Second-order PMCMC for Parameter Inference

Johan Dahlin[†], Fredrik Lindsten[†] and Thomas B. Schön[‡]

UPPSALA UNIVERSITET

Summary

- We propose a PMCMC algorithm that incorporates first- and second-order information into the proposal distribution.
- Enables estimation of parameters in general state space models.
- A large improvement in the initial phase of the algorithm is obtained on a linear Gaussian state space model.

Bayesian parameter inference

We are interested in solving the **parameter inference** problem in **nonlinear state space models**

$$x_{t+1}|x_t \sim f_\theta(x_{t+1}|x_t),$$
$$y_t|x_t \sim h_\theta(y_t|x_t),$$

given a set of observations $\mathcal{D} = \{y_t\}_{t=1}^T$ and where $\theta \in \Theta \subseteq \mathbb{R}^d$ denotes static parameters. The log-posterior distribution is given by

$$\log p(\theta|\mathcal{D}) \propto \log p(\mathcal{D}|\theta) + \log p(\theta) + \text{const.},$$

where $\log p(\theta)$ and $\log p(\mathcal{D}|\theta)$ denote the parameter log-prior and the (often) **intractable log-likelihood function**, respectively.

Zeroth-order PMCMC

This problem can be solved with a **Particle Metropolis-Hastings** (PMH) algorithm, with the acceptance probability

$$\alpha(\theta', \theta) = \min\left\{1, \frac{\widehat{p}(\mathcal{D}|\theta')}{\widehat{p}(\mathcal{D}|\theta)} \frac{p(\theta')}{p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}\right\}, \qquad (1)$$

where $\widehat{p}(\mathcal{D}|\theta)$ denotes the particle estimate of $p(\mathcal{D}|\theta)$. Here, new samples are proposed using e.g. a **Gaussian random walk** proposal

$$\theta' \sim q(\theta'|\theta) = \mathcal{N}(\theta'; \theta, \Sigma_\theta).$$

This is known to scale inefficiently in higher dimensions.

Main idea

Use particle smoothers to estimate the gradient and Hessian of $\log p(\theta|\mathcal{D})$ and to use this information in the PMH proposal.

First- and second-order PMCMC

A second-order Taylor expansion of $\log p(\theta'|\mathcal{D})$ around θ has the form

$$\log p(\theta'|\mathcal{D}) \approx \log p(\theta|\mathcal{D}) + (\theta' - \theta)^\top \nabla \log p(\theta|\mathcal{D}) + \frac{1}{2}(\theta' - \theta)^\top \left[\nabla^2 \log p(\theta|\mathcal{D})\right](\theta' - \theta),$$

which can be rewritten as

$$p(\theta'|\mathcal{D}) \underset{\approx}{\propto} \exp\left[-\frac{1}{2}\Gamma^\top \mathcal{I}(\theta)\Gamma\right], \quad \text{with } \Gamma = \theta' - \theta - \mathcal{I}(\theta)^{-1}\mathcal{S}(\theta),$$

where we introduce $\mathcal{S}(\theta) = \nabla \log p(\theta|\mathcal{D})$ and $\mathcal{I}(\theta) = -\nabla^2 \log p(\theta|\mathcal{D})$.

Guided by the Taylor expansion, we design a proposal using **second-order information** of the form

$$\theta' \sim q(\theta'|\theta) = \mathcal{N}\left(\theta'; \theta + \frac{\epsilon^2}{2}\widehat{\mathcal{I}}^{-1}(\theta)\,\widehat{\mathcal{S}}(\theta), \epsilon^2\widehat{\mathcal{I}}^{-1}(\theta)\right), \qquad (2)$$

where we have introduced the step-size ϵ. By choosing $\widehat{\mathcal{I}}^{-1}(\theta) = \mathbf{I}_d$, we obtain a proposal using only **first-order information**.

Estimation of gradients and Hessians

The estimate of $\mathcal{S}(\theta)$ is obtained using **Fisher's identity**

$$\widehat{\mathcal{S}}(\theta) = \int \nabla_\theta \log p_\theta(x_{1:T}, y_{1:T})\, \widehat{p}_\theta(x_{1:T}|y_{1:T})\, \mathrm{d}x_{1:T},$$

where $\widehat{p}_\theta(x_{1:T}|y_{1:T})$ denotes the empirical distribution obtained from a particle smoother. The estimate of $\mathcal{I}(\theta)$ is obtained similarly using

$$\widehat{\mathcal{I}}(\theta) = \left[\widehat{\mathcal{S}}(\theta)\right]^2 - \int \left[\nabla_\theta \log p_\theta(x_{1:T}, y_{1:T})\right]^2 \widehat{p}_\theta(x_{1:T}|y_{1:T})\, \mathrm{d}x_{1:T} - \int \left[\nabla_\theta^2 \log p_\theta(x_{1:T}, y_{1:T})\right] \widehat{p}_\theta(x_{1:T}|y_{1:T})\, \mathrm{d}x_{1:T},$$

which follows from **Louis' identity**.

