Learning Unknown ODE Models with Gaussian Processes

Markus Heinonen[1,2], Cagatay Yildiz[1], Henrik Mannerström[1], Jukka Intosalmi[1], Harri Lähdesmäki[1]

[1]Aalto University, Finland [2]Helsinki Institute for Information Technology, HIIT, Finland

Aalto University
School of Science

1 Motivation and Model

- We consider multivariate ODEs of the form

$$\dot{\mathbf{x}}(t) = \frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t)) \qquad \mathbf{f} : \mathbb{R}^D \to \mathbb{R}^D$$

- ODE solution $\mathbf{x}(t)$ at time t is computed by

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t \underbrace{\mathbf{f}(\mathbf{x}(\tau))}_{\dot{\mathbf{x}}(\tau)} d\tau$$

☺ We propose **nonparametric** estimation of **arbitrary** vector fields from the data.

☺ We optimize against **full forward simulation** of the system.

- Gaussian process prior over vector field

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, K(\mathbf{x}, \mathbf{x}')), \qquad K(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{D \times D}$$

- Inducing points and vectors

$$Z = (\mathbf{z}_1, \ldots, \mathbf{z}_M)$$
$$U = (\mathbf{u}_1, \ldots, \mathbf{u}_M) = (\mathbf{f}(\mathbf{z}_1), \ldots, \mathbf{f}(\mathbf{z}_M))$$

- Kernel interpolation

$$\mathbf{f}(\mathbf{x}|Z, U) \triangleq \mathbf{K}_\theta(\mathbf{x}, Z)\mathbf{K}_\theta(Z, Z)^{-1} vec(U)$$

2 Previous Work

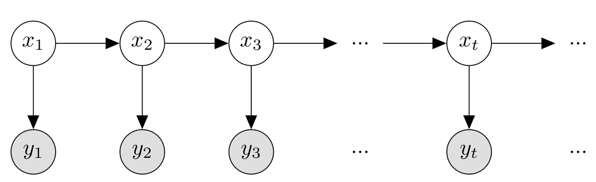
☹ Assumes known parametric form of $\mathbf{f}(\mathbf{x}(t), t)$

$$\textsc{Lotka-Volterra}: \quad \dot{x}_1 = \alpha x_1 - \beta x_1 x_2, \qquad \dot{x}_2 = \delta x_1 x_2 - \gamma x_2$$

