Introduction

The trajectories and time evolution of a cartpole system are characterised and then modelled linearly. Predictions are then compared to reality.

General considerations

The model runs 50 Euler steps over a time step of 0.2 seconds. The system has oscillating variables, and so the Nyquist criterion should be considered: if the system is observed at less than half the sampling rate of its maximum oscillation frequency, aliasing artefacts will occur. Such trajectories would be sharply nonlinear, and so would be difficult to predict with any model, linear or nonlinear. Since the sampling rate is fixed, it is therefore best to stick to low frequencies, and assume that useful controllers will keep the system state within this domain.

For the pole angular velocity:

$$f_s = 5Hz;\ f_\theta = \dot{\theta}/2\pi \implies \dot{\theta} < 5\pi \approx 15.7$$

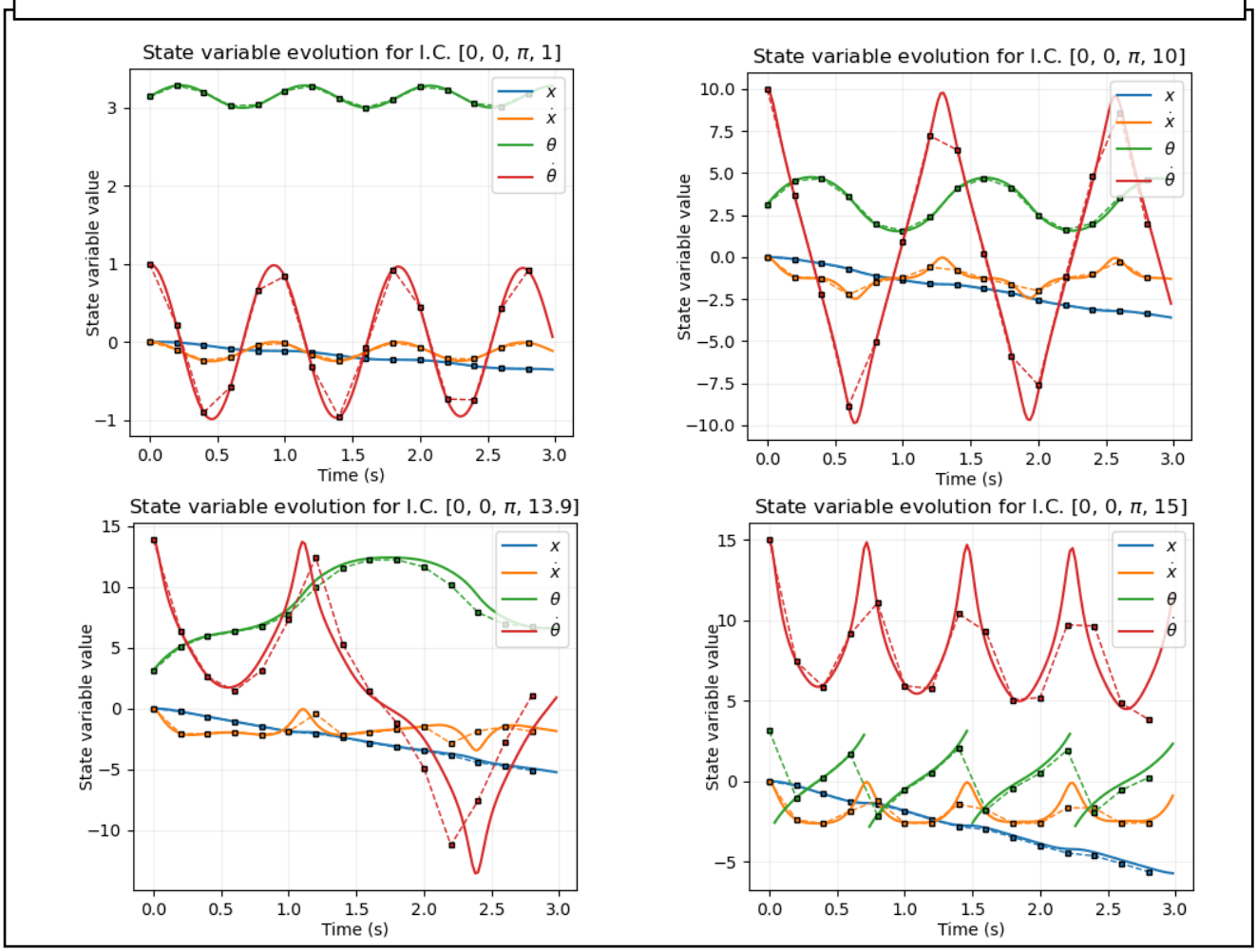
So $\dot{\theta} \in [-15, 15]$ should be satisfactory. Similarly, $\dot{x} \in [-10, 10]$ will be taken. θ is an angle (centred at the unstable equilibrium) and so $\dot{\theta} \in [-\pi, \pi]$. The system dynamics are invariant to shifts in x, so x can be set initially to zero for all trajectories.

