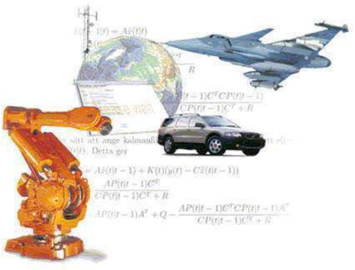
A graph/particle-based method for experiment design in nonlinear systems

IFAC World Congress 2014, Cape Town, South Africa, August 25, 2014.



Johan Dahlin
johan.dahlin@liu.se

Division of Automatic Control,
Linköping University,
Sweden.

This is collaborative work with

Patricio E. Valenzuela (Royal Institute of Technology (KTH), Sweden)
Dr. Cristian R. Rojas (Royal Institute of Technology (KTH), Sweden)
Prof. Thomas B. Schön (Uppsala University, Sweden)

Summary

Aim

Input design for nonlinear state space models.

Methods

Graph theory for input generation.
Sequential Monte Carlo methods.
Convex optimisation.

Contributions

Estimation of the expected information matrix.
Novel method for input design.

Problem formulation

$$x_{t+1} | x_t, u_t \sim f_\theta(x_{t+1}; x_t, u_t),$$
$$y_t | x_t, u_t \sim g_\theta(y_t; x_t, u_t).$$

Task: Design $u^\star_{1:T}$ as a stationary process with:

- finite memory,
- alphabet with finite cardinality.

such that it maximises the expected information

$$\mathcal{I}(\theta) = \mathbb{E}_{y_{1:T}}\Big[\mathcal{S}(\theta)\mathcal{S}(\theta)^\top\Big], \quad \mathcal{S}(\theta) = \nabla \log p_\theta(y_{1:T}).$$

Main idea

We can express the probability mass function of $u^\star_{1:T}$ as

$$p_u(u^\star_{1:T}) \triangleq \operatorname*{argmax}_{p_u \in \mathcal{P}_\mathcal{C}} h\big(\mathcal{I}(p_u)\big),$$

where p_u is a convex combination of the extreme points of $\mathcal{P}_\mathcal{C}$ and $h(\cdot) : \mathbb{R}^{p\times p} \to \mathbb{R}$ denotes a concave function.

Main idea

The optimal pmf has the expected information

$$\mathcal{J}(\gamma) \triangleq \sum_{k=1}^{n_b} \alpha_i \mathcal{I}^{(k)}(\theta), \qquad \text{with } \alpha \in \mathbb{R}_+^{n_b}, \quad \sum_{k=1}^{n_b} \alpha_k \triangleq 1,$$

for $k = 1, \ldots, n_b$. The optimal weights are given by

$$\alpha^\star \triangleq \operatorname*{argmax}_{\alpha} h(\mathcal{J}(\gamma)).$$

Overview of the algorithm

(i) Create a realisation $u_{1:T}^{(k)}$ from each of the n_b basis inputs.

(ii) Estimate the expected information matrix $\widehat{\mathcal{I}}^{(k)}$ for each $u_{1:T}^{(k)}$.

(iii) Compute the optimal weighting $\alpha^\star$ of the basis inputs.

Overview of the algorithm

(i) Create a realisation $u_{1:T}^{(k)}$ from each of the n_b basis inputs.

Using graph theory, see Valenzuela et al. [2013].

(ii) Estimate the expected information matrix $\widehat{\mathcal{I}}^{(k)}$ for each $u_{1:T}^{(k)}$.

Using particle filtering.

(iii) Compute the optimal weighting $\alpha^\star$ of the basis inputs.

Using convex optimisation, see Valenzuela et al. [2013].

