Integrated Pre-Processing for Bayesian Nonlinear System Identification with Gaussian Processes

Roger Frigola

University of Cambridge
Machine Learning Group

12th December 2013

Bayesian Nonparametric Nonlinear System Identification

Bayesian inference:

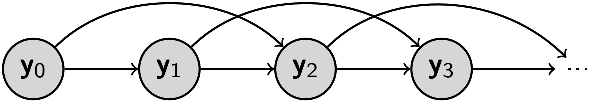
- Formally deals with all sources of uncertainty.
- Avoids overfitting.
- Provides predictions with error bars.

Bayesian nonparametrics:

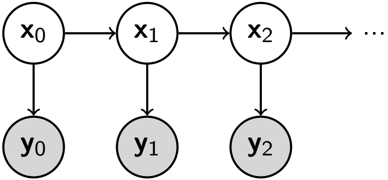
- Highly flexible - doesn't exclude complex models a priori.

Probabilistic Models of Time Series

Auto-regressive model: straightforward but does not incorporate measurement noise.

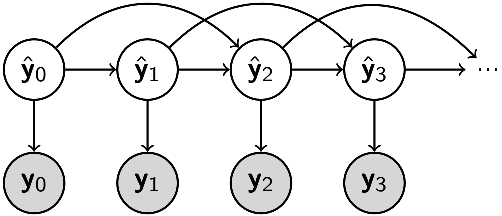


State-space model: incorporates measurement noise but inference for nonlinear models is hard.



Our Model

Non-Markovian model with hidden states with approximate smoothing.



Use your favourite pre-processing step (e.g. low-pass filter) to obtain

$$\hat{y}_{0:T} = h(y_{0:T}, \omega)$$

We chose h to be a cheap point estimate of the posterior over hidden variables

$$p(\hat{y}_{0:T} \mid y_{0:T})$$

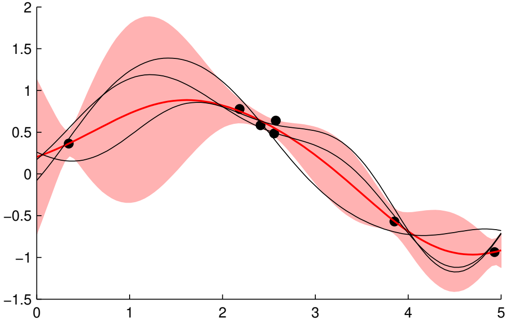
Learning the Model

We have a regression problem where the regressors are the pre-processed signals

$$y_t = f(\hat{y}_{t-1}, \hat{y}_{t-2}, ..., \hat{u}_{t-1}, ...).$$

Find a posterior distribution over $f(\cdot)$ using Gaussian process regression.

Gaussian Process Regression



Use FITC fast Gaussian processes (aka SPGP) [Snelson and Ghahramani, 2006] to scale to hundreds of thousands of data points.

Empirical Bayes procedure for training: jointly maximise the marginal likelihood of the data with respect to the hyper-parameters and ω.

Maximising the Marginal Likelihood

$$(\omega_{\text{Opt}}, \theta_{\text{Opt}}) = \arg\max_{\omega,\theta} \log p(y_{0:T} | \hat{X}(\omega), \theta)$$

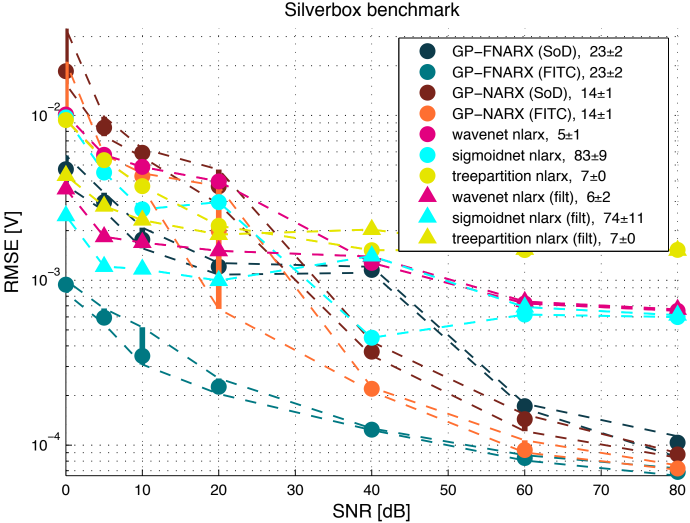
where $\hat{X}(\omega)$ denotes a matrix of filtered regressors.

The marginal likelihood results from integrating analytically the latent variables $f_{0:T}$

$$p(y_{0:T} | \hat{X}(\omega), \theta) = \int \underbrace{p(y_{0:T} | f_{0:T}, \theta)}_{\text{Likelihood}} \underbrace{p(f_{0:T} | \hat{X}(\omega), \theta)}_{\text{GP prior}} \, df_{0:T}$$

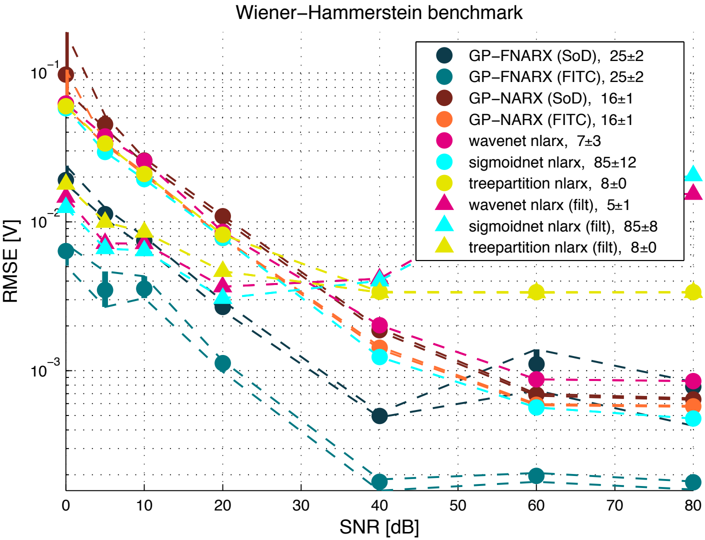
Experiments: Silverbox benchmark

Signals contaminated with different levels of Gaussian iid noise.



Experiments: Wiener-Hammerstein benchmark

Signals contaminated with different levels of Gaussian iid noise.



Conclusions

Practical Bayesian nonparametric nonlinear system identification for $> 10^5$ data points in a few seconds.

From data to model without human intervention.

Empirical Bayes automatically trades off model fit and model complexity.

Deals with measurement noise.

The user can select its own preferred data pre-processing method.