6 September 2022

Learning Invariances in Dynamical System

supervised by Dr Andrew Duncan and Dr Mark van der Wilk

Cheng-Cheng Lao

Dynamical System

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = 0,$$

Dynamical System

$$\frac{d^n x}{dt^n} = F\left(t, x, \frac{dx}{dt}, \ldots, \frac{d^{n-1}x}{dt^{n-1}}\right),$$

$$\begin{cases} \dot{x}_1 = f_1(x_1, \ldots, x_n) \\ \vdots \\ \dot{x}_n = f_n(x_1, \ldots, x_n) \end{cases}$$

Dynamical System

$$\frac{d^n x}{dt^n} = F\left(t, x, \frac{dx}{dt}, \ldots, \frac{d^{n-1}x}{dt^{n-1}}\right),$$

$$\begin{cases} \dot{x}_1 = f_1(x_1, \ldots, x_n) \\ \vdots \\ \dot{x}_n = f_n(x_1, \ldots, x_n) \end{cases}$$

Planetary Evolution; Predator-Prey Dynamics, Protein mechanics, Quantum Mechanics

Gaussian Process (GP)

Definition

GP is a collection of random variables, any finite number of which have a joint Gaussian distribution

Gaussian Process (GP)

$$\begin{bmatrix} f \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix}\right),$$

Gaussian Process (GP)

$$\begin{bmatrix} f \\ f^* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} K(X, X) & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right),$$

$$y = f + \epsilon; \ \epsilon \sim \mathcal{N}(0, \sigma_n^2)$$

Gaussian Process (GP)

$$\begin{bmatrix} f \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix} \right),$$

$$y = f + \epsilon; \ \epsilon \sim \mathcal{N}(0, \sigma_n^2)$$

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2 \mathbb{I} & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix} \right).$$

Gaussian Process (GP)

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2 \mathbb{I} & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix} \right).$$

A very important formula:

$$\text{if } \begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \right),$$

$$x|y \sim \mathcal{N} \left(\mu_x + CB^{-1}(y - \mu_y), A - CB^{-1}C^T \right)$$

Gaussian Process (GP)

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2\mathbb{I} & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix}\right).$$

A very important formula:

$$\text{if } \begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix}\right),$$

$$x|y \sim \mathcal{N}\left(\mu_x + CB^{-1}(y - \mu_y), A - CB^{-1}C^T\right)$$

$$f^*|X, y, X^* \sim \mathcal{N}\left(\overline{f^*}, \text{cov}(f^*)\right)$$

$$\overline{f^*} = K(X^*, X)[K(X, X) + \sigma_n^2\mathbb{I}]^{-1}y$$

$$\text{cov}(f^*) = K(X^*, X^*) - K(X^*, X)[K(X, X) + \sigma_n^2\mathbb{I}]^{-1}K(X, X^*)$$

Gaussian Process (GP)

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2\mathbb{I} & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix}\right).$$

A very important formula:

$$\text{if } \begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix}\right),$$

$$x|y \sim \mathcal{N}\left(\mu_x + CB^{-1}(y - \mu_y), A - CB^{-1}C^T\right)$$

$$f^*|X, y, X^* \sim \mathcal{N}\left(\overline{f^*}, \text{cov}(f^*)\right)$$

$$\overline{f^*} = K(X^*,X)[K(X,X) + \sigma_n^2\mathbb{I}]^{-1}y$$

$$\text{cov}(f^*) = K(X^*,X^*) - K(X^*,X)[K(X,X) + \sigma_n^2\mathbb{I}]^{-1}K(X,X^*)$$

$$\log p(\mathbf{y} \mid X) = -\frac{1}{2}\mathbf{y}^\top \left(K + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{y} - \frac{1}{2}\log\left|K + \sigma_n^2\mathbf{I}\right| - \frac{n}{2}\log 2\pi.$$

Kernel

$$k_{RBF}(r) = \exp\left(-\frac{r^2}{2l^2}\right)$$

$$k_{\text{Matérn}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}r}{\ell}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}r}{\ell}\right)$$

$$k_{periodic\ RBF}(r) = \exp\left(-\frac{2\sin^2\left(\frac{r}{2}\right)}{\ell^2}\right),$$
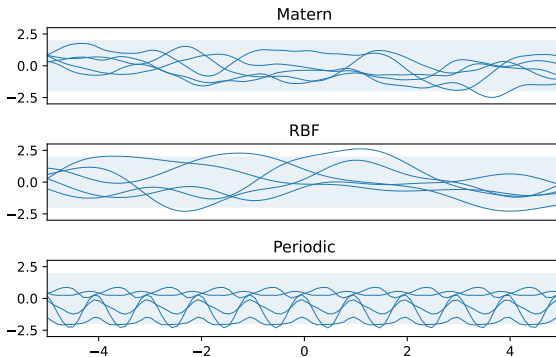
Kernel

Figure 1: Samples from different GP priors of RBF, Matérn and periodic kernel.

GP regression in action
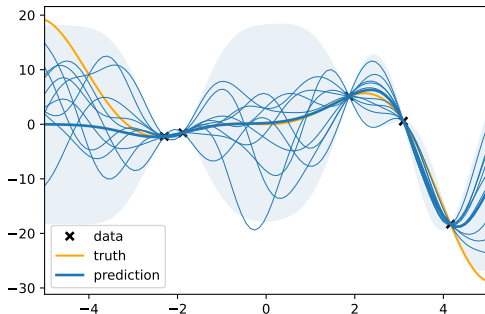
If we would like to fit a function $y = (x + x^2)\sin(x)$.

Figure 2: GP fit of the function $f = (x + x^2)\sin(x)$ with posterior samples, light shaded blue indicates 95% credible interval.

Related Work

Methods	Respect the physics laws	Learn the physics laws	Generalise beyond physics
ODE approach	X	O	O
Symbolic approach	O	O	X
Physics informed ML	O	X	X
Energy conserving NN	O	X	X
GP in dynamical system	X	O	O
Our method	O	O	O

Table 1: Comparing the capabilities of different existing approach to learning invariance in dynamical systems

Invariance Kernel I

We have a general dynamical system with coordinates $\mathbf{p}, \mathbf{q}$, then we will call the dynamics $\frac{d\mathbf{p}}{dt} = a(\mathbf{p}, \mathbf{q})$ and $\frac{d\mathbf{q}}{dt} = v(\mathbf{p}, \mathbf{q})$.

