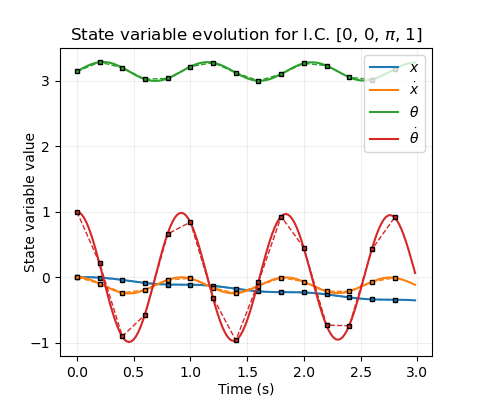
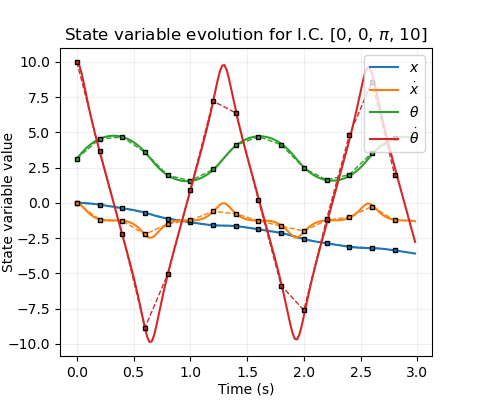
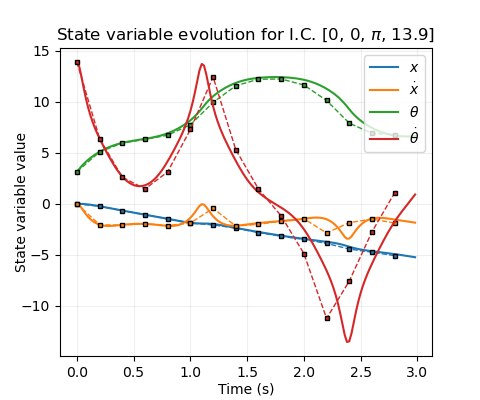
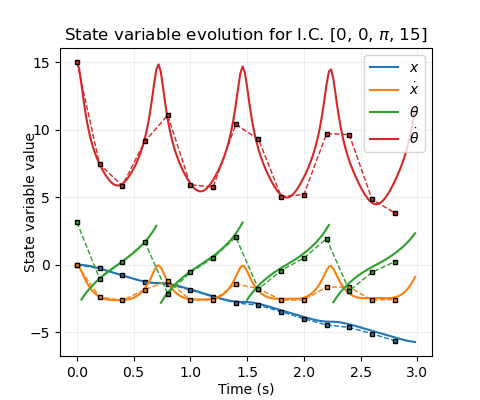
# Introduction

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:



**Fig 1.** Time evolution of system variables with no applied force, starting at the stable equilibrium (, 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.

So should be satisfactory. Similarly, will be taken. is an angle (centred at the unstable equilibrium) and so . The system dynamics are invariant to shifts in , so can be set initially to zero for all trajectories.