☹ Based on matching empirical gradients

$$\mathbf{f}(\mathbf{y}(t_i)) \approx \mathbf{y}(t_{i+1}) - \mathbf{y}(t_i)$$

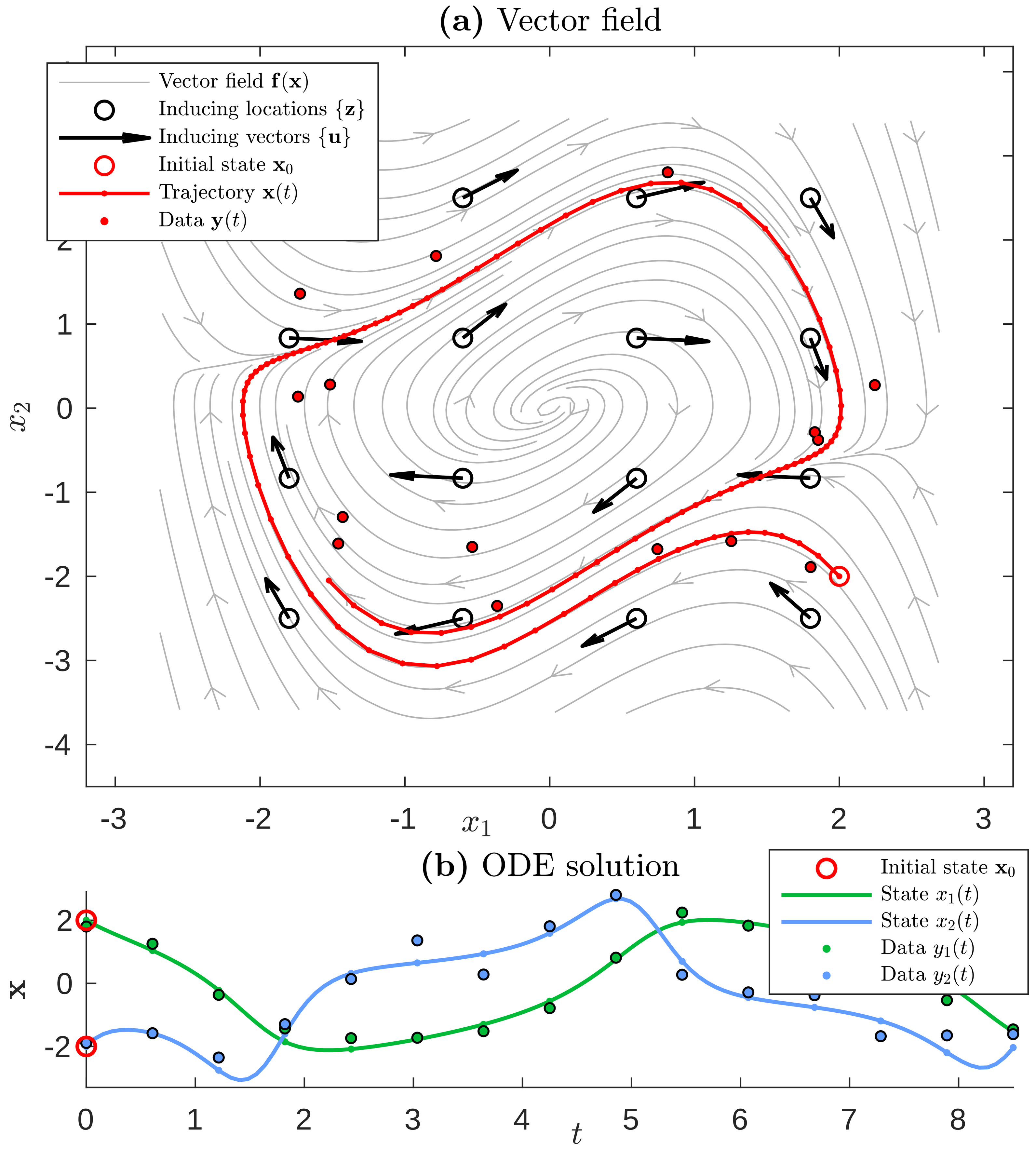
☹ Discrete time systems



Contact: Markus Heinonen - markus.o.heinonen@aalto.fi

Code available at https://github.com/cagatayyildiz/npode

3 Nonparametric ODE Model



4 Likelihood

- Gaussian noise model given observations $Y = \{\mathbf{y}(t_i)\}_{i=1}^N$ at observed time points t_i:

$$\mathbf{y}(t_i) = \mathbf{x}(t_i) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}, \Omega)$$

- Likelihood:

$$p(Y|\mathbf{x}_0, U, \Omega) = \prod_{i=1}^N \mathcal{N}\Big(\mathbf{y}(t_i) \mid \underbrace{\mathbf{x}_0 + \int_0^{t_i} \mathbf{f}_U(\mathbf{x}(\tau)) d\tau}_{\boxed{\mathbf{x}_U(t_i)}}, \Omega\Big)$$

References

[1] P. Kokotovic and J. Heller, "Direct and adjoint sensitivity equations for parameter optimization," *IEEE Transactions on Automatic Control*, 1967.

[2] J. Wang, A. Hertzmann, and D. M. Blei. "Gaussian process dynamical models," *NIPS*, 2006.

[3] A. Damianou, M. K. Titsias, and N. Lawrence. "Variational gaussian process dynamical systems," *NIPS*, 2011.

