Analytic Moment-based Gaussian Process Filtering

Marc Peter Deisenroth

Computational and Biological Learning

UNIVERSITY OF CAMBRIDGE

International Conference on Machine Learning (ICML)
Montreal, Canada

June 15, 2009

joint work with Marco F. Huber[1,2] and Uwe D. Hanebeck[1]

[1] Faculty for Informatics, Universität Karlsruhe (TH), Germany
[2] Fraunhofer Institute for Information and Data Processing, Germany

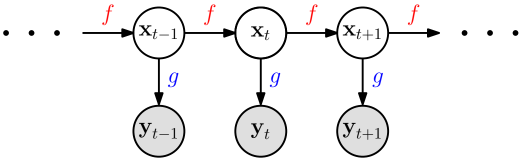
Problem setup

nonlinear state space model, sequential data:

transition function

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \mathbf{w}_t$$

hidden states



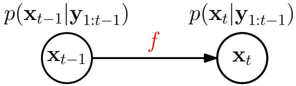
observed variables

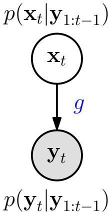
$$\mathbf{y}_t = g(\mathbf{x}_t) + \mathbf{v}_t$$

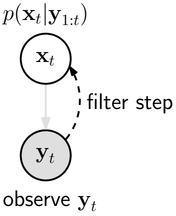
measurement function

objective: compute $p(\mathbf{x}_t|\mathbf{y}_{1:t})$: distribution of hidden state $\mathbf{x}_t$ given observations $\mathbf{y}_1, \ldots, \mathbf{y}_t$ (filter distribution)

Three steps for a filter update





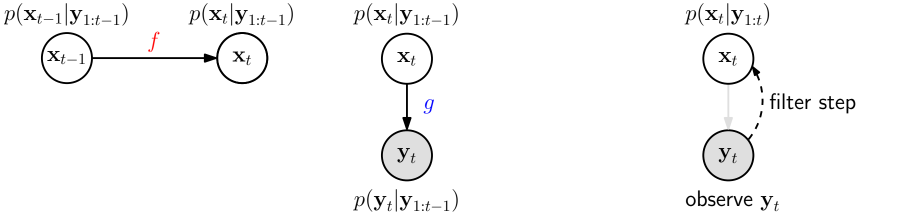


1) predict next hidden state

2) predict observation

3) update hidden state using evidence of new observation

Three steps for a filter update



1) predict next hidden state

2) predict observation

3) update hidden state using evidence of new observation

- transition dynamics f and measurement function g linear ➡ Kalman filter
- here: f and g nonlinear ➡ approximations required
- common assumption: $\mathbf{x}_t, \mathbf{y}_t|\mathbf{y}_{1:t-1}$ are jointly normal
  ➡ filter update (step 3) is a Gaussian conditional

➡ concentrate on predictions in the following

Some approximate algorithms for predictions

basic setup:

- Gaussian input distribution
- predictive distribution is approximated by a Gaussian

Some approximate algorithms for predictions

basic setup:

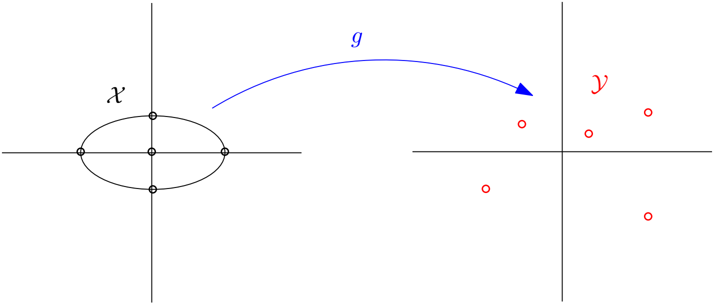
- Gaussian input distribution
- predictive distribution is approximated by a Gaussian

first idea: approximate the function

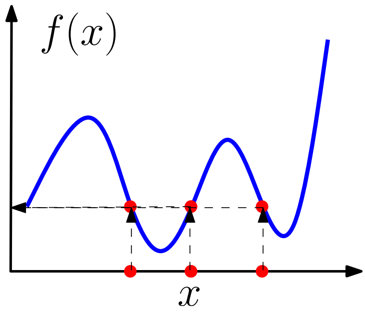
- Extended Kalman Filter (EKF)
  ➡ linearizes the function (Taylor series) and applies Kalman filter
  ➡ predictive distribution: exact for the linearized model
  ➡ requires parametric form of the function (derivatives)
- linear function approximation can be bad

2. Unscented Kalman filter (UKF)

- second idea: "approximating a distribution is often easier than approximating a function" (Julier and Uhlmann, 1997)
- approximate input distribution by finite number of sigma points (deterministically chosen "samples" / "particles")
- predictive distribution: distribution of sigma points after mapping them through original function
- requires a) access to the function, b) noise variance