Estimating the expected information

The expected information can be estimated using

$$\mathcal{I}(\theta) = \mathbb{E}_{y_{1:T}}\left[\mathcal{S}(\theta)\mathcal{S}^{\top}(\theta)\right].$$

The score function can be rewritten using Fisher's identity as

$$\mathcal{S}(\theta) = \int \nabla \log p_\theta(x_{0:T}, y_{1:T}) p_\theta(x_{0:T}|y_{1:T}) \, \mathrm{d}x_{0:T},$$

Estimating the expected information

The expected information can be estimated using

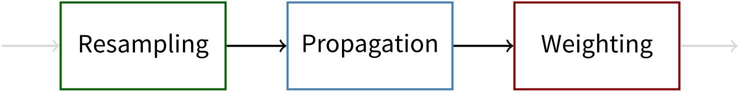
$$\mathcal{I}(\theta) = \mathbb{E}_{y_{1:T}}\left[\mathcal{S}(\theta)\mathcal{S}^{\top}(\theta)\right].$$

The score function can be rewritten using Fisher's identity as

$$\mathcal{S}(\theta) = \sum_{t=1}^{T} \int \xi_\theta(x_t, x_{t-1}) p_\theta(x_{t-1:t} | y_{1:T}) \, \mathrm{d}x_{t-1:t},$$

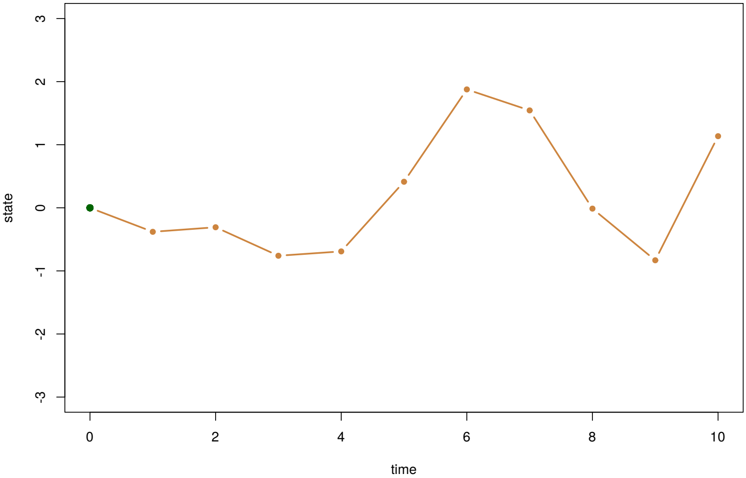
$$\xi_\theta(x_t, x_{t-1}) = \nabla \log f_\theta(x_t | x_{t-1}, u_{t-1}) + \nabla \log g_\theta(y_t | x_t, u_t).$$

Particle filtering: overview

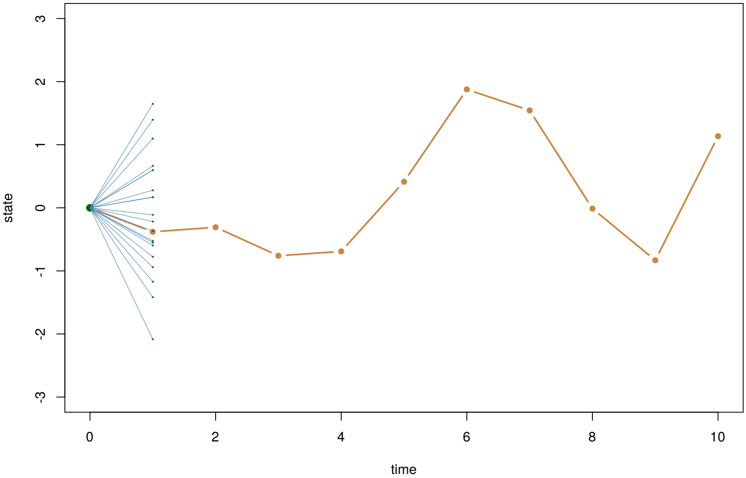


- Resampling: $\mathbb{P}(a_t^{(i)} = j) = \widetilde{w}_{t-1}^{(j)}$ and set $\widetilde{x}_{t-1}^{(i)} = x_{t-1}^{a_t^{(i)}}$.
- Propagation: $x_t^{(i)} \sim f_\theta\big(x_t | \widetilde{x}_{t-1}^{(i)}, u_{t-1}\big)$.
- Weighting: $w_t^{(i)} = g_\theta\big(y_t | x_t^{(i)}, u_t\big)$.

