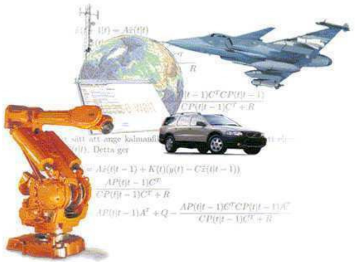
# Estimation of Linear Systems using a Gibbs Sampler



**Adrian Wills**[*]**, Thomas B. Schön**[**]**, Fredrik Lindsten**[**]**, Brett Ninness**[*]

[*]School of EECS,
University of Newcastle, Australia.

[**]Division of Automatic Control,
Linköping University, Sweden

A linear, Gaussian state space (LGSS) model is defined by

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + v_t, \\ y_t &= Cx_t + Du_t + e_t, \end{aligned}$$

where

$$\begin{pmatrix} x_1 \\ v_t \\ e_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_1 & 0 & 0 \\ 0 & Q & S \\ 0 & S^{\mathsf{T}} & R \end{pmatrix} \right).$$

This can equivalently be written as $x_1 \sim \mathcal{N}(x_1 \mid \mu, P_1)$,

$$\begin{pmatrix} x_{t+1} \\ y_t \end{pmatrix} \mid x_t \sim \mathcal{N}\Bigg( \underbrace{\begin{pmatrix} x_{t+1} \\ y_t \end{pmatrix}}_{\xi_t} \mid \underbrace{\begin{pmatrix} A & B \\ C & D \end{pmatrix}}_{\Gamma} \underbrace{\begin{pmatrix} x_t \\ u_t \end{pmatrix}}_{z_t} \underbrace{\begin{pmatrix} Q & S \\ S^{\mathsf{T}} & R \end{pmatrix}}_{\Pi} \Bigg).$$

A more compact formulation of the LGSS model is provided by,

$$\xi_t \mid x_t \sim \mathcal{N}(\xi_t \mid \Gamma z_t, \Pi), \qquad \xi_t = \Gamma z_t + w_t, \qquad w_t \sim \mathcal{N}(0, \Pi),$$
$$x_1 \sim \mathcal{N}(x_1 \mid \mu, P_1). \qquad x_1 \sim \mathcal{N}(x_1 \mid \mu, P_1),$$

where

$$\xi_t \triangleq \begin{pmatrix} x_{t+1} \\ y_t \end{pmatrix}, \quad z_t \triangleq \begin{pmatrix} x_t \\ u_t \end{pmatrix}, \quad \Gamma \triangleq \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad \Pi \triangleq \begin{pmatrix} Q & S \\ S^{\mathsf{T}} & R \end{pmatrix}$$

The parameters are defined as (using set notation)

$$\theta = \{\Gamma, \Pi\}.$$

**Goal:** Identify the LGSS model by computing $p(\theta \mid Y)$, where $\theta = \{\Gamma, \Pi\}$ and $Y \triangleq y_{1:N}$ (assume known initial state).

A **Markov chain Monte Carlo (MCMC)** sampler is a method for generating samples from a target distribution (here $p(\theta \mid Y)$) by simulating a Markov chain with this target distribution as its stationary distribution.

**Key question:** How do we construct a Markov chain with $p(\theta \mid Y)$ as its stationary distribution?

There are several constructive methods available for doing this and the **Gibbs sampler** is one of them.

A *blocked* Gibbs sampler, sampling from $p(\theta, X \mid Y)$ is given by

1. Given $\theta^k$, generate a sample of the state trajectory,

$$X^k \sim p(X \mid Y, \theta^k).$$

2. Then, given $X^k$ generate a sample of the parameters $\theta^{k+1}$,

$$\theta^{k+1} \sim p(\theta \mid X^k, Y).$$

Based on the empirical distribution from the Gibbs sampler $\widehat{p}(\theta \mid Y) = \sum_{m=1}^{M} \frac{1}{M} \delta_{\theta^m}(\theta)$, we can compute the following estimate

$$\mathbf{E}_{p(\theta|Y)}\left[f(\theta)\right] = \int f(\theta) p(\theta \mid Y) \mathrm{d}\theta \approx \frac{1}{M} \sum_{m=1}^{M} f(\theta^m).$$

