Charlie meeting 8th August

Meeting

In today's meeting, I go through Mark and Andrew everything I did so far as well as the thesis draft. In the background section, I need to include stuff at the beginning of our meetings, where the original idea comes from. For example, I should include stuff of Livoulle equation and L(v)=0 that kind of stuff and infidecimal operator etc. Basically how the project build up from the backgrounds. I should also explain just a slice of theory in things like symmetry and invariances (e.g. basic group theory, like his invariance paper) or the readers can refer to some textbook. Learnability of the invariance (as oppose to known) is also a key differentiating point. I should treat the background theory as a "what is known to be true, the reality" and then converts that to a statistical model in later chapters (main part of the thesis).

Also, some sort of restructuring should be considered, like putting examples and how to construct the invariance (specific ones) and experiement results together (i.e. one section just for SHM) Also, for learning polynomial, maybe I should show its training over time and that it gradually converges to the truth. It would also be interesting to see how much data to learn RBF eventually (data efficiency comparison) Maybe I should move the jitter section to the appendix. For damped, I should check that when jitter is small (zero) it is the same as pure invariance; when it is very large, it should be the same as RBF. and make a discussion.

VERY IMPORTANT even when it's damped it still needs to get it right! Maybe need more data? Maybe I need to use mean function, write up for Mark to decide. Maybe we should observe the invariance point using a custom likelihood function.

For 2D learning invariance, an important thing to note is that while we cannot learn some scratch. There are many things we can confirm it is learning the invariance.

1. Grid search (1. fix lengthscales, or 2. fix grids and optimise lengthscales at each grid point)

2. marginal likelihood correlatte with predictive performance

3. Initialise at right solution and let it optimise

Most importantly, good marginal likelihood leads to good prediction is a good thing since we are trying to validate that marginal likelihood is the right objective; recovering the physics is not necessarily a must

Lastly, we discussed about the partially observed system, and we said that we should work on a damped system then we can use the invariance on the latent variable to enforce invariance prior and maybe improve data efficiency. If we imagine a damped mass spring system, where the mass is submerged in a viscous liquid, with some viscoucity coefficient b. Then we will have

$$\frac{dE}{dt} = mv + kx = -bv^2; mv + kx + bv^2 = 0$$

which is again an invariant. Therefore, we observe x, v but does not directly observe the bv^2 term; therefore we will invent a latent variable z to represent

it. Finally our learning problem is that given x, v, and the latent variable z, we need to do inference on the dynamics. We again simply put GP priors on the dynamics f with input space be x, v. We also put a GP prior on z with input space x, v. However, this time, the invariance operator is a function of x, v, f as well as z. We have

$$\begin{pmatrix} f \\ z \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} K_f & 0 \\ 0 & K_z \end{pmatrix}\right)$$

by applying the linear operator L to $\begin{pmatrix} f \\ z \end{pmatrix}$ we have $L[f, z] \sim \mathcal{N}\left(0, LK_{f,z}L^T\right)$, with $K_{f,z}$ being the covariance matrix above and

$$\begin{pmatrix} \begin{pmatrix} f \\ z \end{pmatrix} \\ L[f, z] \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} K_{f,z} & LK_{f,z} \\ K_{f,z}L^T & LK_{f,z}L^T \end{pmatrix}\right)$$

We will then just need to condition again.

We had

$$\begin{pmatrix} f \\ L[f] \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} A & B \\ C & D \end{pmatrix}\right)$$

What I did was instead of conditioning on $L[f] = 0$, I condition on $L[f] = \epsilon$ a learnable constant. So I just implemented a mean function class (inherited from MeanFunction) by $BD^{-1}\epsilon$ using the Gaussian conditional formula.