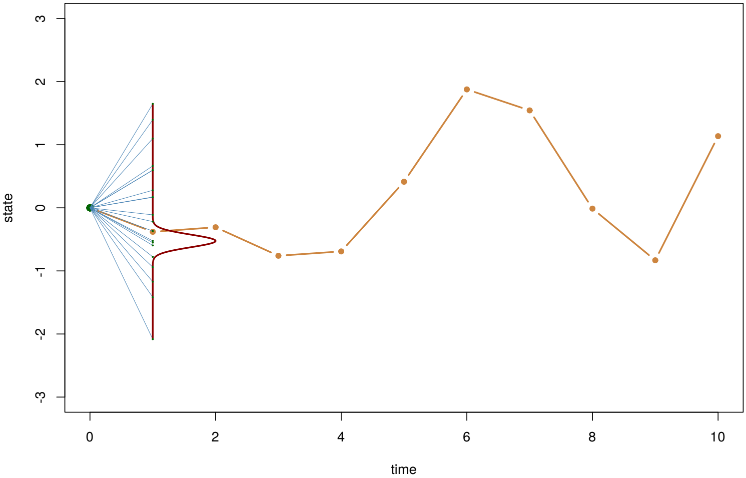
Particle filtering: toy example



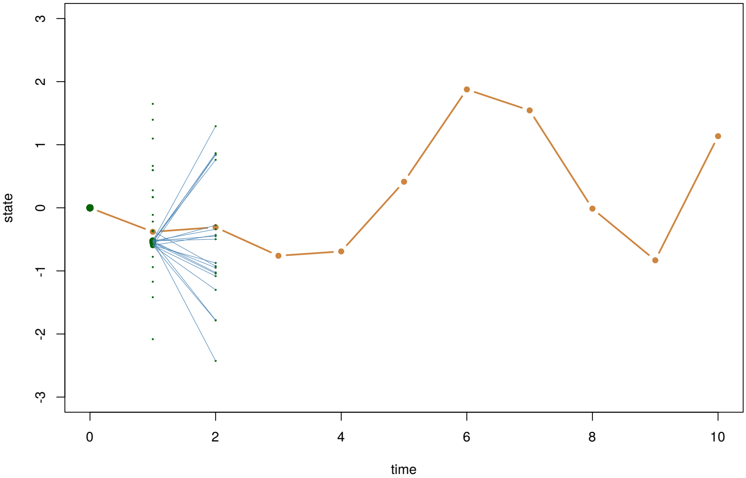
Particle filtering: toy example



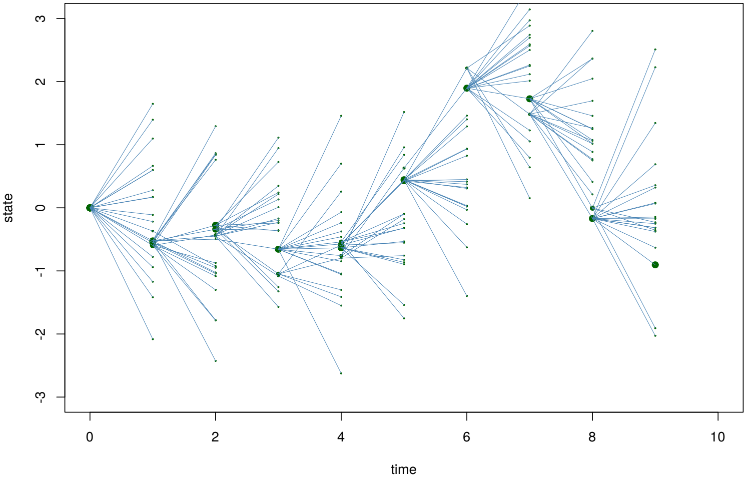
Particle filtering: toy example



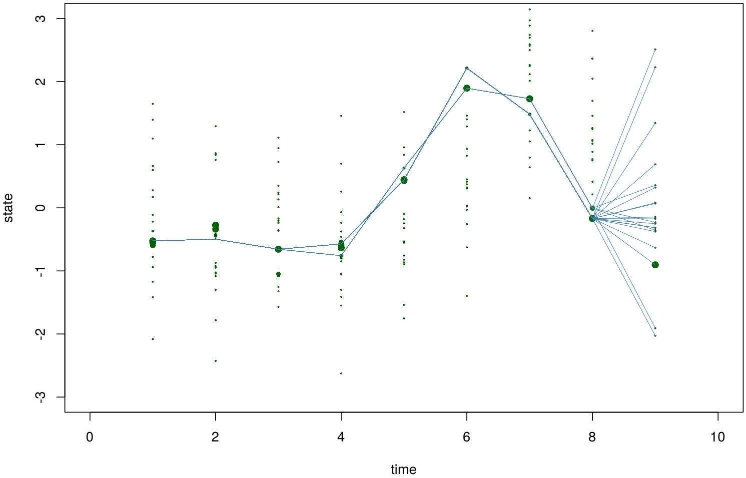
Particle filtering: toy example



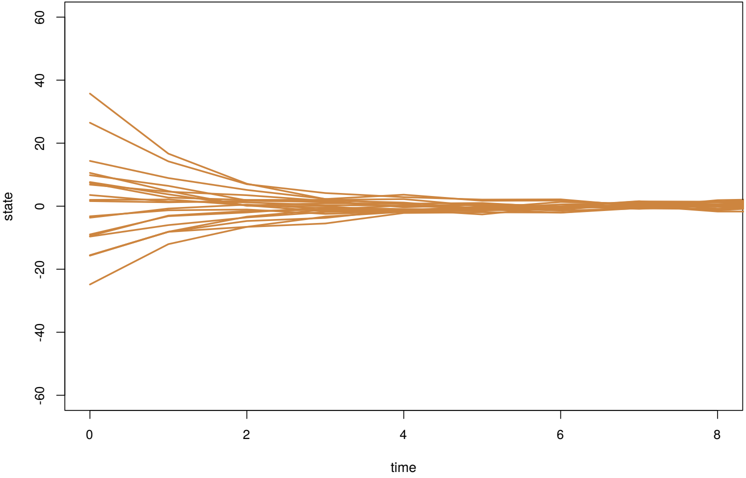
Particle filtering: degeneracy



Particle filtering: degeneracy



Fixed-lag particle smoothing: motivation



Fixed-lag particle smoothing: overview

Given the particle system

$$\left\{ x_t^{(i)}, \widetilde{w}_t^{(i)} \right\}_{i=1}^{N},$$

and the approximation given by

$$\widehat{p}_\theta(\mathrm{d}x_{t:t-1} | y_{1:T}) \approx \widehat{p}_\theta(\mathrm{d}x_{t:t-1} | y_{1:\kappa_t}),$$

where $\kappa_t = \min\{T, t + \Delta\}$, we can compute estimates of the score function by a fixed-lag particle smoother.

Computing the optimal weighting of basis inputs

During iteration m of the algorithm. For each input $k = 1, \ldots, n_b$:

- Generate a system output realisation using $u_{1:T}^{(k)}$.
- Compute L estimates of $\mathcal{S}(\theta)$.
- Estimate the observed information matrix

$$\widehat{\mathcal{I}}_m(\theta) = \frac{1}{L} \sum_{l=1}^{L} \widehat{\mathcal{S}}_l(\theta) \widehat{\mathcal{S}}_l^\top(\theta).$$

Solve the optimisation problem:

$$\widehat{\alpha}_m^\star = \operatorname*{argmax}_{\alpha} h\big(\widehat{\mathcal{J}}_m(\alpha)\big), \quad \text{with} \quad \widehat{\mathcal{J}}_m(\alpha) = \sum_{k=1}^{n_b} \alpha_i \widehat{\mathcal{I}}_m^{(k)}(\theta).$$

Overview of the algorithm

(i) Create a realisation $u_{1:T}^{(k)}$ from each of the n_b basis inputs.

