

THE CONTAINMENT CONDITION AND ADAPFAIL ALGORITHMS

KRZYSZTOF ŁATUSZYŃSKI,* *University of Warwick*

JEFFREY S. ROSENTHAL,** *University of Toronto*

Abstract

This short note investigates convergence of adaptive Markov chain Monte Carlo algorithms, i.e. algorithms which modify the Markov chain update probabilities on the fly. We focus on the containment condition introduced Roberts and Rosenthal (2007). We show that if the containment condition is *not* satisfied, then the algorithm will perform very poorly. Specifically, with positive probability, the adaptive algorithm will be asymptotically less efficient than *any* nonadaptive ergodic MCMC algorithm. We call such algorithms AdapFail, and conclude that they should not be used.

Keywords: Markov chain Monte Carlo; adaptive MCMC; containment condition; ergodicity; convergence rate

2010 Mathematics Subject Classification: Primary 60J05

Secondary 65C05

1. Introduction

Markov chain Monte Carlo (MCMC) algorithms are used to sample from complicated probability distributions. They proceed by simulating an ergodic Markov chain with transition kernel P and stationary distribution of interest, say π . Unlike in the case of independent and identically distributed (i.i.d.) Monte Carlo algorithms, the MCMC output

$$X_0, X_1, \dots, X_n, \dots$$

is a correlated sample. Nevertheless, if the Markov chain is ergodic (i.e. converges in distribution to π), then the asymptotic validity is retained under appropriate conditions (see, e.g. [12], [17]). In particular, for large enough M and the subsampled random variables

$$X_M, X_{2M}, \dots, X_{nM}, \dots \tag{1.1}$$

are approximately independent draws from the target distribution π . For the MCMC-based statistical inference to be reliable, it is essential to design algorithms that mix quickly, i.e. for which the asymptotic i.i.d. property in (1.1) holds with reasonably small M . (Note, however, that for estimation purposes, subsampling is desirable only if the cost of using the sample is substantial compared to the cost of generating samples, otherwise the entire sample should be used; see [5, Section 3.6].)

In a typical MCMC setting, the algorithm is determined by a Markov chain transition kernel P_θ , where $\theta \in \Theta$ is a high dimensional tuning parameter, e.g. the covariance matrix of a

Received 19 June 2013; revision received 23 December 2013.

* Postal address: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.

Email address: latuch@gmail.com

** Postal address: Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada.

Email address: jeff@math.toronto.edu

random walk Metropolis proposal [21], [16], or the vector of random scan Gibbs sampler selection probabilities [11]. Usually the parameter space Θ is large, and for ‘good’ values of θ , the iterates P_θ^n will converge quickly to π as n increases, resulting in small M in (1.1). However, such ‘good’ values are often very difficult to find, and for most values of θ the iterates P_θ^n will converge arbitrary slowly.

Since a good θ is difficult to find manually, the idea of adaptive MCMC was introduced [6], [9] to enable the algorithm to learn ‘on the fly’, and redesign the transition kernel during the simulation as additional information about π becomes available. Thus, an adaptive MCMC algorithm would apply the transition kernel P_{θ_n} for obtaining X_n from X_{n-1} , where the choice of the tuning parameter θ_n at the n th iteration is itself a random variable which may depend on the whole history X_0, X_1, \dots, X_{n-1} and on θ_{n-1} . When using adaptive MCMC, we hope that the adaptive parameter θ_n will settle on ‘good’ values, and that the adaptive algorithm will inherit the corresponding good convergence properties.

Unfortunately, since adaptive algorithms violate the Markovian property they are inherently difficult to analyse theoretically. Whereas the interest in adaptive MCMC is fuelled by some very successful implementations for challenging problems (see, for instance, [1], [7], [8], [14],[19], and [22]), many seemingly reasonable adaptive MCMC algorithms are provably transient or converge to a wrong probability distribution (see, for instance, [2], [3], [10]), and [11]. Thus, the theoretical foundations of adaptive MCMC are a very important topic which is still under active development.

