Quasi-Newton particle Metropolis-Hastings

Johan Dahlin, Fredrik Lindsten and Thomas B. Schön

Main ideas and contribution

- Quasi-Newton update to estimate the Hessian of $\log \pi(\theta)$.
- Incorporate this information into proposal $q(\theta'|\theta, u)$.
- Reduces the need of tedious pilot runs to find ϵ and $\mathcal{P}$.
- Decreases the inefficiency (correlation) in the Markov chain exploring $\pi(\theta)$, which accelerates the algorithm.

Bayesian parameter inference

Parameter inference in state space models (SSMs),

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

is based on the **parameter posterior** given by

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

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

Particle Metropolis-Hastings

We can sample from $\pi(\theta)$ by simulating

$$\theta' \sim q\Big(\theta'\Big|\theta, u\Big) = \mathcal{N}\Big(\theta'; \mu(\theta, u), \Sigma(\theta, u)\Big), \qquad (1)$$

using auxiliary variables u and accept θ' with probability

$$\alpha(\theta', u', \theta, u) = \min\left\{1, \frac{\widehat{\pi}(\theta'|u')}{\widehat{\pi}(\theta|u)} \frac{q(\theta|\theta', u')}{q(\theta'|\theta, u)}\right\},$$

where $\widehat{\pi}(\theta|u)$ denotes a particle estimate of $\pi(\theta)$.

Proposal	$\mu(\theta, u)$	$\Sigma(\theta, u)$
PMH0	θ	$\epsilon_0^2\,\mathcal{P}^{-1}$
PMH1	$\theta + \frac{1}{2}\epsilon_1^2\,\mathcal{P}^{-1}\widehat{\mathcal{G}}(\theta\|u)$	$\epsilon_1^2\,\mathcal{P}^{-1}$
PMH2	$\theta + \frac{1}{2}\epsilon_2^2\,\widehat{\mathcal{H}}^{-1}(\theta\|u)\,\widehat{\mathcal{G}}(\theta\|u)$	$\epsilon_2^2\widehat{\mathcal{H}}^{-1}(\theta\|u)$

Table: Proposals with the gradient $\mathcal{G}(\theta) = \nabla \log \pi(\theta)$, negative Hessian $\mathcal{H}(\theta) = -\nabla^2 \log \pi(\theta)$ and pre-conditioning matrix $\mathcal{P}$.

