System Identification in Gaussian Process Dynamical Systems

UNIVERSITY OF CAMBRIDGE

Computational and Biological Learning

Ryan Turner, Marc Peter Deisenroth, and Carl Edward Rasmussen

Abstract

- Inference and learning (system identification) in GP state-space models
- EM for learning parameters of GP dynamics and measurement models
- Referred to as Gaussian process inference and learning (GPIL)

1 Introduction

Consider

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \boldsymbol{\epsilon}_t\,, \quad \mathbf{y}_t = g(\mathbf{x}_t) + \boldsymbol{\nu}_t\,, \tag{1}$$

- GP models for both the transition function f and measurement function g
- No ground truth observations of the latent states $\mathbf{x}_t$
- Handles system noise, measurement noise, and model uncertainty

Main contributions:

- Tractable algorithm for approximate inference (smoothing) in GP state-space models.
- Learn parameters of the GP models for f and g without ground-truth observations $\mathbf{x}_t$ of the latent states.

2 Pseudo Training Sets

Use pseudo training sets like in sparse GP approximations, Fig 1 and 2.

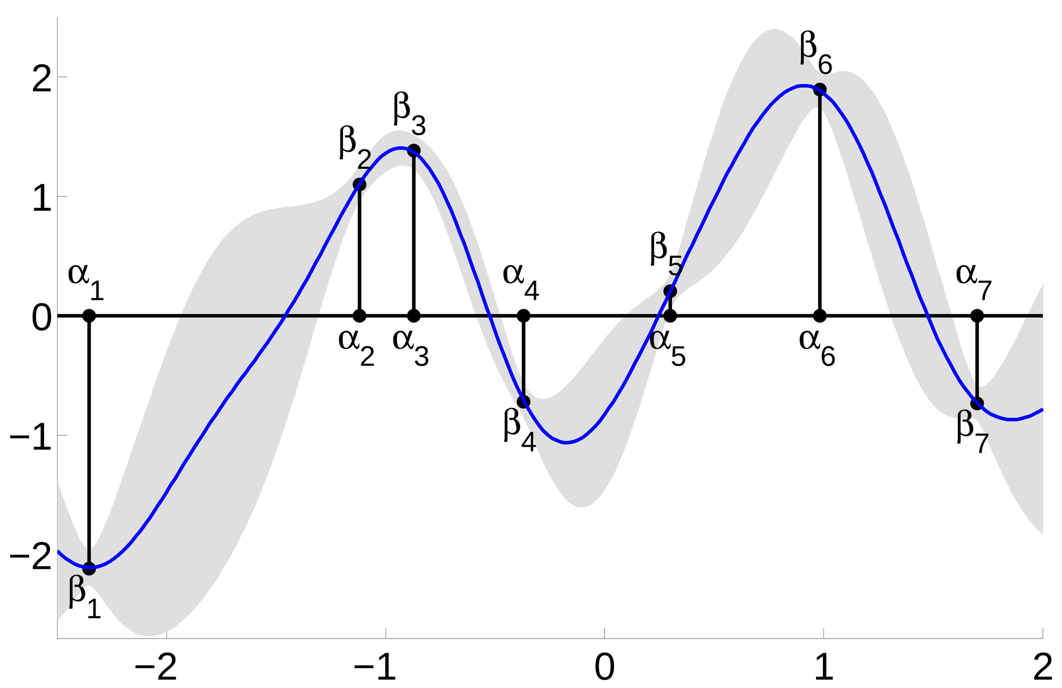

FIGURE 1: An example of a function predicted by a set of support points. The $\boldsymbol{\alpha}_i$ are the pseudo training inputs, while the $\boldsymbol{\beta}_i$ are the pseudo training targets. The shaded area represents the 95% confidence region around the expected function value (blue, solid).