Invariance Kernel I

We have a general dynamical system with coordinates $\mathbf{p}, \mathbf{q}$, then we will call the dynamics $\frac{d\mathbf{p}}{dt} = a(\mathbf{p}, \mathbf{q})$ and $\frac{d\mathbf{q}}{dt} = v(\mathbf{p}, \mathbf{q})$. We will have

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

Invariance Kernel I

We have a general dynamical system with coordinates $\mathbf{p}, \mathbf{q}$, then we will call the dynamics $\frac{d\mathbf{p}}{dt} = a(\mathbf{p}, \mathbf{q})$ and $\frac{d\mathbf{q}}{dt} = v(\mathbf{p}, \mathbf{q})$. We will have

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

We will then put a GP prior on $\mathbf{f}$ so that

$$\mathbf{f} \sim \mathcal{GP}(m, K)$$

Invariance Kernel II

$$X \equiv \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} q_{11} & q_{21} & \ldots & q_{d1} & p_{11} & p_{21} & \ldots & p_{d1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{1n} & q_{2n} & \ldots & q_{dn} & p_{1n} & p_{2n} & \ldots & p_{dn} \end{pmatrix}.$$

Invariance Kernel II

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

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

Invariance Kernel II

$$K = \mathrm{Cov}(\mathbf{f}(X), \mathbf{f}(X')) =$$

$$\begin{pmatrix} K_{a_1}(X, X') & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \dots & \ddots & \vdots \\ 0 & \dots & K_{a_d}(X, X') & \dots & 0 \\ \vdots & \ddots & \vdots & K_{v_1}(X, X') & \vdots \\ 0 & \dots & 0 & \dots & K_{v_d}(X, X') \end{pmatrix},$$

where each K_f is an RBF kernel

Invariance Kernel III

$$\mathcal{L}[E] \equiv \frac{dE}{dt} = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial t} + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} \frac{\partial q_i}{\partial t}$$

$$= \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$$

Invariance Kernel III

$$\mathcal{L}[E] \equiv \frac{dE}{dt} = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial t} + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} \frac{\partial q_i}{\partial t}$$

$$= \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$$

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

Invariance Kernel III

$$\mathcal{L}[E] \equiv \frac{dE}{dt} = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial t} + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} \frac{\partial q_i}{\partial t}$$

$$= \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$$

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

$$\mathbf{f}(X) | L[\mathbf{f}(X_L)] = 0 \sim \mathcal{N} \left(\mathbf{0}_{2nd}, \left(K - LK(LKL^T)^{-1}KL^T \right) \right)$$

Learning Invariance

$$\mathcal{L}[E] = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q})$$

Learning Invariance

$$\mathcal{L}[E] = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q})$$

- One dimensional:

$$L[\mathbf{f}] = f(p)a + g(q)v$$

Learning Invariance

$$\mathcal{L}[E] = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q})$$

- One dimensional:

$$L[\mathbf{f}] = f(p)a + g(q)v$$

- Two dimensional:

$$L[\mathbf{f}] = f_1(q_1, q_2, p_1, p_2)a_1 + f_2(q_1, q_2, p_1, p_2)a_2$$
$$+ g_1(q_1, q_2, p_1, p_2)v_1 + g_2(q_1, q_2, p_1, p_2)v_2$$

Damped System- Approximate Invariance

we have $L[\mathbf{f}(X_L)] = \epsilon$, where $\epsilon \sim \mathcal{N}(m_L, \sigma_L^2)$

Damped System- Approximate Invariance

we have $L[\mathbf{f}(X_L)] = \epsilon$, where $\epsilon \sim \mathcal{N}(m_L, \sigma_L^2)$

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

Damped System- Latent Dynamics

$$\mathcal{L}[E] + z = \frac{dE}{dt} + z = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial t} + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} \frac{\partial q_i}{\partial t} + z =$$

$$\sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i + z = L_{\gamma}[\mathbf{f}_{\gamma}] = 0$$

Damped System- Latent Dynamics

$$\mathcal{L}[E] + z = \frac{dE}{dt} + z = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial t} + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} \frac{\partial q_i}{\partial t} + z =$$

$$\sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i + z = L_\gamma[\mathbf{f}_\gamma] = 0$$

$$\begin{pmatrix} \mathbf{f}(X) \\ z(X) \\ L_\gamma[\mathbf{f}(X_L), z(X_L)] \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0}_{3nd} \\ \mathbf{0}_\ell \end{pmatrix}, \begin{pmatrix} K & L_\gamma K \\ K L_\gamma^T & L_\gamma K L_\gamma^T \end{pmatrix} \right),$$

Imperial College London

Experiments

1. Data Generation
2. Evaluation Methods
3. Implementation Techicalities

Simple Harmonic Motion (SHM)

$$\frac{d^2x}{dt^2} = -\frac{kx}{m}$$

$$x = A\sin(\omega_0 t + \phi)$$

Simple Harmonic Motion (SHM)
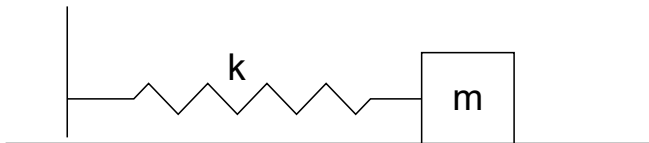
$$\frac{d^2x}{dt^2} = -\frac{kx}{m}$$

$$x = A\sin(\omega_0 t + \phi)$$

Simple Harmonic Motion (SHM)

$$\frac{d^2x}{dt^2} = -\frac{kx}{m}$$

$$x = A\sin(\omega_0 t + \phi)$$
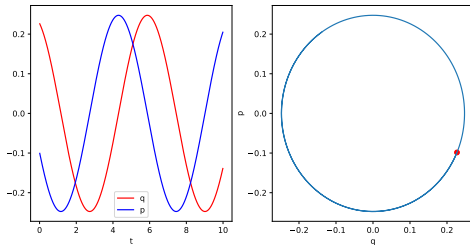
SHM Invariance Kernel- I

