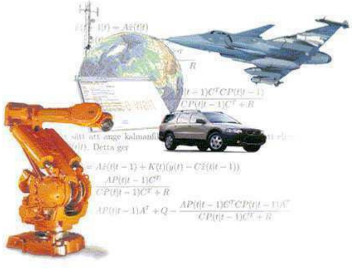
Learning Wiener models

– starting from the model

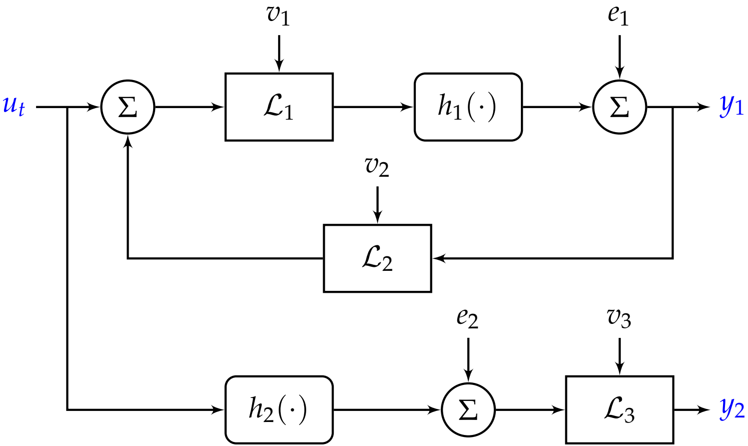


Thomas B. Schön

Division of Automatic Control
Linköping University
Sweden

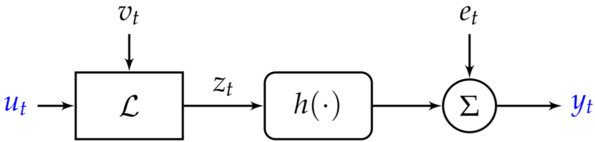
Joint work with (alphabetical order): **Michael I. Jordan** (UC Berkeley), **Fredrik Lindsten** (Linköping University), **Lennart Ljung** (Linköping University), **Brett Ninness** (University of Newcastle, Australia) and **Adrian Wills** (University of Newcastle, Australia).

Block-oriented nonlinear models are dynamic models consisting of interactions of linear dynamic models and static nonlinear models.



The Wiener model

We will study one notable member of the class of block-oriented nonlinear models, the **Wiener model**.



A Wiener model is a linear dynamical model ($\mathcal{L}$) followed by a static nonlinearity $(h(\cdot))$.

Learning problem: Find $\mathcal{L}$ and $h(\cdot)$ based on $\{u_{1:T}, y_{1:T}\}$.

Wiener model – explicit state space equations

Linear Gaussian state space (LGSS) model:

$$x_{t+1} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} + v_t, \qquad v_t \sim \mathcal{N}(0, Q),$$
$$z_t = C x_t.$$

Static nonlinearity:

- Parametric: $y_t = h(z_t, \beta) + e_t, \quad e_t \sim \mathcal{N}(0, R)$.
- Non-parametric: $y_t = h(z_t) + e_t, \quad e_t \sim \mathcal{N}(0, R)$.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Example of simplifying assumptions

Most of the existing work deals with special cases of the general problem. Typical restrictions imposed are:

- The nonlinearity $h(\cdot)$ is assumed to be invertible.
- The measurement noise e_t is absent.
- The LGSS model is deterministic (v_t is absent).
- The LGSS model is stochastic, but v_t is assumed white.

In the models and solutions provided here we do **not** have to make any of these assumptions.

Outline

1. Parametric model ($y_t = h(z_t, \beta) + e_t$)

2. Semiparametric models ($y_t = h(z_t) + e_t$)
   - The LGSS model is parametric, but the nonlinearity is modelled using a Bayesian nonparametric model (Gaussian process).
   - Data driven model: Same as above, but a sparseness prior (ARD) is placed on the dynamics.

Linear Gaussian state space (LGSS) model:

$$
\begin{aligned}
x_{t+1} &= \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} + v_t, & v_t &\sim \mathcal{N}(0, Q), \\
z_t &= C x_t, & & \\
y_t &= h(z_t, \beta) + e_t, & e_t &\sim \mathcal{N}(0, R).
\end{aligned}
$$

Maximum Likelihood (ML) amounts to solving,

$$\widehat{\theta}^{\text{ML}} = \arg\max_{\theta} \log p_\theta(y_{1:T})$$

where the log-likelihood function is given by

$$\log p_\theta(y_{1:T}) = \sum_{t=1}^{T} \log p_\theta(y_t \mid y_{1:t-1})$$

Where is the challenge?

