Quasi-Newton particle Metropolis-Hastings

Johan Dahlin, Fredrik Lindsten and Thomas B. Schön

Main idea and results

- Use quasi-Newton update to estimate the Hessian of the log-posterior.
- Enables implementation of manifold MALA proposals that only requires estimates of the gradient of the log-posterior.
- Improves mixing and reduces the need of tedious pilot runs.

Bayesian parameter inference

Parameter inference in state space models (SSMs),

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

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 and likelihood, respectively.

Particle Metropolis-Hastings

We can sample from $\pi(\theta)$ by simulating a **Gaussian random walk**

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

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 filter-based estimate of $\pi(\theta)$.

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

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

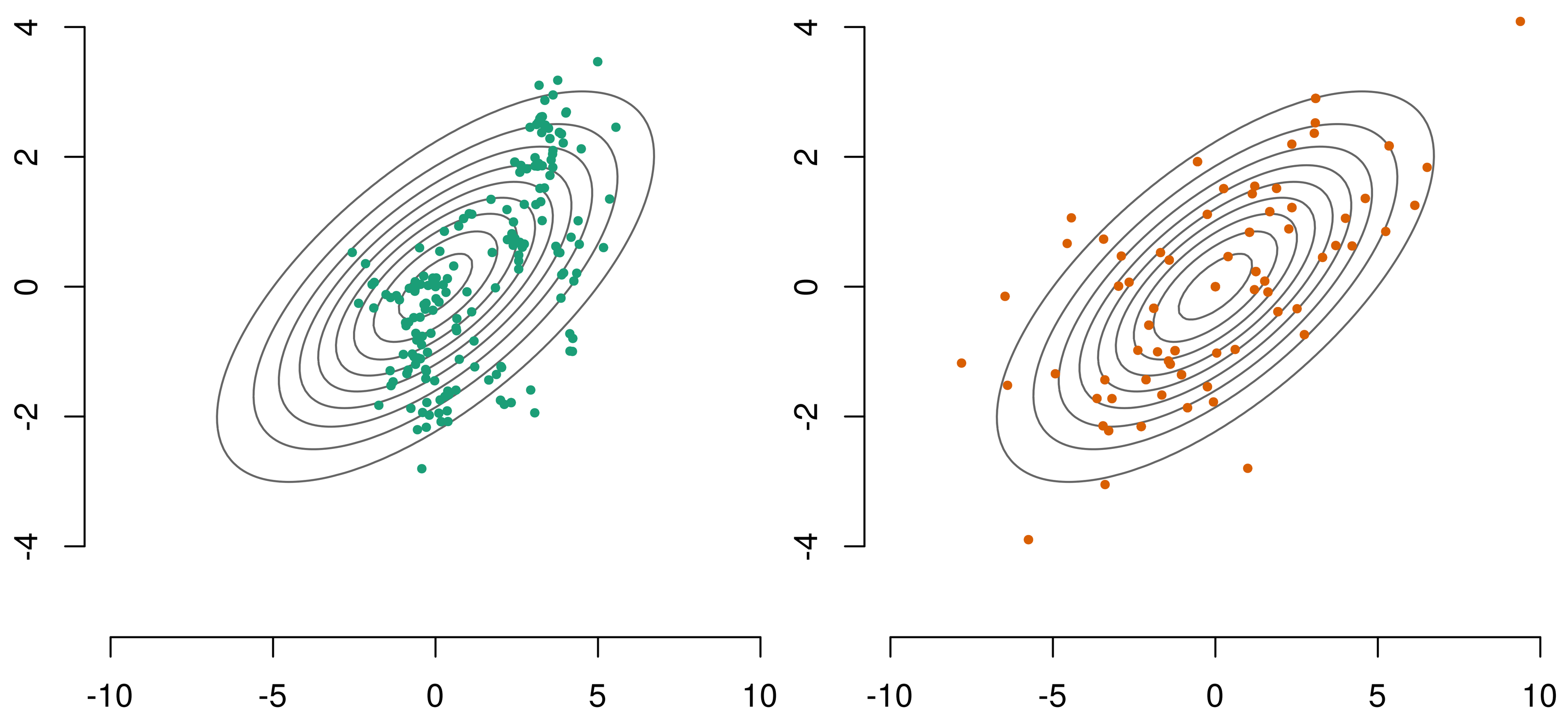


Figure: Taking the covariance of $\pi(\theta)$ into account is important when the posterior is non-isotropic. Compare the exploration when $\mathcal{P} = \epsilon_0^2 \mathbf{I}_p$ (left) and $\mathcal{P} = \mathcal{H}(\theta)$ (right).

Quasi-Newton update for Hessian

The estimate of $\mathcal{H}(\theta)$ is obtained by iterating

$$\begin{aligned} B_{l+1}^{-1}(\theta') &= \big(\mathbf{I}_p - \rho_l s_l g_l^\top\big) B_l^{-1} \big(\mathbf{I}_p - \rho_l g_l s_l^\top\big) + \rho_l s_l s_l^\top, \text{ with} \\ \rho_l^{-1} &= g_l^\top s_l, \\ s_l &= \theta_{I(l)} - \theta_{I(l-1)}, \\ g_l &= \widehat{\mathcal{G}}(\theta_{I(l)}|u_{I(l)}) - \widehat{\mathcal{G}}(\theta_{I(l-1)}|u_{I(l-1)}), \end{aligned}$$

over $l \in \{1, 2, \ldots, M-1\}$ with $I(l) = k - l$, i.e. over the $M-1$ previous states of the Markov chain, after which we obtain $\widehat{H}(\theta'|u') = -B_M(\theta')$.