# Rollouts

## Time-trajectories

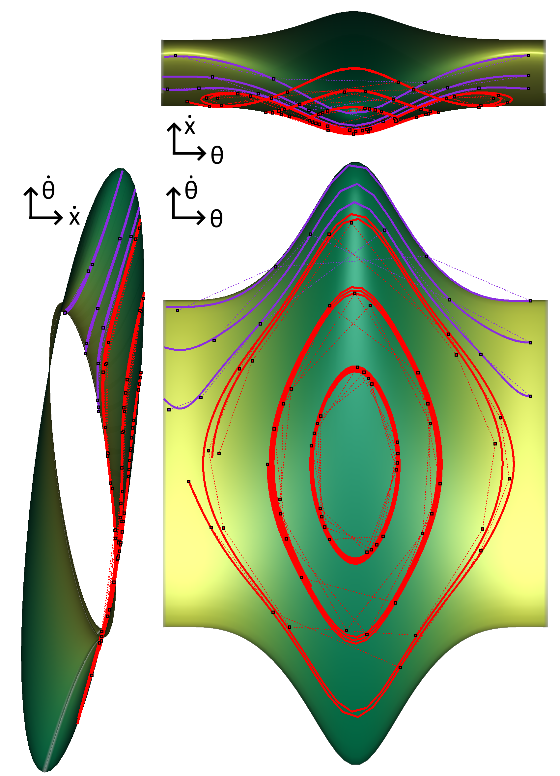
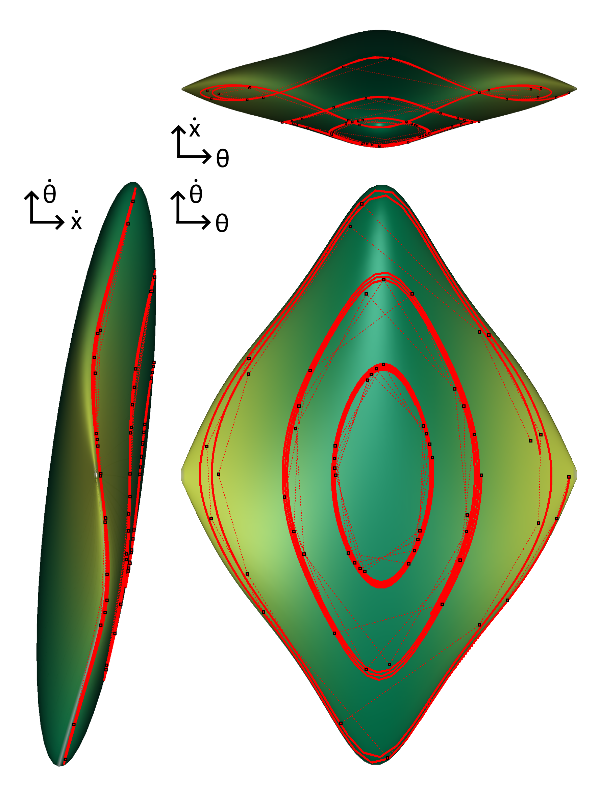
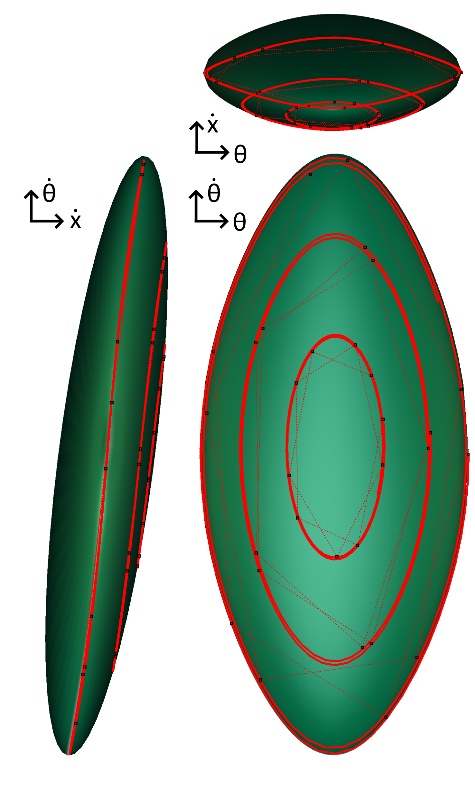
There are two principal trajectories of the system: *bound*and *unbound*. Bound trajectories occur when the pole oscillates with insufficient energy to reach . In the bound, low-energy limit (fig 1a), the pole oscillates sinusoidally about . The cart’s velocity couples to this, oscillating in phase with the but with a much smaller amplitude. However, it remains of an opposite sign to the initial angular velocity, so that the cart position “drifts” monotonically. Conversely, unbound trajectories have sufficiently high energy that the pole repeatedly “flips”, causing to monotonically increase (sans remapping) (fig 1d). Between these extremes, the behaviour stays broadly similar either side of the critical transition energy (, defining zero potential at [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 [.

**Fig 2.** Orthographic views of trajectories in [-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.



**c)**

**a)**

**b)**

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

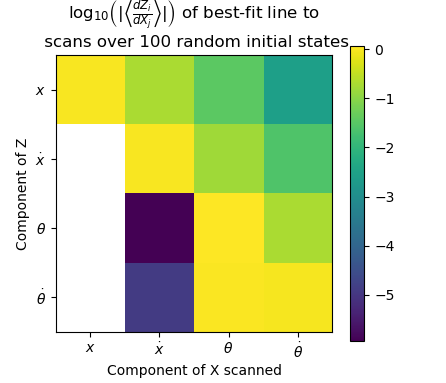
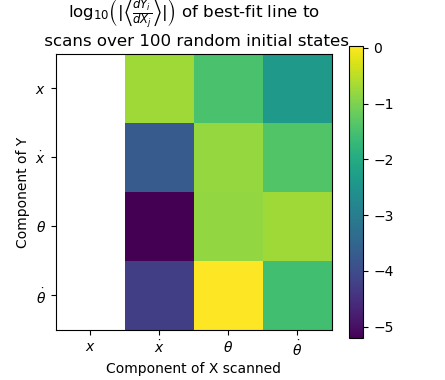
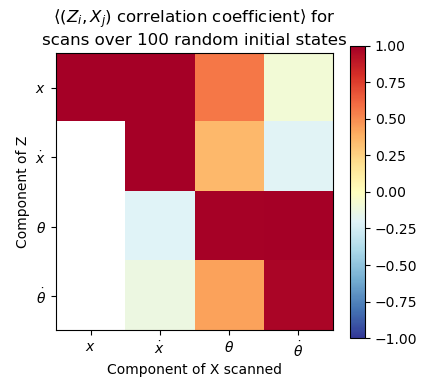
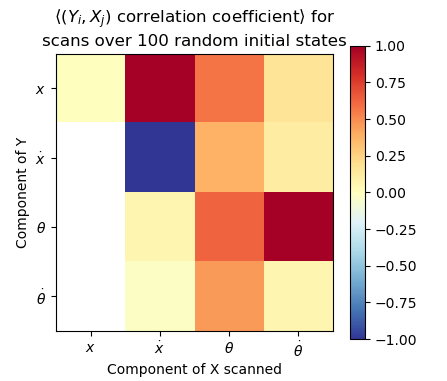
# Changes of State