Two challenges:

1. The one-step prediction pdf $p_\theta(y_t \mid y_{1:t-1})$ has to be computed,
2. In solving the optimisation problem

$$\widehat{\theta}^{\text{ML}} = \arg\max_{\theta} \; \log p_\theta(y_{1:T})$$

the derivatives $\frac{\partial}{\partial\theta} p_\theta(y_t \mid y_{1:t-1})$ are useful.

The **Expectation Maximisation (EM)** algorithm together with a **Particle Smoother (PS)** provides a systematic way of dealing with both of these challenges.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Rather than studying

$$\arg\max_{\theta} \; \log p_\theta(y_{1:T}),$$

the EM algorithm is concerned with

$$\arg\max_{\theta} \; \log p_\theta(y_{1:T}, x_{1:T}).$$

The **key idea** underlying EM is to consider the joint likelihood function of the observed variables $y_{1:T}$ and the latent variables $x_{1:T}$.

EM splits the original problem

$$\arg\max_{\theta} \; p_\theta(y_{1:N})$$

into two manageable (and closely linked) subproblems:

1. Compute a conditional expectation

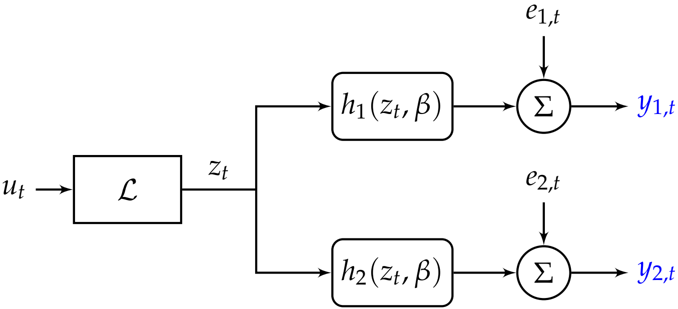
$$\begin{aligned}\mathcal{Q}(\theta, \theta_k) &= \mathbf{E}_{\theta_k} \left\{ \log p_\theta(x_{1:T}, y_{1:T}) \mid y_{1:T} \right\} \\ &= \int \log p_\theta(x_{1:T}, y_{1:T}) p_{\theta_k}(x_{1:T} \mid y_{1:T}) \mathrm{d}x_{1:T}\end{aligned}$$

R. Douc, A. Garivier, E. Moulines, and J. Olsson. **Sequential Monte Carlo smoothing for general state space hidden Markov models**. *Annals of Applied Probability*, 21(6):2109 2145, 2011.

2. Solve a maximisation problem

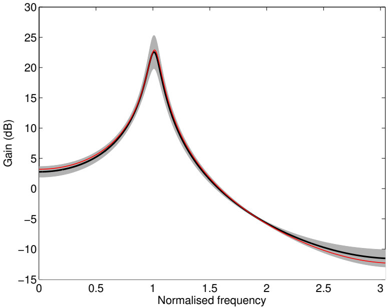
$$\theta_{k+1} = \arg\max_{\theta} \; \mathcal{Q}(\theta, \theta_k)$$

Parametric model – blind Wiener learning (I/III)



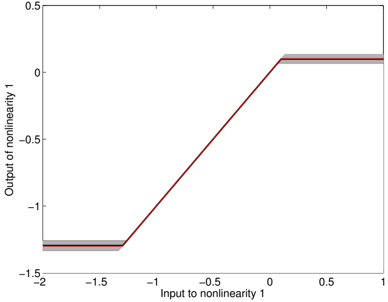
Learning problem: Find $\mathcal{L}$ and β, r_1, r_2 based on $\{y_{1,1:T}, y_{2,1:T}\}$.

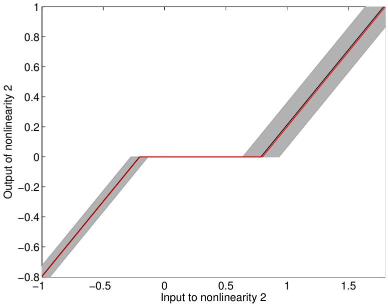
- Second order LGSS model with complex poles.
- Employ the EM-PS with $N = 100$ particles.
- EM-PS was terminated after just 100 iterations.
- Results obtained using $T = 1000$ samples.
- The plots are based on 100 realisations of data.
- Nonlinearities (dead zone and saturation) shown on next slide.



