Imperial College London

Department of Mathematics

Learning Invariances in Dynamical System

Cheng-Cheng Lao

CID: 01353756

Supervised by Dr. Andrew Duncan and Dr. Mark van der Wilk

July 6, 2022

Submitted in partial fulfilment of the requirements for the MSc in Statistics of Imperial College London

The work contained in this thesis is my own work unless otherwise stated.

Signed: Cheng-Cheng Lao Date: July 6, 2022

Abstract

ABSTRACT GOES HERE

Acknowledgements

ANY ACKNOWLEDGEMENTS GO HERE

1 Introduction

The introduction section goes here[1].

[1] Tip: write this section last.

2 Background

Here we will cover the background knowledge required to understand the remaining thesis, including theortical foundation of Gaussian Process, dynamical systems and invariances.

Background chapter.

2.1 Gaussian Process

Gaussian Process (GP)

Section content goes here.

2.1.1 My subsection

A subsection.

My subsubsection

A subsubsection.

2.2 Dynamical Systems

A dynamical system is simply a system that evolves with time under some kind of rules.

2.3 Invariances

Invariances are functions of the states that describe a dynamical system, and is unchanged throughout the evolution of the system over time. An example would be from the field of physics, the conservation of energy, which will be unchanged throught out the trajectories of the system.

2.4 Figures

It is better to create figures in a vector-based format, such as PDF.

2.5 Tables

Here is an example of a table

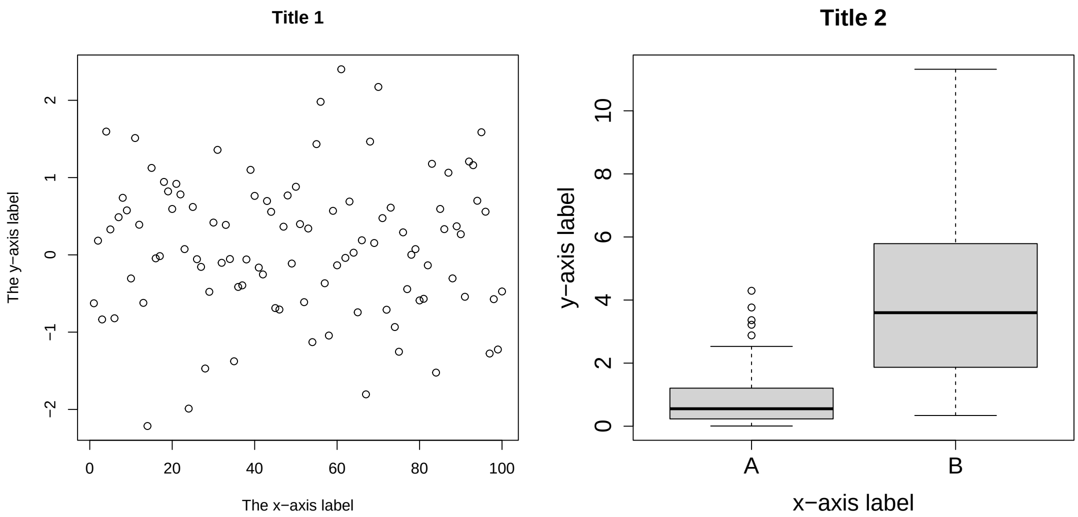


Figure 2.1: Remember to make fonts in figures large enough (compare the two figures.)

z	P$(Z < z)$
1.281	0.900
1.645	0.950
1.960	0.975
2.326	0.990
2.576	0.995

Table 2.1: Partial table showing values of z for P$(Z < z)$, where Z has a standard normal distribution.

2.6 Referencing sources, sections and items

A good book on the bootstrap is Efron and Tibshirani (1994), although the idea appeared in an earlier paper (Efron, 1979).

Note that to make the references appear, you will need to compile the bibtex, otherwise you may just see question marks where the references should be.

2.6.1 Referencing sections, results and equations

Theorem 2.7.2 is proved in Section 2.7; see Equation (2.2).

Referencing tables and figures

When labelling figures and tables, it is important that the label command `\label{LABELNAME}` comes **after** the caption command. See Table 2.1 and Figure 2.1 above.