## Formulation

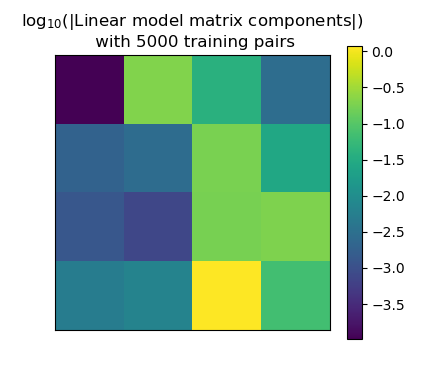
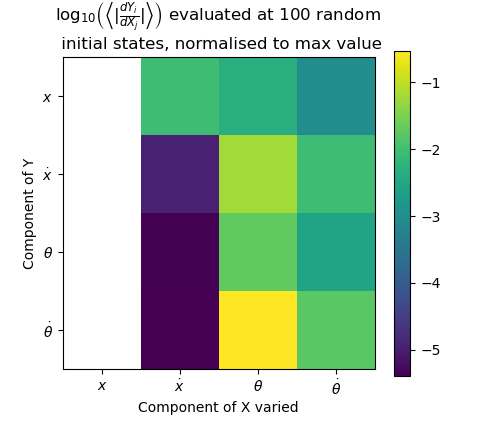
Suppose is the current state of the system. After one step of time evolution (one “action”), let be the next state, and define . The goal of modelling this system is to learn the mapping , or equivalently .

## General Properties of and

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 . 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).



**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 and ; **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.*



Subtracting the initial state gives 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 , 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 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 , and establish that dependence on is indeed weak except of course for . However, a better picture of nonlinear mappings should be obtained first.

## Understanding in Physical Terms

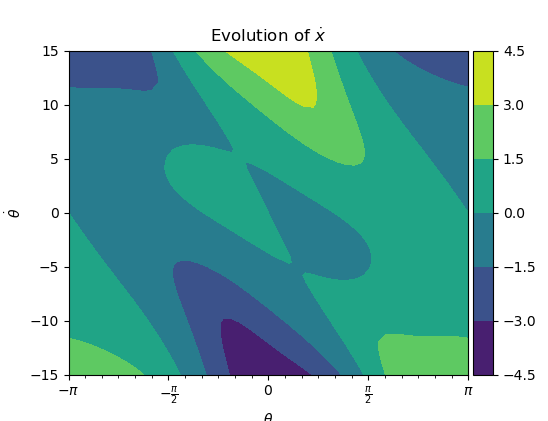
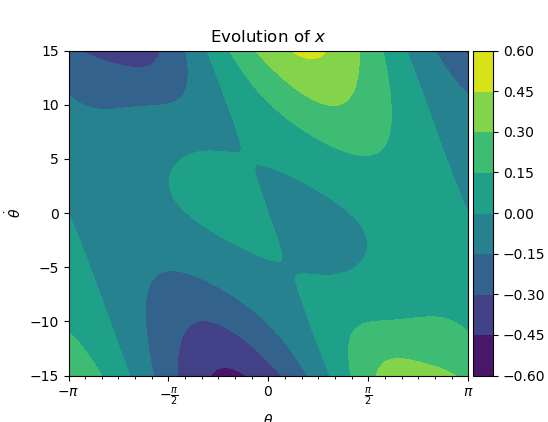
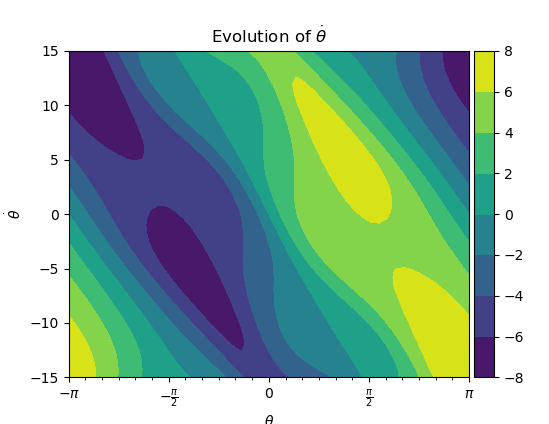
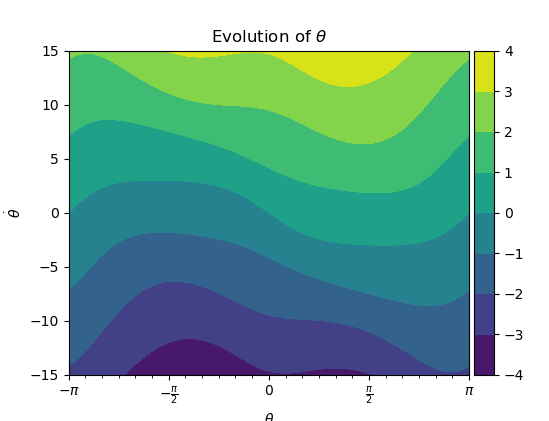
**Fig 4.** Lines represent the flow of information and causal changes between state variables over an infinitesimally short timestep. Solid lines indicate that the dependence is important on short timescales; dashed lines indicate that the only dependence is via the weak effects of friction.

