Contrasting Discrete and Continuous Time Methods for Bayesian System Identification

UNIVERSITY OF CAMBRIDGE

Talay M Cheema (tmc49@cam.ac.uk) Carl Edward Rasmussen

computational &biological learning laboratory

Background

- Goal: infer the dynamics function f_{DT} or f_{CT} given measurements y (system identification)
- Many physical or biological systems are modelled in continuous time (CT) as stochastic differential equations
- But learning in CT can be hard (high computational cost, numerical errors), so it is tempting to model in discrete time (DT)
- We model $f \sim \mathcal{GP}(m, k)$ and approximate the posterior $p(f|y)$, so we can propagate meaningful uncertainty estimates for downstream tasks, such as active learning
- If we do this, there is a mismatch between the prior and model – we investigate how problematic this is

$$x_{t+\delta} = f_{DT}(x_t) + L_{DT}\kappa_t \quad \text{or} \quad dx_t = f_{CT}(x_t)\, dt + L_{CT}\, d\beta_t$$
$$y_i = g(x_{t_i}) + \rho_i$$

- The latent state is $x_t \in \mathbb{R}^D$ over time $t \in \mathbb{R}$, measurements $\{y_i \in \mathbb{R}^\Delta\}_{i=1}^n$ at some particular times $\{t_i\}_{i=1}^n$
- Each $\kappa_t \sim \mathcal{N}(0, I)$ independently, each $\rho_i \sim \mathcal{N}(0, R)$ independently, and β_t is standard D-dimensional Brownian motion

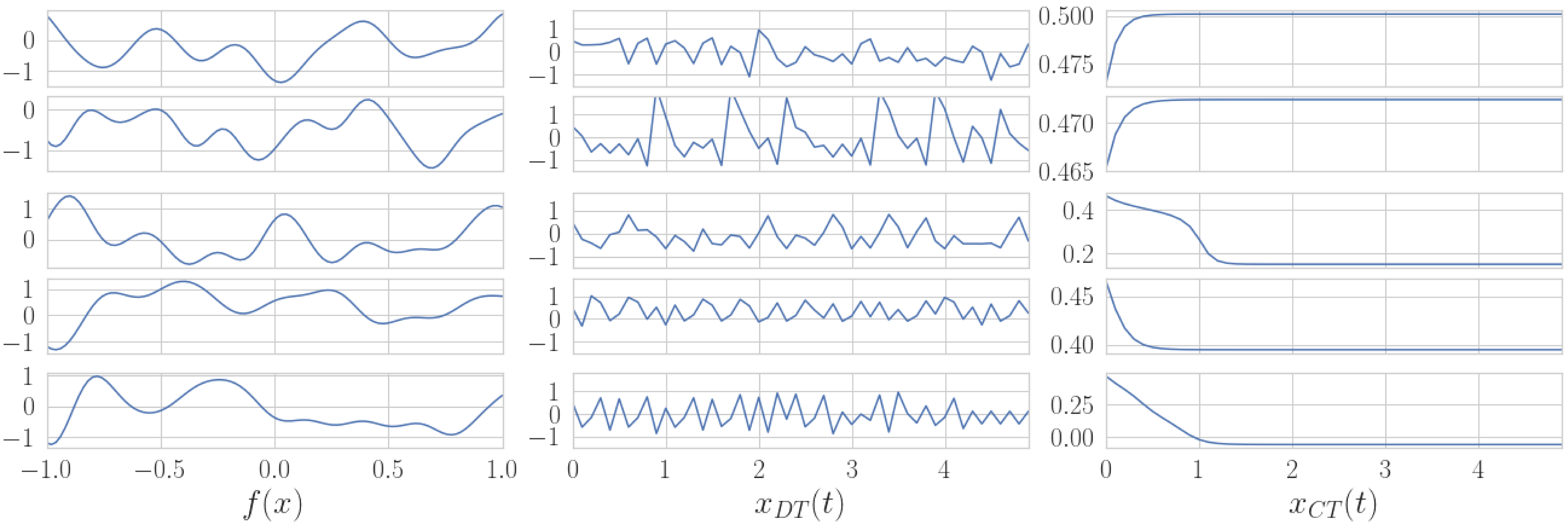

Figure 1: Prior samples from a GP (left) and corresponding latent trajectories from a DT and CT GPSSM.

