UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 124

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 124

ON THE CONVERGENCE OF UNCONSTRAINED ADAPTIVE MARKOV CHAIN MONTE CARLO ALGORITHMS

MATTI VIHOLA

JYVÄSKYLÄ
2010

UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 124

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 124

ON THE CONVERGENCE OF UNCONSTRAINED ADAPTIVE MARKOV CHAIN MONTE CARLO ALGORITHMS

MATTI VIHOLA

To be presented, with the permission of the Faculty of Mathematics and Science
of the University of Jyväskylä, for public criticism in Auditorium H 320,
on March 6th, 2010, at 12 o'clock noon.

JYVÄSKYLÄ
2010

Editor: Pekka Koskela
Department of Mathematics and Statistics
P.O. Box 35 (MaD)
FI–40014 University of Jyväskylä
Finland

ISBN 978-951-39-3809-3
ISSN 1457-8905

University Printing House
Jyväskylä 2010

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor Professor Eero Saksman for supporting me throughout my PhD studies. Without his wonderful guidance and ideas, this work might have never completed. Sincere thanks also to my reviewers, Professor Christophe Andrieu (University of Bristol) and Professor Esa Nummelin (University of Helsinki).

Everyone in the Department of Mathematics and Statistics deserves thanks; I have really enjoyed the good working atmosphere. Especially, I want to thank Professor Stefan Geiss and Professor Antti Penttinen for many interesting discussions, support and encouragement. Special thanks also to MSc Heikki Seppälä for enjoyable mathematical and non-mathematical discussions.

For financial support, I gratefully acknowledge the Academy of Finland (project nos. 110599 and 201392), the Finnish Academy of Science and Letters, Vilho, Yrjö and Kalle Väisälä Foundation, the Finnish Centre of Excellence in Analysis and Dynamics Research and the Finnish Graduate School in Stochastics and Statistics.

Warmest thanks to my parents Tapio and Pirjo and to my dear family Leena and Laura, for everything.

Jyväskylä, January 2010

Matti Vihola

LIST OF INCLUDED ARTICLES

This dissertation consists of an introductory part and the following articles.

[A] E. Saksman and M. Vihola
On the ergodicity of the adaptive Metropolis algorithm on unbounded domains
To appear in *Annals of Applied Probability*.
Preprint arXiv:0806.2933.

[B] M. Vihola
On the stability and ergodicity of an adaptive scaling Metropolis algorithm
Submitted.
Preprint arXiv:0903.4061.

[C] M. Vihola
Can the adaptive Metropolis algorithm collapse without the covariance lower bound?
Submitted.
Preprint arXiv:0911.0522.

In the introductory part, the articles are referred to as [A], [B] and [C], whereas the other references are numbered as [1], [2], ...

The author of this dissertation has actively taken part in the research of the joint article [A].

INTRODUCTION

The Markov chain Monte Carlo (MCMC) stochastic integration method, in particular the Metropolis-Hastings algorithm [17, 23], applies very generally in practical problems. The algorithm involves, however, a very complicated and delicate parameter: the proposal distribution. The choice of the proposal distribution strongly affects the efficiency of the method, and therefore determines its practical value. It may be difficult to come up with a good proposal, especially in high dimensions. Often, the proposal is a result of several trial runs and hours of manual work.

This work deals with adaptive MCMC algorithms, aiming to learn the proposal distribution automatically during the simulation. That is, the proposal distribution is tuned based on the simulated history of the chain. The goal is to end up with a good proposal ensuring efficient simulation. There have been a number of previous attempts to make the Metropolis-Hastings algorithm include some sort of adaptivity [36]. Most of the previous approaches have suffered from complicated constructions and limited applicability. This work focuses on a so called *non-Markovian adaptation* within MCMC [14], which has attracted increasing popularity in the recent years.

