Charlie meeting 30th May

This meeting we start implementing some ideas. We have two main sources of information, trajectory and invariances. We will see if there is a difference between conditioning on invariances or not. We model $v(\mathbf{x})$ and $a(\mathbf{x})$ as two independent GP prior, with kernel K_v, K_a. We denote X_g as the grid points of invariances conditioning, X^* as our test points and X as our trajectory data. Since the invariance of a simple harmonic oscillator is conservation of energy, $\dot{x}a + xv = 0$ We will have

$$
\begin{pmatrix} v(X) \\ v(X^*) \\ a(X) \\ a(X^*) \\ \dot{x}_g a(X_g) + x_g v(X_g) \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \right.
$$

$$
\begin{pmatrix} K_v(X,X) + \sigma^2 \mathcal{I} & K_v(X,X^*) & 0 & 0 & x_g * K_v(X,X_g) \\ K_v(X^*,X) & K_v(X^*,X^*) & 0 & 0 & x_g * K_v(X^*,X_g) \\ 0 & 0 & K_a(X,X) + \sigma^2 \mathcal{I} & K_a(X,X^*) & \dot{x}_g * K_a(X,X_g) \\ 0 & 0 & K_a(X^*,X) & K_a(X^*,X^*) & \dot{x}_g * K_a(X^*,X_g) \\ x_g * K_v(X_g,X) & x_g * K_v(X_g,X^*) & \dot{x}_g * K_a(X_g,X) & \dot{x}_g * K_a(X_g,X^*) & \dot{x}_g * K_a(X_g,X_g) * \dot{x}_g^T + x_g * K \end{pmatrix}
$$

Here, $*$ means element wise multiplication Using the Gaussian condtional formula again

$$
p\left(\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right) = \mathcal{N} \left(\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)
$$

$$
p\left(\mathbf{x}_1 \mid \mathbf{x}_2 \right) = \mathcal{N} \left(\mathbf{x}_1; \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \Sigma_{22}^{-1} \left(\mathbf{x}_2 - \boldsymbol{\mu}_2 \right), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right)
$$

This time, for $v(X*)|v(X) = \mathcal{D}, a(X_g, \dot{X}_g)\dot{X}_g + v(X_g, \dot{X}_g)X_g = 0$

we have $\Sigma_{11} = K_v(X^*,X^*), \Sigma_{12} = [K_v(X^*,X), x_g * K_v(X^*,X_g)], \Sigma_{21} = \begin{bmatrix} K_v(X,X^*) \\ x_g * K_v(X_g,X^*) \end{bmatrix}, \Sigma_{22} = \begin{bmatrix} K_v(X,X) + \sigma^2 \mathcal{I} & x_g * K_v(X,X_g) \\ x_g * K_v(X_g,X) & \dot{x}_g * K_a(X_g,X_g) * \dot{x}_g^T + x_g * K_v(X_g,X_g) * x_g^T \end{bmatrix}$,

Therefore, we have

$$
p(v(X^*)|v(X) = \mathcal{D}, a(X_g, \dot{X}_g)*\dot{x}_g + v(X_g, \dot{X}_g)*x_g = 0) \sim \mathcal{N} \left(\Sigma_{12} \Sigma_{22}^{-1} \begin{bmatrix} \mathcal{D} \\ 0 \end{bmatrix}, \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right)
$$

Similarly, for acceleartion, $a(X*)|a(X) = \mathcal{D}, a(X_g, \dot{X}_g)\dot{X}_g + v(X_g, \dot{X}_g)X_g = 0$

we have $\Sigma_{11} = K_a(X^*,X^*)$, $\Sigma_{12} = [K_a(X^*,X), \dot{x}_g * K_a(X^*,X_g)]$, $\Sigma_{21} = \begin{bmatrix} K_a(X,X^*) \\ \dot{x}_g * K_a(X_g,X^*) \end{bmatrix}, \Sigma_{22} = \begin{bmatrix} K_a(X,X) + \sigma^2 \mathcal{I} & \dot{x}_g * K_a(X,X_g) \\ \dot{x}_g * K_a(X_g,X) & \dot{x}_g * K_a(X_g,X_g) * \dot{x}_g^T + x_g * K_v(X_g,X_g) * x_g^T \end{bmatrix}$,

