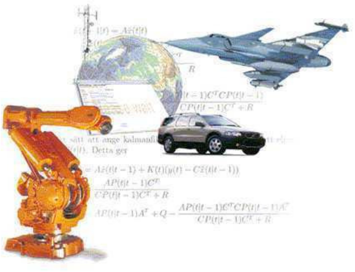
Hierarchical Bayesian ARX models for robust inference



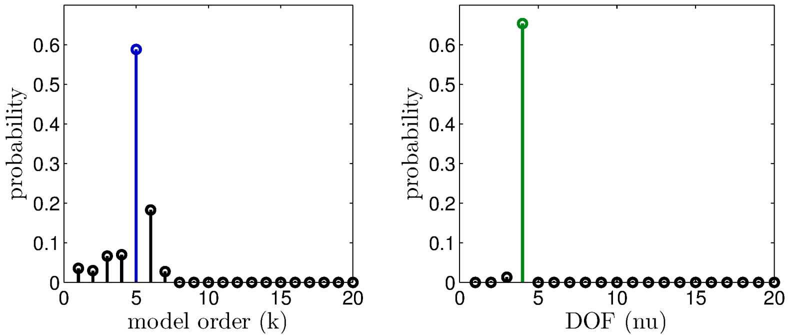
Johan Dahlin, Fredrik Lindsten,
Thomas B. Schön, Adrian Wills

Division of Automatic Control,
Linköping University

School of EECS,
University of Newcastle, Australia

Summary

- **Aim**: Robustly infer parameters in ARX processes with Student's t-distributed innovations.
- **Purpose**: Evaluate the practical use of **Reversible Jump-MCMC** (and ARD priors) in System Identification.
- **Method**: **Bayesian modelling** with conjugate priors and algorithms based on RJ-MCMC.
- **Results**: Good performance on simulated random ARX systems with Student's t innovations as well as on real EEG data.



probability
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
model order (k)
probability
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
DOF (nu)

The ARX model

An **autoregressive exogenous** (ARX) model of orders $\mathbf{n} = (\mathbf{n_a}, \mathbf{n_b})$ is defined by

$$y_t + \sum_{t=1}^{\mathbf{n_a}} a_i y_{t-i} = \sum_{t=1}^{\mathbf{n_b}} b_i u_{t-i} + \mathbf{e_t}.$$

Two practical problems using the least square (LS) solution are:

- The correct model order n is **often unknown** or does not exist.
- The observed data could be **non-Gaussian**.

Two special features of our model:

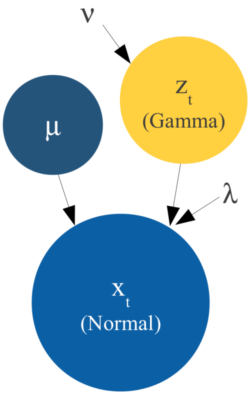
- The **excitation noise** is modelled as Student's t distributed.
- An **automatic order selection** by two different methods:
  - incorporating the system orders in the posterior distribution.
  - applying a sparseness prior (ARD) over the ARX coefficients.

We are using a hierarchical structure for the **excitation noise**

$$
\begin{aligned}
p_{z_t,\lambda}(x_t) &= \mathcal{N}(x_t; 0, (z_t\lambda)^{-1}), \\
p_\nu(z_t) &= \mathcal{G}(\nu/2, \nu/2),
\end{aligned}
$$

where the prior distributions of the **hyperparameters** $\{\lambda, \nu\}$ are chosen as (vague) Gamma distributions

$$
\begin{aligned}
p(\nu) &= \mathcal{G}(\nu; \alpha_\nu, \beta_\nu), \\
p(\lambda) &= \mathcal{G}(\lambda; \alpha_\lambda, \beta_\lambda).
\end{aligned}
$$



The **automatic order selection** consists of choosing a suitable model in the model set

$$\mathcal{M}_n : \quad y_t = (\varphi_t^n)^\top \theta^n + e_t,$$

for $n = \{1,1\}, \{1,2\}, \ldots, \{n_{\max}, n_{\max}\}$ and where φ_t^n denotes a vector of known inputs and outputs.

We use a **non-informative** uniform prior over the model hypotheses

$$p(n) = \begin{cases} 1/n_{\max}^2 & \text{if } n_a, n_b \in \{1, \ldots, n_{\max}\} \\ 0 & \text{otherwise} \end{cases}.$$

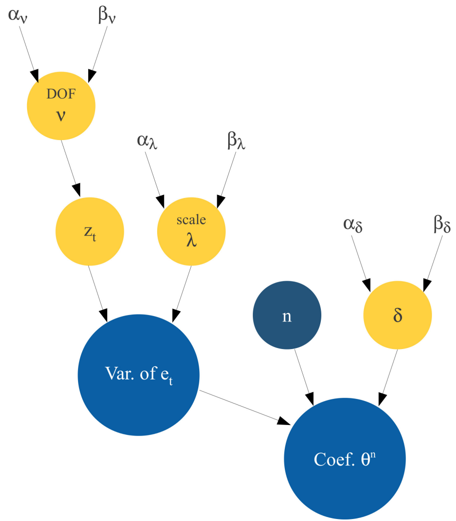
The **distribution of the model coefficients** (given the model order) is assumed to be

$$\begin{aligned} p(\theta^n|n,\delta) &= \mathcal{N}\left(\theta^n; 0, \delta^{-1} I_{n_a+n_b}\right), \\ p(\delta) &= \mathcal{G}\left(\delta; \alpha_\delta, \beta_\delta\right), \end{aligned}$$

where we used conjugate priors to obtain an analytic expression.