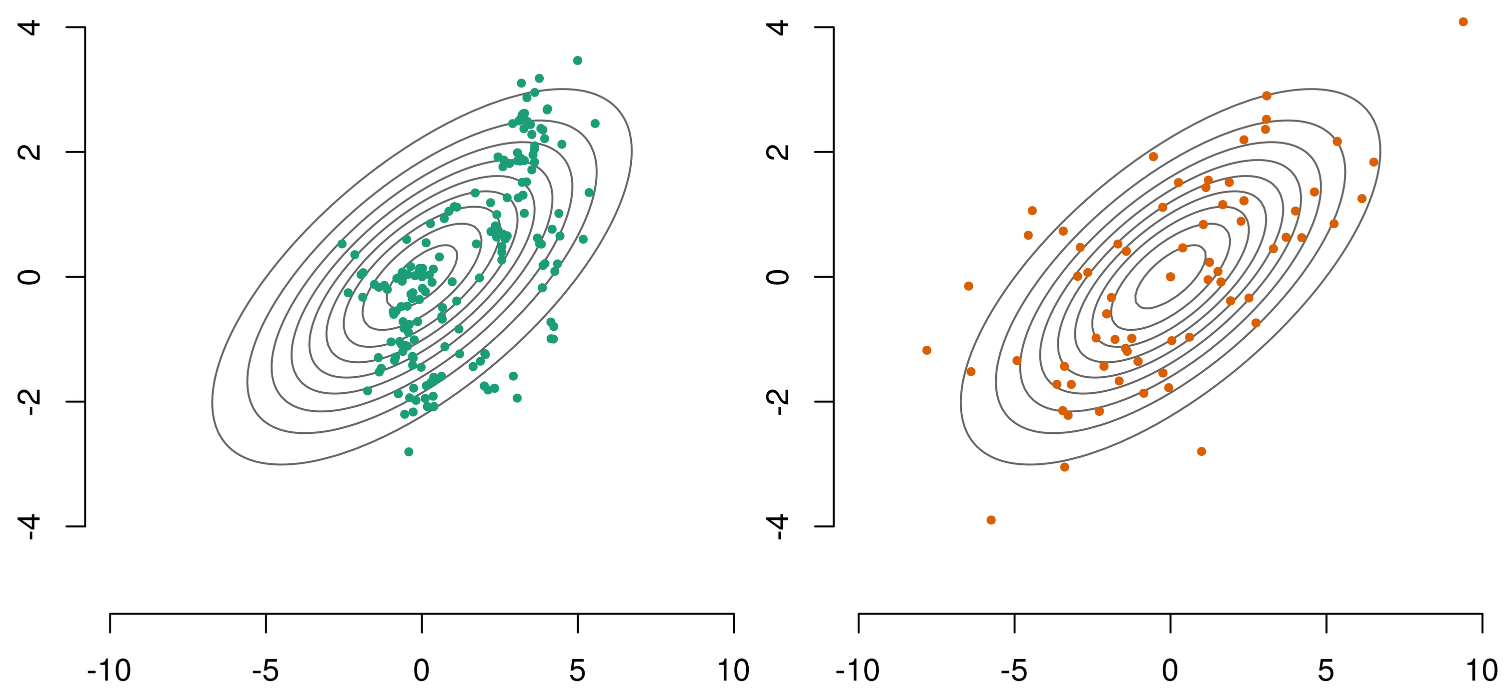


Figure: Incorporating $\mathcal{H}(\theta)$ into $q(\theta'|\theta, u)$ is important when the posterior is non-isotropic. Compare the exploration when $\mathcal{P} = \epsilon_0^2 \mathbf{I}_p$ (left) and $\mathcal{P} = \mathcal{H}(\theta)$ (right).

Particle smoothing for estimating $\mathcal{G}(\theta)$

We can estimate $\mathcal{G}(\theta)$ using the **Fisher identity**

$$\widehat{\mathcal{G}}(\theta|u) = \int \nabla \log p_\theta(x_{1:T}, y_{1:T})\, \widehat{p}_\theta^N(x_{1:T}|y_{1:T}, u)\, \mathrm{d}x_{1:T},$$

$$\nabla \log p_\theta(x_{1:T}, y_{1:T}) = \sum_{t=1}^{T} \Big[\nabla \log f_\theta(x_t|x_{t-1}) + \nabla \log g_\theta(y_t|x_t)\Big],$$

which can be approximated using the particle system

$$u = \left[\left\{x_t^{(i)}, a_t^{(i)}\right\}_{i=1}^{N}\right]_{t=1}^{T}$$

which can be obtained from any particle smoother.

Quasi-Newton update for estimating $\mathcal{H}(\theta)$

We can then estimate $\mathcal{H}(\theta)$ using $\widehat{\mathcal{G}}(\theta|u)$ by iterating

$$B_{l+1}^{-1}(\theta') = \big(\mathbf{I}_p - \rho_l s_l g_l^\top\big) B_l^{-1} \big(\mathbf{I}_p - \rho_l g_l s_l^\top\big) + \rho_l s_l s_l^\top, \qquad (2)$$

$$s_l = \theta_{I(l)} - \theta_{I(l-1)}, \qquad g_l = \widehat{\mathcal{G}}(\theta_{I(l)}|u_{I(l)}) - \widehat{\mathcal{G}}(\theta_{I(l-1)}|u_{I(l-1)}),$$

over $l \in \{1, 2, \ldots, M-1\}$ with $I(l) = k - l$ and $\rho_l^{-1} = g_l^\top s_l$.

The qPMH2 proposal is obtained from (1) by

$$\mu(\theta, u) = \theta_{k-M}, \qquad \Sigma(\theta, u) = \widehat{\mathcal{H}}^{-1}(\theta|u) = -B_M^{-1}(\theta).$$

This follows from a lag-M dependency introduced into the proposal $q(\theta'|\theta, u)$ by using (2).