We had $\mathcal{L}[E] \equiv \dfrac{dE}{dt} = \displaystyle\sum_{i=1}^{d} \dfrac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \dfrac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$

SHM Invariance Kernel- I

We had $\mathcal{L}[E] \equiv \dfrac{dE}{dt} = \displaystyle\sum_{i=1}^{d} \dfrac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \dfrac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$

$$E = \dfrac{kq^2}{2} + \dfrac{mp^2}{2}$$

SHM Invariance Kernel- I

We had $\mathcal{L}[E] \equiv \dfrac{dE}{dt} = \displaystyle\sum_{i=1}^{d} \dfrac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \dfrac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$

$$E = \frac{kq^2}{2} + \frac{mp^2}{2}$$

So we have

$$L[\mathbf{f}] = mpa + kvp = 0$$

SHM Invariance Kernel- I

We had $\mathcal{L}[E] \equiv \dfrac{dE}{dt} = \displaystyle\sum_{i=1}^{d} \dfrac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \dfrac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$

$$E = \frac{kq^2}{2} + \frac{mp^2}{2}$$

So we have

$$L[\mathbf{f}] = mpa + kvp = 0$$

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

SHM Invariance Kernel- II

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

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

SHM Invariance Kernel- II

$$A = K(X, X), B = \begin{pmatrix} K_a(X, X_L) \\ K_v(X, X_L) \end{pmatrix} \odot \begin{pmatrix} mP_L \\ kQ_L \end{pmatrix}, C = B^T,$$

$$D = K_a(X_L, X_L) \odot m^2(p_L \otimes p_L) + K_v(X_L, X_L) \odot k^2(q_L \otimes q_L),$$

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

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

SHM Invariance Kernel- III

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

SHM Invariance Kernel- III

$$D_{ij} = \mathrm{Cov}(L[\mathbf{f}(X_L)], L[\mathbf{f}(X_L)])_{ij}$$
$$= \mathrm{Cov}(mp_{L,i}a(q_{L,i}, p_{L,i}) + kq_{L,i}v(q_{L,i}, p_{L,i}), mp_{L,i}a(q_{L,i}, p_{L,i}) + kq_{L,i}v($$
$$= m^2 p_{L,i}p_{L,j}K_{RBF,a}(\mathbf{x}_{L,i}, \mathbf{x}_{L,j}) + k^2 q_{L,i}q_{L,j}K_{RBF,v}(\mathbf{x}_{L,i}, \mathbf{x}_{L,j})$$

Learning Invariance

$$\mathcal{L}[E] \equiv \frac{dE}{dt} = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$$

$$L[\mathbf{f}] = f(p)a + g(q)v$$

Learning Invariance

$$L[\mathbf{f}] = f(p)a + g(q)v$$

$$A = K(X, X), B = \begin{pmatrix} K_a(X, X_L) \\ K_v(X, X_L) \end{pmatrix} \odot \begin{pmatrix} f(P_L) \\ g(Q_L) \end{pmatrix}, C = B^T,$$

$$D = K_a(X_L, X_L) \odot (f(p_L) \otimes f(p_L)) + K_v(X_L, X_L) \odot (g(q_L) \otimes g(q_L)),$$

Learning Invariance

$$L[\mathbf{f}] = f(p)a + g(q)v$$

$$A = K(X, X), B = \begin{pmatrix} K_a(X, X_L) \\ K_v(X, X_L) \end{pmatrix} \odot \begin{pmatrix} f(P_L) \\ g(Q_L) \end{pmatrix}, C = B^T,$$

$$D = K_a(X_L, X_L) \odot (f(p_L) \otimes f(p_L)) + K_v(X_L, X_L) \odot (g(q_L) \otimes g(q_L)),$$

$$f(P_L) = \begin{pmatrix} f(p_{L,1}) & \cdots & f(p_{L,\ell}) \\ \vdots & \text{repeats n rows} & \vdots \\ f(p_{L,1}) & \cdots & f(p_{L,\ell}) \end{pmatrix},$$

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

Results for SHM

Method	RBF	Known Invariance	Learnt Invariance
Log Marginal Likelihood	67.67	82.00	79.24
MSE	0.0950	0.0017	0.0027

Table 2: SHM performance.

Results for SHM
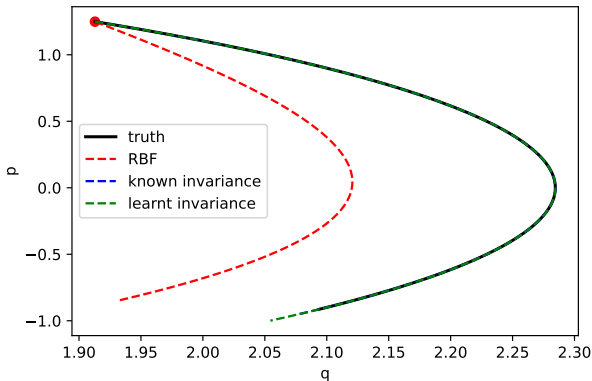
Figure 3: One SHM predicted trajectory.
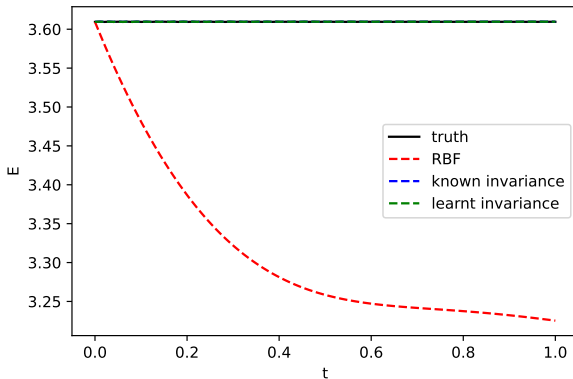
Results for SHM

Figure 3: The energy along the trajectory.

