Accelerating delayed-acceptance Markov chain Monte Carlo algorithms

Samuel Wiqvist, Centre for Mathematical Sciences, Lund University, Sweden

samuel.wiqvist@matstat.lu.se http://www.maths.lu.se/staff/samuel-wiqvist/

Joint work with Umberto Picchini (University of Gothenburg) and Julie Lyng Forman (University of Copenhagen), see [3].

Introduction

- We build upon the delayed-acceptance scheme (DA-MCMC) [1] and develop an accelerated delayed-acceptance scheme (ADA-MCMC) [3].
- As a case study we introduce a novel double-well potential SDE for modelling of protein folding data (reaction coordinate path).

Conclusions

- Our ADA-MCMC algorithm is particularly useful in settings where it is a computationally demanding task to evaluate or approximate the likelihood.
- The acceleration obtained via ADA-MCMC is problem dependent. Higher accelerations are obtained for computationally challenging problems.
- We obtain a 2-6 folds reduction in number of evaluations of the expensive likelihood compared to DA-MCMC for our challenging case study.
- Inference results with the ADA-MCMC algorithm and standard MCMC algorithms are similar.
- ADA-MCMC can be adapted to target a generic distribution $p(x)$ outside the Bayesian framework.

Delayed-acceptance MCMC

- DA-MCMC (due to [1]) is a known strategy to deal with expensive likelihood functions that utilizes a computationally cheap *surrogate* model $\tilde{L}(\theta)$ of the likelihood $L(\theta)$.

Pseudo-code

1. Generate $\tilde{L}(\theta^\star)$ and $\tilde{L}(\theta^{r-1})$ from *some* fast to simulate surrogate model.
2. First stage:

$$\alpha_1 = \min\big(1, \frac{\tilde{L}(\theta^\star) g(\theta^{r-1} \mid \theta^\star)\pi(\theta^\star)}{\tilde{L}(\theta^{r-1}) g(\theta^\star \mid \theta^{r-1})\pi(\theta^{r-1})}\big).$$

   Reject $\theta^\star$ with probability $1-\alpha_1$, otherwise go to second stage.
3. Second stage:

$$\alpha_2 = \min(1, \frac{L(\theta^\star)\tilde{L}(\theta^{r-1})}{L(\theta^{r-1})\tilde{L}(\theta^\star)}).$$

   Accept $\theta^\star$ with probability α_2. Update chain.

Accelerated delayed-acceptance MCMC

- The DA-MCMC algorithm is governed by the four values: $L(\theta^\star)$, $L(\theta^{r-1})$, $\tilde{L}(\theta^\star)$, and $\tilde{L}(\theta^{r-1})$, which can be arranged in for cases:

$$\begin{aligned}
&1)\ \tilde{L}(\theta^\star) > \tilde{L}(\theta^{r-1}) \text{ and } L(\theta^\star) > L(\theta^{r-1}),\\
&2)\ \tilde{L}(\theta^\star) < \tilde{L}(\theta^{r-1}) \text{ and } L(\theta^\star) < L(\theta^{r-1}),\\
&3)\ \tilde{L}(\theta^\star) > \tilde{L}(\theta^{r-1}) \text{ and } L(\theta^\star) < L(\theta^{r-1}),\\
&4)\ \tilde{L}(\theta^\star) < \tilde{L}(\theta^{r-1}) \text{ and } L(\theta^\star) > L(\theta^{r-1}).
\end{aligned}$$

- Assume we are in case 1: we can then split the acceptance region (see Figure 1):

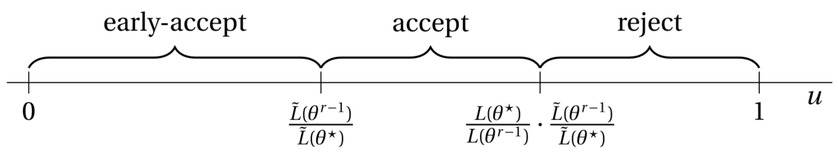


Figure 1: Acceptance regions for case 1.

- In case 1 we have a possibility for fast acceptance (early-accept) using the cheap surrogate only.
- Similar analyses can be done for the other three cases.
- But how to know which case of the four to assume for a new proposal $\theta^\star$? We introduce the selection models $s_{1,3}(\cdot)$ and $s_{2,4}(\cdot)$, that are used to select a case for a new proposal.