Rollouts

Time-trajectories

There are two principal trajectories of the system: *bound* and *unbound*. Bound trajectories occur when the pole oscillates with insufficient energy to reach $\theta = 0$. In the bound, low-energy limit (fig 1a), the pole oscillates sinusoidally about $\theta = \pi$. The cart's velocity couples to this, oscillating in phase with the $\dot{\theta}$ but with a much smaller amplitude. However, it remains of an opposite sign to the

Fig 1. Time evolution of system variables with no applied force, starting at the stable equilibrium ($\theta = \pi$), with varying angular velocities. **Solid lines**: Illustrative evolution using a timestep of 0.02 s. **Dashed lines & markers**: Sampled values with the actual timestep (0.2 s). Note that the nature of the Euler scheme causes these pairs to diverge when friction is significant.



initial angular velocity, so that the cart position "drifts" monotonically. Conversely, unbound trajectories have sufficiently high energy that the pole repeatedly "flips", causing θ to monotonically increase (sans remapping) (fig 1d). Between these extremes, the behaviour stays broadly similar either side of the critical transition energy ($E = 0$, defining zero potential at $\theta = 0$ [A1]), and the monotonic cart "drifting" remains, though the waveforms change shape in a nontrivial way (fig 1b). In all cases, friction gradually causes the energy to diminish and the oscillation amplitudes to reduce, so that bound states decay to zero and unbound states eventually become bound (fig 1c). This effect of friction can be largely decoupled from the broader system dynamics, especially when a controller is present to offset the energy loss.

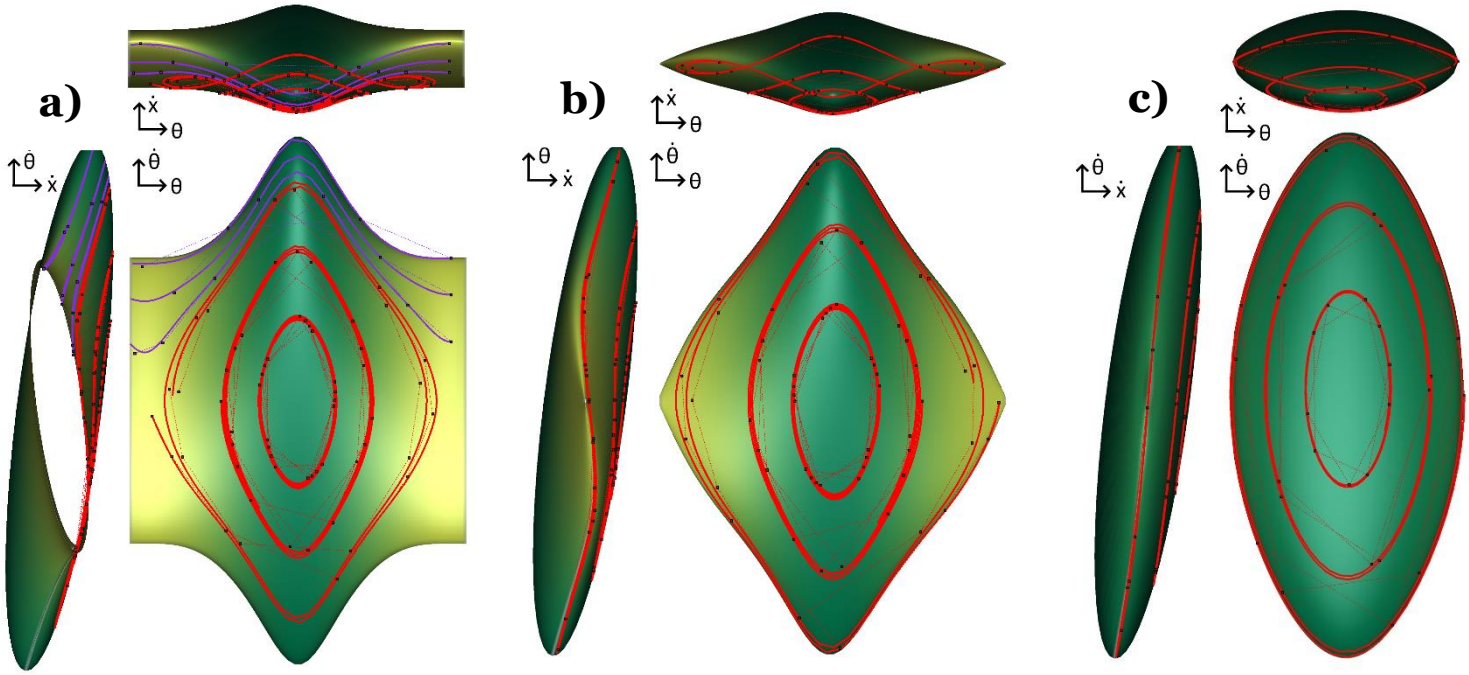
Adding a nonzero cart velocity leaves the pole dynamics mostly unchanged (see next section), and shifts in initial cart position are trivial.

Phase portraits

Since the cart position is trivial to deduce given its velocity history, and the timescale of energy loss due to friction is much greater than the timescale of the cart dynamics, the a qualitative understanding of the state-space trajectories of this system can be gained entirely via energy embedded in the 3d space formed by $[\dot{x}, \theta, \dot{\theta}]$.

This consists of ellipses in the $\dot{x}$-$\dot{\theta}$ plane modulated in size and tilt angle by the value of θ. When $E > 0$, a continuous, periodic tube is formed. It splits into disjoint surfaces at $E = 0$, which shrink into ellipsoids as E reduces to its minimal value. Over longer time periods, trajectories migrate onto lower-energy isosurfaces until they become bound and oscillations decay.