Such non-Markovian algorithms are typically quite easy to formulate and implement in practice. Their analysis, however, is not as straightforward as traditional (non-adaptive) algorithms. There are even examples of intuitively 'reasonable' non-Markovian adaptation schemes that are invalid in the sense that the computed averages do not converge to the correct value [29]. There have been substantial theoretical advances in this field after the seminal work [14], with different types of arguments [1, 2, 5, 6, 29]. This work focuses on relaxing assumptions on the algorithms used for adaptation, in particular allowing unconstrained, fully adaptive mechanisms. There is a strong emphasis on finding conditions to ensure the ergodicity of some adaptive MCMC algorithms that are verifiable in practical applications.

The rest of the introductory part is organised as follows. Section 1 introduces the MCMC method, and in particular the random walk Metropolis algorithm, which plays a central role in this work. Section 2 formulates a fairly general framework for adaptive MCMC and describes two generally applicable algorithms, that are analysed in detail in the included articles. Section 3 outlines the previous ergodicity results for adaptive MCMC in the literature, and summarises the main contributions of this work. Section 4 contains some concluding discussion on the main results and on some future research directions, and Section 5 gives a summary of the included articles.

1. Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a family of methods to construct a (typically time-homogeneous) Markov chain $(X_n)_{n\geq 1}$ evolving in the Euclidean space $\mathbb{R}^d$ such that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x \qquad \text{almost surely,} \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a Borel-measurable function of interest and $\pi : \mathbb{R}^d \to [0,\infty)$ is a probability density.[1]

Section 1.1 starts by introducing some basic concepts of Markov chains and the related notations. Section 1.2 continues with the celebrated Metropolis-Hastings algorithm, originating from the seminal 1953 physics article by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller [23], later generalised by Hastings [17]. The random walk Metropolis algorithm [23] is discussed in more detail, as it plays the central role in the rest of this work. Section 1.3 gives some conditions that ensure a strong law of large numbers (1) holds. Section 1.4 introduces a stronger concept of geometric ergodicity, that guarantees a certain 'mixing speed' for the chain. Geometric ergodicity is a key concept in the analysis of the adaptive MCMC algorithms introduced in Section 2.

1.1. Markov Chains.

Let $\mathbb{X} \subset \mathbb{R}^d$ be a Borel set, and denote by $\mathcal{B}(\mathbb{X})$ the Borel subsets of $\mathbb{X}$. The $\mathbb{X}$-valued random variables $X_1, X_2, \ldots, X_n$ form a *Markov chain*, if

$$\mathbb{P}\left(X_n \in A \mid X_1, \ldots, X_{n-1}\right) = \mathbb{P}\left(X_n \in A \mid X_{n-1}\right)$$

almost surely for all $A \in \mathcal{B}(\mathbb{X})$ [34, Chapter VIII]. The Markov chains on $\mathbb{X}$ are determined by so called transition kernels [24, 26].

Definition 1.1. A mapping $P : \mathbb{X} \times \mathcal{B}(\mathbb{X}) \to [0,1]$ is a *transition kernel*[2], given that $P(x, \cdot\,)$ is a probability measure on $\mathbb{X}$ for each $x \in \mathbb{X}$ and $P(\cdot\,, A)$ is measurable for every $A \in \mathcal{B}(\mathbb{X})$.

In particular, suppose that μ is a probability measure on $\mathbb{X}$, and let $(P_n)_{n\geq 2}$ be transition kernels. There is a unique probability measure $\mathbb{P}$ defined on $(\mathbb{X}^\infty, \mathcal{B}(\mathbb{X}^\infty))$ such that $\mathbb{P}(X_1 \in A) = \mu(A)$ for any $A \in \mathcal{B}(\mathbb{X})$, and for $n \geq 2$

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{A_1} \mu(\mathrm{d}x_1) \int_{A_2} P_2(x_1, \mathrm{d}x_2) \cdots \int_{A_n} P_n(x_{n-1}, \mathrm{d}x_n)$$

[1] The MCMC methods also apply in a more general state space setting. The introductory part of this work is written considering $\mathbb{R}^d$ for expository reasons, and because the main contribution of this work lies in analysing practical algorithms evolving on $\mathbb{R}^d$.