Results for SHM

RBF GP Posterior of a   RBF GP Posterior of v
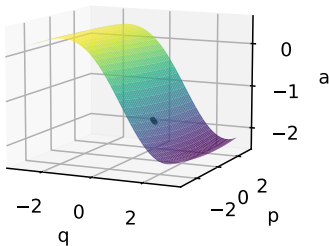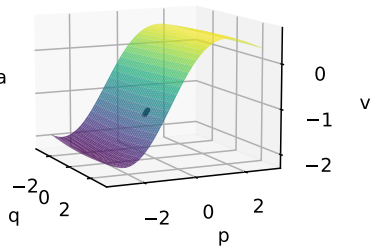
Results for SHM

Invariance GP Posterior of a   Invariance GP Posterior of v
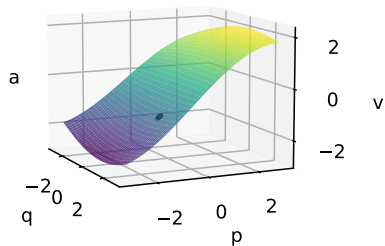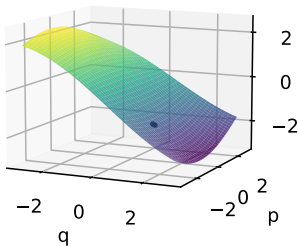
Results for SHM
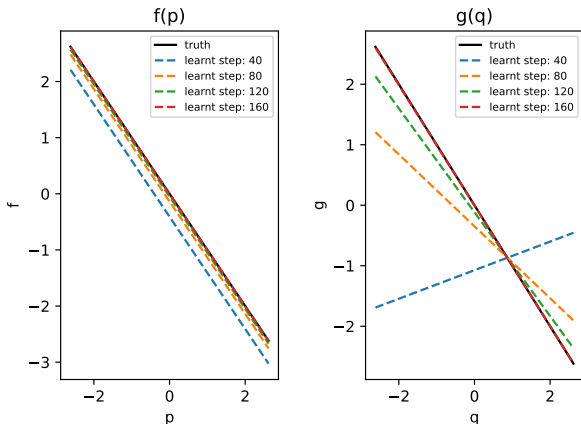
Figure 3: Learnt invariance for SHM.

Pendulum

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

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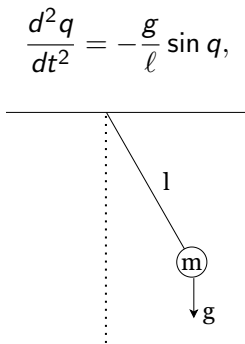
$$\frac{d^2 q}{dt^2} = -\frac{g}{\ell} \sin q,$$

Figure 4: A pendulum is a simple system that is nonlinear.

Pendulum

$$\frac{d^2 q}{dt^2} = -\frac{g}{\ell} \sin q,$$
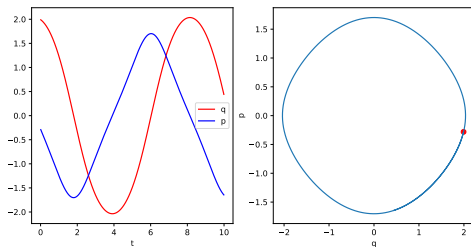
Figure 4: Example trajectory of pendulum.

Pendulum Invariance

$$\mathcal{L}[E] \equiv \frac{dE}{dt} = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$$

Pendulum Invariance

$$\mathcal{L}[E] \equiv \frac{dE}{dt} = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$$

$$E = \frac{m\ell^2 p^2}{2} + mg\ell(1 - \cos q)$$

Pendulum Invariance

$$\mathcal{L}[E] \equiv \frac{dE}{dt} = \sum_{i=1}^{d} \frac{\partial E}{\partial p_i} a_i(\mathbf{p}, \mathbf{q}) + \sum_{i=1}^{d} \frac{\partial E}{\partial q_i} v_i(\mathbf{p}, \mathbf{q}) = L[\mathbf{f}]$$

$$E = \frac{m\ell^2 p^2}{2} + mg\ell(1 - \cos q)$$

$$L[\mathbf{f}] = \ell p a + g(\sin q)v = 0$$

Pendulum Invariance

$$L[\mathbf{f}] = \ell p a + g(\sin q)v = 0$$

$$B = \begin{pmatrix} K_a(X, X_L) \\ K_v(X, X_L) \end{pmatrix} \odot \begin{pmatrix} \ell P_L \\ g \sin(Q_L) \end{pmatrix},$$

$$D = K_a(X_L, X_L) \odot \ell^2(p_L \otimes p_L) + K_v(X_L, X_L) \odot g^2(\sin(q_L) \otimes \sin(q_L)),$$

Pendulum Invariance

$$L[\mathbf{f}] = \ell p a + g(\sin q)v = 0$$

$$B = \begin{pmatrix} K_a(X, X_L) \\ K_v(X, X_L) \end{pmatrix} \odot \begin{pmatrix} \ell P_L \\ g\sin(Q_L) \end{pmatrix},$$

$$D = K_a(X_L, X_L) \odot \ell^2(p_L \otimes p_L) + K_v(X_L, X_L) \odot g^2(\sin(q_L) \otimes \sin(q_L)),$$

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

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

Results for Pendulum

Method	RBF	Known Invariance	Learnt Invariance
Log Marginal Likelihood	299.12	331.66	325.76
MSE	0.0021	0.0009	0.0006

Table 2: Pendulum performance.

Results for Pendulum
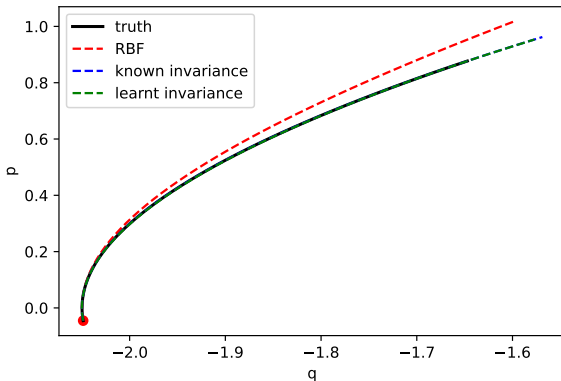
Figure 5: Pendulum predicted trajectory.