Each state variable’s evolution depends fairly strongly on and . Variables only have a very weak dependence on , except of course . The variables are all completely independent of .

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 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.

## Nonlinear Properties of

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 except the first can be characterised in a contour plot against and . For the evolution of cart position, the relationship with will be simply excluding friction, so with this in mind, a contour plot against and with 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 , the broad behaviour will be correct (ignoring the “wiggles”), and for the evolution of , a linear model trained from data in the central diagonal region would roughly characterise the local contours. However, the full range of needs to be modelled, so a linear model cannot be expected to model the oscillatory relationship perpendicular to these contours well. For and 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 , which definitely cannot. However, if 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.



**Fig 5.** Contour plots of state variable evolutions as functions of and .

# Linear model

## Formulation

For each component of , assume the dependence on is linear. As discussed, the constant term is then zero, and for each ,

Where the coefficients of are to be learned by least-squares minimisation over a large number of pairs. This gives an equation of the form where is a matrix of stacked vectors and is the corresponding values. Each such equation can then be optimised in a least-squares sense by standard algorithms and the resulting vectors stacked in into the matrix to give .

**Fig 6.** Frobenius norm of matrix used as a metric for convergence, plotted for collections of ( pairs of varying sizes (with randomly drawn within the variable ranges). 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.

## Convergence

A good number of random samples to generate is (fig. 6). The magnitudes of elements of a matrix generated 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 datapoints for this section [A3].

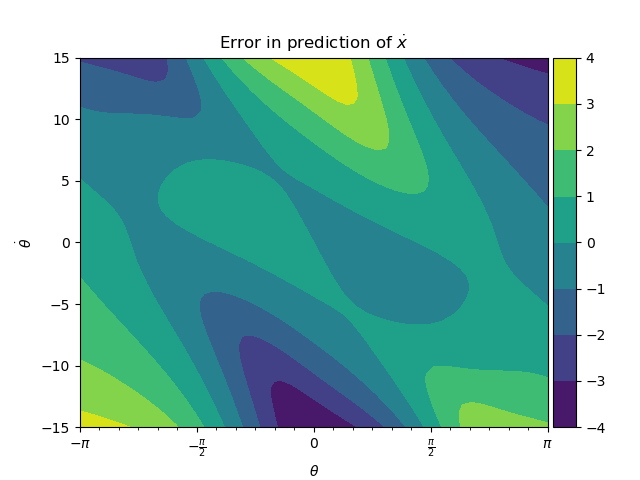
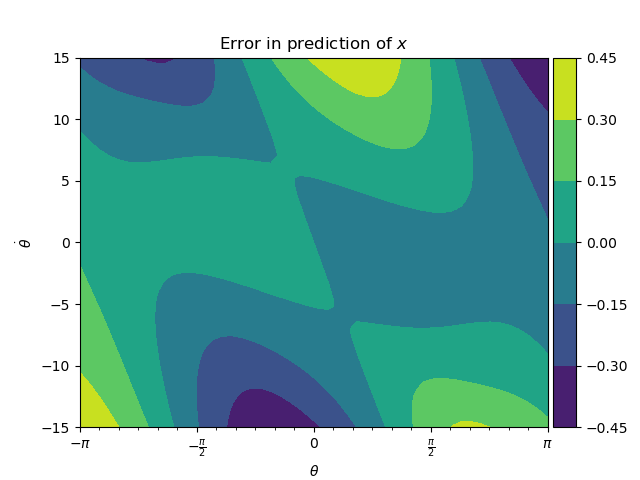
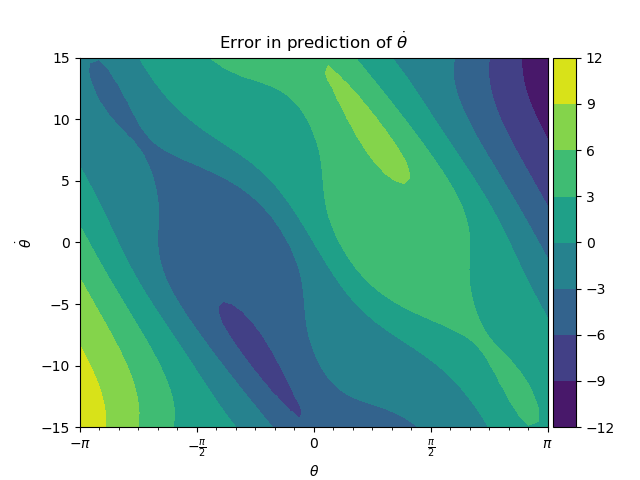
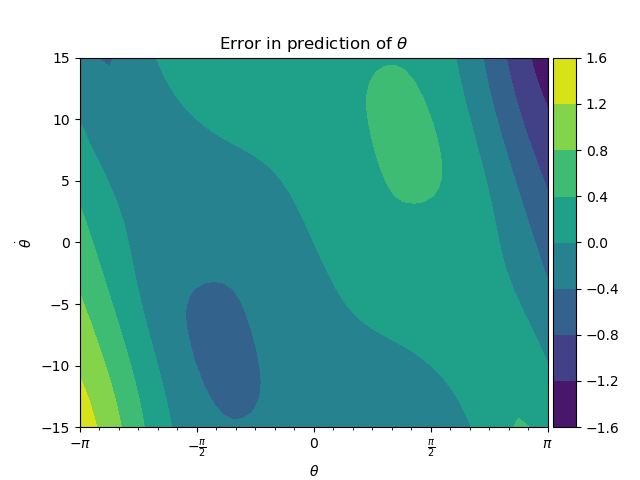
#### Expected vs target next step