One general and relatively simple approach to analysing adaptive MCMC algorithms was presented in [18], which showed that the two properties of *diminishing adaptation* and *containment* were sufficient to guarantee that an adaptive MCMC algorithm would converge asymptotically to the correct target distribution (at some rate). While the diminishing adaptation property is fairly standard and can be easily controlled by the user, the containment property is more subtle and can be challenging to verify (see, for example, [3]). This leads to the question of how important or useful the containment condition actually is, especially since it is known (see, for example, [4]) that containment is not a necessary condition for the ergodicity of an adaptive MCMC algorithm.

The purpose of this short note is to show that if containment does not hold, then the adaptive algorithm will perform very poorly. Specifically, with positive probability the adaptive algorithm will be asymptotically less efficient than *any* nonadaptive MCMC algorithm. Here, efficiency is understood as the total variation distance convergence time (see [13] for a different concept of efficiency in terms of asymptotic variance in the central limit theorem). In effect, the approximate i.i.d. property in (1.1) will be violated for *any* finite M . We call such algorithms AdapFail, and conclude that they should not be used. In particular, we argue that the containment condition is actually a reasonable condition to impose on adaptive MCMC algorithms, since without it they perform so poorly as to be unusable.

This paper is structured as follows. In Section 2 we define and characterise the class of AdapFail algorithms. In Section 3 we relate the AdapFail property to the containment condition. In Section 4 we present a very simple example to illustrate our results.

2. The class of AdapFail algorithms

We first introduce necessary notation (see, for example, [12], [17], and [18] for a more complete development related to Markov chains and adaptive MCMC). Let P_θ , parametrized by $\theta \in \Theta$, be a transition kernel of a Harris ergodic Markov chain on $(\mathcal{X}, \mathcal{F})$ with stationary distribution π . Thus, for all $x \in \mathcal{X}$ and $\theta \in \Theta$ we have $\lim_{n \rightarrow \infty} \|P_\theta^n(x, \cdot) - \pi(\cdot)\| = 0$, where

$\|v(\cdot) - \mu(\cdot)\| := \sup_{A \in \mathcal{F}} |v(A) - \mu(A)|$ is the usual total variation norm. We shall also use the ‘ ε convergence time function’ $M_\varepsilon: \mathcal{X} \times \Theta \rightarrow \mathbb{N}$ defined as

$$M_\varepsilon(x, \theta) := \inf\{n \geq 1: \|P_\theta^n(x, \cdot) - \pi(\cdot)\| \leq \varepsilon\}.$$

Let $\{(X_n, \theta_n)\}_{n=0}^\infty$ be a corresponding adaptive MCMC algorithm, where X_n is updated from X_{n-1} using P_{θ_n} for some Θ -valued random variable θ_n (which might depend on the chain’s history and on θ_{n-1}). For the adaptive algorithm, denote the marginal distribution at time n by

$$A^{(n)}((x, \theta), B) := \mathbb{P}(X_n \in B \mid X_0 = x, \theta_0 = \theta),$$

and say that the algorithm is *ergodic* for starting values x and θ if

$$\lim_{n \rightarrow \infty} \|A^{(n)}((x, \theta), \cdot) - \pi(\cdot)\| = 0.$$

Similarly, let the ‘ ε convergence time function’ for the adaptive case be

$$M_\varepsilon^A(x, \theta) := \inf\{n \geq 1: \|A^{(n)}((x, \theta), \cdot) - \pi(\cdot)\| \leq \varepsilon\}.$$

In both cases the function $M_\varepsilon(x, \theta)$ has the same interpretation: it is the number of iterations that the algorithm must take to be within ε of stationarity.

We are now ready to define the class of AdapFail algorithms.

Definition 2.1. Let $\{(X_n, \theta_n)\}_{n=0}^\infty$ evolve according to the dynamics of an adaptive MCMC algorithm \mathcal{A} , with starting values $X_0 = x^*$ and $\theta_0 = \theta^*$. We say that $\mathcal{A} \in \text{AdapFail}$ if there is $\varepsilon_{AF} > 0$ such that

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > M \mid X_0 = x^*, \theta_0 = \theta^*) =: \delta_{AF} > 0. \tag{2.1}$$

Remark 2.1. Intuitively, (2.1) says that the convergence times of the adaptive algorithm will be greater than any fixed value M , i.e. that the algorithm will converge arbitrarily slowly and, thus, perform so poorly as to be unusable.