Results for Pendulum
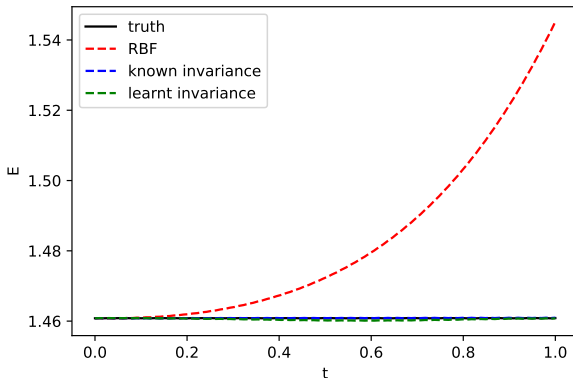
Figure 5: The energy along the trajectory.

Results for Pendulum

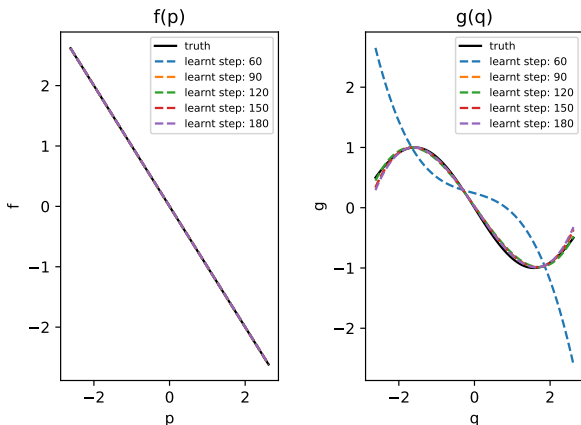Invariance GP Posterior of a   Invariance GP Posterior of v
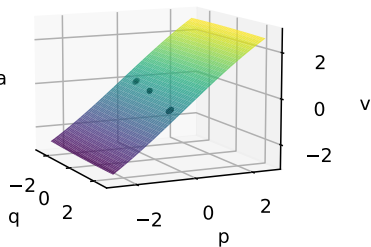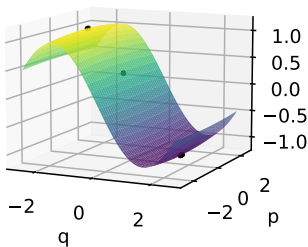
Results for Pendulum

Figure 5: Learnt invariance for pendulum.

Damped Systems

$$\frac{d^2q}{dt^2} + 2\gamma\frac{dq}{dt} + \omega_0^2 q = 0; \quad \frac{d^2q}{dt^2} + 2\gamma\frac{dq}{dt} + \omega_0^2 \sin q = 0,$$

Damped Systems
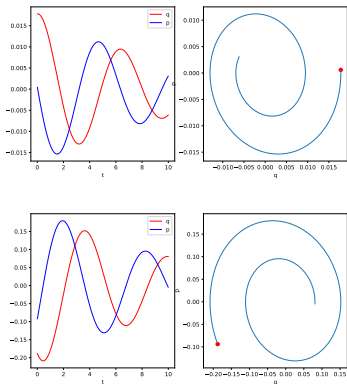
Figure 6: Example trajectories of damped systems, with damping factor $\gamma = 0.1$.

Damped SHM

Approximate Invariance

$$\mathbf{f}(X)|L[\mathbf{f}(X_L)] = 0 \sim \mathcal{N}(-B(D + \sigma_L^2\mathbb{I})^{-1}\mathbf{m}_\ell, A - B(D + \sigma_L^2\mathbf{I})^{-1}C),$$

Damped SHM

Latent Dynamics

$$L_\gamma[\mathbf{f}, z] = \frac{dE}{dt} + z = mpa + kqv + z = 0,$$

and we obtain

$$\begin{pmatrix} \begin{pmatrix} \mathbf{f}(X) \\ z(X) \end{pmatrix} \\ L_\gamma[\mathbf{f}(X_L), z(X_L)] \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0}_{3n} \\ \mathbf{0}_\ell \end{pmatrix}, \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right),$$

Damped SHM

$$A = K(X, X) \text{ with } K = \begin{pmatrix} K_{RBF,f} & 0 \\ 0 & K_{RBF,z} \end{pmatrix} C = B^T$$

.

$$B = \begin{pmatrix} K_{RBF,a}(X, X_L) \\ K_{RBF,v}(X, X_L) \\ K_{RBF,z}(X, X_L) \end{pmatrix} \odot \begin{pmatrix} mP_L \\ kQ_L \\ 1 \end{pmatrix},$$

$$D = K_{RBF,a}(X_L, X_L) \odot m^2(p_L \otimes p_L) + K_{RBF,v}(X_L, X_L) \odot k^2(q_L \otimes q_L)$$

$$+K_{RBF,z}(X_L, X_L)$$

Damped SHM

If we have damped SHM, we knew that $E = \frac{mp^2}{2} + \frac{kq^2}{2}$, then as before $\frac{dE}{dt} = mpa + kqv$. Using our knowledge of $p = v = \frac{dq}{dt}$, we have $\frac{dE}{dt} = mva + kvq = v(ma + kq)$. From the differential equation $\frac{d^2q}{dt^2} + 2\gamma\frac{dq}{dt} + \frac{kq}{m} = 0$, we can write it as $m\frac{dp}{dt} + 2m\gamma v + kq = 0$ so that $ma + 2m\gamma v + kq = 0$ or $ma + kq = -2m\gamma v = -bv$. Finally, we obtain

$$\frac{dE}{dt} = v(-bv) = -bv^2$$

As a result, from equation **??**, we identify $z = bv^2 = bp^2$, and therefore we would expect its posterior to be quadratic in p and constant in q. Simliar result can be derived for damped pendulum, where we again expect z to be quadratic in p and independent of q.