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:

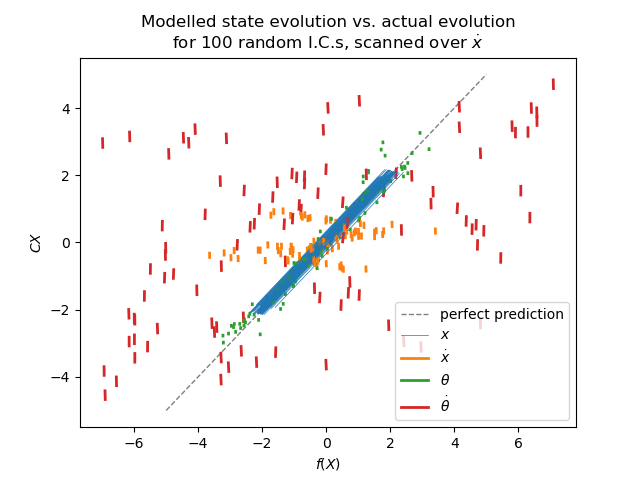
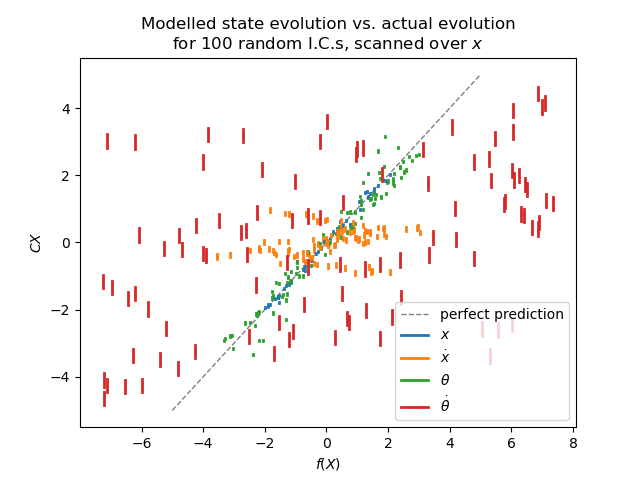
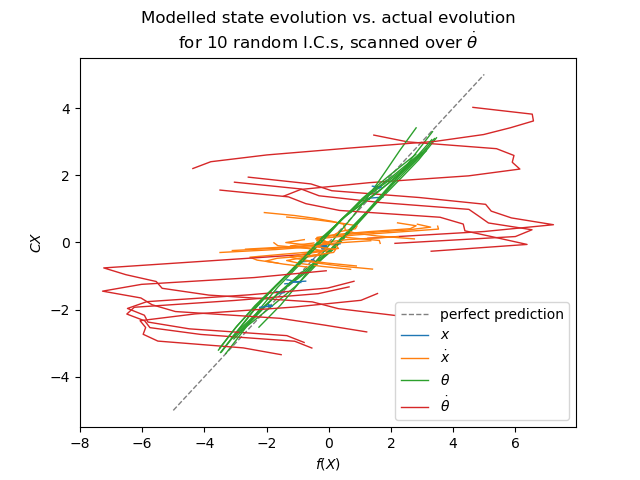
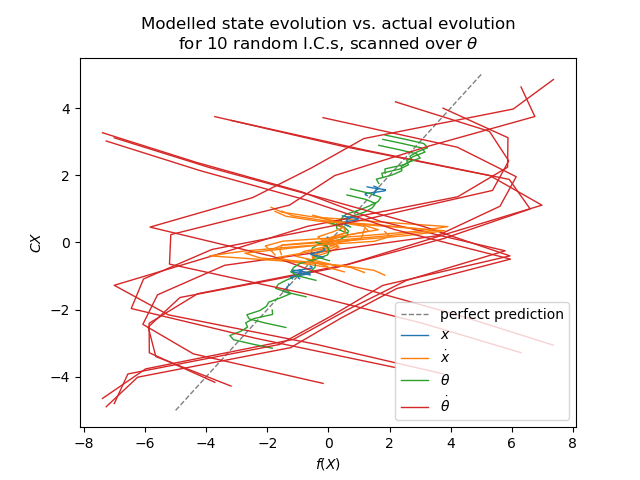
* Predictions of and are in general very good, while predictions of the velocities are much poorer.
* The model correctly demonstrates independence of and on and (hence their lines are points).
* The velocities are slightly more sensitive to and than the positions, but are still fairly independent.
* and 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 and , showing how well the linear model fits to fig. 8. Predictions are poor in general, and get worse as and increase. Unfortunately, since the stable equilibrium is at , the system will always end up in this region of state space.