**Task:** Generate a sample $x_{1:N}$ from $p(x_{1:N} \mid y_{1:N})$.

An efficient way of doing this can be found by noting that

$$p(x_{1:N+1} \mid y_{1:N}) = p(x_{N+1} \mid y_{1:N}) \prod_{t=1}^{N} \underbrace{p(x_t \mid x_{t+1}, y_{1:t})}_{\text{backwards kernel}}$$

and then employing the following strategy

- Sample $x_{N+1} \sim p(x_{N+1} \mid y_{1:N})$.
- Sample $x_N \sim p(x_N \mid x_{N+1}, y_{1:N})$.
- $\vdots$
- Sample $x_1 \sim p(x_1 \mid x_2, y_1)$.

The backward kernel can be computed by noting that

$$p(x_t \mid x_{t+1}, y_{1:t}) = \frac{p(x_{t+1} \mid x_t)p(x_t \mid y_{1:t})}{p(x_{t+1} \mid y_{1:t})} = \mathcal{N}\left(x_t \mid \mu_t, M_t\right),$$

where

$$\mu_t = M_t \left( A^{\mathsf{T}} Q^{-1}(x_{t+1} - Bu_t) + P_{t|t}^{-1}\widehat{x}_{t|t} \right),$$
$$M_t = P_{t|t} - P_{t|t}A^{\mathsf{T}}\left(AP_{t|t}A^{\mathsf{T}} + Q\right)^{-1} AP_{t|t}.$$

Here, $\widehat{x}_{t|t}$ and $P_{t|t}$ are provided by the Kalman filter.

This is a so called **backwards simulator**.

(See the paper for details and a square root implementation)

# Sampling the parameters (I/III)

**Task:** Generate a sample $\theta$ from $p(\theta \mid x_{1:N+1}, y_{1:N})$.

As usual we have

$$\underbrace{p(\theta \mid x_{1:N+1}, y_{1:N})}_{\text{posterior}} \propto \underbrace{p(x_{1:N+1}, y_{1:N} \mid \theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

Recall that $\theta = \{\Gamma, \Pi\}$, where $\Gamma$ and $\Pi$ are random matrices.

We will make use of a conjugate prior, which for this model is given by the **matrix normal inverse Wishart ($\mathcal{MNIW}$)** distribution.

The likelihood ($X \triangleq \{x_1, \dots, x_{N+1}\}, Y \triangleq \{y_1, \dots, y_N\}$)

$$p(X, Y \mid \theta) = \prod_{t=1}^{N} \mathcal{N}\left(\xi_t \mid \Gamma z_t, \Pi\right)$$

$$= \frac{1}{(2\pi)^{Nd/2} \left|\Pi\right|^{N/2}} \exp\left(-\frac{1}{2}\operatorname{Tr}\left(\Pi^{-1}\sum_{t=1}^{N}(\xi_t - \Gamma z_t)(\xi_t - \Gamma z_t)^{\mathsf{T}}\right)\right),$$

can using

$$\Xi \triangleq \left(\xi_1, \xi_2, \dots, \xi_N\right), \qquad Z \triangleq \left(z_1, z_2, \dots, z_N\right),$$

be written as

$$p(X, Y \mid \theta) = \frac{1}{(2\pi)^{\frac{Nd}{2}} \left|\Pi\right|^{\frac{N}{2}}} \exp\left(-\frac{1}{2}\operatorname{Tr}\left((\Xi - \Gamma Z)^{\mathsf{T}}\Pi^{-1}(\Xi - \Gamma Z) I\right)\right)$$

$$= \mathcal{MN}\left(\Xi \mid \Gamma Z, I, \Pi\right).$$

The $\mathcal{MNIW}$ distribution makes use of

$$p(\Gamma, \Pi) = p(\Gamma \mid \Pi)p(\Pi),$$

and it place

- a matrix normal distribution (generalisation of the multivariate Normal to the matrix case) on $\Gamma \mid \Pi$,
- and an inverse Wishart distribution on $\Pi$ (generalisation of the inverse Gamma to the matrix case).