Fig 2. Orthographic views of trajectories in $[\dot{x}, \theta, \dot{\theta}]$-space with initial energies equal to 1, 0, and -2 for a), b), c) respectively, superimposed onto energy isosurfaces at those values. **Red lines**: Bound trajectories. **Purple lines:** Unbound. **Solid lines:** Illustrative, with 0.02 s timestep. **Dotted lines & markers:** 0.2 s timestep.



Changes of State

Formulation

Suppose $\boldsymbol{X}$ is the current state of the system. After one step of time evolution (one "action"), let $\boldsymbol{Z}$ be the next state, and define $\boldsymbol{Y} = \boldsymbol{Z} - \boldsymbol{X}$. The goal of modelling this system is to learn the mapping $\boldsymbol{g}: \mathbb{R}^4 \to \mathbb{R}^4, \boldsymbol{g}(\boldsymbol{X}) = \boldsymbol{Z}$, or equivalently $\boldsymbol{f}: \mathbb{R}^4 \to \mathbb{R}^4, \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{Y}$.

General Properties of $\boldsymbol{f}$ and $\boldsymbol{g}$

Over the relatively small timestep T=0.2 s, variables should generally evolve in a relatively simple way. In fact, many pairs of input-output variables have approximately linear relationships for $\boldsymbol{g}(\boldsymbol{X})$. This is demonstrable by initialising the system in random states, scanning alternately over the full ranges of state variables, and calculating correlation coefficients with the output state variables. Figure 3a displays the mean of such values over 100 initial states – clearly a number of matrix elements imply linearity, at least for particular settings of other parameters. In particular, since each variable does not change much in a timestep, it is correlated strongly to itself, and the gradient of its best fit line is approximately equal to 1 (fig. 3b).

Subtracting the initial state gives $\boldsymbol{f}(\boldsymbol{X})$ and removes these diagonal correlations – though the others remain (fig. 3c,d)– as well as removing constant terms. If we fit a linear model to $\boldsymbol{f}(\boldsymbol{X})$, we can hope to capture these particular relationships, hoping that the gradients do not change significantly over the range of other parameter settings.

The nonlinear parameter dependencies of $\boldsymbol{f}$ which have been obscured by looking for linear relationships can be approximately found by evaluating the gradients at random points, and summing their absolute values over many samples (fig 3e). With this in mind, we can entirely exclude dependence on x, and establish that dependence on $\dot{x}$ is indeed weak except of course for x. However, a better picture of nonlinear mappings should be obtained first.