Remark 2.2. In our experience, the inner limit in (2.1) will typically exist, so that $\limsup_{n \rightarrow \infty}$ can be replaced by $\lim_{n \rightarrow \infty}$ there (and similarly in the related expressions below). However, without assuming specific details about the type of adaptation used, we are unable to make conclusive statements about what conditions guarantee this.

Remark 2.3. For the probabilities in (2.1) to make sense, the function M_ε^A needs to be measurable. This follows from [15, Appendix]. Moreover, if the inner limit in (2.1) is denoted as $\delta_{AF}(M)$ then this sequence is positive and nonincreasing as a function of M , and will thus converge to δ_{AF} as $M \rightarrow \infty$.

Remark 2.4. To obtain the approximate i.i.d. property of the $\{X_n\}$ in (1.1), we want the distribution of $X_{(n+1)M}$ conditionally on the value of X_{nM} to be within ε of the stationary measure, i.e.

$$\|\mathcal{L}(X_{(n+1)M} \mid X_{nM}) - \pi\| \leq \varepsilon. \tag{2.2}$$

Being an AdapFail algorithm means that for any fixed $0 < \varepsilon \leq \varepsilon_{AF}$ and some fixed $\delta_{AF} > 0$, we are infinitely often in a regime where (2.2) is violated for *any* finite M , with probability at least δ_{AF} , further illustrating its poor performance.

The following two results shed additional light on the AdapFail class.

Proposition 2.1. *Any ergodic nonadaptive MCMC algorithm P_θ is not in AdapFail.*

Proof. For a nonadaptive chain, the quantity M_ε^A in (2.1) becomes M_ε and $\theta^* = \theta$. For arbitrary $\varepsilon > 0$ and $\delta > 0$, we shall show that $\delta_{AF} < 2\delta$, from which it follows that $\delta_{AF} = 0$. Indeed, first find n_0 such that $\|P_\theta^{n_0}(x^*, \cdot) - \pi(\cdot)\| < \delta$, and then find M_0 such that $\pi(\{x : M_\varepsilon(x, \theta) \leq M_0\}) > 1 - \delta$. Then for every $n \geq n_0$ and every $M \geq M_0$, we can write

$$\mathbb{P}(M_\varepsilon(X_n, \theta) > M \mid X_0 = x^*) \leq \delta + \pi(\{x : M_\varepsilon(x, \theta) > M\}) < 2\delta.$$

The result follows.

Theorem 2.1. *For an algorithm \mathcal{A} the following conditions are equivalent.*

- (i) *It hold that $\mathcal{A} \in \text{AdapFail}$.*
- (ii) *There are $\varepsilon > 0$ and $\delta > 0$ such that for all $x \in \mathcal{X}, \theta \in \Theta$, and $K > 0$,*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(M_\varepsilon^A(X_n, \theta_n) > K M_\varepsilon(x, \theta) \mid X_0 = x^*, \theta_0 = \theta^*) \geq \delta.$$

- (iii) *There are $\varepsilon > 0$ and $\delta > 0$ such that for all $\theta \in \Theta, K > 0$, and $y^* \in \mathcal{X}$,*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(M_\varepsilon^A(X_n, \theta_n) > K M_\varepsilon(Y_n, \theta) \mid X_0 = x^*, \theta_0 = \theta^*, Y_0 = y^*) \geq \delta,$$

where $\{Y_n\}$ is a Markov chain which follows the dynamics P_θ and is independent of the adaptive process $\{X_n\}$.

Moreover, in (ii) and (iii) we can take $\delta = \delta_{AF}$.