Bode plot of estimated mean (black), true system (red) and the result for all 100 realisations (gray).

Parametric model – blind Wiener learning (III/III)





Estimated mean (black), true static nonlinearity (red) and the result for all 100 realisations (gray).

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Semiparametric model 1 – known model order

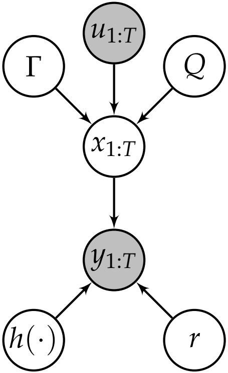
First step towards a fully data driven model, the order of the LGSS model is assumed known.

Parameters: $\theta = \{\Gamma, Q, r, h(\cdot)\}$.

Bayesian model specified by priors

- Conjugate priors for $\Gamma = [A \; B]$, Q and r,
  - $p(\Gamma, Q) =$ Matrix-normal inverse-Wishart
  - $p(r) =$ inverse-Wishart
- Gaussian process prior on $h(\cdot)$,

$$h(\cdot) \sim \mathcal{GP}(z, k(z, z')).$$



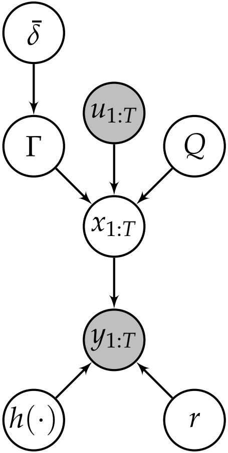
Everything is learned from the data, by introducing the possibility to switch specific model components on and off.

Parameters: $\theta = \{\Gamma, Q, \bar{\delta}, r, h(\cdot)\}$.

Bayesian model specified by priors

- Sparseness prior (ARD) on $\Gamma = [A\ B]$,
  - $p(\Gamma \mid \bar{\delta}) = \prod_{j=1}^{n_x+n_u} \mathcal{N}(\gamma_j \mid 0, \delta_j^{-1} I_{n_x+n_u})$
  - $\bar{\delta} = \{\delta_j\}_{j=1}^{n_x+n_u}, \quad p(\delta_j) = \text{Gam}(\delta_j; a, b)$
- Inverse-Wishart prior on Q and r
- Gaussian process prior on $h(\cdot)$,

$$h(\cdot) \sim \mathcal{GP}(z, k(z, z')).$$



Aim: Devise a Gibbs sampler targeting $p(\theta, x_{1:T} \mid y_{1:T})$.

PG-BS for Wiener system identification:

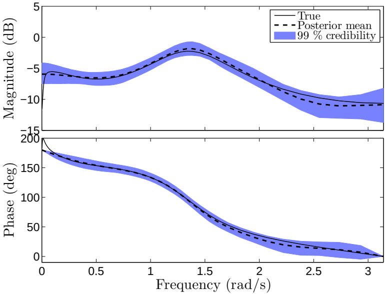
- PF-BS
  - Run a conditional PF, targeting $p(x_{1:T} \mid \theta, y_{1:T})$;
  - Run a backward simulator to sample $x_{1:T}^{\star}$;
- Draw
  - $\{\Gamma^{\star}, Q^{\star}, r^{\star}\} \sim p(\Gamma, Q, r \mid h, x_{1:T}^{\star}, y_{1:T})$, if MNIW prior, or;
  - $\{\Gamma^{\star}, \bar{\delta}^{\star}, Q^{\star}, r^{\star}\} \sim p(\Gamma, \bar{\delta}, Q, r \mid h, x_{1:T}^{\star}, y_{1:T})$, if ARD prior;
- Draw $h^{\star} \sim p(h \mid r^{\star}, x_{1:T}^{\star}, y_{1:T})$.