[2] Also known as *transition probability kernel*, *Markov transition kernel* or *transition probability*.

for all $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{X})$ [26]. It holds that

$$\mathbb{P}\,(\,X_n \in A \mid X_1, \ldots, X_{n-1}) = \mathbb{P}\,(\,X_n \in A \mid X_{n-1}) = P_n(X_{n-1}, A)$$

almost surely for all $n \geq 2$ and $A \in \mathcal{B}(\mathbb{X})$. Often, the initial variable is chosen so that $X_1 \equiv x_1 \in \mathbb{X}$, corresponding to the initial measure $\mu(A) = \mathbb{1}_A(x_1)$ where $\mathbb{1}_A$ denotes the characteristic function of the set A defined as $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ if $x \notin A$. The chain is *homogeneous* if there is a transition kernel P such that $P_n = P$ for all $n \geq 2$.

Let $f : \mathbb{X} \to \mathbb{R}$ be a Borel measurable function and μ any probability measure on $\mathbb{X}$. The transition kernel P maps μ to another probability measure $\mu P(A) := \int_{\mathbb{X}} \mu(\mathrm{d}x) P(x, A)$. If f is integrable with respect to μ the notation $\mu(f) := \int_{\mathbb{X}} \mu(\mathrm{d}x) f(x)$ is used. Likewise, if f is integrable with respect to each $P(x, \cdot\,)$, then P maps f to a new function $Pf(x) := \int_{\mathbb{X}} P(x, \mathrm{d}y) f(y)$. Moreover, a transition kernel maps any transition kernel to a transition kernel, so one may define inductively

$$P^n(x, A) = \int_{\mathbb{X}} P(x, \mathrm{d}y) P^{n-1}(y, A)$$

for $n \geq 2$. If the chain is homogeneous one has $\mathbb{P}\,(\,X_{m+n} \in A \mid X_m) = P^n(X_m, A)$ almost surely for any $n, m \geq 1$.

Suppose $V : \mathbb{X} \to [1, \infty)$ is a measurable function. The *V-total variation* norm for any finite signed measure μ is defined as

$$\|\mu\|_V := \sup_{f : |f| \leq V} \mu(f)$$

where the supremum is taken over all measurable f. The special case $V \equiv 1$ induces the *total variation* norm

$$\|\mu\| := \|\mu\|_1 = \sup_{A \in \mathcal{B}(\mathbb{X})} \mu(A) - \inf_{A \in \mathcal{B}(\mathbb{X})} \mu(A).$$

For a function f, the *V-norm* is defined through

$$\|f\|_V := \sup_{x \in \mathbb{X}} \frac{|f(x)|}{V(x)}.$$

1.2. **Metropolis-Hastings Algorithms.** Let us consider first the algorithmic construction of the Metropolis-Hastings process. Suppose that $q : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ is a *proposal density*, that is, it is Borel measurable and defines a probability density $q(x, \cdot\,)$ for each $x \in \mathbb{R}^d$. Let $X_1 \equiv x_1$ where $x_1 \in \mathbb{R}^d$ is an arbitrary starting point within the support of the *target distribution* π, that is, $\pi(x_1) > 0$. For $n = 2, 3, \ldots$, iterate the following steps:

(M1) simulate $Y_n \sim q(X_{n-1}, \cdot\,)$, and
(M2) with probability $\alpha(X_{n-1}, Y_n)$ the proposal is accepted and $X_n = Y_n$; otherwise the proposal is rejected and $X_n = X_{n-1}$.