Prediction problems in the UKF



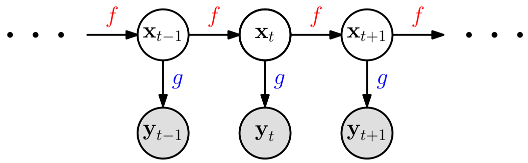
- does not preserve the exact predictive mean/covariance
- predictive distribution can be overconfident! (or too cautious)

3. Approximation in function space

transition function

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \mathbf{w}_t$$

hidden states



observed variables

$$\mathbf{y}_t = g(\mathbf{x}_t) + \mathbf{v}_t$$

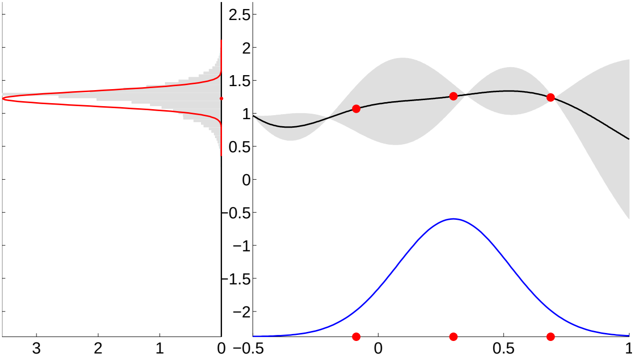
measurement function

approximate the function

- use Gaussian processes (GPs) to model the transition function f and the measurement function g
- no parametric model required, more flexible than linearization (EKF)
- can be used if the "true functions" f and g are no longer accessible
- additional source of uncertainty: model uncertainty

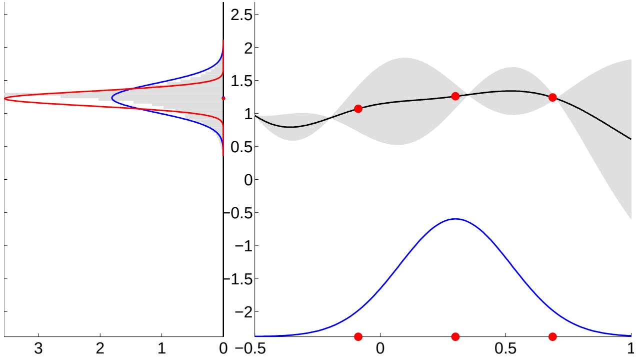
GP filters

1. GP-UKF (Ko and Fox, 2007–2009): predict by squashing sigma points through GP model (combine UKF with GPs); sample average of model uncertainty

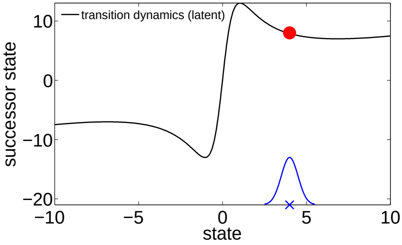


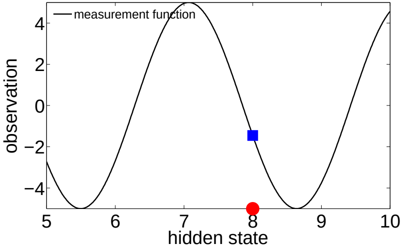
GP filters

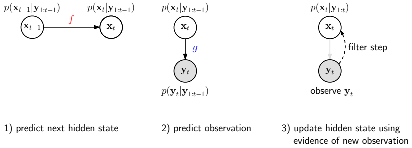
1. GP-UKF (Ko and Fox, 2007–2009): predict by squashing sigma points through GP model (combine UKF with GPs); sample average of model uncertainty
2. GP-ADF (our work): compute exact mean and covariance of predictive distribution (exact moment matching) according to Quiñonero-Candela et al. (2003); integrate out model uncertainty



Single filter step

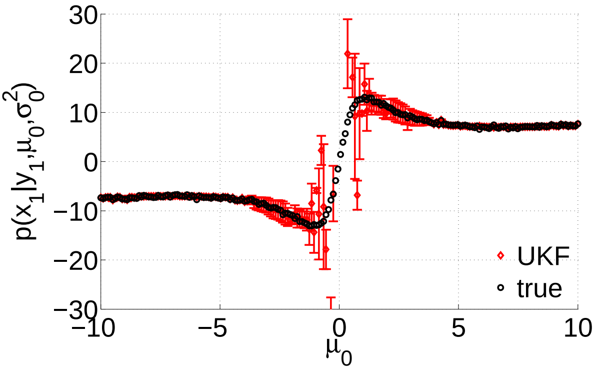






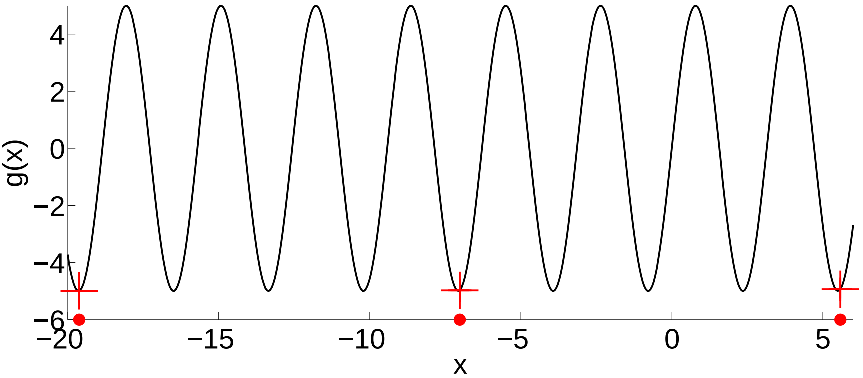
1. pick initial distribution
2. draw sample and map it through transition function f to ●
3. map ● through measurement function g
4. observe ■
   ➡ infer ● from initial distribution and ■