We have introduced two Bayesian semiparametric models:

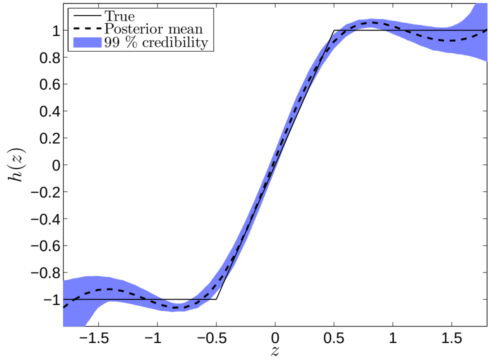
- Assume the model order of the LGSS model to be known and employ conjugate priors (MNIW).
- Fully data driven model where everything is learned from data, made possible via the ARD prior.

We have a learning algorithm (PG-BS) for each model.

- Bayesian semiparametric model with conjugate prior (MNIW).
- 6^{th} order LGSS model and a saturation.
- Using $T = 1000$ measurements.
- Employ the PG-BS sampler with $N = 15$ particles.
- Run 15000 MCMC iterations, discard 5000 as burn-in.



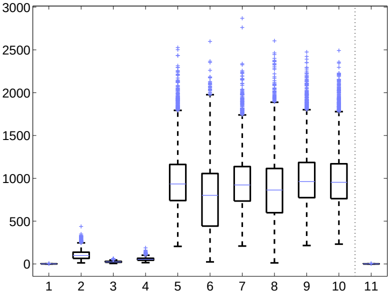
True Bode diagram of the linear system (black), estimated Bode diagram (dashed black) and 99% credibility interval (blue).



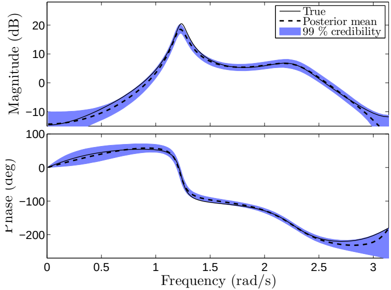
True static nonlinearity (black), estimated posterior mean (dashed black) and 99% credibility interval (blue).

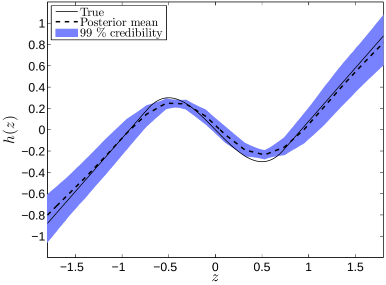
Data driven learning (I/II)

- Bayesian semiparametric model with ARD prior.
- 4^{th} order LGSS model.
- Using $T = 1000$ measurements.
- Employ the PG-BS sampler with $N = 15$ particles.
- Run 15000 MCMC iterations, discard 5000 as burn-in.



ARD precision parameters $\delta_j = 1, \ldots, 11$. The rightmost box plot corresponds to the input signal.

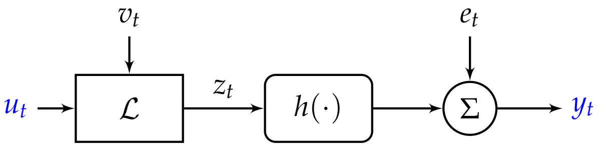




Bode diagram of the 4th-order linear system. Estimated mean, true system and 99% credibility intervals (blue).

Estimated mean (thick black), 99% credibility intervals (blue) and the true static nonlinearity (non-monotonic) (black).

Conclusions



- The Wiener model offeres a “controlled” venture into the realm of nonlinear dynamical models.
- Note the “characterization of the uncertainty”.
- Three cases studied
  - Fully parametric (ML – EM-PS)
  - Semiparametric (Bayes – PG-BS)
  - Data driven semiparametric (Bayes – PG-BS)

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Learning Wiener models – references

- Maximum Likelihood approach (EM-PS)

Adrian Wills, Thomas B. Schön, Lennart Ljung and Brett Ninness. **Identification of Hammerstein-Wiener Models**. *Automatica*, 2012. (Accepted for publication)

Thomas B. Schön, Adrian Wills and Brett Ninness. **System Identification of Nonlinear State-Space Models**. *Automatica*, 47(1):39-49, January 2011.

- Bayesian approach (PG-BS)

Fredrik Lindsten, Thomas B. Schön and Michael I. Jordan. **Data driven Wiener system identification**. *Automatica*, 2012 (submitted).

Christophe Andrieu, Arnaud Doucet and Roman Holenstein, **Particle Markov chain Monte Carlo methods**, *Journal of the Royal Statistical Society: Series B*, 72:269-342, 2010.

MATLAB code is available from `www.control.isy.liu.se/~lindsten/code.html`