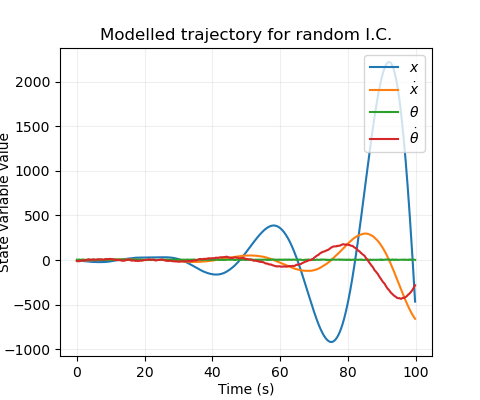
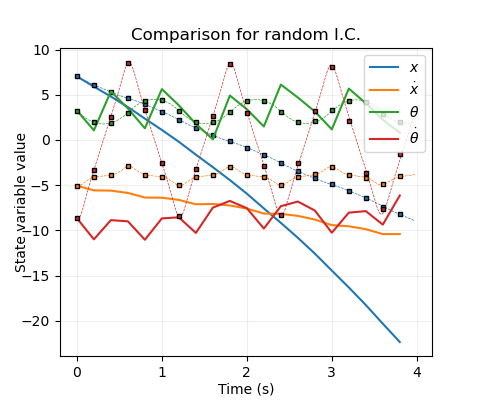
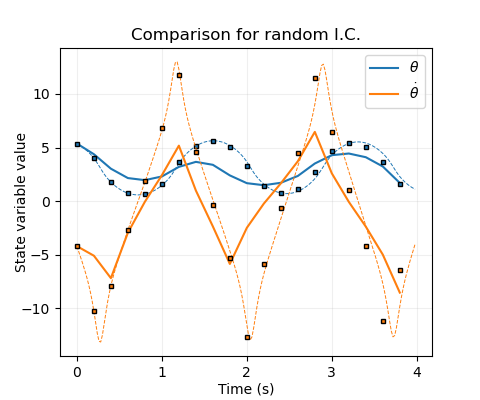
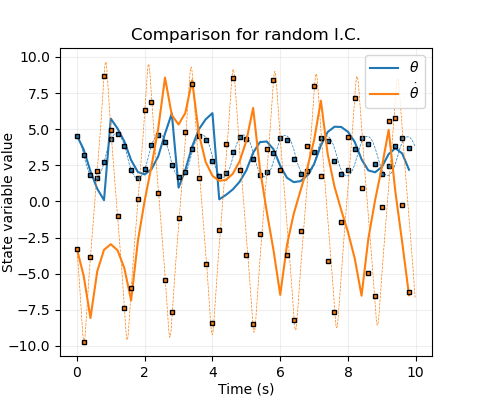
**Fig 8.** Contour plots of state variable prediction errors as functions of and .



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

## 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.



**Fig 5.** Contour plots of state variable evolutions as functions of and .

There are some situations when the model performs better. Some matrices, including the previously fitted one with 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.

# Nonlinear Fit

## Outline

To fit a good regression model to the update function, nonlinear kernel functions are required. For simplicity, diagonal-covariance Gaussians were chosen. Using a subset of training datapoints as kernel centres makes this correspond to a sparse Gaussian process. The solution can be regularised with hyperparameter to aid numerical stability.