Proof. It is immediate (i) implies (ii) when $\delta = \delta_{AF}$ and $\varepsilon = \varepsilon_{AF}$. To verify (ii) implies (iii), fix $\delta^* > 0$ and using monotonicity of the total variation distance (see [17]) take n_0 such that $\|P_\theta^n(y^*, \cdot) - \pi(\cdot)\| \leq \delta^*$ for every $n > n_0$. Next, find M_0 such that $\pi(\mathcal{X}_{M_0}) > 1 - \delta^*$, where $\mathcal{X}_{M_0} = \{x : M_{\varepsilon_{AF}}(x, \theta) \leq M_0\}$. Then for fixed θ, K , and y^* , compute

$$\begin{aligned} & \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K M_{\varepsilon_{AF}}(Y_n, \theta) \mid X_0 = x^*, \theta_0 = \theta^*, Y_0 = y^*) \\ &= \int_{\mathcal{X}} \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K M_{\varepsilon_{AF}}(x, \theta) \mid X_0 = x^*, \theta_0 = \theta^*) P_\theta^n(y^*, dx) \\ &\geq \int_{\mathcal{X}} \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K M_{\varepsilon_{AF}}(x, \theta) \mid X_0 = x^*, \theta_0 = \theta^*) \pi(dx) - \delta^* \\ &\geq \int_{\mathcal{X}_{M_0}} \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K M_0 \mid X_0 = x^*, \theta_0 = \theta^*) \pi(dx) - \delta^* \\ &\geq (1 - \delta^*) \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K M_0 \mid X_0 = x^*, \theta_0 = \theta^*) - \delta^*. \end{aligned}$$

Consequently,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K M_{\varepsilon_{AF}}(Y_n, \theta) \mid X_0 = x^*, \theta_0 = \theta^*, Y_0 = y^*) \\ & \geq (1 - \delta^*) \limsup_{n \rightarrow \infty} \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K M_0 \mid X_0 = x^*, \theta_0 = \theta^*) - \delta^*. \end{aligned}$$

Since δ^* was arbitrary, (iii) follows from (ii) with $K = KM_0$ and $\delta = \delta_{AF}$. To verify that (iii) implies (i), note that $M_{\varepsilon_{AF}}(Y_n, \theta) \geq 1$, so (iii) gives

$$\limsup_{n \rightarrow \infty} \mathbb{P}(M_{\varepsilon_{AF}}^A(X_n, \theta_n) > K \mid X_0 = x^*, \theta_0 = \theta^*) > \delta_{AF} \quad \text{for every } K > 0.$$

The result follows by taking $K \rightarrow \infty$.

Remark 2.5. Condition (iii) has the interpretation that if we run the adaptive algorithm $\{X_n\}$ and a nonadaptive $\{Y_n\}$ independently on two computers side by side, and monitor the ε convergence time of both algorithms, then as the simulation progresses, the ε convergence time of the adaptive algorithm will infinitely often be bigger by an arbitrarily large factor K , with probability at least δ , i.e. $\{X_n\}$ will be arbitrarily worse than $\{Y_n\}$ (no matter how bad the tuning parameters θ and starting point Y_0 for $\{Y_n\}$ are).

3. Relation to the containment condition

The following condition was introduced in [18] as a tool to analyse adaptive MCMC algorithms.

Definition 3.1. (*Containment condition.*) The algorithm \mathcal{A} with starting values $X_0 = x^*$ and $\theta_0 = \theta^*$ satisfies containment if, for all $\varepsilon > 0$, the sequence $\{M_\varepsilon(X_n, \theta_n)\}_{n=0}^\infty$ is bounded in probability.

It is augmented by the usual requirement of diminishing adaptation.

Definition 3.2. (*Diminishing adaptation.*) The algorithm \mathcal{A} with starting values $X_0 = x^*$ and $\theta_0 = \theta^*$ satisfies diminishing adaptation, if $\lim D_n = 0$ as $n \rightarrow \infty$ in probability, where

$$D_n := \sup_{x \in \mathcal{X}} \|P_{\theta_{n+1}}(x, \cdot) - P_{\theta_n}(x, \cdot)\|.$$

Containment has been extensively studied in [3] and [18] and verified for large classes of adaptive MCMC samplers (compare also [11] and [19]). Together with diminishing adaptation, it guarantees ergodicity. As illustrated in the next section, it is *not* a necessary condition. However, it still turns out to be an appropriate condition to require, due to the following result.