Damped SHM Results

Method	RBF	Known (Approx)	Learnt (Approx)	Known (Latent)	Learnt (Latent)
Log Marginal Likelihood	636	647	646	649	653
MSE	0.00142	0.00130	0.00134	0.00091	0.00101

Table 2: Damped SHM performance. We can see the approximate invariance is no longer significantly better than RBF, while the latent dynamics model is much better.
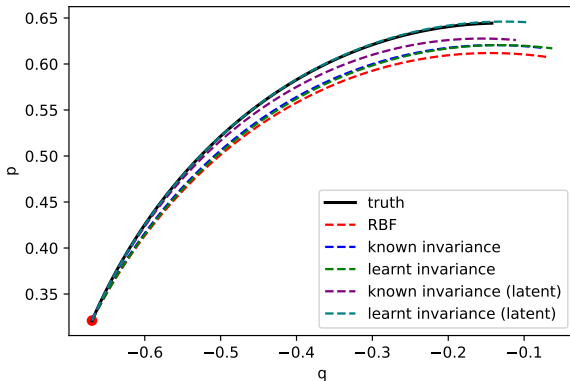
Table 3: Damped SHM performance.

Damped SHM Results

Figure 7: Damped SHM predicted trajectory.
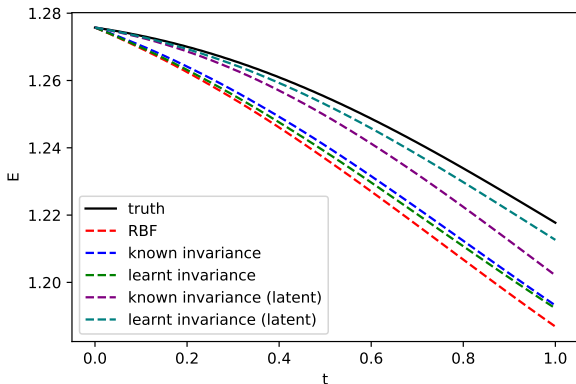
Damped SHM Results
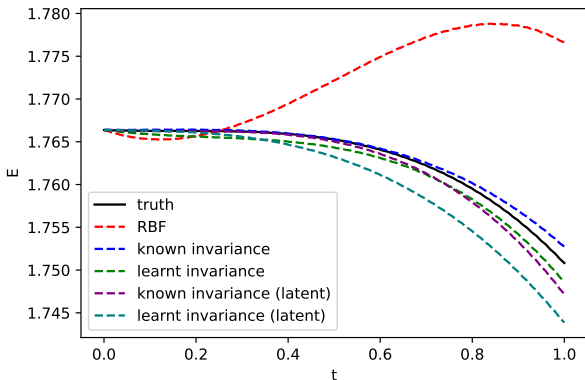

Figure 7: The energy along the trajectory.
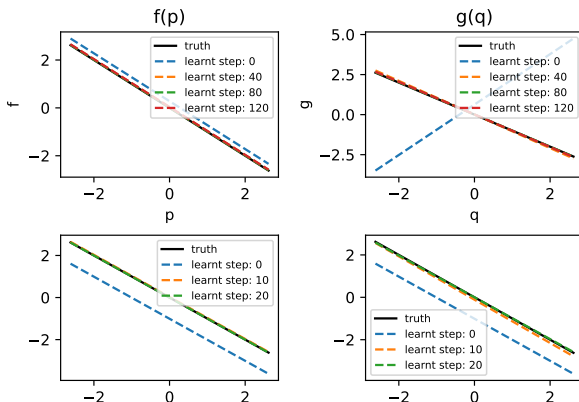
Damped SHM Results

Figure 7: Learnt invariance for damped SHM.

Damped SHM Results
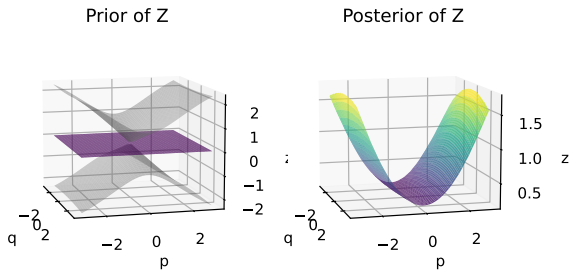
Figure 7: Latent variable distribution.

Damped pendulum Results

Method	RBF	Known (Approx)	Learnt (Approx)	Known (Latent)	Learnt (Latent)
Log Marginal Likelihood	516	525	525	548	522
MSE	0.0012	0.0008	0.0008	0.0008	0.0008

Table 2: Damped pendulum performance. We can see the approximate invariance is no longer significantly better than RBF, while the latent dynamics model is much better.

Table 3: Damped pendulum performance.

Damped pendulum Results
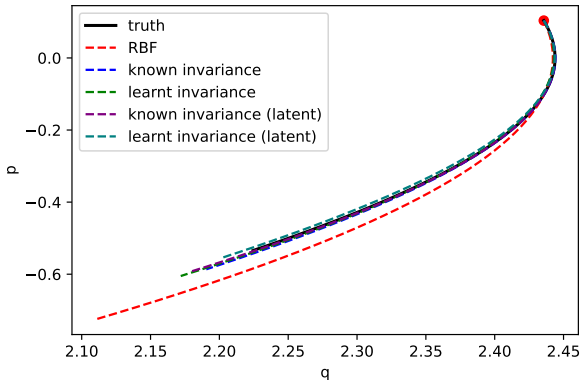
Figure 8: Damped pendulum predicted trajectory.

Damped pendulum Results

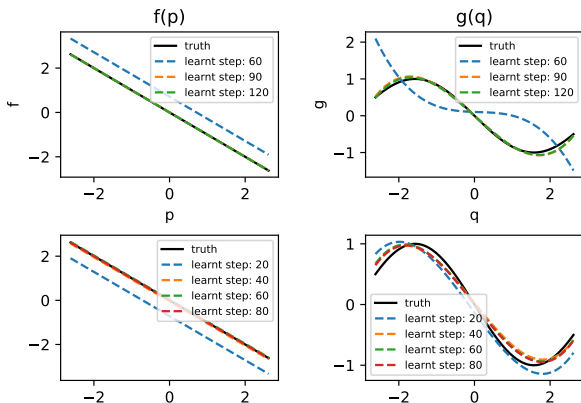Figure 8: The energy along the trajectory.

Damped pendulum Results

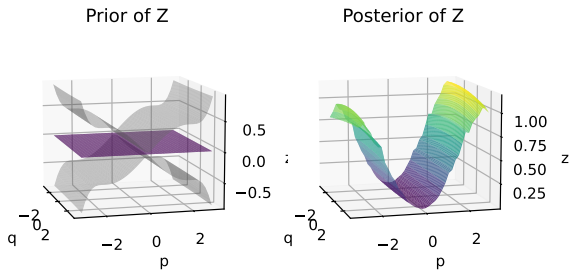Figure 8: Learnt invariance for damped pendulum.