(See the paper for details and a square root implementation)

Iterate the following

1. Given $\theta^k$, generate a sample of the state trajectory,

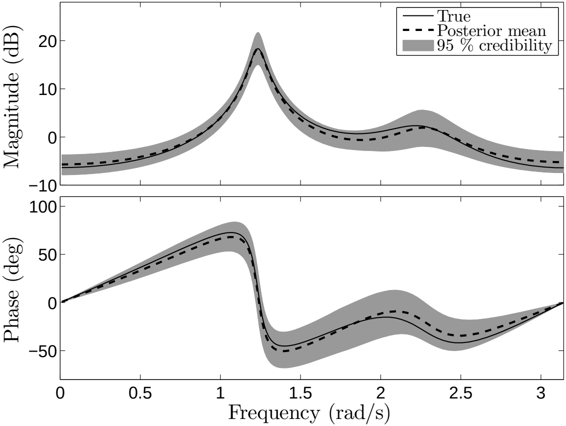
$$X^k \sim p(X \mid Y, \theta^k).$$

2. Then, given $X^k$ generate a sample of the parameters $\theta^{k+1}$,

$$\theta^{k+1} \sim p(\theta \mid X^k, Y).$$

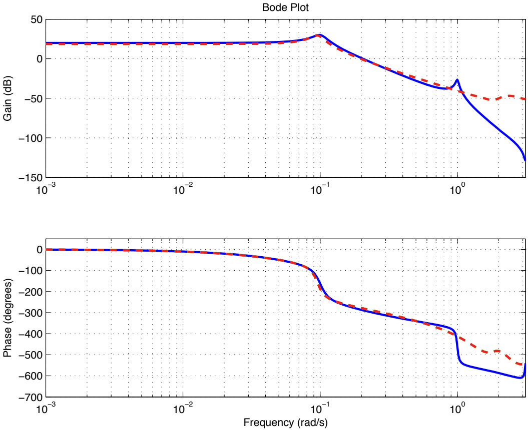
to obtain an empirical approximation of the posterior distribution

$$\widehat{p}(\theta \mid Y) = \sum_{m=1}^{M} \frac{1}{M} \delta_{\theta^m}(\theta).$$


True
Posterior mean
95 % credibility
Magnitude (dB)
20
10
0
−10
Phase (deg)
100
50
0
−50
0
0.5
1
1.5
2
2.5
3
Frequency (rad/s)

# Numerical example 2 (I/III)

In this example we are concerned with a 6th order system shown in the Bode plot below (solid blue).



As a means of utilising the empirical distribution

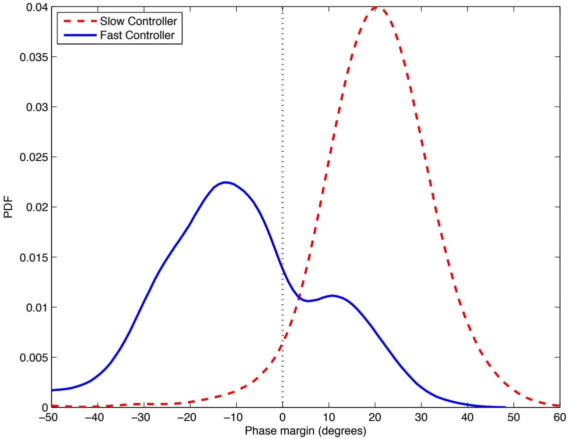
$$\widehat{p}(\theta \mid Y) = \sum_{m=1}^{M} \frac{1}{M} \delta_{\theta^m}(\theta)$$

provided by the Gibbs sampler, we first consider the conditional mean estimate of the transfer function $G(z) = C(zI - A)^{-1}B + D$ (dashed red).

Based on this estimate, we designed two controllers,

1. A “slow” controller with a nominal phase margin of $\varphi = 22^\circ$.
2. A “fast” controller with a nominal phase margin of $\varphi = 14^\circ$.

The pdf's of the phase margin for the given data set $p(\varphi \mid Y)$.



# Conclusions

- Derived a Gibbs sampler to compute $p(\theta \mid Y)$ for LGSS models.
- Numerically robust square root implementations provided.
- Opens up for uncertainty descriptions of nonstandard objects.

---

- The same strategy can be employed for nonlinear systems! See presentation tomorrow if you are interested:

  Fredrik Lindsten, Thomas B. Schön and Michael I. Jordan, A semiparametric Bayesian approach to Wiener system identification. At $16.50 - 17.10$ in Meeting Studio 201 A/B .

- New PhD course, for more information,
  `www.control.isy.liu.se/~schon/courses.html`