

If $F_{x'} \sim \mathcal{N}(\mu_{x'}, k_{x',x'})$, $\mathcal{H}(F_{x'}) = \frac{1}{2}\log\left(2\pi e\, k_{x',x'}\right)$

In the Gaussian case, the variance $k_{x',x'}$ characterise the entropy



Practical Question

In this context, given the following:

$F(x) = \sin(x)$, $x \in \mathcal{X}$

How to choose the Dataset $\mathcal{D}$, given a fixed budget n ?

Hypothesis

- $\mathcal{D} \sim \text{Uniform}(\mathcal{X}^n)$?
- $\mathcal{D}$ maximises $\mathcal{H}(\mathcal{D})$? (Note this is equivalent to the previous one if $\mathcal{X}$ is bounded)
- $\mathcal{D} = \{(x_k, F(x_k))\}_{k=1}^{n}$, where $x_{k+1} = F(x_k)$ (a trajectory) ?

References

Bayesian Statistics and Gaussian Process

- C. E. Rasmussen, C. K. I. Williams - *Gaussian Processes for Machine Learning*, MIT Press (2006)
- M. Raissi, G. E. Karniadakis - *Machine Learning of Linear Differential Equations using Gaussian Processes*, Journal of Computational Physics (2017)

Information Theory

- T. M. Cover, J. A. Thomas - *Elements of Information Theory*, Wiley (2006)
- M. Mézard, A. Montanari - *Information, Physics, and Computation*, Wiley (2009)

Reproducing Kernel Hilbert Space

- A. Berlinet, C. Thomas-Agnan - *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Springer (2004)