Particle-Metropolis Hastings with second-order proposals

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

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Summary

- We propose a PMH algorithm that incorporates first- and second-order information into the proposal distribution.
- The information is obtained from a fixed-lag smoother using the particle system obtained from the likelihood estimation.
- Large improvements in the initial phase of the algorithm are obtained on a linear Gaussian and stochastic volatility model.

Bayesian parameter inference

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

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

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

$$\log p(\theta|y_{1:T}) \propto \log p(y_{1:T}|\theta) + \log p(\theta) + \text{const.},$$

where $\log p(\theta)$ and $\log p(y_{1:T}|\theta)$ denote the log-prior and the (often) **intractable log-likelihood function**, respectively.

Zeroth-order PMH (PMMH)

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

$$\alpha(\theta',\theta) = \min\left\{1, \frac{\widehat{p}(y_{1:T}|\theta')}{\widehat{p}(y_{1:T}|\theta)} \frac{p(\theta')}{p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}\right\},$$

where $\widehat{p}(y_{1:T}|\theta)$ denotes the unbiased estimate of $p(y_{1:T}|\theta)$ obtained from a particle filter. 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 a particle smoother to estimate the gradient and Hessian of $\log p(\theta|y_{1:T})$ and use this information in the PMH proposal.

First- and second-order PMH

Consider a **Laplace approximation** of $\log p(\theta'|y_{1:T})$ around θ,

$$\begin{aligned}\log p(\theta'|y_{1:T}) \approx{}& \log p(\theta|y_{1:T}) \\ &+ (\theta'-\theta)^\top \nabla_\theta \log p(\theta|y_{1:T}) \\ &+ \frac{1}{2}(\theta'-\theta)^\top \left[\nabla_\theta^2 \log p(\theta|y_{1:T})\right](\theta'-\theta).\end{aligned}$$

By standard manipulations, we arrive at the posterior approximation

$$\begin{aligned}p(\theta'|y_{1:T}) &= \mathcal{N}(\theta'; \theta + \mathcal{G}_T(\theta), \mathcal{W}_T(\theta)), \text{ with} \\ \mathcal{W}_T^{-1}(\theta) &\triangleq \mathcal{I}_T(\theta) + \nabla_\theta^2 \log p(\theta), \\ \mathcal{G}_T(\theta) &\triangleq \mathcal{W}_T(\theta)\left[\mathcal{S}_T(\theta) + \nabla_\theta \log p(\theta)\right],\end{aligned}$$

where $\mathcal{S}_T(\theta) \triangleq \nabla_\theta \log p_\theta(y_{1:T})$ and $\mathcal{I}_T(\theta) \triangleq -\nabla_\theta^2 \log p_\theta(y_{1:T})$.

This approximation suggests the **second-order proposal**,

$$q\left(\theta'|\theta, \mathcal{S}_T(\theta), \mathcal{I}_T(\theta)\right) = \mathcal{N}\left(\theta'; \theta + \frac{\Gamma}{2}\mathcal{G}_T(\theta), \Gamma\mathcal{W}_T(\theta)\right),$$

by introducing a diagonal matrix of step lengths Γ, defined by the user. By choosing $\mathcal{W}_T(\theta) = \mathbb{I}_{d\times d}$, we obtain a **first-order proposal**.