Contrasting priors

- Compare prior samples from the DT and CT models in Figure 1
- Trajectories from CT models do not cross over, but from DT models they may
- In 1D, this means CT trajectories all converge to an equilibrium, whereas DT trajectories may have much more varied behaviour
- These types of differences disappear if f_{DT} is a diffeomorphism
- But we do not know a good way of cosnstructing diffeomorphic priors without essentially creating a CT system!
- If each co-ordinate of f_{CT} depends on only a few elements of x, this structure is generally lost when computing the equivalent DT transition (harder to infer causal structures – this would persist even if f_{DT} were a diffeomorphism)

Contrasting learnt posteriors

- To show this matters in the full learning problem, we fit data from a van der Pol oscillator (Figure 2)
- The approximate posterior for f and x are optimised with respect to the standard variational lower bound on the log marginal likelihood
- In moderate observation noise (left hand panels) both models learn fairly well, though the CT model does slightly better
- In high observation noise (right hand panels) both struggle, but the CT model is more robust
- Note that the DT model ends up with a self-intersecting trajectory – which is not possible in CT
- Using the correct prior can be significant in challenging scenarios, but if we could construct a diffeomorphic DT prior, that would be a competetive alternative

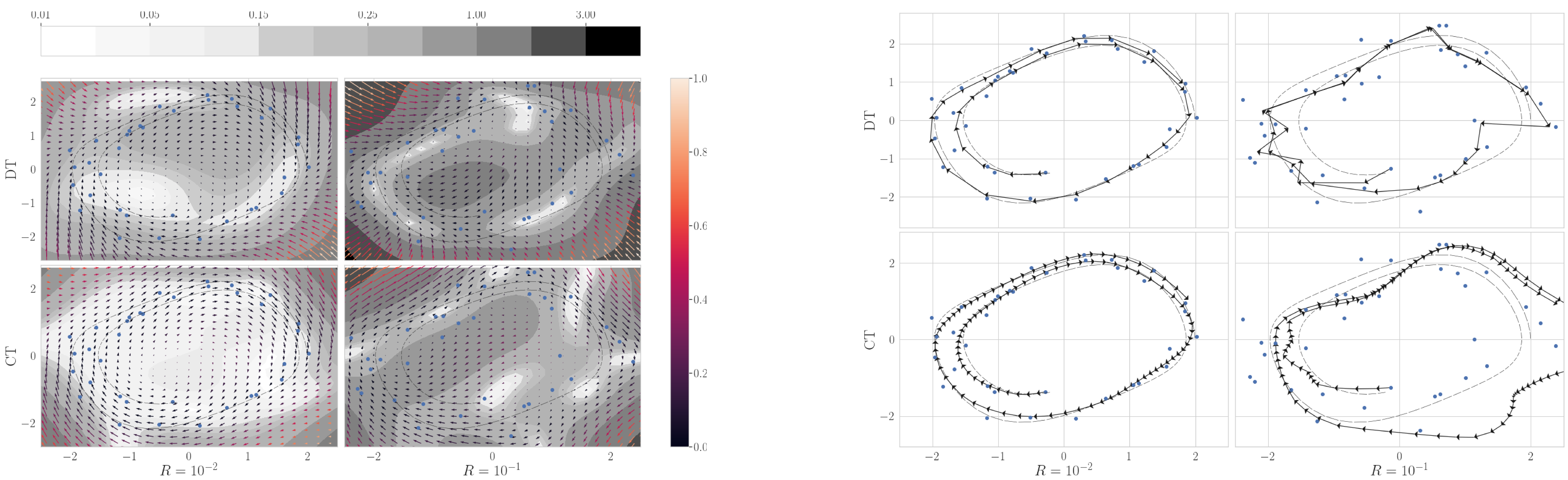

Figure 2: Learnt posteriors. Left: learnt transition functions as arrows, coloured according to standard deviation. The background shading is the RMSE between the learnt and grountruth function. Right: latent mean trajectory. In both, the groundruth latent trajectory is faint and dashes, and the measurements are blue dots.