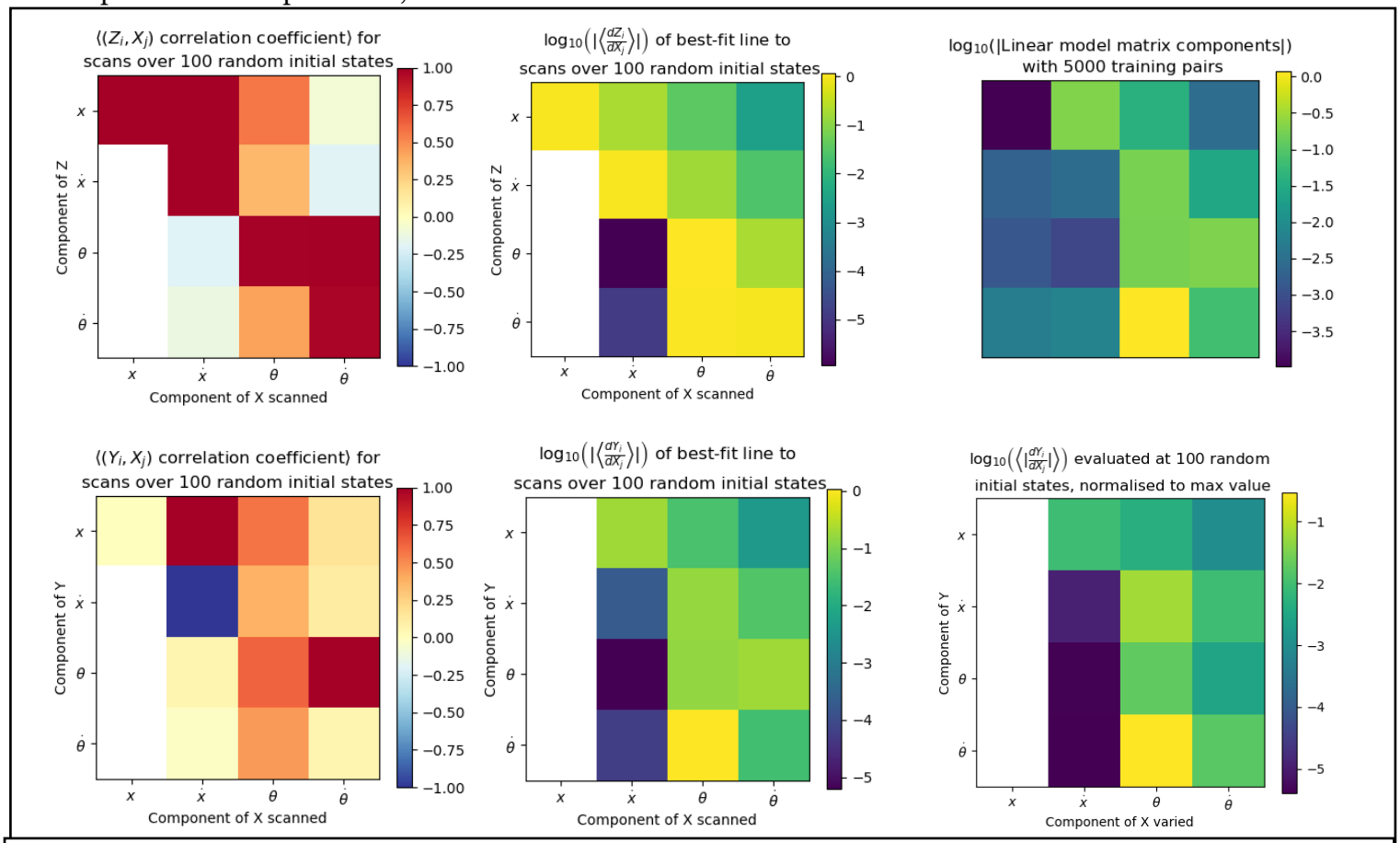


Fig 3. Colour maps of: **a), b)** Mean Pearson correlation coefficient between change in function value and initial state scanned over full range for 100 random initial states for $\boldsymbol{g}$ and $\boldsymbol{f}$; **c), d)** Log magnitude of mean fitted gradient for corresponding scans; **e)** Components of the linear model matrix fit to 5000 random (Y, X) data pairs; **f)** Log mean magnitudes of gradient components evaluated at 100 random initial states. *White-coloured values are those with magnitudes < 1e-12, and so considered to be zero.*

Understanding in Physical Terms

Each state variable's evolution depends fairly strongly on θ and $\dot{\theta}$. Variables only have a very weak dependence on $\dot{x}$, except of course x. The variables are all completely independent of x.

These traits can be understood by considering the dynamical equations of the system [A2], where causal connections for an infinitesimal timestep are depicted schematically in fig. 4. There is a general flow of information downward through the figure, with $\dot{x}$ only affecting the angular variables and itself through the weak effects of friction. During the 50 Euler iterations per timestep, the information flow is diffused through the network, resulting in the causality matrix of Figure 3e. In this sense, the angular variables completely drive the dynamics, and so it is important to take care to model them well.

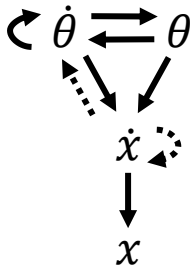


Fig 4. Lines represent the flow of information and causal changes between state variables over an infinitesimally short timestep. Solid lines indicate that the dependence is important on short timescales; dashed lines indicate that the only dependence is via the weak effects of friction.

Nonlinear Properties of $\boldsymbol{f}$

Although friction is important on longer time scales, the goal is to actively control the system, and so the shorter-term dynamics are the only important aspects to model. Therefore all of the components of $\boldsymbol{f}(\boldsymbol{X})$ except the first can be characterised in a contour plot against θ and $\dot{\theta}$. For the evolution of cart position, the relationship with $\dot{x}$ will be simply $\frac{df_1}{d\dot{x}} = 0.2\dot{x}$ excluding friction, so with this in mind, a contour plot against θ and $\dot{\theta}$ with $\dot{x} = 0$ is again sufficient for understanding. These plots are shown in fig. 5. It is clear that for these mappings, a linear function will be insufficient to approximate the entire landscape. For the evolution of θ, the broad behaviour will be correct (ignoring the "wiggles"), and for the evolution of $\dot{\theta}$, a linear model trained from data in the central diagonal region would roughly characterise the local contours. However, the full range of θ needs to be modelled, so a linear model