2.6.2 Quoting sources

If you wish to quote a source, be sure to use quotation marks and cite the reference. The `\usequote` command is useful here:

"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way - in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only." (Dickens, 1859)

2.7 Definitions, theorems and examples

The following environments are supported: Definition, Theorem, Proof, Proposition, Lemma, Remark, Example.

Definition 2.7.1. The **variance** of a random variable X is defined as

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2]. \tag{2.1}$$

Theorem 2.7.2. Given a random variable X, over all values $a \in \mathbb{R}$,

$$\min_{a\in\mathbb{R}} \mathrm{E}[(X - a)^2] = \mathrm{E}[(X - \mathrm{E}[X])^2]. \tag{2.2}$$

Proof. Starting with the left-hand side,

$$\begin{aligned}
\mathrm{E}[(X - a)^2] &= \mathrm{E}[(X - \mathrm{E}[X] + \mathrm{E}[X] - a)^2] \\
&= \mathrm{E}[(X - \mathrm{E}[X])^2] + 2\mathrm{E}[(X - \mathrm{E}[X])\,(\mathrm{E}[X] - a)] + \mathrm{E}[(\mathrm{E}[X] - a)^2] \\
&= \mathrm{E}[(X - \mathrm{E}[X])^2] + (\mathrm{E}[X] - a)^2 \\
&\geq \mathrm{E}[(X - \mathrm{E}[X])^2],
\end{aligned}$$

since $\mathrm{E}[X]$ is a real number and $(\mathrm{E}[X] - a)^2 \geq 0$, and the third line follows from linearity of expectation:

$$\mathrm{E}[(X - \mathrm{E}[X])\,(\mathrm{E}[X] - a)] = (\mathrm{E}[X] - a)\,\mathrm{E}[(X - \mathrm{E}[X])] = (\mathrm{E}[X] - a)\,(\mathrm{E}[X] - \mathrm{E}[X]) = 0,$$

since $\mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X]$, which proves the result. □

Remark 2.7.3. This theorem shows that that the minimum of the quantity $\mathrm{E}[(X-a)^2]$ is equal to $\mathrm{Var}(X)$. In some sense, this makes the variance a natural measure of dispersion if we are taking the metric to be the squared deviation of X.

Lemma 2.7.4 (Stein's Lemma)**.** Let $X \sim \mathrm{N}(\mu, \sigma^2)$, and let g be a differentiable function satisfying $\mathrm{E}[|g'(X)|] < \infty$. Then

$$\mathrm{E}[g(X)(X - \mu)] = \sigma^2 \mathrm{E}[g'(X)].$$

Proposition 2.7.5 (Popoviciu's inequality)**.** Suppose that the random variable X is known to only take values in the bounded range $[a, b]$. Then

$$\mathrm{Var}[X] \leq \frac{(b - a)^2}{4}.$$

Example 2.7.6. Suppose $X \sim \mathrm{Bern}(p)$, for some $p \in [0, 1]$. Then, since $X \in \{0, 1\}$, X is bounded between 0 and 1 and so $\mathrm{Var}[X] \leq \frac{1}{4}$.

3 Invariance Kernel

In this chapter we will start with general theoretical construction of invariance kernel given the knowlwedge of the invariance of the system. We will then apply the construction to various systems, namely linear and nonlinear system in both one and two dimensions.

3.1 General Construction

If we are given a dynamical system of dimension d with variables $\mathbf{q}, \mathbf{p}$, where $\mathbf{q} = (q_1, q_2, \ldots, q_d)$ is the vector of positional coordinates, and $\mathbf{p} = (p_1, p_2, \ldots, p_d)$ is the vector of velocity coordinates of the states of the dynamical system. We are interested to predict the future trajectories of the state of the system. Therefore, we would like to know the time derivative of these coordinates so we can update them according to Euler's integrator. These time derivatives are referred to as the dynamics of a dynamical system, which governs the evolution of the dynamical system, and is a function of the coordinates $\mathbf{q}$ and $\mathbf{p}$ For our systems, we will denote the dynamics of $\mathbf{p}$ as $\frac{d\mathbf{p}}{dt} = \mathbf{a}(\mathbf{q}, \mathbf{p})$, and that of $\mathbf{q}$ as $\frac{d\mathbf{q}}{dt} = \mathbf{v}(\mathbf{q}, \mathbf{p})$. Once we have the dynamics, we can integrate up to obtain the future trajectories. Notation wise, we will collect the two dynamics term and call them

$$\mathbf{f}(\mathbf{q}, \mathbf{p}) = \begin{pmatrix} \mathbf{a}(\mathbf{q}, \mathbf{p}) \\ \mathbf{v}(\mathbf{q}, \mathbf{p}) \end{pmatrix}$$

