# Efficient Gaussian Processes for data-driven decision making

**Vincent Dutordoir**

Department of Engineering
University of Cambridge

Report submitted to be registered for the PhD Degree
*First-Year-Report*

Jesus College

August 2021

I would like to dedicate this thesis to my loving parents . . .

## Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Vincent Dutordoir
August 2021

# Acknowledgements

I would like to thank my advisor Dr. Carl Henrik Ek, and my supervisor Professor Zoubin Ghahramani for their guidance and support during the first year of my PhD degree. Their advice has greatly shaped the form of this thesis, and at the same time it has also been a pleasure to work with them.

# Abstract

This is where you write your abstract ...

# Table of contents

**1 Introduction** 1

1.1 Contributions and Layout of this Report . . . . . . . . . . . . . . . . . . . . . . . . 2

**2 Theoretical Framework** 3

2.1 Gaussian Processes and Exact Bayesian Inference . . . . . . . . . . . . . . . . . . 3

2.2 Approximate Inference in Sparse Gaussian Processes . . . . . . . . . . . . . . . . 3

2.3 Deep Gaussian Processes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.3.1 GPflux – a library for deep Gaussian processes . . . . . . . . . . . . . . . . 3

2.3.2 Interdomain Inducing Variables . . . . . . . . . . . . . . . . . . . . . . . . 3

2.3.3 Example . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.4 Kernels . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

**3** 5

**4 Sparse Gaussian Processes with Spherical Harmonic Features** 7

# Chapter 1

# Introduction

As the world we live in grows ever more interconnected and complex, making good decisions becomes increasingly difficult. Heterogenous transportation systems in large cities need to be optimised, global supply chains must be operated such thay they are as efficient and reliable as possible, and connected smart agents, such as self-driving cars, need to adhere to a policy that benefits the overall cause. Humans can typically make good decisions when the number of covariates is small, but suffer quickly when they have to consider thousands of influencing factors, or have to take into account the impact of their earlier decisions on future ones. Data-driven decision-making is a general framework that studies this problem. The aim of data-driven decision-making is to use past and current information to build a model of the environment that can be used to reason about future decisions and their impact.

A crucial building block for data-driven decision-making is a *model*, which tries to explain the given data. uncertainty many explanations for the data aleatoric epistemic

1. Data-driven Decision-making

2. The data can be explained by many models

3. Models that represent uncertainty

4. Statistical learning theory: Emperical risk minimisation

   Supervised machine learning setting $x \in \mathcal{X}$ and $y \in \mathbb{R}$, and a dataset $\mathcal{D} = \{x_i, y_i\}_i^N$

   General problem:

   $$\operatorname*{argmin}_f \sum_i L(f(x_i), y_i) + ||f|| \tag{1.1}$$

5. No Free Lunch Theorem

6.

7. Bayesian Linear Regression: $f(x) = w^\top \phi(x)$

8. Parametric models
9. Probabilistic machine learning: Bayes Rule
10. Kernel methods

## 1.1 Contributions and Layout of this Report

This report represents my learning and the research that I conducted during the first year of my PhD degree. Most notably, we developed a novel sparse approximation for (deep) Gaussian processes based on the decomposition of the kernel in Spherical harmonics. In chapter 2 we cover the necessary theoretical background

**Chapter 3** In this chapter we introduce a new class of inter-domain variational GPs where data is mapped onto the unit hypersphere in order to use spherical harmonic representations. The inference scheme is comparable to Variational Fourier Features, but it does not suffer from the curse of dimensionality, and leads to diagonal covariance matrices between inducing variables. This enables a speed-up in inference, because it bypasses the need to invert large covariance matrices. The experiments show that our model is able to fit a regression model for a dataset with 6 million entries two orders of magnitude faster compared to standard sparse GPs, while retaining state of the art accuracy.

The content of this chapter is largely based on:

Vincent Dutordoir, Nicolas Durrande, and James Hensman [2020]. "Sparse Gaussian Processes with Spherical Harmonic Features". In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*,

with the exception of the algorithm for computing the spherical harmonics in high dimensions.

**Chapter 3** Following up on the previous chapter, we use the decomposition of zonal kernels to design an interdomain inducing variable that mimics the behaviour of activation functions is neural network layers.

The content of this chapter is largely based on:

Vincent Dutordoir, James Hensman, Mark van der Wilk, Carl Henrik Ek, Zoubin Ghahramani, and Nicolas Durrande [2021]. "Deep Neural Networks as Point Estimate for Deep Gaussian Processes". In: *submission to NeurIPS*.

**Chapter 4**

# Chapter 2

# Theoretical Framework

## 2.1 Gaussian Processes and Exact Bayesian Inference

## 2.2 Approximate Inference in Sparse Gaussian Processes

variational inference (VI), where the problem of Bayesian inference is cast as an optimization problem—namely, to maximize a lower bound of the logarithm of the marginal likelihood.

## 2.3 Deep Gaussian Processes

### 2.3.1 GPflux – a library for deep Gaussian processes

### 2.3.2 Interdomain Inducing Variables

### 2.3.3 Example

## 2.4 Kernels

1.
2. RKHS
3. Mercer Decomposition
4.

# Chapter 3

# Chapter 4

# Sparse Gaussian Processes with Spherical Harmonic Features

# References

Vincent Dutordoir, Nicolas Durrande, and James Hensman (2020). "Sparse Gaussian Processes with Spherical Harmonic Features". In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*.

Vincent Dutordoir, James Hensman, Mark van der Wilk, Carl Henrik Ek, Zoubin Ghahramani, and Nicolas Durrande (2021). "Deep Neural Networks as Point Estimate for Deep Gaussian Processes". In: *submission to NeurIPS*.