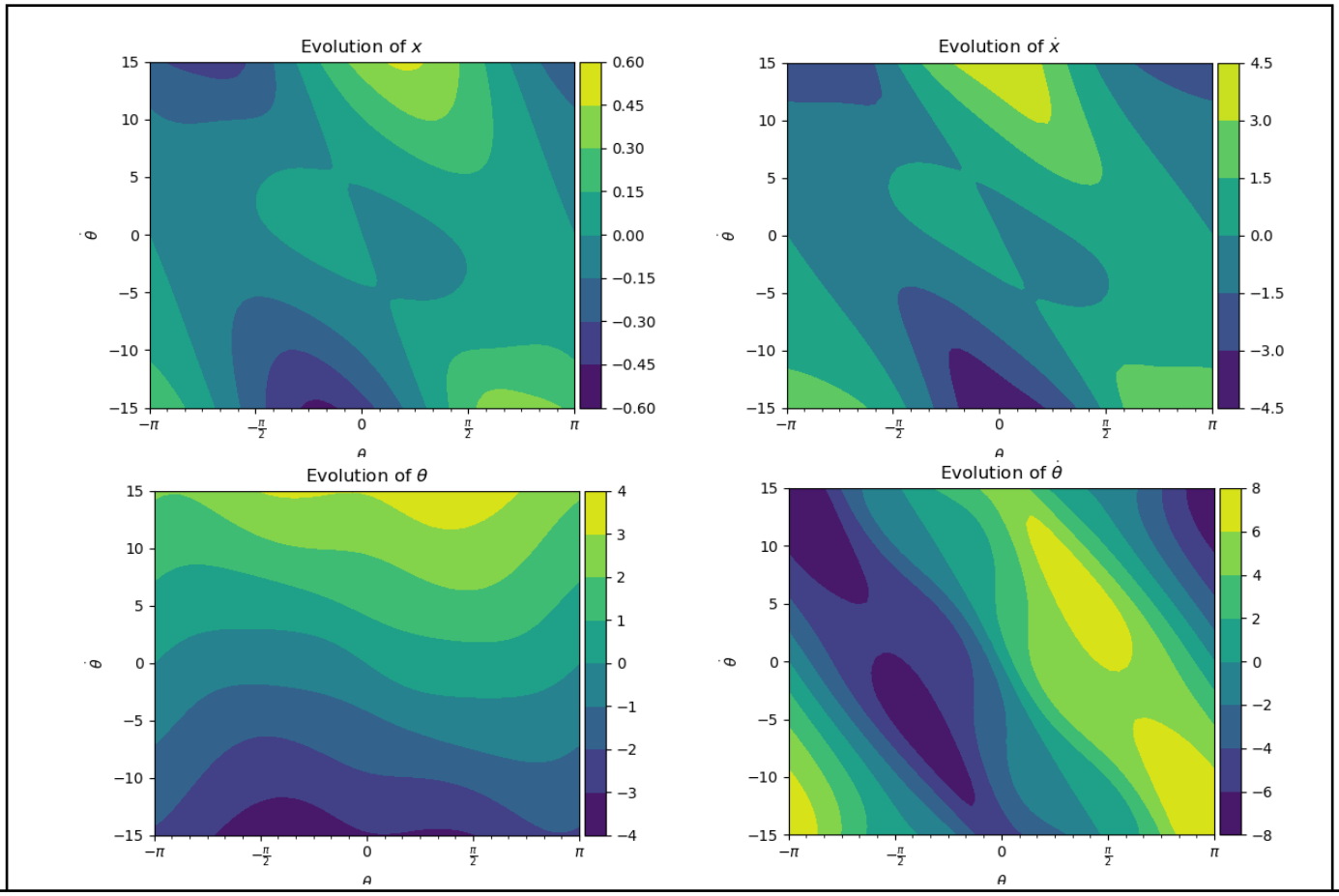


Fig 5. Contour plots of state variable evolutions as functions of θ and $\dot{\theta}$.

cannot be expected to model the oscillatory relationship perpendicular to these contours well. For x and $\dot{x}$, the function is near zero in the majority of the landscape – which can be reproduced trivially with a linear model – but there are hills and basins at large $\dot{\theta}$, which definitely cannot. However, if $\dot{\theta}$ is kept sufficiently small, an approximately zero function could work moderately well, excluding the shallower central wavelet-like variation which corresponds to the minor couple oscillations in fig. 1.

Linear model

Formulation

For each component Y_i of $\boldsymbol{Y}$, assume the dependence on $\boldsymbol{X}$ is linear. As discussed, the constant term is then zero, and for each Y_i,

$$Y_i = \boldsymbol{c}_i^T \boldsymbol{X} = \boldsymbol{X}^T \boldsymbol{c}_i$$

Where the coefficients of $\boldsymbol{c}_i$ are to be learned by least-squares minimisation over a large number of $(\boldsymbol{X}, \boldsymbol{Y})$ pairs. This gives an equation of the form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ where $\boldsymbol{A}$ is a matrix of stacked $\boldsymbol{X}^T$ vectors and $\boldsymbol{b}$ is the corresponding Y_i values. Each such equation can then be optimised in a least-squares sense by standard algorithms and the resulting $\boldsymbol{c}_i$ vectors stacked in into the matrix $\boldsymbol{C}$ to give $\boldsymbol{Y} = \boldsymbol{C}\boldsymbol{X}$.

Convergence

A good number N of random $\boldsymbol{X}$ samples to generate is $N = 500$ (fig. 6). The magnitudes of elements of a matrix generated $N = 5000$ is shown in fig. 3e, and its elements broadly line up with expectations from fig. 3d. The nonzero coefficients of the cart position can be attributed to the fact the matrix has not fully converged.

Single-step effectiveness of model

A model was trained with $N = 5000$ datapoints for this section [A3].

Expected vs target next step

To visualise how well the model matches the real system evolution over a single step, predictions can be plotted on the y-axis against the actual evolutions on the x-axis in pairs. The closer the points to the line y=x, the better the prediction. If an initial state is then scanned over, this forms a line which should ideally lie entirely near y=x. Such plots are shown in fig 7. They illustrate several features of the model:

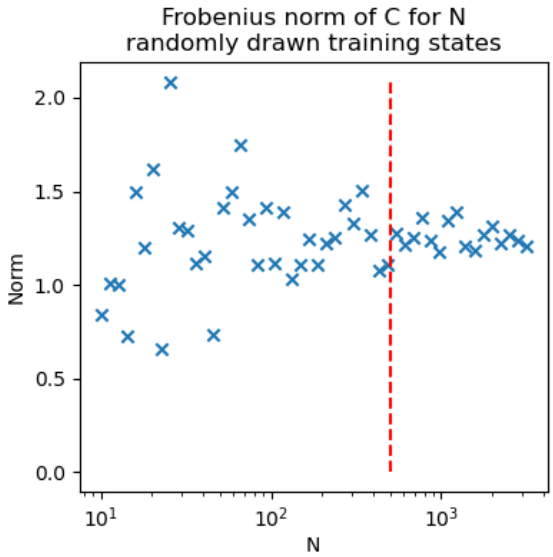


Fig 6. Frobenius norm of $\boldsymbol{C}$ matrix used as a metric for convergence, plotted for collections of $(\boldsymbol{X}, \boldsymbol{Y})$ pairs of varying sizes (with randomly drawn $\boldsymbol{X}$ within the variable ranges). $N = 500$ represents a good tradeoff of optimisation speed versus accuracy. The model is expected to perform fairly poorly, so a little noise is not much of a problem.