Damped pendulum Results

Figure 8: Latent variable distribution.

Two-dimensional SHM

$$\begin{cases} \frac{d^2 q_1}{dt^2} = -\frac{k}{m} q_1 \\ \frac{d^2 q_2}{dt^2} = -\frac{k}{m} q_2 \end{cases}$$

Two-dimensional SHM

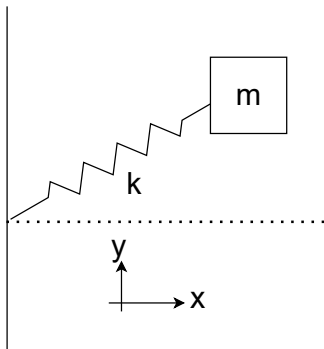$$\begin{cases} \frac{d^2 q_1}{dt^2} = -\frac{k}{m} q_1 \\ \frac{d^2 q_2}{dt^2} = -\frac{k}{m} q_2 \end{cases}$$

Figure 9: Two-dimensional mass-spring system.

Two-dimensional SHM Invariance

$$E = \frac{m(p_1^2 + p_2^2)}{2} + \frac{k(q_1^2 + q_2^2)}{2}$$

Two-dimensional SHM Invariance

$$E = \frac{m(p_1^2 + p_2^2)}{2} + \frac{k(q_1^2 + q_2^2)}{2}$$

$$L[\mathbf{f}] = \frac{dE}{dt} = mp_1 a_1 + mp_2 a_2 + kq_1 v_1 + kq_2 v_2 = 0.$$

Two-dimensional SHM Invariance

$$L[\mathbf{f}] = \frac{dE}{dt} = mp_1 a_1 + mp_2 a_2 + kq_1 v_1 + kq_2 v_2 = 0.$$

$$K(X, X') = \begin{pmatrix} K_{a_1}(X, X') & 0 & 0 & 0 \\ 0 & K_{a_2}(X, X') & 0 & 0 \\ 0 & 0 & K_{v_1}(X, X') & 0 \\ 0 & 0 & 0 & K_{v_2}(X, X') \end{pmatrix}.$$

Two-dimensional SHM Invariance

$$L[\mathbf{f}] = \frac{dE}{dt} = mp_1 a_1 + mp_2 a_2 + kq_1 v_1 + kq_2 v_2 = 0.$$

$$K(X, X') = \begin{pmatrix} K_{a_1}(X, X') & 0 & 0 & 0 \\ 0 & K_{a_2}(X, X') & 0 & 0 \\ 0 & 0 & K_{v_1}(X, X') & 0 \\ 0 & 0 & 0 & K_{v_2}(X, X') \end{pmatrix}.$$

$$\begin{pmatrix} \mathbf{f}(X) \\ L[\mathbf{f}(X_L)] \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0_{4n} \\ 0_\ell \end{pmatrix}, \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right),$$

Two-dimensional SHM Invariance

$$\begin{pmatrix} \mathbf{f}(X) \\ L[\mathbf{f}(X_L)] \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0_{4n} \\ 0_\ell \end{pmatrix}, \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right),$$

$$A = K(X, X'), B = \begin{pmatrix} K_{a_1} \\ K_{a_2} \\ K_{v_1} \\ K_{v_2} \end{pmatrix} \odot \begin{pmatrix} mP_{1,L} \\ mP_{2,L} \\ kQ_{1,L} \\ kQ_{2,L} \end{pmatrix}, C = B^T$$

$$D = K_{a_1} m^2 \odot (p_{1,L} \otimes p_{1,L}) + K_{a_2} m^2 \odot (p_{2,L} \otimes p_{2L})$$
$$+ K_{v_1} k^2 \odot (q_{1,L} \otimes q_{1,L}) + K_{v_2} k^2 \odot (q_{2,L} \otimes p_{2,L})$$

Learning Invariance

$$L[\mathbf{f}] = f_1(p_1, p_2, q_1, q_2)a_1 + f_2(p_1, p_2, q_1, q_2)a_2$$

$$+g_1(p_1, p_2, q_1, q_2)v_1 + g_2(p_1, p_2, q_1, q_2)v_2$$

Learning Invariance

$$L[\mathbf{f}] = f_1(p_1, p_2, q_1, q_2)a_1 + f_2(p_1, p_2, q_1, q_2)a_2$$

$$+g_1(p_1, p_2, q_1, q_2)v_1 + g_2(p_1, p_2, q_1, q_2)v_2$$

1. Compare random invariance to the theortically correct one as well as the known form in terms of marginal likelihood and MSE.

2. Find the correlation between the marginal likelihood and predictive performance, which is expect to be positive

3. Allow the polynomial coefficients to be optimised from the theoretical value.

Two-dimensional SHM Results

Method	RBF	Known	Learnt
Log Marginal Likelihood	430.62	478.70	475.42
MSE	0.0271	0.0035	0.0035

Table 2: Two-dimensional SHM Invariance performance.
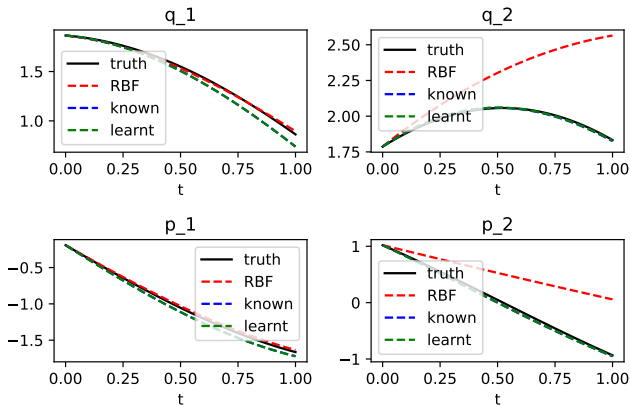
Two-dimensional SHM Results

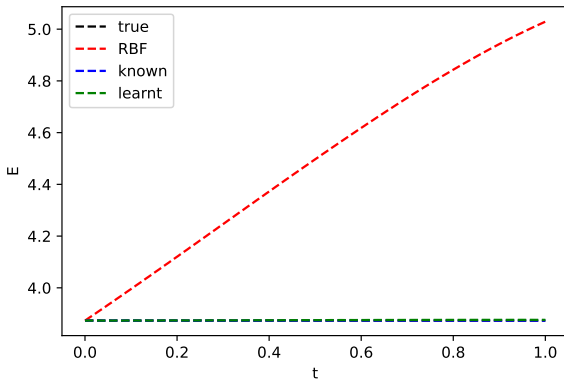Figure 10: Two-dimensional SHM prediction.

