Particle Metropolis-Hastings using Langevin Dynamics

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

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Summary

- We propose an approach to incorporate gradient information into the proposal distribution for estimating parameters in nonlinear state space models.
- This work combines recent advances in Particle MCMC and Sequential MC with well-known results from Hamiltonian MCMC.

Bayesian inference in SSM

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

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

given a set of observations $y_{1:T} = \{y_t\}_{t=1}^T$ and where θ denotes the parameters. The Bayesian approach to this problem requires the computation of the parameter posterior distribution given by

$$p(\theta|y_{1:T}) = \frac{p(y_{1:T}|\theta)p(\theta)}{p(y_{1:T})},$$

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

This problem can be solved by using **Particle Marginal Metropolis-Hastings**. Let the proposal and target distributions be denoted by $q(\theta'|\theta'')$ and $\pi(\theta) = \widehat{p}(y_{1:T}|\theta)p(\theta)$, respectively. Then, the acceptance probability is given by

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

where $\widehat{p}(y_{1:T}|\theta)$ denotes the estimated likelihood function obtained from a particle filter.

New samples are often generated via a **Gaussian random walk**

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

which is known to scale inefficiently in higher dimensions and might result in a long burn-in period.

Incorporating gradient information

Main idea: Use Sequential Monte Carlo to estimate the score function (the gradient of the target distribution) and to incorporate this information into the proposal.

This idea results in sampling form a proposal of the form

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

for some user-defined step length ϵ. Here, $\widehat{S}_T(\theta'')$ denotes an unbiased estimate of the score function obtained using **Fisher's identity** by

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

where $\widehat{p}_{\theta''}(x_{1:T}|y_{1:T})$ denotes the empirical distribution obtained from any particle smoother. A proposal distribution on the form (2) is commonly referred to as following a **Langevin dynamics**.