- Predictions of θ and x are in general very good, while predictions of the velocities are much poorer.
- The model correctly demonstrates independence of θ and x on x and $\dot{x}$ (hence their lines are points).
- The velocities are slightly more sensitive to θ and x than the positions, but are still fairly independent.
- θ and $\dot{\theta}$ correctly have a large effect on state variables, but their predictions are fairly poor.

Contour plots

The difference between the predicted and actual evolutions can be plotted as a contour map against θ and $\dot{\theta}$, showing how well the linear model fits to fig. 8. Predictions are poor in general, and get worse as θ and $\dot{\theta}$ increase. Unfortunately, since the stable equilibrium is at $\theta = \pi$, the system will always end up in this region of state space.

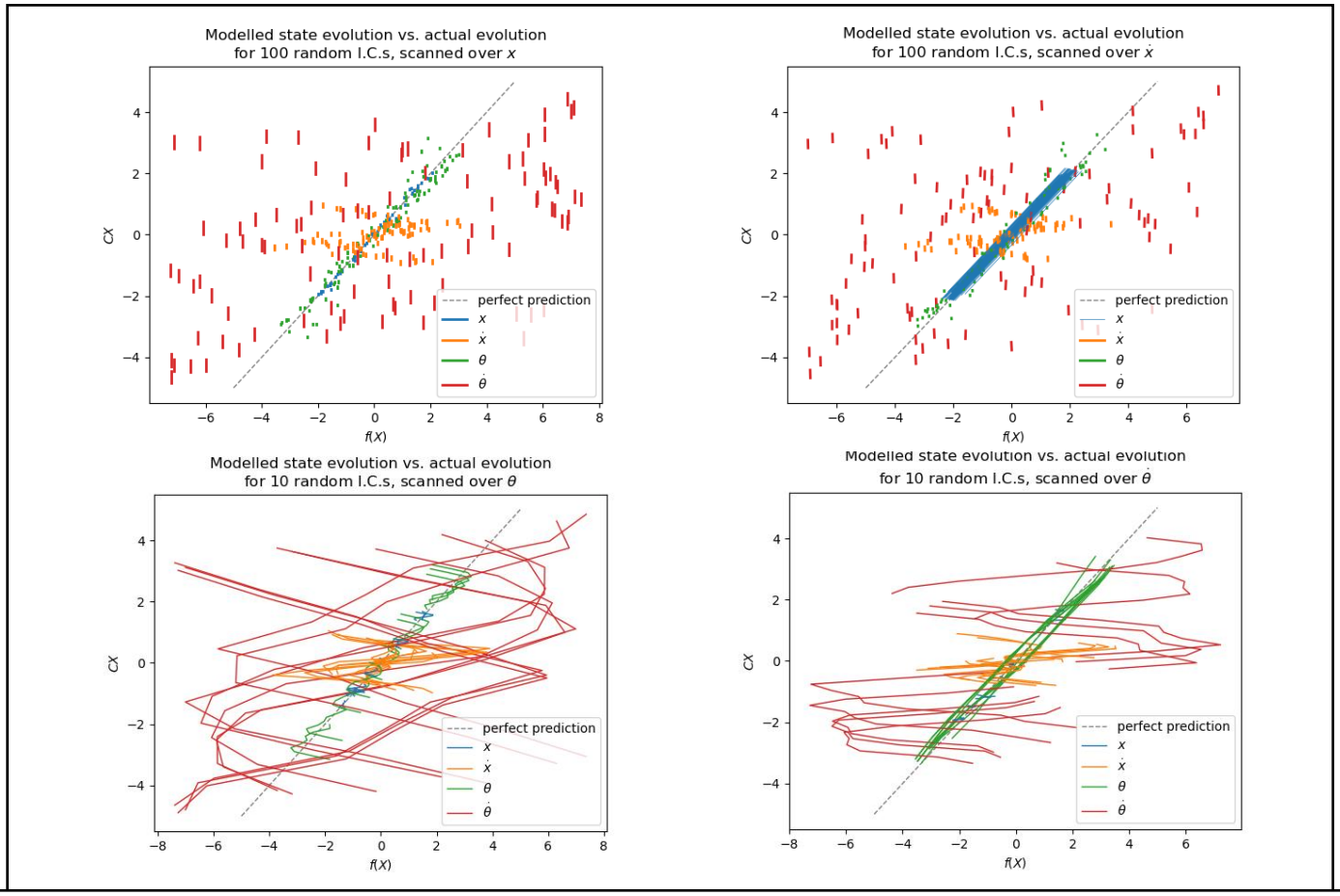


Fig 7. Comparison of state evolutions, initialised with random states, scanned over state variables.