Estimation of gradients and Hessians

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

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

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

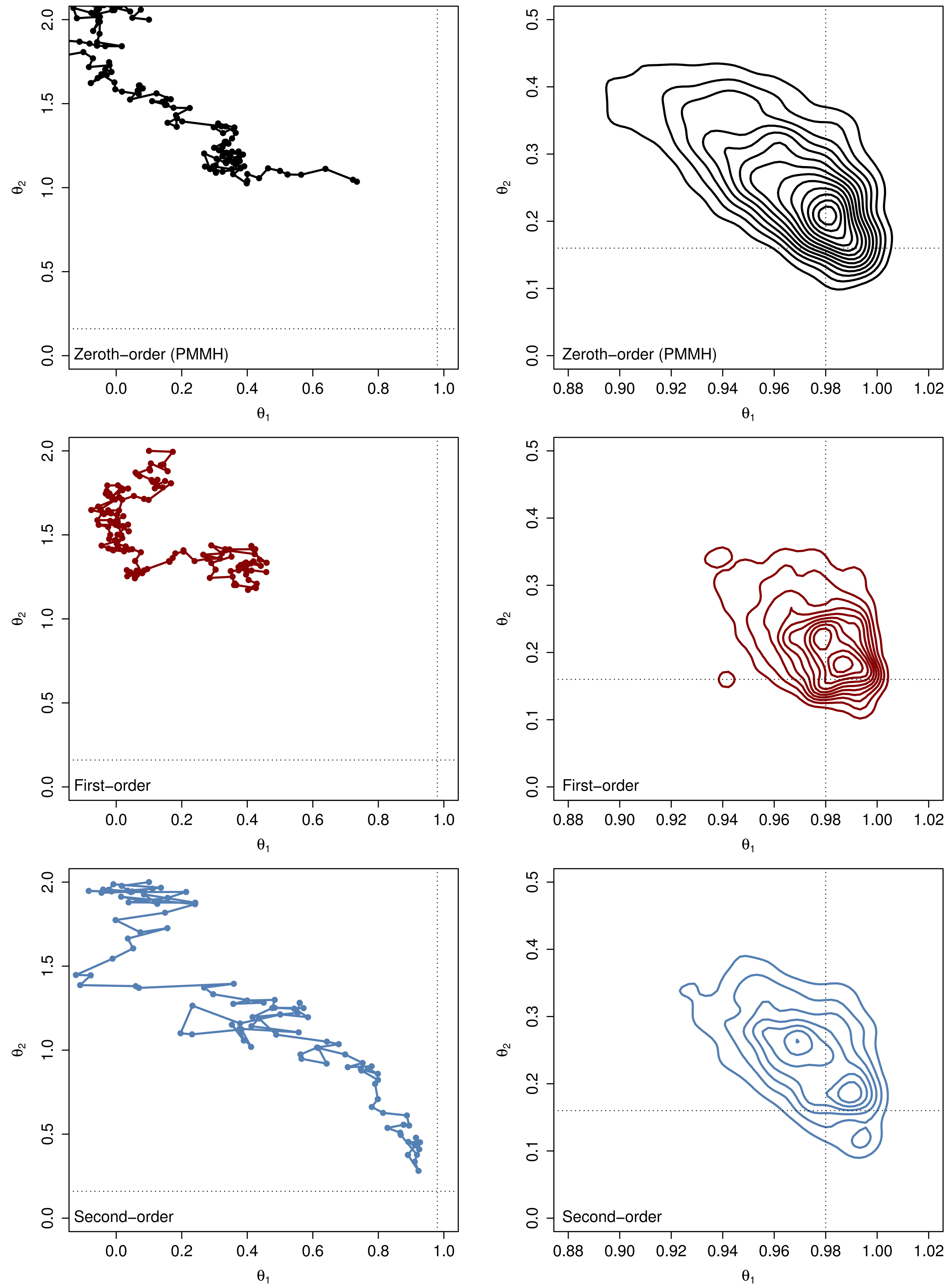
$$\begin{aligned}\widehat{\mathcal{I}}_T(\theta) = \left[\widehat{\mathcal{S}}_T(\theta)\right]^2 &- \int \left[\nabla_\theta \log p_\theta(x_{1:T}, y_{1:T})\right]^2 \widehat{p}_\theta(x_{1:T}|y_{1:T})\, \mathrm{d}x_{1:T} \\ &- \int \left[\nabla_\theta^2 \log p_\theta(x_{1:T}, y_{1:T})\right] \widehat{p}_\theta(x_{1:T}|y_{1:T})\,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

which follows from **Louis' identity**.

Illustration: Stochastic Volatility

$$x_{t+1}|x_t \sim \mathcal{N}\left(x_{t+1}; \boldsymbol{\theta}_1 x_t, \boldsymbol{\theta}_2^2\right),$$
$$y_t|x_t \sim \mathcal{N}\left(y_t; 0, 0.65^2\exp(x_t)\right),$$

with parameters $\boldsymbol{\theta}^\star = \{\boldsymbol{\theta}_1^\star, \boldsymbol{\theta}_2^\star\} = \{0.98, 0.16\}$. We use $T = 250$ time steps, $N = 1\,000$ particles and $M = 10\,000$ MCMC iterations.



More information and source code

http://work.johandahlin.com/