For simplicity of notation, the dependence on $\mathbf{q}, \mathbf{p}$ is now implicit. We will put independent GP prior on $\mathbf{a}$ and $\mathbf{v}$ since there are no reason we should assume they are correlated. For the choice of kernel, we chose the standard smooth kernel, the squared exponential kernel, or RBF, and we denote this kernel to be K_{RBF}. We will also denote any set of coordinates of length n and dimension d as

$$X = \begin{pmatrix} q_{11} & q_{21} & \cdots & q_{d1} & p_{11} & p_{21} & \cdots & p_{d1} \\ q_{12} & q_{22} & \cdots & q_{d2} & p_{12} & p_{22} & \cdots & p_{d2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{1n} & q_{2n} & \cdots & q_{dn} & p_{1n} & p_{2n} & \cdots & p_{dn} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}.$$

Without loss of generality, we will choose to stack the dynamics vertically in $\mathbf{f}$; for example in a d dimensional system, we have

$$\mathbf{f}(X) = \begin{pmatrix} a_1(\mathbf{x}_1) \\ \vdots \\ a_1(\mathbf{x}_n) \\ a_2(\mathbf{x}_1) \\ \vdots \\ a_2(\mathbf{x}_n) \\ \vdots \\ a_d(\mathbf{x}_1) \\ \vdots \\ a_d(\mathbf{x}_n) \\ v_1(\mathbf{x}_1) \\ \vdots \\ v_1(\mathbf{x}_n) \\ v_2(\mathbf{x}_1) \\ \vdots \\ v_2(\mathbf{x}_n) \\ \vdots \\ v_d(\mathbf{x}_1) \\ \vdots \\ v_d(\mathbf{x}_n) \end{pmatrix}$$

For our naive independent RBF GP prior, we have

$$K = \mathrm{Cov}(\mathbf{f}(X), \mathbf{f}(X')) = \begin{pmatrix} K_{RBF,a_1}(X,X') & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \dots & \ddots & \vdots \\ 0 & \dots & K_{RBF,a_d}(X,X') & \dots & 0 \\ \vdots & \ddots & \vdots & K_{RBF,v_1}(X,X') & \vdots \\ 0 & \dots & 0 & \dots & K_{RBF,v_d}(X,X') \end{pmatrix},$$

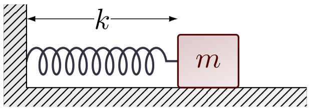
with all the off diagonal terms being zero block matrix because of the independence prior assumption. This naive kernel will be our baseline to be compared to throughout the whole project. Now we can start considering the invariance. If we assume the invariance is true throughout the input space $\mathbb{R}^{2d}$, then we can condition the GP on the a grid of points, which we referred to invariance points, X_L, and assume it has length ℓ, and we will call the invariance constraints L. The form of L will depend on the system as well as X_L, and examples will be described in the following sections. Since the invariance function will always be a linear function on the dynamics, if we apply this linear transformation L on $\mathbf{f}(X_L)$, $L(\mathbf{f}(X_L))$ will again be a GP with a transformed kernel. Therefore, we have

$$\begin{pmatrix} \mathbf{f}(X) \\ L[\mathbf{f}(X_L)] \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0_{2nd} \\ 0_\ell \end{pmatrix}, \begin{pmatrix} K & LK \\ KL^T & LKL^T \end{pmatrix} \right)$$

3.2 1D System

3.2.1 Linear

We will first examine one of the most simple dynamical system, an 1D simple harmonic motion (SHM). An example would be a mass spring system as shown in figure with mass m and spring constant k.