which has mean

$$
[K_a(X^*,X), \dot{x}_g * K_a(X^*,X_g)] \begin{bmatrix} K_a(X,X) + \sigma^2 \mathcal{I} & \dot{x}_g * K_a(X,X_g) \\ \dot{x}_g * K_a(X_g,X) & \dot{x}_g * K_a(X_g,X_g) * \dot{x}_g^T + x_g * K_v(X_g,X_g) * x_g^T \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{D} \\ 0 \end{bmatrix}
$$

and covariance

$$
K_a(X^*,X^*) -
$$

$$
[K_a(X^*,X), \dot{x}_g * K_a(X^*,X_g)] \begin{bmatrix} K_a(X,X) + \sigma^2 \mathcal{I} & \dot{x}_g * K_a(X,X_g) \\ \dot{x}_g * K_a(X_g,X) & \dot{x}_g * K_a(X_g,X_g) * \dot{x}_g^T + x_g * K_v(X_g,X_g) * x_g^T \end{bmatrix}^{-1} \begin{bmatrix} K_a(\\ \dot{x}_g * K \end{bmatrix}
$$

I mainly just showed the results for acceleration from now onwards since ve-cloity is pretty much the same as we model them as independent. I performed the computation manually (manually invert matrix and do matrix multiplica-tion) because I don't know how to fit it into GPflow. I first used a grid as invaraince points. My data is simulated by just a simple sine curve for position and cosine curve for velocity. I then took the finite time difference to find the acceleartion and vecloity as targets. I took 20 by 20 invarince points from -5 to 5. I have also calculate the