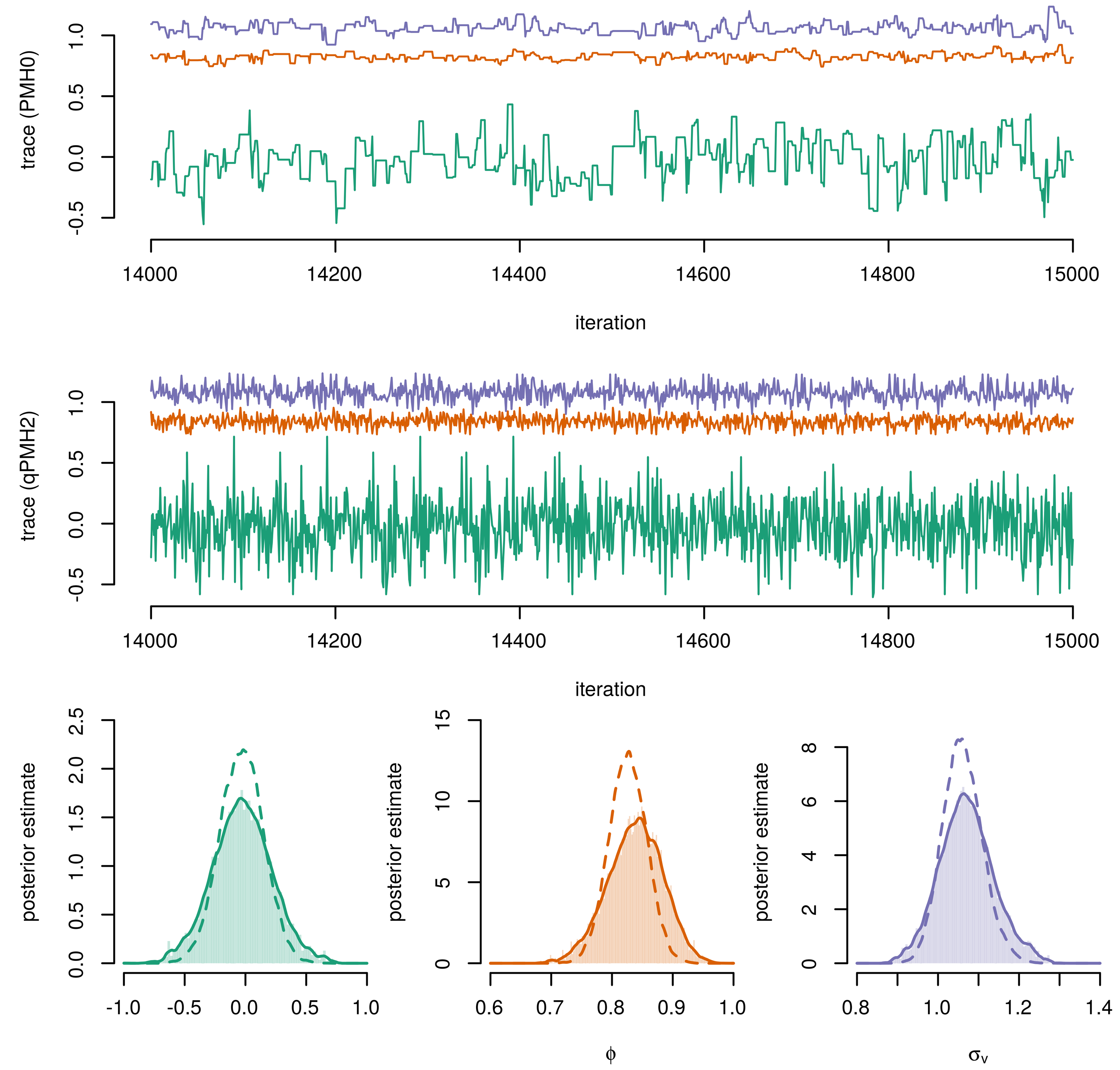
Numerical illustration

Consider the linear Gaussian SSM with $\theta = \{\mu, \phi, \sigma_e\}$,

$$x_t|x_{t-1} \sim \mathcal{N}\Big(x_t; \mu + \phi(x_{t-1} - \mu), \sigma_v^2\Big), \qquad y_t|x_t \sim \mathcal{N}\Big(y_t; x_t, 0.1^2\Big).$$

We simulate $T = 250$ obs. with $\theta = \{0.20, 0.80, 1.0\}$ and compare the inefficiency factors (IFs) when estimating $\pi(\theta)$.

Alg.	Acc. rate	min IF	max IF
PMH0	0.28	12.13	13.71
PMH1	0.78	11.28	14.50
qPMH2	0.55	**3.00**	**3.01**



Paper and source code

Available at `http://work.johandahlin.com/`

LINKÖPING UNIVERSITY
UNIVERSITY OF CAMBRIDGE
UPPSALA UNIVERSITET

LINKÖPING UNIVERSITY
Department of Electrical Engineering