Two-dimensional SHM Results

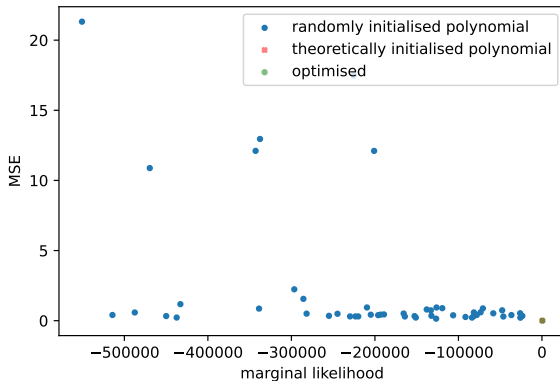Figure 10: Two-dimensional SHM energy.

Two-dimensional SHM Results

Figure 10: Two-dimensional SHM learnt invariance.

Double Pendulum

$$\begin{cases} \dfrac{d^2q_1}{dt^2} = \dfrac{-g(2m_1+m_2)\sin q_1 - m_2 g \sin(q_1-2q_2) - 2\sin(q_1-q_2)m_2\big(p_2^2 l_2 + p_1^2 l_1 \cos(q_1-q_2)\big)}{l_1(2m_1+m_2-m_2\cos(2q_1-2q_2))} \\[2ex] \dfrac{d^2q_2}{dt^2} = \dfrac{2\sin(q_1-q_2)\big(p_1^2 l_1(m_1+m_2) + g(m_1+m_2)\cos q_1 + p_2^2 l_2 m_2 \cos(q_1-q_2)\big)}{l_2(2m_1+m_2-m_2\cos(2q_1-2q_2))} \end{cases}$$
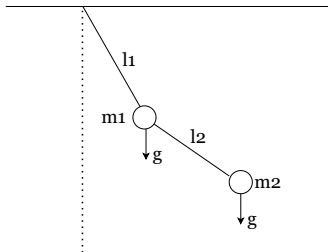
Double Pendulum

$$\begin{cases} \dfrac{d^2 q_1}{dt^2} = \dfrac{-g(2m_1+m_2)\sin q_1 - m_2 g \sin(q_1-2q_2) - 2\sin(q_1-q_2)m_2\left(p_2^2 l_2 + p_1^2 l_1 \cos(q_1-q_2)\right)}{l_1(2m_1+m_2-m_2\cos(2q_1-2q_2))} \\ \dfrac{d^2 q_2}{dt^2} = \dfrac{2\sin(q_1-q_2)\left(p_1^2 l_1(m_1+m_2) + g(m_1+m_2)\cos q_1 + p_2^2 l_2 m_2 \cos(q_1-q_2)\right)}{l_2(2m_1+m_2-m_2\cos(2q_1-2q_2))} \end{cases}$$

Figure 11: Double Pendulum.

Double Pendulum Results

Method	RBF	Known	Learnt
Log Marginal Likelihood	783.46	838.41	869.09
MSE	0.0040	0.0004	0.0018

Table 2: Double pendulum Invariance performance.
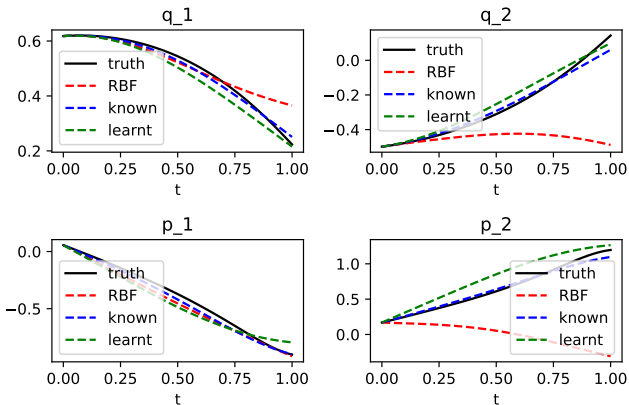
Double Pendulum Results

Figure 12: Double pendulum prediction.

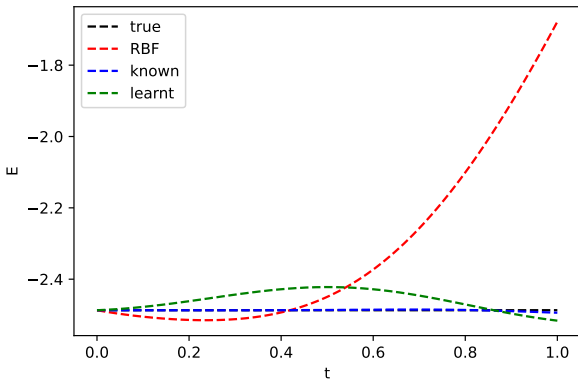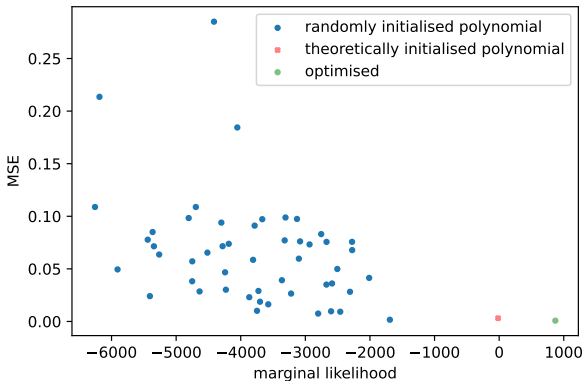# Double Pendulum Results

Figure 12: Double pendulum energy.

Double Pendulum Results

Figure 12: Double pendulum learnt invariance.