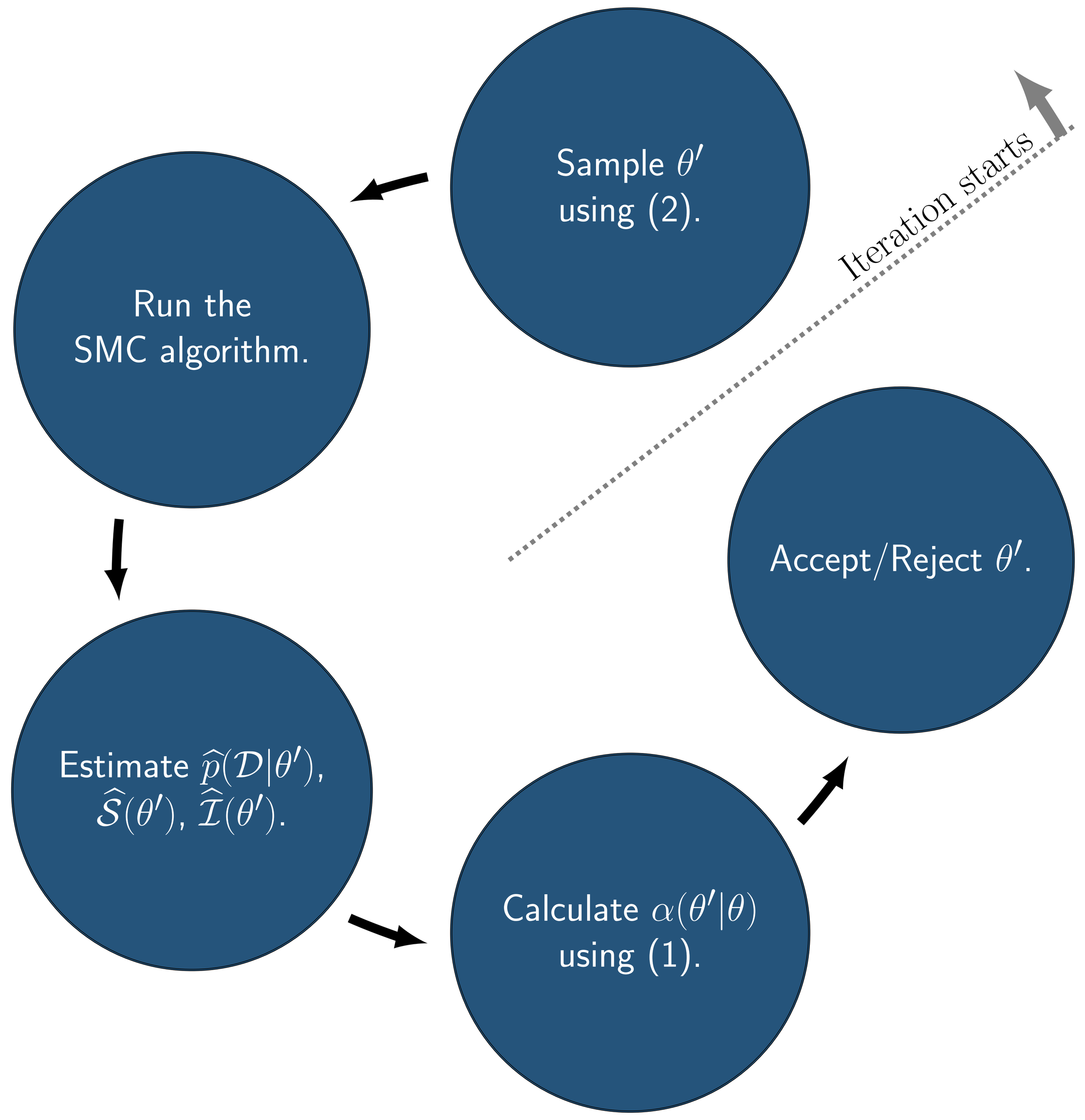
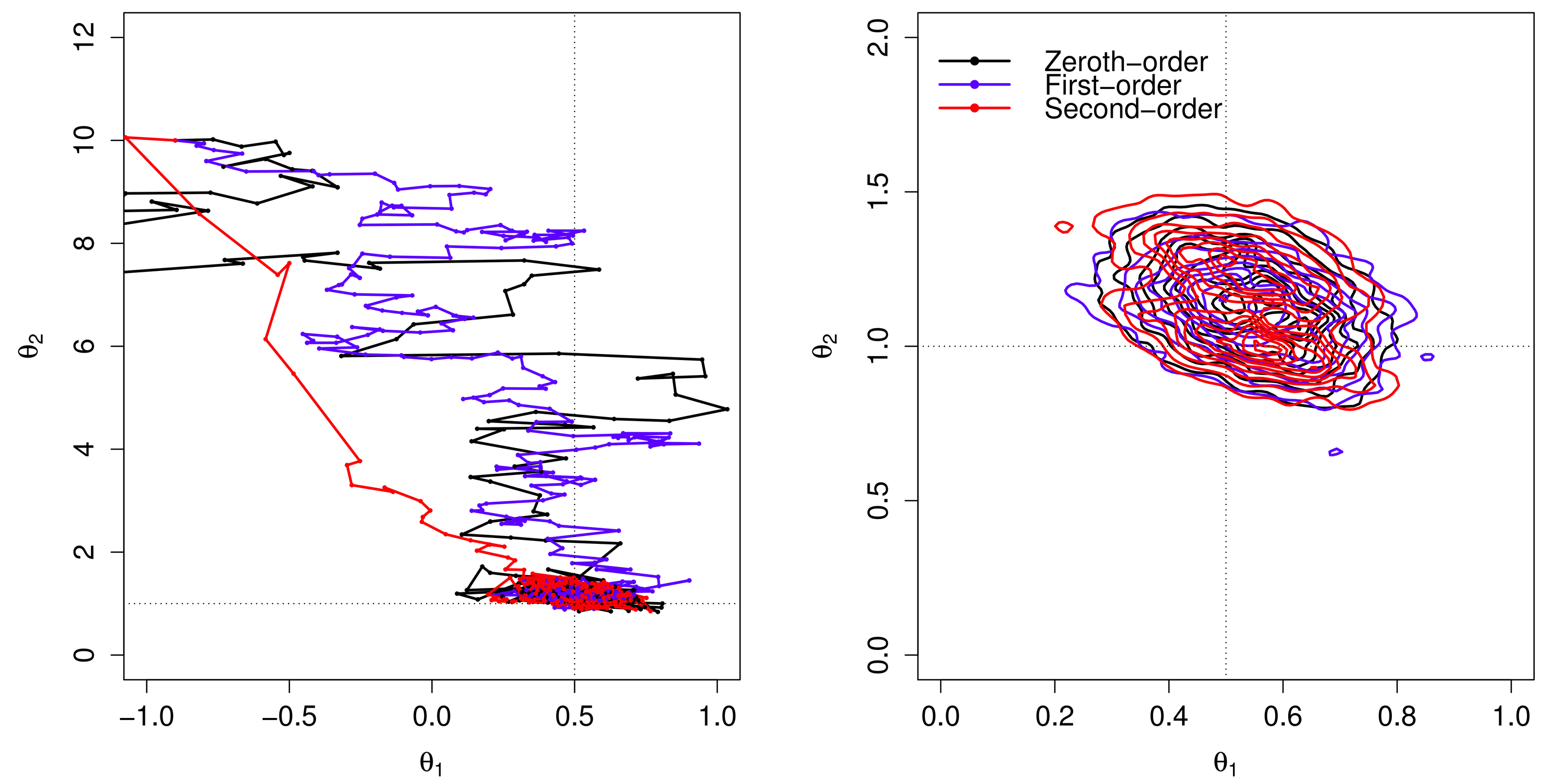


Figure: An iteration of the second-order PMH algorithm.

Example: Linear Gaussian model

$$x_{t+1}|x_t \sim \mathcal{N}\left(x_{t+1}; \theta_1 x_t, \theta_2^2\right),$$
$$y_t|x_t \sim \mathcal{N}\left(y_t; x_t, 0.1^2\right),$$

with true parameters $\theta^\star = \{\theta_1^\star, \theta_2^\star\} = \{0.5, 1.0\}$. We use $T = 100$ time steps, $N = 1\,000$ particles and $M = 30\,000$ MCMC iterations.



More information and source code

http://users.isy.liu.se/rt/johda87/

[†]{johan.dahlin,lindsten}@isy.liu.se, [‡]thomas.schon@it.uu.se