The defining equation is

$$m\frac{d^2q}{dt^2} = -kq$$

, where q is the displacement. And the analytical solution would be of the form $q = A\sin(\omega_0 t + \phi)$, where $\omega = \frac{k}{m}$ and A, ϕ depends on the initial condition, which dictates the amplitude and phase of the motion. For this case, we have

$$\mathbf{f}(X) = \begin{pmatrix} a(X) \\ v(X) \end{pmatrix}$$

We also have the energy, $E = \frac{kq^2}{2} + \frac{mp^2}{2}$, Therefore, to obtain our invariance L, we use the conservation of energy $\frac{dE}{dt} = 0$ so we finally have

$$L(a, v) = kqv + mpa = 0$$

In our system, since the only parameters that control the periodicity of the oscillation is k and m, but they only come in as the ratio between them. For simplicity, I will assume $k = m = 1$ so $\omega_0 = 1$. While L is not able to be into a matrix form, it is an linear operator as shown below. If we have

$$X_L = \begin{pmatrix} q_{L,1} & p_{L,1} \\ \vdots & \vdots \\ q_{L,\ell} & p_{L,\ell} \end{pmatrix} = \begin{pmatrix} \vdots & \vdots \\ q_L & p_L \\ \vdots & \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{L,1} \\ \vdots \\ \mathbf{x}_{L,\ell} \end{pmatrix},$$

then we have

$$L([\mathbf{f}(X_L)]) = \begin{pmatrix} p_{L,1}a(q_{L,1},p_{L,1}) + q_{L,1}v(q_{L,1},p_{L,1}) \\ \vdots \\ p_{L,\ell}a(q_{L,\ell},p_{L,\ell}) + q_{L,\ell}v(q_{L,\ell},p_{L,\ell}) \end{pmatrix}.$$

Combine with original GP prior assumption, we will have

$$\begin{pmatrix} \mathbf{f}(X) \\ L([\mathbf{f}(X_L)]) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0_{2nd} \\ 0_\ell \end{pmatrix}, \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right),$$

where

$$A = K(X,X), B = \begin{pmatrix} K_{RBF,a} \\ K_{RBF,v} \end{pmatrix} \odot \begin{pmatrix} P_L \\ Q_L \end{pmatrix}, C = B^T, D = K_{RBF,a} \odot (p_L \otimes p_L) + K_{RBF,v} \odot (q_L \otimes q_L),$$

where $\odot$ is the element wise product and $\otimes$ is the Kronecker product so that

$$P_L = \begin{pmatrix} p_{L,1} & \cdots & p_{L,\ell} \\ \vdots & \text{repeats n rows} & \vdots \\ p_{L,1} & \cdots & p_{L,\ell} \end{pmatrix}, Q_L = \begin{pmatrix} q_{L,1} & \cdots & q_{L,\ell} \\ \vdots & \text{reqeats n rows} & \vdots \\ q_{L,1} & \cdots & q_{L,\ell} \end{pmatrix}$$

and we have

$$p_L \otimes p_L = \begin{pmatrix} p_{L,1}^2 & p_{L,1}p_{L,2} & \cdots & p_{L,1}p_{L,\ell} \\ \vdots & \vdots & \vdots & \vdots \\ p_{L,\ell}p_{L,1} & p_{L,\ell}p_{L,2} & \cdots & p_{L,\ell}^2 \end{pmatrix}, q_L \otimes q_L = \begin{pmatrix} q_{L,1}^2 & q_{L,1}q_{L,2} & \cdots & q_{L,1}q_{L,\ell} \\ \vdots & \vdots & \vdots & \vdots \\ q_{L,\ell}q_{L,1} & q_{L,\ell}q_{L,2} & \cdots & q_{L,\ell}^2 \end{pmatrix},$$

These matrices can be obtained if we try to compute the covariance manually. For B, we wish to calculate

$$\begin{aligned} B_{ij} &= \mathrm{Cov}(\mathbf{f}(X), L[\mathbf{f}(X_L)])_{ij} \\ &= \mathrm{Cov}(\mathbf{f}(X)_i, L\mathbf{f}(X_L)_j) \\ &= \begin{cases} \mathrm{Cov}(a(q_i,p_i), p_{L,j}a(q_{L,j},p_{L,j}) + q_{L,j}v(q_{L,j},p_{L,j})) & i \le n \\ \mathrm{Cov}(v(q_i,p_i), p_{L,j}a(q_{L,j},p_{L,j}) + q_{L,j}v(q_{L,j},p_{L,j})) & i > n \end{cases} \\ &= \begin{cases} K_{RBF,a}(\mathbf{x}_i, \mathbf{x}_{L,j})p_{L,j} & i \le n \\ K_{RBF,v}(\mathbf{x}_i, \mathbf{x}_{L,j})q_{L,j} & i > n \end{cases}, \end{aligned}$$