invariance prior conditioning the test points just on the invariance but not the trajectory data. The result is as follows (this is for the highest marginal likelihood, length scale 1, likelihood variance 0.01, and invariance density 10): The first two are normal GPs, the blackdots are the data and the grey surface is the uncertainity. For contours, the ring in the middle is the data. The last four are invariance GPs (prior and posterior), and the red dots are where I put the conditioning invariance points.
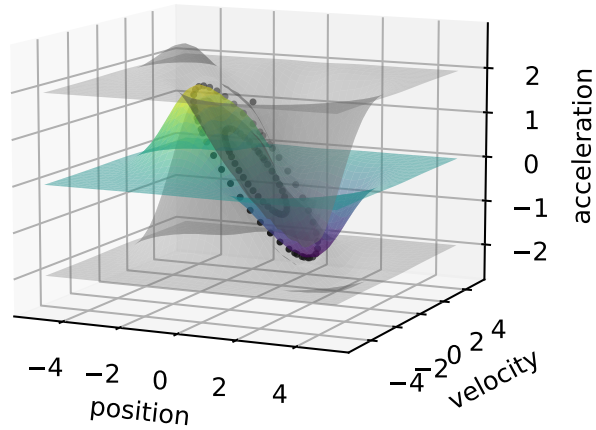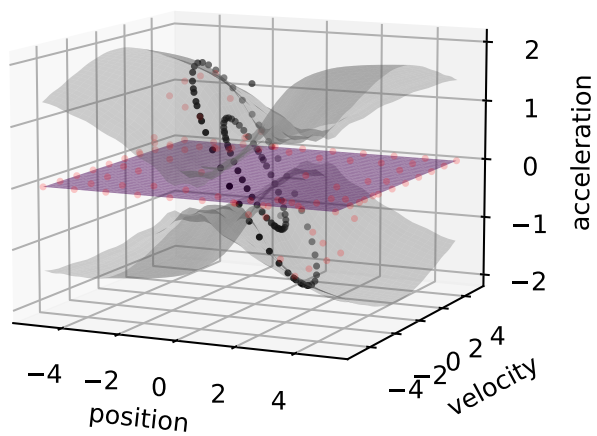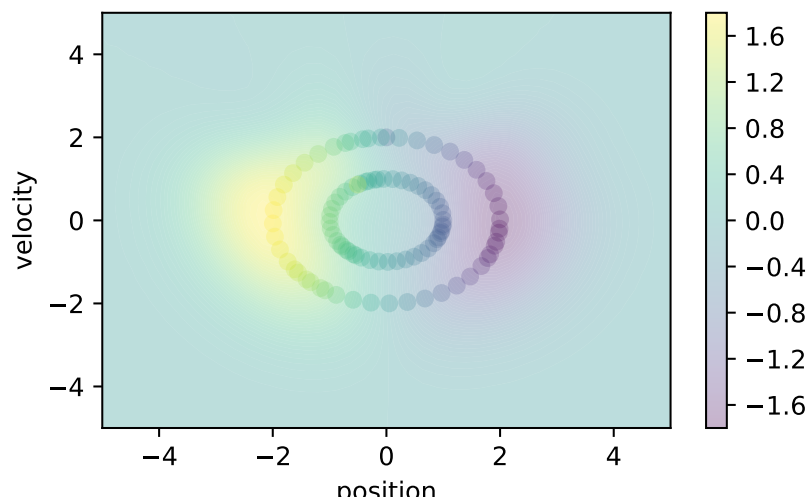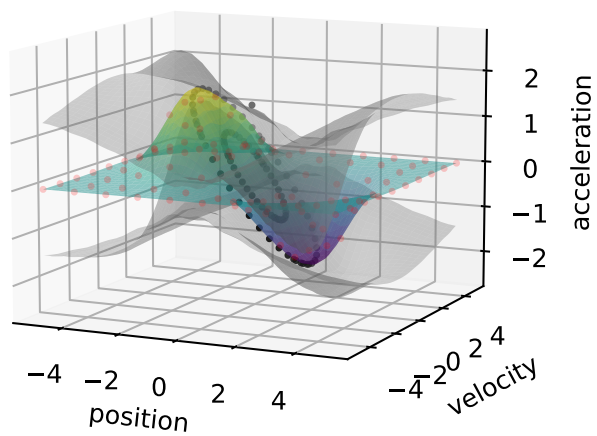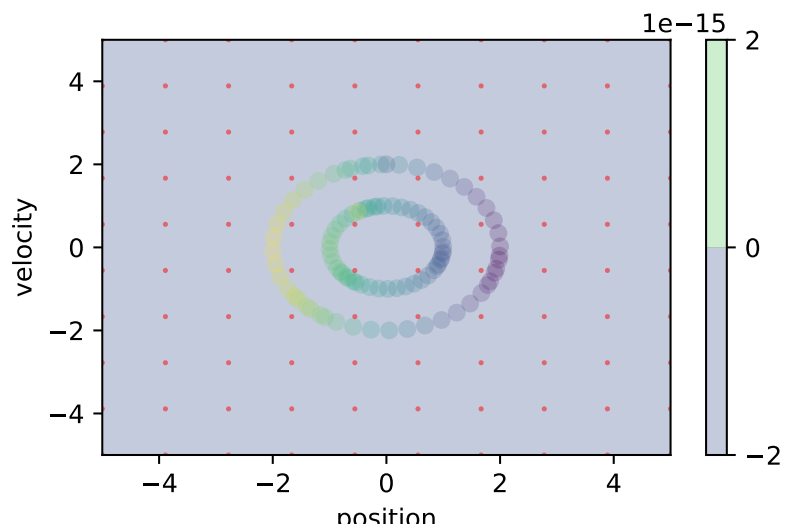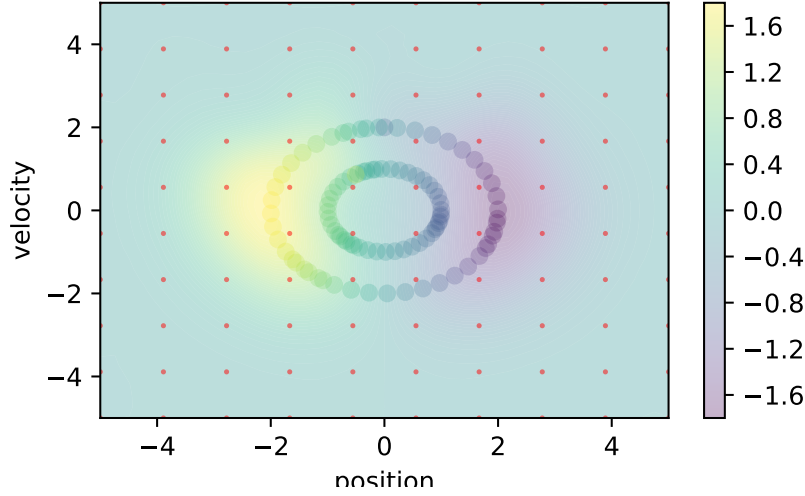
It does not seem to change the mean function, but the variance seems to have reduced, is that a positive result? The analytical solution of acceleration should be equal to $-x$ so we should have a flat plane in the -45 degrees direction. I calculated the marginal likelihood as below, please check if the calculation is correct. I have tried increase the density of invariance and the result varied accordingly, so I suppose this is a "hyperparamter" to tune too.