Using graph theory, see Valenzuela et al. [2013].

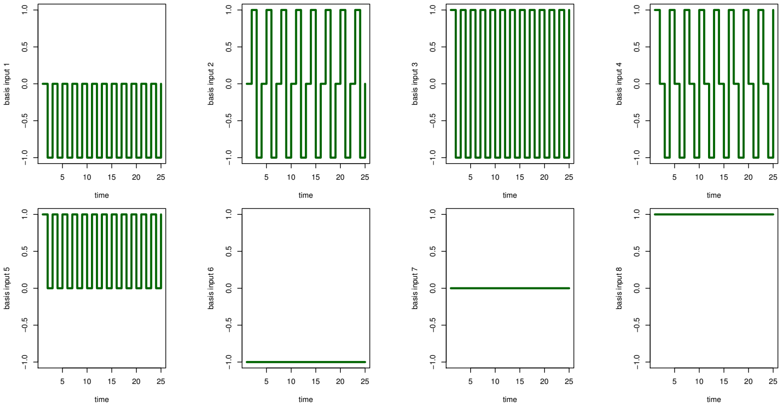
(ii) Estimate the expected information matrix $\widehat{\mathcal{I}}^{(k)}$ for each $u_{1:T}^{(k)}$.

Using particle filtering.

(iii) Compute the optimal weighting $\alpha^\star$ of the basis inputs.

Using convex optimisation, see Valenzuela et al. [2013].

Results: basis inputs



Results: linear Gaussian model

$$x_{t+1}|x_t, u_t \sim \mathcal{N}(x_{t+1}; \boldsymbol{\phi} x_t + u_t, \boldsymbol{\sigma}^2),$$
$$y_t|x_t, u_t \sim \mathcal{N}(y_t; x_t, 0.1^2),$$

with true parameters $\boldsymbol{\theta} = \{\boldsymbol{\phi}, \boldsymbol{\sigma}\} = \{0.5, 0.1\}$ and $T = 100$.

Input	log det$(\widehat{\mathcal{I}})$	tr$(\widehat{\mathcal{I}}^{-1})$
Optimal (det)	$20.67(0.01)$	$1.51 \cdot 10^{-4}(5.18 \cdot 10^{-7})$
Optimal (tr)	$20.82(0.01)$	$1.32 \cdot 10^{-4}(4.45 \cdot 10^{-7})$
Binary	$\mathbf{20.91}(0.01)$	$\mathbf{1.21 \cdot 10^{-4}}(4.51 \cdot 10^{-7})$
Uniform	$19.38(0.01)$	$5.32 \cdot 10^{-4}(2.12 \cdot 10^{-6})$