Theorem 3.1. *Assume the diminishing adaptation is satisfied. Then the containment condition does not hold for \mathcal{A} if and only if $\mathcal{A} \in \text{AdapFail}$.*

Proof. The proof utilises a construction similar to the coupling proof of [18, Theorem 1] (see also [20]). First, by the diminishing adaptation property, for any fixed $\delta_c > 0$, $\varepsilon_c > 0$, and integer $M \geq 1$, we can choose n large enough such that

$$\mathbb{P}\left(\bigcup_{k=1}^M \left\{D_{n+k} > \frac{\varepsilon_c}{2M^2}\right\}\right) \leq \frac{\delta_c}{2}.$$

Now, on the set $\bigcap_{k=1}^M \{D_{n+k} \leq (\varepsilon_c/2M^2)\}$ for transition kernels $P_{\theta_n}, P_{\theta_{n+1}}, \dots, P_{\theta_{n+M}}$, by the triangle inequality we have

$$\begin{aligned} \sup_{x \in \mathcal{X}} \left\| \left(\prod_{k=0}^M P_{\theta_{n+k}} \right) (x, \cdot) - P_{\theta_n}^M (x, \cdot) \right\| &\leq \sum_{k=1}^M \sup_{x \in \mathcal{X}} \left\| \left(\left(\prod_{i=0}^{k+1} P_{\theta_{n+i}} \right) P_{\theta_n}^{M-k-1} \right) (x, \cdot) \right. \\ &\quad \left. - \left(\left(\prod_{i=0}^k P_{\theta_{n+i}} \right) P_{\theta_n}^{M-k} \right) (x, \cdot) \right\| \\ &\leq \sum_{k=1}^M (k+1) \frac{\varepsilon_c}{2M^2} \\ &= \frac{M+1}{4M} \varepsilon_c \\ &< \frac{\varepsilon_c}{2}. \end{aligned} \tag{3.1}$$

Consequently, we conclude that for large enough n ,

$$\mathbb{P} \left(\text{LHS of (3.1)} < \frac{\varepsilon_c}{2} \right) > 1 - \frac{\delta_c}{2}. \tag{3.2}$$

For the ‘only if’ part of the theorem, note that if containment does not hold, then for the adaptive algorithm in question, there is $\varepsilon_c > 0$ and $\delta_c > 0$ such that for all M, n_0 , there exists $n > n_0$ such that

$$\mathbb{P}(M_{\varepsilon_c}(X_n, \theta_n) > M) > \delta_c. \tag{3.3}$$

By (3.2), we obtain for all M, n_0 , there exists $n > n_0$ such that

$$\mathbb{P}(M_{\varepsilon_c/2}^A(X_n, \theta_n) > M) > \frac{\delta_c}{2},$$

which implies the AdapFail condition with $\varepsilon_{AF} \geq \varepsilon_c/2$ and $\delta_{AF} \geq \delta_c/2$.

The proof for the ‘if’ part of the theorem is essentially the same. From (2.1) and (3.2), we obtain (3.3) with $\varepsilon_c \geq \varepsilon_{AF}/2$ and $\delta_c \geq \delta_{AF}/2$.

Remark 3.1. Without assuming diminishing adaptation, Theorem 3.1 does not hold. For example, if P is a fixed ergodic Markov chain, and I is the identity kernel (which does not move at all), then the adaptive scheme which simply alternates between P and I converges well (at half-speed compared to P) and is not in AdapFail. However, this scheme violates containment, since if θ_1 is the adaptive parameter corresponding to I then $M_\varepsilon(x, \theta_1) = \infty$.

4. A very simple example

In this section we analyse a very simple example of an adaptive algorithm to illustrate our results about AdapFail.