The advantage with this update is that it only requires estimates of $\mathcal{G}(\theta)$, which can be obtained by the **Fisher identity**

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

approximated using the particle system $u = \Big(\{x_t^{(i)}, a_t^{(i)}\}_{i=1}^N\Big)_{t=1}^T$.

The quasi-Newton update induces a lag-M dependency in the proposal. To obtain a valid MCMC we view the resulting algorithm as a standard MCMC for the M-fold product of the target (Zhang and Sutton, 2011),

$$\boldsymbol{\pi}(\theta_{k,1:M}) = \prod_{i=1}^{M} \pi(\theta_{k,i}).$$

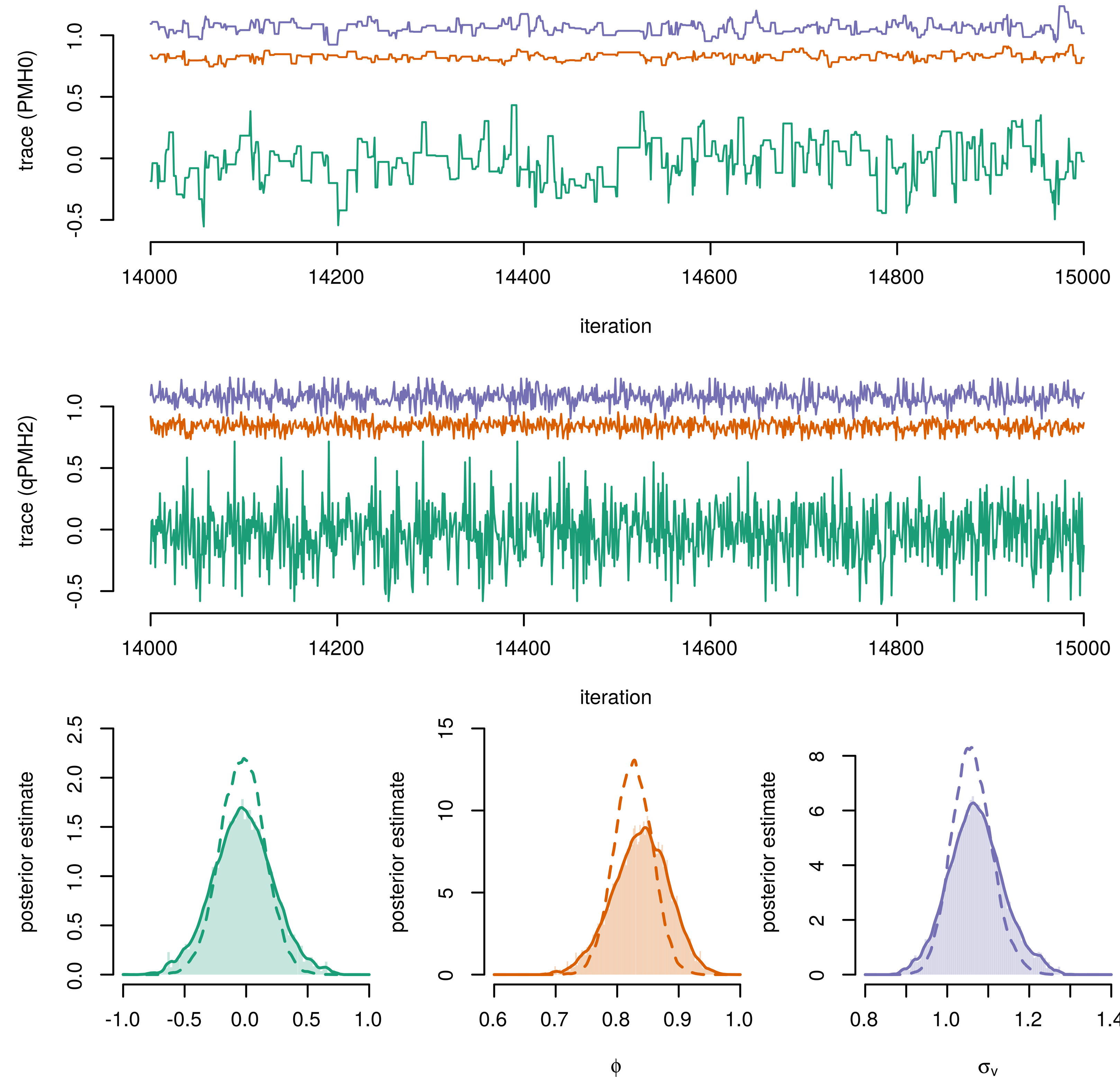
Numerical illustration

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

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

We simulate $T = 250$ obs. with $\theta = \{0.20, 0.80, 1.0\}$ and compute the integrated autocorrelation times (IACTs) when estimating $\pi(\theta)$.

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



More information and source code are available at `http://work.johandahlin.com/`.

LINKÖPING UNIVERSITY
UNIVERSITY OF CAMBRIDGE
UPPSALA UNIVERSITET

References

Dahlin, Lindsten and Schön, **PMH using gradient and Hessian information.** Statistics and Computing 25(1), pp 81-92, Springer, 2015.

Dahlin, Lindsten and Schön, **Quasi-Newton PMH.** Proceedings of the 17th IFAC Symposium on SYSID, Beijing, China, October 2015.

Zhang and Sutton, **Quasi-Newton methods for MCMC.** Advances in NIPS 24, 2011.