5 Sensitivities

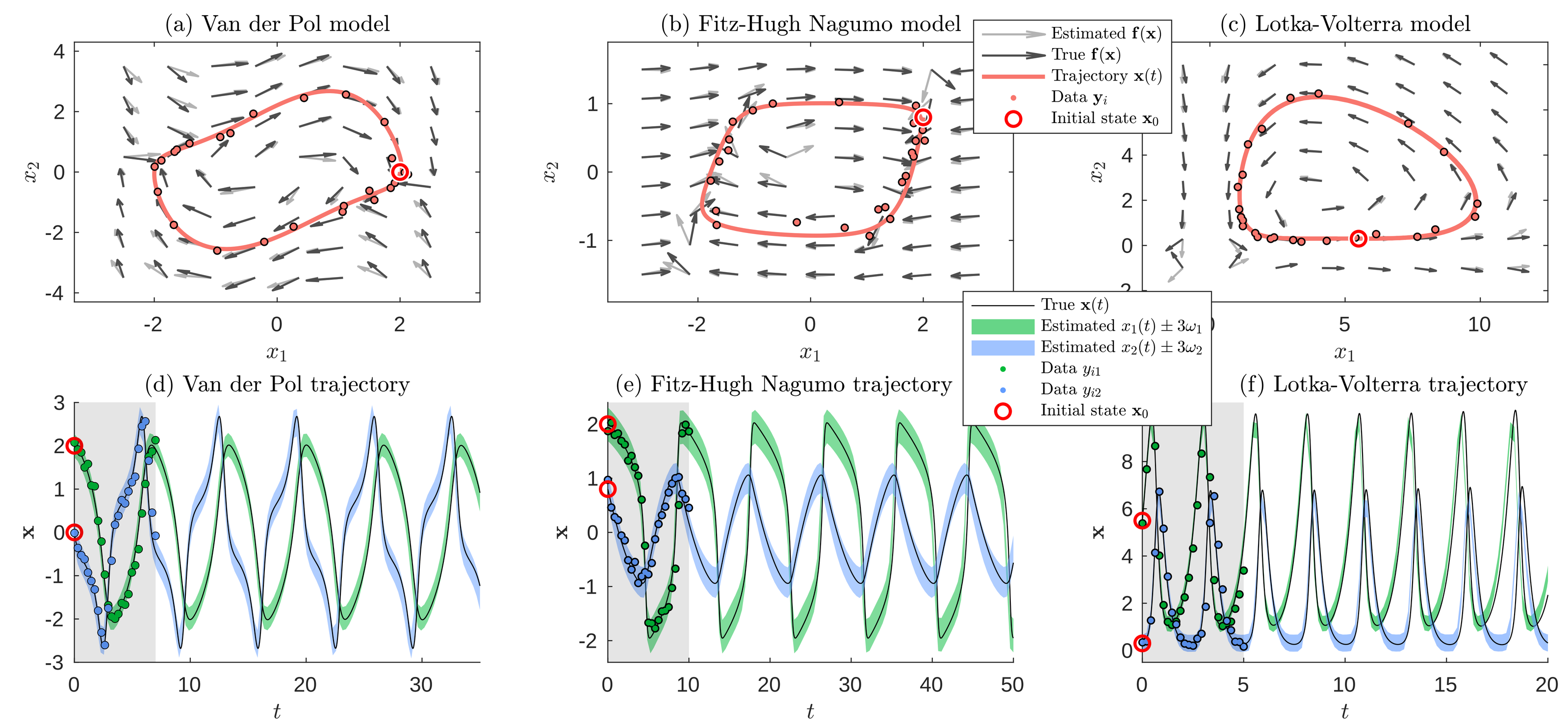
Derivative of the likelihood requires $\boxed{\frac{d\mathbf{x}_U(t)}{dU}} \equiv S(t)$.

$$\dot{S}(t) = \frac{d}{dt}\frac{d\mathbf{x}(t,U)}{dU} = \frac{d}{dU}\frac{d\mathbf{x}(t,U)}{dt} = \frac{d\mathbf{f}(\mathbf{x}(t,U),U)}{dU}$$

Total derivative of the right hand side gives [1]

$$\overbrace{\frac{d}{dt}\frac{d\mathbf{x}(t,U)}{dU}}^{\dot{S}(t)} = \overbrace{\frac{\partial\mathbf{f}(\mathbf{x}(t,U),U)}{\partial\mathbf{x}}}^{J(t)}\overbrace{\frac{d\mathbf{x}(t,U)}{dU}}^{S(t)} + \overbrace{\frac{\partial\mathbf{f}(\mathbf{x}(t,U),U)}{\partial U}}^{R(t)}$$

6 Simulated Oscillators



7 50-dimensional CMU Walking Data

50-dimensional walking sequences are projected to 3D manifold by PCA, where inference is performed. GPDM[2] and VGPLVM[3] optimizes the latent representation.