and hence we have the form above. For D, we have

$$\begin{aligned} D_{ij} &= \mathrm{Cov}(L[\mathbf{f}(X_L)], L[\mathbf{f}(X_L)])_{ij} \\ &= \mathrm{Cov}(p_{L,i}a(q_{L,i},p_{L,i}) + q_{L,i}v(q_{L,i},p_{L,i}), p_{L,i}a(q_{L,i},p_{L,i}) + q_{L,i}v(q_{L,i},p_{L,i})) \\ &= p_{L,i}p_{L,j}K_{RBF,a}(\mathbf{x}_{L,i}, \mathbf{x}_{L,j}) + q_{L,i}q_{L,j}K_{RBF,v}(\mathbf{x}_{L,i}, \mathbf{x}_{L,j}) \end{aligned}$$

using the bilnear property of the covariance operator and the fact that v and a are independent. Since we assume invariance on these invariance points, we will condition on $L([\mathbf{f}(X_L)]) = 0$. Now we can simply use the Gaussian conditional formula to obtain the Schur Complement

$$\mathbf{f}(X)|L[\mathbf{f}(X_L)] = 0 \sim \mathcal{N}(0_{2n}, A - BD^{-1}C),$$

we will then call the covariance part our Invariance Kernel for 1D SHM, K_L.

3.2.2 Non Linear

The story is pretty much the same for nonlinear system, it is just the fitting would be expected to be more difficult. A simple nonlinear system in every day life is a simple pendulum as shown in figure below. The governing equation is

$$\frac{d^2q}{dt^2} = -\frac{g}{\ell}\sin q,$$

where q is the angle of displacement this time. The nonlinear dynamics bit occurs because of the sine term, which will complicate things slightly. However, since $\sin x \approx x$ at small angle, this system is approximately linear under small displacement. For simplicity, I will again set $g = \ell = 1$. There is no analytical solution to this nonlinear problem. This time we have energy, $E = \frac{m\ell^2p^2}{2} + mg\ell(1 - \cos q)$, and by setting the time derivative to 0, we have

$$L(a, v) = \frac{dE}{dt} = m\ell^2pa + mg\ell\sin qv = 0.$$

If we cancel out the common term $m\ell$ since their product cannot be zero, we have

$$L(a, v) = \ell pa + g\sin qv = 0$$

Most of the terms are unchanged from the linear case. However, this time,

$$B = \begin{pmatrix} K_{RBF,a} \\ K_{RBF,v} \end{pmatrix} \odot \begin{pmatrix} P_L \\ \sin(Q_L) \end{pmatrix}, D = K_{RBF,a} \odot (p_L \otimes p_L) + K_{RBF,v} \odot (\sin(q_L) \otimes \sin(q_L)),$$

where

$$\sin(Q_L) = \begin{pmatrix} \sin(q_{L,1}) & \dots & \sin(q_{L,\ell}) \\ \vdots & \text{reqeats n rows} & \vdots \\ \sin(q_{L,1}) & \dots & \sin(q_{L,\ell}) \end{pmatrix},$$

$$\sin(q_L) \otimes \sin(q_L) = \begin{pmatrix} \sin(q_{L,1})^2 & \sin(q_{L,1})\sin(q_{L,2}) & \dots & \sin(q_{L,1})\sin(q_{L,\ell}) \\ \vdots & \vdots & \vdots & \vdots \\ \sin(q_{L,\ell})\sin(q_{L,1}) & \sin(q_{L,\ell})\sin(q_{L,2}) & \dots & \sin(q_{L,\ell})^2 \end{pmatrix},$$

derived in almost exactly the same way as the linear case.

3.3 Damped System

3.4 2D System

3.4.1 Linear

3.4.2 Non Linear

4 Learning Invariance

5 Experiments

6 Conclusion

Conclusion goes here.

References

Charles Dickens. *A Tale of Two Cities*. Chapman & Hall, 1859.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1):1–26, 1979.

Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.