Example 4.1. Consider the toy example from [4] with state space $\mathcal{X} = \{0, 1\}$ and stationary distribution $\pi = (\frac{1}{2}, \frac{1}{2})$, with Markov transition kernels

$$P_\theta = \begin{pmatrix} 1 - \theta & \theta \\ \theta & 1 - \theta \end{pmatrix}.$$

Suppose the n th iteration of the Markov chain uses kernel P_{θ_n} (independent of the chain’s past history), where $\theta_n > 0$ and $\sum_n \theta_n = \infty$ but $\theta_n \rightarrow 0$ (e.g. $\theta_n = 1/n$). Since the θ_n converge,

clearly diminishing adaptation is satisfied. On the other hand, as $\theta \rightarrow 0$, $M_\epsilon(x, \theta) \rightarrow \infty$. Hence, this adaptive algorithm does *not* satisfy containment. So, by the above theorems, this algorithm converges more slowly than any fixed nonadaptive algorithm. But since $\sum_n \theta_n = \infty$, this algorithm is still ergodic [4]. We thus have a (very simple) example of an adaptive algorithm which is ergodic, but is nevertheless in AdapFail and has very poor convergence properties. (A similar result presents itself if instead $\theta_n \rightarrow 1$ with $\sum_n (1 - \theta_n) = \infty$.)

Acknowledgements

KŁ acknowledges funding from CRISM and other grants from EPSRC. JSR acknowledges funding from NSERC of Canada. We thank Gersende Fort and Gareth O. Roberts for helpful discussions.

References

- [1] ANDRIEU, C. AND THOMS, J. (2008). A tutorial on adaptive MCMC. *Statist. Comput.* **18**, 343–373.
- [2] ATCHADÉ, Y. F. AND ROSENTHAL, J. S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11**, 815–828.
- [3] BAI, Y., ROBERTS, G. O. AND ROSENTHAL, J. S. (2011). On the containment condition for adaptive Markov chain Monte Carlo algorithms. *Adv. Appl. Statist.* **21**, 1–54.
- [4] FORT, G., MOULINES, E. AND PRIOURET, P. (2011). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.* **39**, 3262–3289.
- [5] GEYER, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7**, 473–483.
- [6] GILKS, W. R., ROBERTS, G. O. AND SAHU, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *J. Amer. Statist. Assoc.* **93**, 1045–1054.
- [7] GIORDANI, P. AND KOHN, R. (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models. *J. Business Econom. Statist.* **26**, 66–77.
- [8] GRIFFIN, J. E., ŁATUSZYŃSKI, K. AND STEEL, M. F. J. (2014). Individual adaptation: an adaptive MCMC scheme for variable selection problems. Submitted.
- [9] HAARIO, H., SAKSMAN, E. AND TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7**, 223–242.
- [10] ŁATUSZYŃSKI, K. (2012). A path stability condition for adaptive MCMC. In preparation.
- [11] ŁATUSZYŃSKI, K., ROBERTS, G. O. AND ROSENTHAL, J. S. (2013). Adaptive Gibbs samplers and related MCMC methods. *Ann. Appl. Prob.* **23**, 66–98.
- [12] MEYN, S. AND TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.
- [13] MIRA, A. AND GEYER, C. J. (1999). Ordering Monte Carlo Markov chains. Tech. Rep. No. 632, School of Statistics, U. of Minnesota, April 1999. Available at: <http://eco.uninsubria.it/webdocenti/amira/papers.html>
- [14] RICHARDSON, S., BOTTOLO, L. AND ROSENTHAL, J. S. (2011). Bayesian models for sparse regression analysis of high dimensional data. In *Bayesian Statistics 9*, Oxford University Press, pp. 539–568.
- [15] ROBERTS, G. O. AND ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Prob.* **2**, 13–25.
- [16] ROBERTS, G. O. AND ROSENTHAL, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16**, 351–367.
- [17] ROBERTS, G. O. AND ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Prob. Surveys* **1**, 20–71.
- [18] ROBERTS, G. O. AND ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Prob.* **44**, 458–475.
- [19] ROBERTS, G. O. AND ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graphical Statist.* **18**, 349–367.
- [20] ROBERTS, G. O. AND ROSENTHAL, J. S. (2013). A note on formal constructions of sequential conditional couplings. *Statist. Prob. Lett.* **83**, 2073–2076.
- [21] ROBERTS, G. O., GELMAN, A. AND GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Prob.* **7**, 110–120.
- [22] SOLONEN, A. *et al.* (2012). Efficient MCMC for climate model parameter estimation: parallel adaptive chains and early rejection. *Bayesian Anal.* **7**, 715–736.