Single filter step



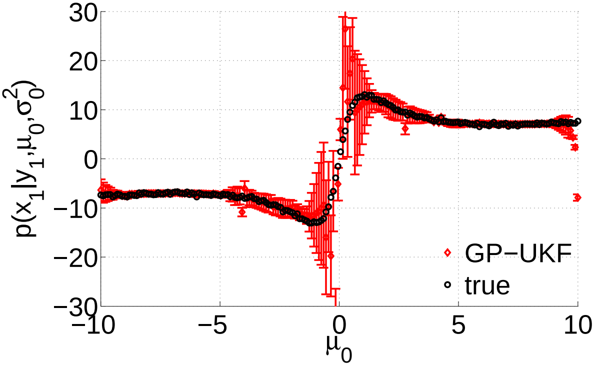
- given an initial state distribution $p(x_0) = \mathcal{N}\left(\mu_0,\, \sigma_0^2 = 0.5^2\right)$ compute $p(x_1)$ given an observation y_1

Single filter step



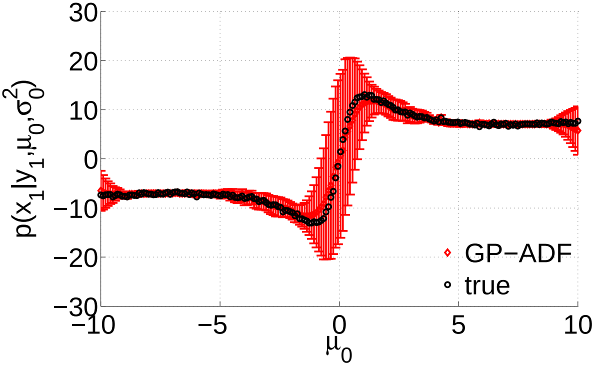
- given an initial state distribution $p(x_0) = \mathcal{N}\left(\mu_0,\, \sigma_0^2 = 0.5^2\right)$ compute $p(x_1)$ given an observation y_1

Single filter step



- given an initial state distribution $p(x_0) = \mathcal{N}\left(\mu_0, \sigma_0^2 = 0.5^2\right)$ compute $p(x_1)$ given an observation y_1

Single filter step



- given an initial state distribution $p(x_0) = \mathcal{N}\left(\mu_0,\, \sigma_0^2 = 0.5^2\right)$ compute $p(x_1)$ given an observation y_1

Wrap-up

summary

- GP-ADF: coherent filter algorithm for nonlinear state space models
- transition dynamics and measurement model are described by GPs
- prediction and filtering can be done analytically
- consistent predictions in contrast to UKF, GP-UKF
- current limitation: requires access to hidden state to train GP models

Wrap-up

summary

- GP-ADF: coherent filter algorithm for nonlinear state space models
- transition dynamics and measurement model are described by GPs
- prediction and filtering can be done analytically
- consistent predictions in contrast to UKF, GP-UKF
- current limitation: requires access to hidden state to train GP models

current projects:

- extension to smoothing
- parameter learning ➡ no need for ground truth observations in latent space

Wrap-up

summary

- GP-ADF: coherent filter algorithm for nonlinear state space models
- transition dynamics and measurement model are described by GPs
- prediction and filtering can be done analytically
- consistent predictions in contrast to UKF, GP-UKF
- current limitation: requires access to hidden state to train GP models

current projects:

- extension to smoothing
- parameter learning ➡ no need for ground truth observations in latent space

acknowledgement

- ICML student scholarship

References

- Simon J. Julier and Jeffrey K. Uhlmann.
  A New Extension of the Kalman Filter to Nonlinear Systems.
  In *Proceedings of AeroSense: 11th Symposium on Aerospace/Defense Sensing, Simulation and Controls*, pages 182–193, Orlando, FL, USA, 1997.

- Jonathan Ko and Dieter Fox.
  GP-BayesFilters: Bayesian Filtering Using Gaussian Process Prediction and Observation Models.
  In *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3471–3476, Nice, France, September 2008.

- Jonathan Ko and Dieter Fox.
  Gp-BayesFilters: Bayesian Filtering using Gaussian Process Prediction and Observation Models.
  Autonomous Robots, 2009.

- Joaquin Quiñonero-Candela, Agathe Girard, Jan Larsen, and Carl E. Rasmussen.
  Propagation of Uncertainty in Bayesian Kernel Models—Application to Multiple-Step Ahead Forecasting.
  In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003)*, volume 2, pages 701–704, April 2003.