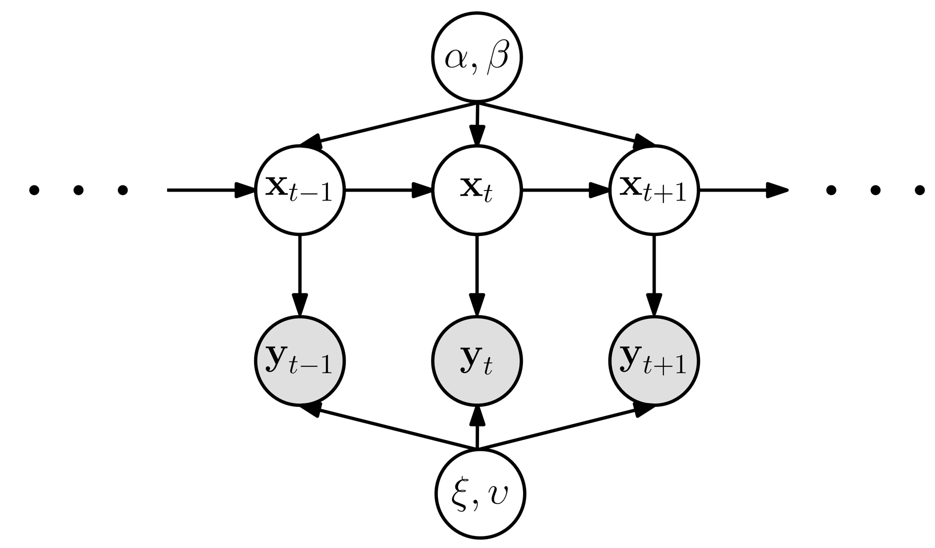

FIGURE 2: The free parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\xi}, \boldsymbol{\upsilon}$ serve as a pseudo training set for $\mathcal{GP}_f$ and $\mathcal{GP}_g$, respectively. $\mathcal{GP}_f$ and $\mathcal{GP}_g$ are *not* full GPs, but rather sparse GPs that impose the condition $\mathbf{x}_{t+1} \perp \mathbf{x}_{t-1} | \mathbf{x}_t, \boldsymbol{\alpha}, \boldsymbol{\beta}$ during predictions.

3 System Identification with EM

- Inference (E-step): we extend the filtering algorithm of [1] to smoothing in GP state-space models
- Learning (M-Step): we seek the parameters Θ that maximize the likelihood lower bound $Q = \mathbb{E}_{\mathbf{X}}[\log p(\mathbf{X}, \mathbf{Y}|\Theta)]$

$$Q = \mathbb{E}_{\mathbf{X}} \log p(\mathbf{x}_1|\Theta) + \sum_{t=2}^{T} \underbrace{\log p(\mathbf{x}_t|\mathbf{x}_{t-1}, \Theta)}_{\text{Transition}} + \sum_{t=1}^{T} \underbrace{\log p(\mathbf{y}_t|\mathbf{x}_t, \Theta)}_{\text{Measurement}}\,. \tag{2}$$

4 Results

Synthetic data: Tested method on sinusoidal dynamics function. The results were produced using a pseudo training set of size $N = 50$, $T = 100$ training observations and 10,000 test observations.

- Accurately learns the true dynamics and measurement model, Fig 3 and 5

Real data: We used historical snowfall data in Whistler, BC, Canada The models were trained on two years of data; the GPIL used a pseudo training set of size $N = 15$; Tested on 35 years of test data.

- Learns close-to-linear dynamics and hinge function measurement model, Fig 4 and 5