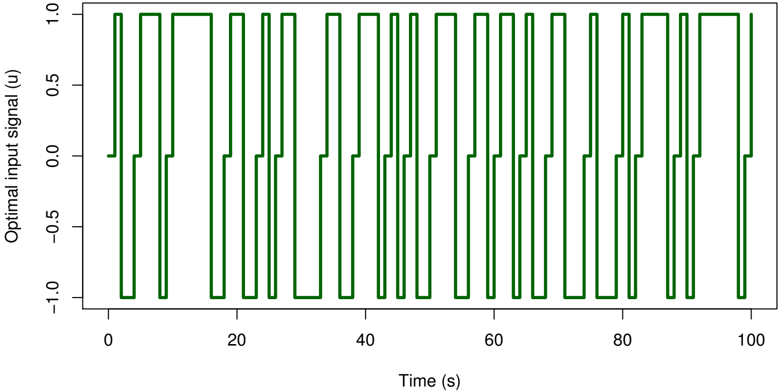
Results: nonlinear model

$$x_{t+1}|x_t, u_t \sim \mathcal{N}\Big(x_{t+1}; \boldsymbol{\theta_1} x_t + \frac{x_t}{\boldsymbol{\theta_2} + x_t^2} + u_t, 0.1^2\Big),$$
$$y_t|x_t, u_t \sim \mathcal{N}\Big(y_t; \frac{1}{2}x_t + \frac{2}{5}x_t^2, 0.1^2\Big),$$

with true parameters $\boldsymbol{\theta} = \{0.7, 0.6\}$ and $T = 100$.

Input	log det$(\widehat{\mathcal{I}})$
Optimal (det)	$\mathbf{25.34}(0.01)$
Binary	$24.75(0.01)$
Uniform	$24.38(0.01)$

Results: nonlinear model



Conclusions

Contributions

Improved parameter information compared with standard inputs.

Methods

Convex combination of basis inputs.
Fixed-lag particle smoothing and Fisher's identity.
Monte Carlo estimators.

Future work

More efficient information matrix estimation (new paper).
Further evaluation of the proposed method.
Robust input design.
Bayesian input design.

Thank you for your attention!

Questions, comments and suggestions are most welcome.

Two more presentations on nonlinear identification.

Thursday August, 28 at 10:00 in Brian Anderson.

The paper and more information are found at: `http://work.johandahlin.com/`.

Fixed-lag particle smoothing (cont.)

Assume that

$$p_\theta(x_t|y_{1:T}) \approx p_\theta(x_t|y_{1:\kappa_t}), \qquad \kappa_t = \min\{T, t+\Delta\},$$

for some $0 \leq \Delta \leq T$. It follows that

$$\widehat{p}_\theta(x_{t-1:t}|y_{1:T}) = \sum_{i=1}^{N} \widetilde{w}_{\kappa_t}^{(i)} \delta_{\widetilde{x}_{t-1:t,\kappa_t}^{(i)}}(\mathrm{d}x_{t-1:t})$$

which can be used to estimate the gradient and Hessian information about the log-target.

Score estimation using the FL smoother

The score can be estimated using *Fisher's identity* given by

$$\nabla_\theta \log p_\theta(y_{1:T})\big|_{\theta=\theta'} = \int \nabla_\theta \log p_\theta(x_{1:T}, y_{1:T}) p_{\theta'}(x_{1:T}|y_{1:T}) \mathrm{d}x_{1:T}$$
$$\approx \int \nabla_\theta \log p_\theta(x_{1:T}, y_{1:T}) \widehat{p}_{\theta'}(x_{1:T}|y_{1:T}) \mathrm{d}x_{1:T}$$

We also know that

$$\nabla_\theta \log p_\theta(x_{1:T}, y_{1:T}) = \sum_{t=1}^{T} \underbrace{\left[\nabla_\theta \log f_\theta(x_t|x_{t-1}) + \nabla_\theta \log g_\theta(y_t|x_t)\right]}_{\triangleq \xi(x_t, x_{t-1})},$$

which gives

$$\nabla_\theta \log p_\theta(y_{1:T})\big|_{\theta=\theta'} \approx \sum_{t=1}^{T} \sum_{i=1}^{N} \widetilde{w}_{\kappa_t}^{(i)} \xi(\widetilde{x}_{t-1,\kappa_t}^{(i)}, \widetilde{x}_{t,\kappa_t}^{(i)}).$$

P. E. Valenzuela, C. R. Rojas, and H. Hjalmarsson. Optimal input design for dynamic systems: a graph theory approach. In *Proceedings of the IEEE Conference on Decision and Control (CDC)*, Florence, Italy, dec 2013.