Since Gaussians are better for fitting mean-zero fluctuations than linear trends, it’s best to fit the nonlinear model to the residuals of the linear model. (Note: This requires enforcing zero -dependence in the matrix to prevent periodicity artefacts.)

and , the number of training data points and the number of basis function centres, can be separately varied to give a trade-off of model complexity vs. fit, and the six hyperparameters , , , , can be optimised for given data.

## Hyperparameter Optimisation

#### Choosing initial values

First , data points were obtained with a Sobol sequence over the fitting region (same bounds as before).

From the plots of (.), rough length scale variations can be selected as where the first two are already well fit by the linear model (so do we need to evaluate them? Speed up computation?). will worsen the fit when it is large but shouldn’t have much effect when it is small, as long as it’s large enough for numerical stability. Fit can be evaluated by generating new random data (not from the Sobol sequence to prevent systematic biases), and calculating root-mean-squared error.. This is validated in the plots of fig (.).

#### Numerical optimisation

The mean-squared error function does not appear complicated. Starting with the minima indicated in fig(.), the Nelder-Mead algorithm was employed

**Fig. (.). Scans through variables given N=2048, M=512.**

## Amount of data vs Fit.

and should be independent of the model’s optimal hyperparameters – the real function’s length scales don’t depend on what data is available.

Start with scale params = std devs

Set N, M = 1024, 512, scan through lambda.

Show error decreases monotonically with M, N, with max reasonable computation time at M=1024.

Show asymptote of increasing N with M=1024.

Scan through scale params.

Rationalise approx. locations.

Do local search for 3 with minima – use Nelder-Mead method?

Comment that lambda/error w/M, N plots remain mostly unchanged.

Decide that when going for different performances, should use N = 4M

## Assessing single-step fit.

Single-variable scans.

Scatter plots.

## Assessing rollout

A few matching trajectories.

Explain scoring system.

Show score vs M/N. Explain that main issue is with transition to bound, reference trajectory.

# Linear controller

## Visualisation

Simplified picture. What if we just care about theta and theta dot? What if we just care about x and x dot?

Strategy:

1. Optimise for x, x dot.

2. Build controller on top for theta, theta dot.

1a: global scans for 0.5, 0.5, 1000, 1000

1b: Clearly fixed value for xdot. What relative value for x?

Only consider small displacements or force will max out and time window will be large (assume smaller perturbations than this). Should still expect to stabilise for larger initial displacements, just take longer. Therefore also just consider small x dot loss scale. X loss scale then determines relative punishment of velocity. Doesn’t matter that much tbh.

Optimise?

Strategy:

Optimise only caring about cart position.

Optimise only caring about angle position.

If we can allow a wide range of values… Split into two timescales? (first swing up to theta, then start controlling x & x dot, and finally start to bring the timescales together?) -> explore longer loss window to care about cart position

Strategy final: Broad idea – start with simple problem which encompasses hard problem, and restrict solutions until solves hard problem. Make problem as easy as possible so solutions can be found. Once solution(s) found, optimise. Can then make problem harder and try to optimise in same region.

Start with simple problem of balancing upright.

Objectives:  
Find broad region of P-space in which to expect good minima.

Set loss parameters, test ICs, and number of runs so that the function is smooth enough and with small enough noise (from #runs) to optimise on easily, but still has a good optimum.

In each case start with loss parameter large to discover initial solutions then reduce it.

Theta, theta dot plot:

* Find minimum, with weak loss function and very small initial condition range.
* Set loss scale parameters and initial condition range to give a large and wide basin of convergence.
* (Check robustness and evolution of basin by increasing IC range – it gets narrower.)
* (Range turns out to roughly be 10-20, 0-5)

Including x dot:

* Should only restrict existing range.
* Start with weak loss function and small IC range. Scan to give set of contour plots. Find rough region of best convergence. [.1 and 5 gives good basins]
* (Check robustness/how basin changes by increasing IC range.)
* Zero full basin roughly -.5-4.0, 10-20, 1-4 (IC up to .2ish for all)
* Zero main part of basin roughly 0-2, 10-20, 1-3

Including x, searching within zero score region.

* Set theta dot = 1.8 + .5x dot
* As above.
* Broad basin then -2 to 2 for x
* Narrow basin…

zero velocity IC, then with range for robustness.

Extend this to include x, which shouldn’t be tricky.

Optimisation:

* Simulated annealing.
* Callback function: when < maybe 0.1, increase the IC range by 0.001\*P\_RANGE or similar.

Testing basin of attraction automation:

* Given optimum, set IC range to 0 and then increase single variable.
* Do this for each variable.
* Plot lines. Set threshold at 99% success rate.
* Given thresholds, increase all IC param ranges linearly proportional to threshold value and do a final plot.