The full collection of **unknown parameters** is

$$\eta = \{\theta^n, n, \delta, z_{1:T}, \lambda, \nu\}.$$



α_ν
β_ν
DOF ν
α_λ
β_λ
z_t
scale λ
α_δ
β_δ
n
δ
Var. of e_t
Coef. θ^n

- A discrete **Laplace distribution** is used as the proposal distribution for the model order

$$\mathbf{J}(\mathbf{n}'|\mathbf{n}) \propto \exp(-\ell|n' - n|), \quad \text{if } 1 \leq n' \leq n_{\max}.$$

- The acceptance probability is (using the Candidate's identity)

$$\rho_{nn'} = \min\left\{1, \frac{p(n'|z_{s+1:T}, \lambda, \delta, D_T)}{p(n|z_{s+1:T}, \lambda, \delta, D_T)} \frac{\mathbf{J}(\mathbf{n}|\mathbf{n}')}{\mathbf{J}(\mathbf{n}'|\mathbf{n})}\right\}.$$

- Note that $\rho_{nn'}$ is independent of θ^n, i.e. we can decide the model order before sampling coefficients.

- Draw new model order and coefficients (**Reversible jump step**)

$$\{n', \theta^{n'}\} | z_{s+1:T}, \lambda, \delta, D_T,$$

- Draw new coefficient variance (**Gibbs step**)

$$\delta' | \theta^{n'}, n'.$$

- Draw new innovation latent variable, innovation scale parameter and innovation DOF parameter (**Gibbs steps/MH step**)

$$z'_{s+1:T} | \theta^{n'}, n', z, \lambda, \nu, D_T, \quad \lambda' | \theta^{n'}, n', z'_{s+1:T}, D_T, \quad \nu' | z'_{s+1:T}.$$

Markov Chain Monte Carlo (cont.)

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Results

Three different straightforward numerical illustrations are presented to compare the proposed methods with the naive LS method:

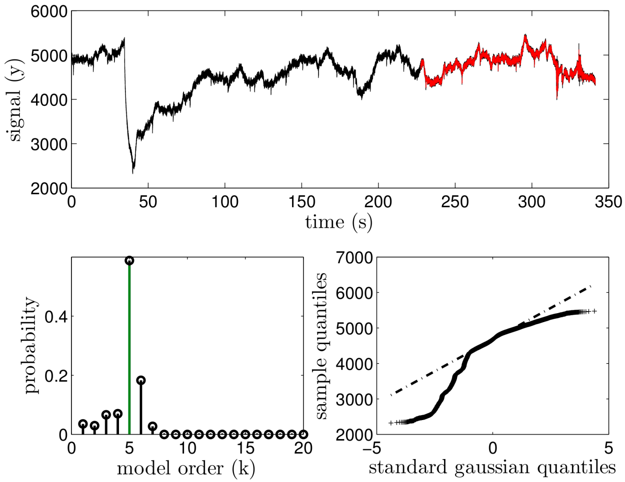
- **Large-scale simulation studies on random ARX systems**.
- Case-studies of ARX systems with missing data and outliers. (see paper for details)
- **Analysis of real EEG data**.

Method	mean	CI
LS	77.51	[77.21 77.81]
RJ-MCMC	78.24	[77.95 78.83]
ARD-MCMC	77.73	[77.47 78.06]

Table: The average and 95% confidence intervals (CI) for the model fit (in percent) from experiments with $25,000$ random ARX models.

- Significant difference between using RJ-MCMC and LS.
- RJ-MCMC seems to perform better than the algorithm based on an ARD sparseness prior. (see paper for details)

Results - EEG data



Summary

- **Aim**: Robustly infer parameters in ARX processes with Student's t-distributed innovations.
- **Purpose**: Evaluate the practical use of **Reversible Jump-MCMC** (and ARD priors) in System Identification.
- **Method**: **Bayesian modelling** with conjugate priors and algorithms based on RJ-MCMC.
- **Results**: Good performance on simulated random ARX systems with Student's t innovations as well as on real EEG data.

Thank you for your attention!

Questions, suggestions, or comments?

To download the code and data from this paper, please visit:
`http://www.control.isy.liu.se/~johda87`
and click on Software.