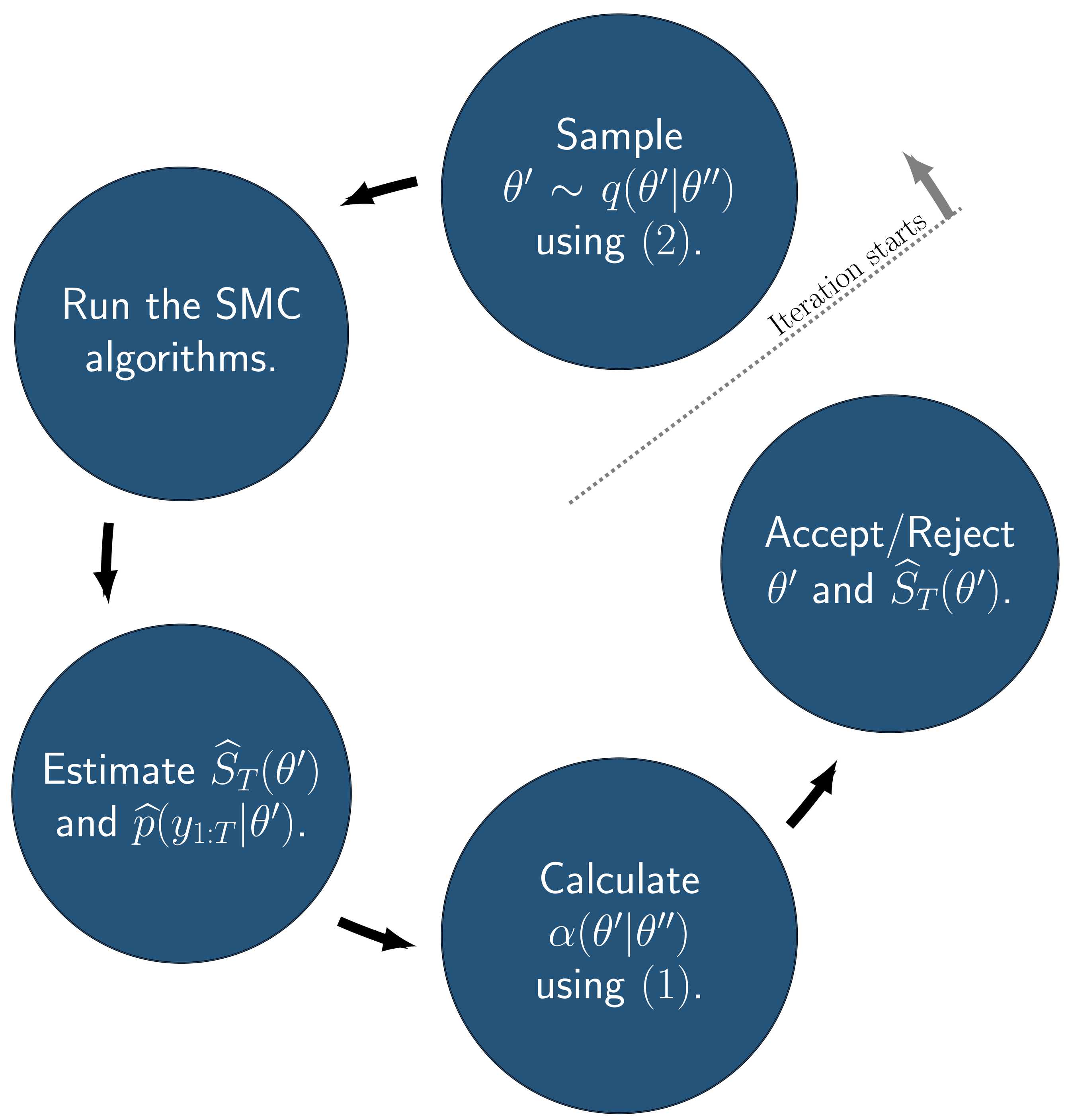


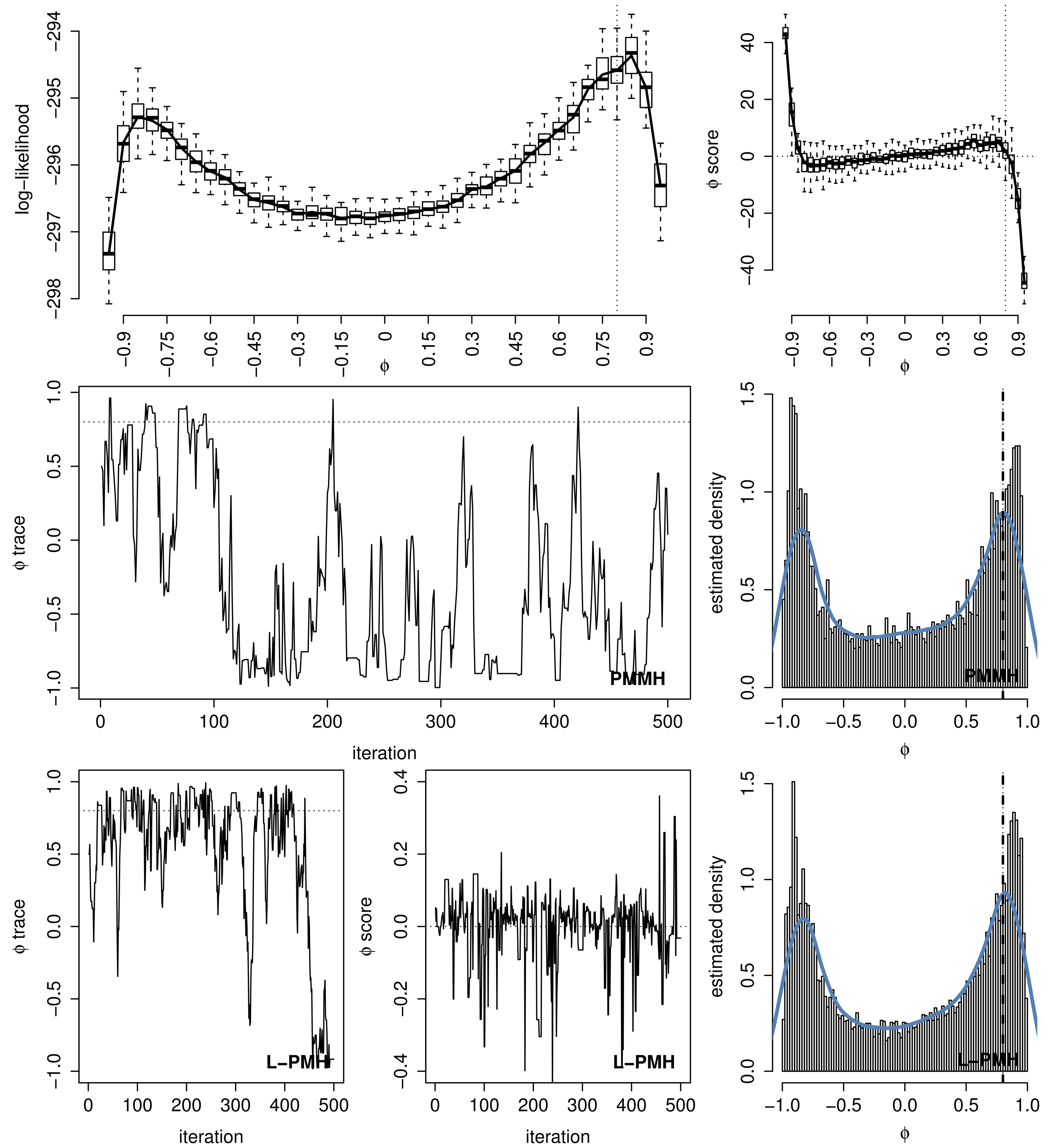
Figure: The steps in one iteration of the L-PMH algorithm.

Example: Stochastic volatility model

Consider the Hull-White model with the form

$$x_{t+1}|x_t \sim \mathcal{N}\left(x_{t+1}; \phi x_t, \sigma^2\right),$$
$$y_t|x_t \sim \mathcal{N}\left(y_t; 0, \beta^2 \exp(x_t)\right),$$

with the true parameters $\theta^\star = \{\phi^\star, \sigma^\star, \beta^\star\} = \{0.8, 0.2, 0.7\}$. To estimate the parameter ϕ we use $T = 500$ time steps with the known initial state $x_0 = 0$ and $N = 1000$ particles.



More information and source code

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

[†]{johan.dahlin,lindsten,schon}@isy.liu.se