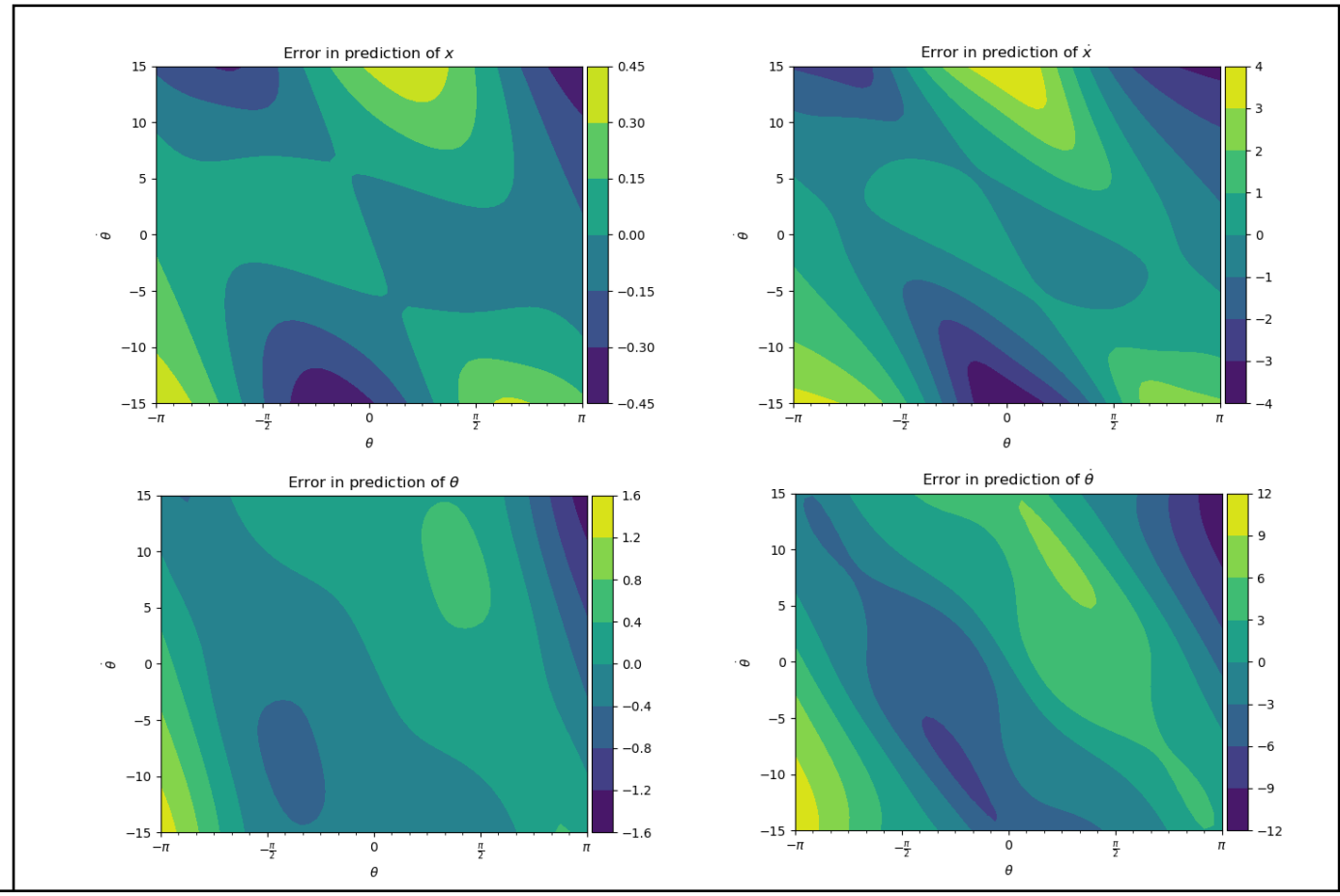


Fig 8. Contour plots of state variable prediction errors as functions of θ and $\dot{\theta}$.

Trajectories of model

In general, fitted models perform poorly. The state variables diverge from real trajectories within a few timesteps and sometimes immediately. The state variables (excluding the angle) eventually diverge to infinity, driven by the cart velocity, rather than decaying to zero, although this often takes a long time.

There are some situations when the model performs better. Some matrices, including the previously fitted one with $N = 5000$, reproduce simple oscillations qualitatively, even if they rapidly become out of phase and often have different frequencies. One particular pathology, which manifests even more strongly if the time step is decreased so that more evolutions are required per cycle (despite the simpler modelling task for a single evolution), is that the model unpredictably alternates between the bound and unbound states, ending up eventually in the bound state, in contradiction to reality.

Conclusions

A linear model is insufficient to explain the evolution of the cartpole system, due to periodicity-induced nonlinearities in the angular variables which drive the system. A nonlinear model would be required to fit the evolution function better.