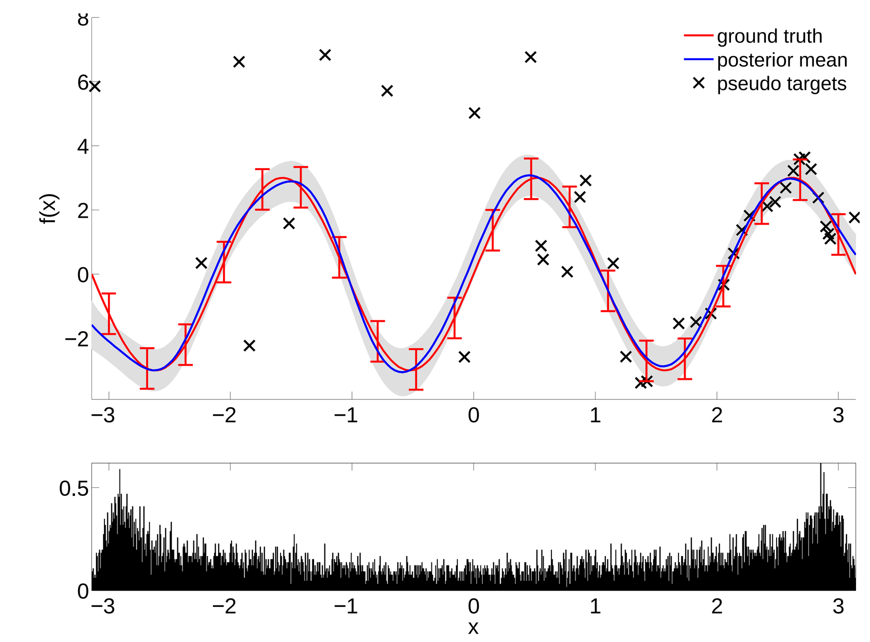

FIGURE 3: Sinusoid data: True (red) and learned (blue) transition function with histogram of the inputs x_t in the test set.

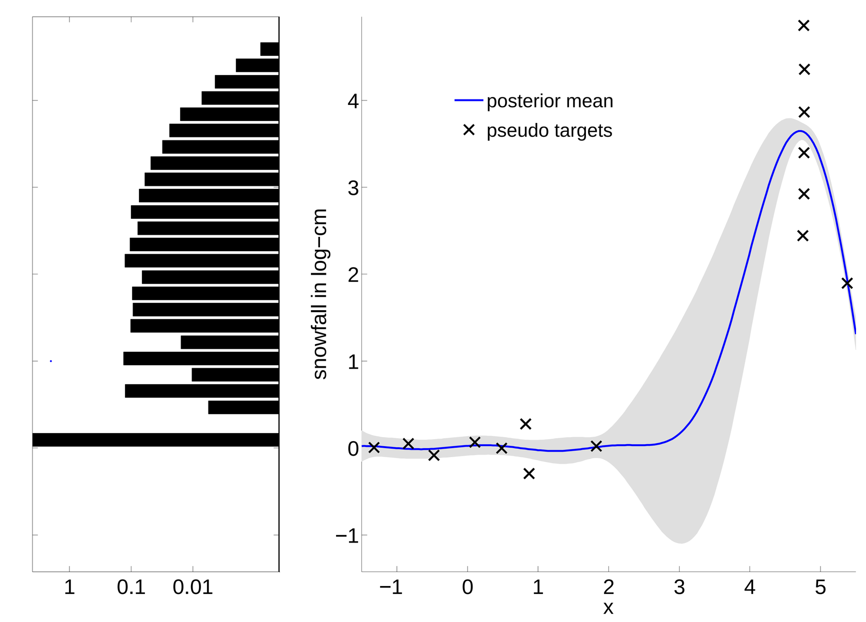

FIGURE 4: Snowfall data: Learned measurement GP (right) and log histogram of the observations (left) during testing, real data set.

FIGURE 5: GPIL and six other methods. Negative log likelihood (NLL) per data point and the RMSE are shown.

	Method	NLL synth.	RMSE synth.	NLL snow	RMSE snow
general	TIM	2.21	2.18	1.47	1.007
	Kalman	2.07	1.91	1.29	0.783
	ARGP	1.01	0.66	1.25	0.793
	NDFA	2.20	2.18	14.6	1.06
	GPIL ⋆	**0.917**	**0.654**	**0.684**	**0.769**
requires	UKF	4.55	2.19	1.84	0.938
prior	EKF	1.23	0.67	1.46	0.905
knowledge	GP-UKF	6.15	2.06	3.03	0.884

References

[1] M. P. Deisenroth, M. F. Huber, and U. D. Hanebeck. Analytic moment-based Gaussian process filtering. In *26th ICML*, pp. 225–232, Montreal, Canada, 2009. Omnipress.