In all cases, when the invariance kernel is valid, (determinant ≥ 0), the invariance performace are almost all better than the regular version.

The log marginal likelihood is worked out to be

$$p(y|X,\theta) = \int p(y|f,X)p(f_{inv}|X,\theta)df$$

$$= \int \mathcal{N}(y|f,\sigma^2 I)p(f(X)|\text{invaraince},\theta)df$$

$$= \int \mathcal{N}(y|f,\sigma^2 I)\mathcal{N}(f_{inv}|0, K(X,X) - x_g * K(X,X_g)[\dot{x}_g K_a \dot{x}_g^T + x_g K_v x_g^T]^{-1} x_g * K(X_g,X))$$

$$= \mathcal{N}(y|0, \sigma^2 I + K(X,X) - x_g * K(X,X_g)[\dot{x}_g K_a \dot{x}_g^T + x_g K_v x_g^T]^{-1} x_g * K(X_g,X))$$

Therefore, the log marginal likelihood will be

$$\log p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{K_\theta}^{-1}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K_\theta}| + \text{ const}$$

, where K_θ will be the covariance derived in the previous equation. Therefore, we note that the grid of invariance is also a hyperparamter to be integrated over.

When I am changing my parameters, sometimes the marginal likelihood would be not defined due to the determinant of the kernel being negative, which

shouldn't be if it is a valid covaraince matrix. So how do we to make sure the new kernel is still positive definite or a valid covariance matrix?

To implement the kernel in GPflow, I need to find the covariance of two matrix X, X1 and X2, condition on the invaraince in closed form. This will require me to find the joint distribution of X1, X2 condition on the invariance; therefore this time, I will need to find the covariance function between X1 and X2 and conditioning. I have

$$\begin{pmatrix} f(X1) \\ f(X2) \end{pmatrix} | \text{invariance} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K_{inv}(X1, X1) & K_{inv}(X1, X2) \\ K_{inv}(X2, X1) & K_{inv}(X2, X2) \end{pmatrix} \right)$$

At the same time, we have the normal condition formula as before with the covariance equal to

$$\begin{bmatrix} K_a(X1, X1) & K_a(X1, X2) \\ K_a(X2, X1) & K_a(X2, X2) \end{bmatrix} -$$
$$\begin{bmatrix} \dot{x}_g * K_a(X1, X_g) \\ \dot{x}_g * K_a(X2, X_g) \end{bmatrix} \left[\dot{x}_g * K_a(X_g, X_g) * \dot{X}_g^T + x_g * K_v(X_g, X_g) * X_g^T \right]^{-1} [\dot{x}_g * K_a(X1, X_g), \dot{x}_g * K_a(X2, X_g)]$$

and similarly for vecloity, I am actually not sure if we can write it as a general kernel, since after I condition the joint distribution of X1, X2 on invariance, I will get something that I am not sure how to put it into the closed form.

To discuss

1. is it a good result?

2. is it possible to extend it to the whole phase space

3. how to write the invariance kernel analytically

4. what's the next step