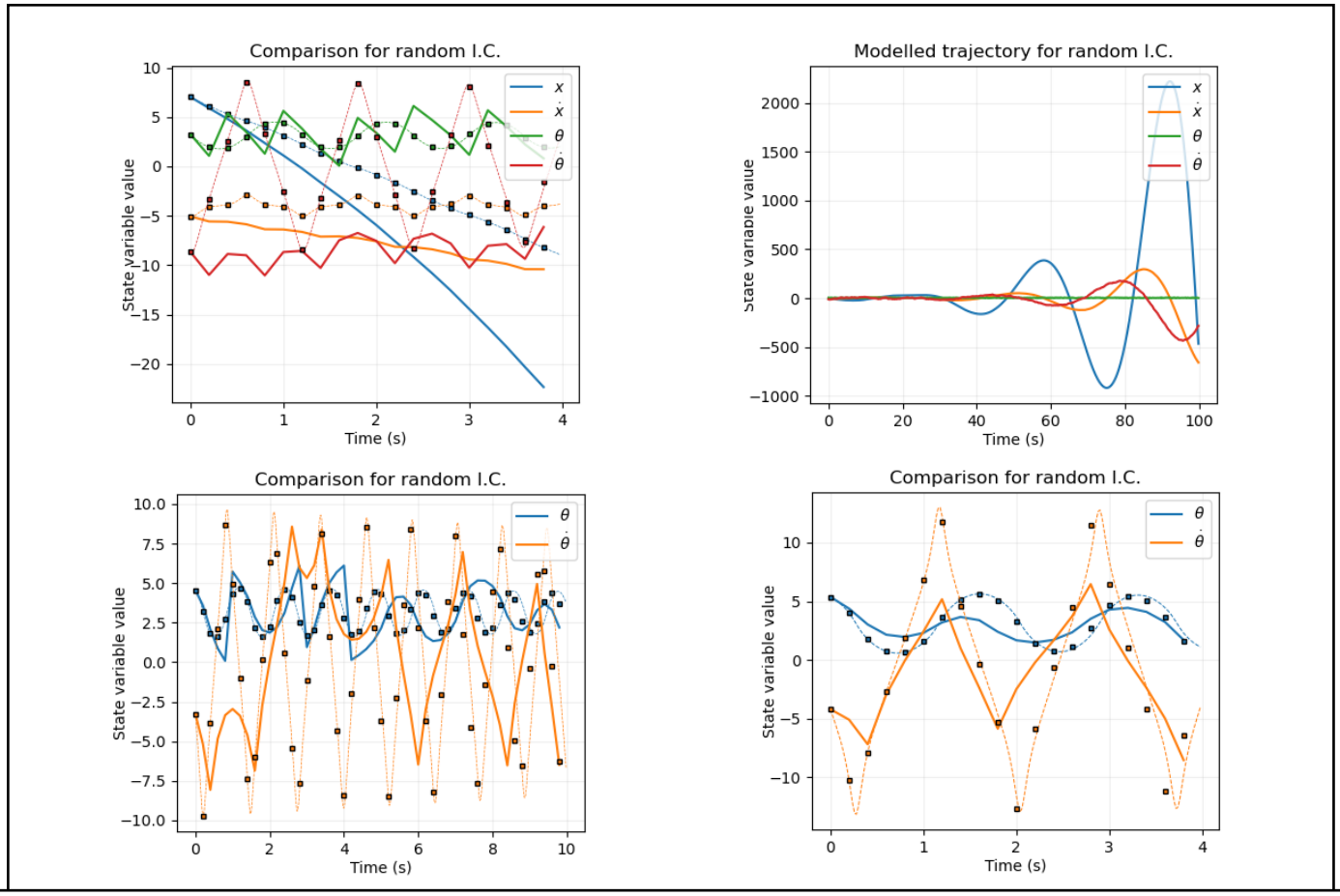


Fig 5. Contour plots of state variable evolutions as functions of θ and $\dot{\theta}$.

Appendix

A1: Energy definition

Let T be kinetic energy, V be potential energy with zero reference at $\theta = 0$, and E be total energy.

$$T = \frac{1}{2}(M+m)\dot{x}^2 + \frac{1}{2}mL\dot{x}\dot{\theta}\cos\theta + \frac{1}{6}mL^2\dot{\theta}^2$$

$$V = \frac{1}{2}mgL(\cos\theta - 1)$$

and $E = T + V$, so

$$E = \frac{1}{2}\dot{x}^2 + \frac{1}{8}\dot{x}\dot{\theta}\cos\theta + \frac{1}{48}\dot{\theta}^2 + \frac{g}{8}(\cos\theta - 1)$$

A2: Equations of motion

Defining $\mu = 4(m+M) - 3m\cos^2\theta$, the equations of motion are

$$\frac{d}{dt}\begin{bmatrix} x \\ \dot{x} \\ \theta \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \frac{1}{\mu}\left(2mL\dot{\theta}^2\sin\theta - 3mg\cos\theta\sin\theta + 4F - 4\mu_x\dot{x} + \frac{6}{L}\mu_\theta\dot{\theta}\cos\theta\right) \\ \dot{\theta} \\ \frac{3}{\mu}\left(-m\dot{\theta}^2\cos\theta\sin\theta + 2\frac{(m+M)}{mL}(mg\sin\theta) - \frac{2}{L}F\cos\theta + \frac{2}{L}\mu_x\dot{x}\cos\theta + 4\frac{(m+M)}{mL}\mu_\theta\dot{\theta}\right) \end{bmatrix}$$

$$\frac{d}{dt}\begin{bmatrix} x \\ \dot{x} \\ \theta \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \frac{1}{4-1.5\cos^2\theta}\left(.5\dot{\theta}^2\sin\theta - 1.5g\cos\theta\sin\theta + 0.001(12\dot{\theta}\cos\theta - 4\dot{x}) + 4F\right) \\ \dot{\theta} \\ \frac{3}{4-1.5\cos^2\theta}\left(-.5\dot{\theta}^2\cos\theta\sin\theta + 4g\sin\theta + 0.001(4\dot{x}\cos\theta - 16\dot{\theta}) - 4F\cos\theta\right) \end{bmatrix}$$

13: Correlation matrix with N=5000

```
[[4.08391096e-04  2.00806862e-01  4.18320139e-02  2.62501158e-03]
 [-3.52547831e-03  6.42016490e-03  2.06380111e-01  2.61166600e-02]
 [-1.28462041e-03  1.28783926e-03  1.80439712e-01  1.97702879e-01]
 [-1.32082079e-02  1.29174134e-02  1.25855381e+00 -6.06674820e-02]]
```