Pseudo-code

1. Generate $\tilde{L}(\theta^\star)$ and $\tilde{L}(\theta^{r-1})$ from *some* fast to simulate surrogate model.
2. First stage:

$$\alpha_1 = \min\big(1, \frac{\tilde{L}(\theta^\star) g(\theta^{r-1} \mid \theta^\star)\pi(\theta^\star)}{\tilde{L}(\theta^{r-1}) g(\theta^\star \mid \theta^{r-1})\pi(\theta^{r-1})}\big).$$

   Reject $\theta^\star$ with probability $1-\alpha_1$, otherwise go to second stage.
3. Second stage:
   - 3.1. If $\tilde{L}(\theta^\star) > \tilde{L}(\theta^{r-1})$:
     - i. Select case 1 or 3 according to the model $s_{1,3}(\theta^\star)$.
     - ii. Run the accelerated delayed-acceptance scheme for the selected case.
   - 3.2. Else:
     - i. Select case 2 or 4 according to the model $s_{2,4}(\theta^\star)$.
     - ii. Run the accelerated delayed-acceptance scheme for the selected case.

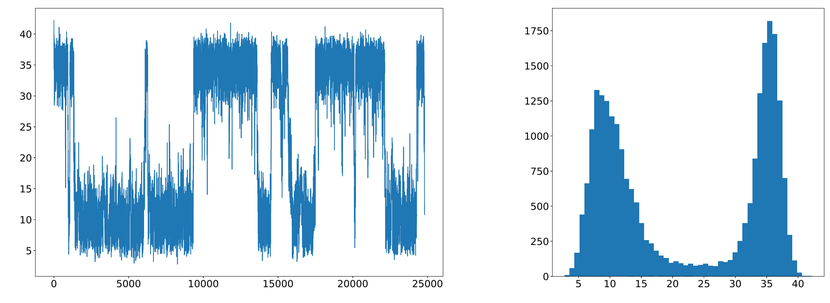
Comparing DA-MCMC and ADA-MCMC

- DA-MCMC targets the exact posterior, while ADA-MCMC targets an approximate posterior distribution. The approximation in ADA-MCMC is introduced by the probabilistic error of selecting a case for a new proposal, and by sometimes accepting proposals only based on the surrogate model.
- Unlike DA-MCMC, in ADA-MCMC we can sometimes skip completely the evaluation of the expensive $L(\theta)$, even at the second stage.

This poster was created using the *LaTeX Portrait Poster Template*

Case study: modeling of protein folding data

- We are interested in developing a statistical model for protein folding data (reaction coordinate path, sample size $n = 2.5 \times 10^4$, see Figure 2).



(a) Data. **(b)** Marginal distribution of the data..

Figure 2: Protein folding data.

- The novel double-well potential stochastic differential equation (DWP-SDE) model in (1) is a model with additive red noise where the stationary distribution is modeled with the potential function V in (2).

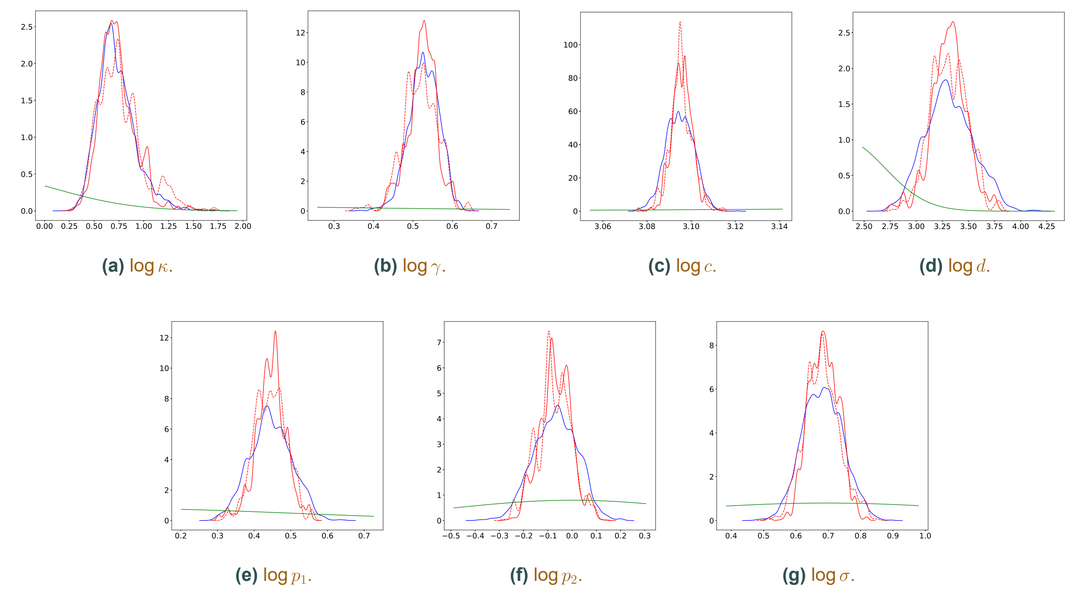
$$\begin{cases} Z_t = X_t + Y_t. \\ dX_t = -\nabla V(X_t)\,dt + \sigma\, dW_t^X. \\ dY_t = -\kappa Y_t\,dt + \sqrt{\kappa\gamma^2}\, dW_t^Y. \end{cases} \tag{1}$$