Need to optimise on the model data!

COMPARE NONLINEAR MODEL @ with np.dot.

DISCUSSION:

A potential optimally-fast search could involve:

Optimise over bad model and easy ICs.

After termination, improve model and widen ICs.

# Linear Control

## Outline

#### Problem formulation

The simplest way to control the cart is with a force that is a linear function of the state variables:

To find an effective policy, a score must be given to each set of policy parameters based on how well it stabilises the cart. This is defined as the loss function

With scale parameters for each state vector component, a window of iterations, and target state . Although a linear controller is not expected to stabilise initial states far from the equilibrium, there is no harm in also introducing periodicity to the loss function in the same way as for the nonlinear model.

The policy should ideally be:

1. **Robust:** stabilise over a wide range of initial states.
2. **Fast:** stabilise reasonably quickly.
3. **Bounded:** Avoid visiting extreme states, i.e. states outside the modelling range.

The loss of a *policy* should then be defined as evaluated over a relatively small number of iterations (to ensure goal 2), evaluated as an average value over a number of random perturbed initial states (to ensure goal 1). If the trajectory diverges outside of the modelling range, the loss should return its maximum value of 1 (satisfying goal 3). Since restricting the time allowed for convergence reduces robustness, was set to 20. (Note that the nonlinear model also only agrees with the real system for a few cycles). – inform with this.

#### Strategy

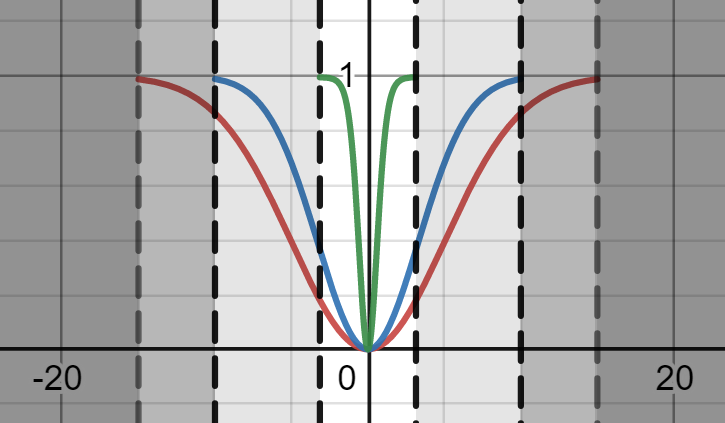
The global minimum basin in -space giving good policies is likely to be very small, possibly hard to find, and one of multiple local minima, even when loss is evaluated using the true system dynamics. When using the modelled dynamics, whose deviations from the true evolution are nontrivial, the loss landscape will be more complex and the global minimum may not coincide with the previous. These problems will reduce in severity as the fit converges (with high numbers of basis functions).

In a real-world situation, the goal is then to find a fit good enough that it can be used to obtain a robust linear policy, while having few enough basis functions that evaluating the function during optimisation is computationally feasible. Here, this will be explored:

1. Optimising using the true dynamics to find ideal policies
2. Optimising using modelled dynamics for various model sizes
3. Analysing quality and robustness of policies obtained using these models.

## True dynamics optimisation

This is still a hard problem with a small region of good policies and large region of terrible policies. To find the former,

* Loss scales should be as large as possible (to encourage a more gradually-sloping loss landscape, assisting with locating minima) while being small enough to punish bad policies. Scales were chosen to have when each component is at its range limit and the others are zero ().
* The problem was split into easier problems, the solutions of which were nested supersets of the overall solution: the penalties for and were originally neglected, before being reintroduced (in that order).

Noise

Main things to investigate:

Gaussian plus observations

Gaussian plus forces

## Linear model

#### Convergence with noise

As N increases, stabilises. Regardless of noise amount, N~512 marks approximate convergence.

However, by plotting against noise fraction, and standard deviation of the data itself (typical variance amount):

Nonlinear Controller.

Brainstorm…

3-phase control:

* Drive to the correct energy contour,
  + Some energy in the cart though
  + But can use energy \_in pole\_ as guide/or roughly use energy?
* Stabilise about the pole,
* Migrate back to the start position.
* Assuming IC range is below E=+small amount, no need for third step here.

Simple -> hard problem:

* Start without caring about cart position or velocity loss
* Then introduce cart velocity loss
* Finally introduce position loss

Superpose on linear controller?

* It takes over when E close to zero and near-upright. Could have some handing-over function (linear controller) + (nonlinear controller). Where =current value of loss function or similar.

# Appendix

NOTE – updated equations still missing a factor of L given to gabor. Corrected here, I think.

## A1: Energy definition

Let be kinetic energy, be potential energy with zero reference at , and be total energy.

and , so

## A2: Equations of motion

Defining the equations of motion are

## 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]]