$$V(x, c, d, A, g, p_1, p_2) = \frac{1}{2}\Big|\frac{1}{2}|x-c|^{p_1} - d + gx\Big|^{p_2} + \frac{1}{2}Ax^2. \tag{2}$$

- The likelihood function is intractable for our DWP-SDE model and therefore we used pseudo-marginal MCMC algorithms to estimate the parameters of the model.
- In this case study we employ as Gaussian process (GP) based surrogate model, similar to the GP model used in [2].

Inference results

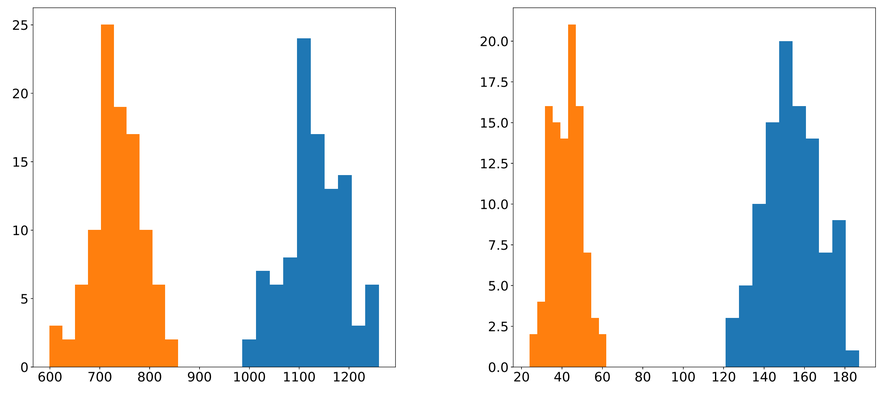
- The marginal posteriors in Figure 3 show that we obtain similar inference results when we compare ADA-GP-MCMC to standard algorithms (Markov chain within Metropolis (MCWM), and DA-GP-MCMC).
- The speed-up analyses in Figure 4 (where we run ADA-GP-MCMC and DA-GP-MCMC 100 times to evaluate the performance of the algorithms) show that we indeed obtain an acceleration for ADA-GP-MCMC, and that we obtain a reduction in the number of particle filter evaluations in the second stage compared to DA-GP-MCMC.



(a) $\log \kappa$. **(b)** $\log \gamma$. **(c)** $\log c$. **(d)** $\log d$.

(e) $\log p_1$. **(f)** $\log p_2$. **(g)** $\log \sigma$.

Figure 3: Marginal posteriors. MCWM (blue solid line), DA-GP-MCMC (red solid line), and ADA-GP-MCMC (red dashed line). Priors are denoted with green lines.



(a) Histogram of run-times (secounds.). **(b)** Number of particle filter evaluations in second stage.

Figure 4: Speed up analysis of 100 repetitions. DA-GP-MCMC in blue, ADA-GP-MCMC in orange.

References

[1] J A. Christen and C. Fox. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical statistics*, 14(4):795–810, 2005.

[2] C. C Drovandi, M. T Moores, and R. J Boys. Accelerating pseudo-marginal MCMC using Gaussian processes. *Computational Statistics & Data Analysis*, 118:1–17, 2018.

[3] S. Wiqvist, U. Picchini, and J. Forman. Accelerating delayed-acceptance markov chain monte carlo algorithms. *arXiv preprint arXiv:1806.05982*, 2018.