The notation $Y \sim q(x, \cdot)$ above is read that Y follows, independently, the distribution with the density $q(x, \cdot)$. That is, Y_n follows, conditional on X_{n-1}, the distribution with the density $q(X_{n-1}, \cdot)$. The acceptance probability α is defined as the Metropolis-Hastings ratio

$$\alpha(x,y) := \begin{cases} \min\left\{1, \dfrac{\pi(y)}{\pi(x)}\dfrac{q(y,x)}{q(x,y)}\right\}, & \text{if } \pi(x)q(x,y) > 0 \text{ and} \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

Observe that, given $\pi(x_1) > 0$, the latter case is never encountered, almost surely, and $\mathbb{P}(\pi(X_n) > 0) = 1$ for all $n \geq 1$.

In mathematical terms, the variables $(X_n)_{n\geq 1}$ in the Metropolis-Hastings algorithm form a time-homogeneous Markov chain with initial state $x_1 \in \mathbb{X} := \operatorname{supp}(\pi) := \{x \in \mathbb{R}^d : \pi(x) > 0\}$ and having the following transition kernel M_q

$$M_q(x, A) := \mathbb{1}_A(x)\left(1 - \int_{\mathbb{R}^d} \alpha(x,y)q(x,y)\mathrm{d}y\right) + \int_A \alpha(x,y)q(x,y)\mathrm{d}y. \tag{3}$$

As observed above, $\mathbb{P}(X_n \in \mathbb{X}) = 1$ for all $n \geq 1$, and therefore one can assume that X_n are $\mathbb{X}$-valued, without loss of generality. The notation π is used also for the probability measure defined by $\pi(A) := \int_A \pi(x)\mathrm{d}x$.

This work deals mostly with symmetrically defined q. In particular, suppose (by slight abuse of notation) that q is a symmetric probability density on $\mathbb{R}^d$, and let $q(x,y) = q(y,x) = q(y-x)$ for all x and y. This construction is often referred to as the *random walk Metropolis algorithm*, as the original formulation [23] was of this form. In this case, the acceptance probability reduces to $\alpha(x,y) := \min\{1, \pi(y)/\pi(x)\}$ for all $\pi(x) > 0$.

1.3. **Validity of Random Walk Metropolis.** The analysis on the behaviour of Metropolis-Hastings algorithms rely on the theory of the general state-space Markov chains; see, for example the monographs by Nummelin [26] and Meyn and Tweedie [24]. Due to the special structure of the Metropolis-Hastings kernel, the density π is automatically invariant under M_q, that is, $\pi P = \pi$, since by definition

$$\pi(x)\alpha(x,y)q(x,y) = \pi(y)\alpha(y,x)q(y,x)$$

for all x and y in $\mathbb{R}^d$.

Some additional assumptions are, however, required on the proposal density q. A homogeneous Markov chain on $\mathbb{X}$ with a transition kernel P is *π-irreducible* if for any $x \in \mathbb{X}$ and $A \subset \mathcal{B}(\mathbb{X})$ with $\pi(A) > 0$ there is an integer $n = n(x, A) \geq 1$ such that $P^n(x, A) > 0$.

For example, consider a random walk Metropolis algorithm with a compactly supported proposal q and let $R > 0$ be sufficiently large so that $\operatorname{supp}(q)$ is contained in the ball $B(0, R) := \{x \in \mathbb{R}^d : |x| \leq R\}$. Consider a target distribution π with a support $\operatorname{supp}(\pi) = A \cup B$ where A and B are disjoint. If the distance $\operatorname{dist}(A, B) := \inf\{|x - y| : x \in A, y \in B\} > R$, the

algorithm is 'trapped' on the set where it is started. For example, if $x_1 \in A$, then $\mathbb{P}(X_n \in A) = 1$ for all $n \geq 1$. This chain is *reducible*.

Essentially, excluding the above case, a strong law of large numbers always holds for the random walk Metropolis algorithm.

Theorem 1.2. *Assume that q is a symmetric probability density bounded away from zero on compact sets. Then, for a function f satisfying $\pi(|f|) < \infty$, the strong law of large numbers* (1) *holds for the random walk Metropolis chain with the transition kernel M_q.*

Proof. Let us check that M_q is irreducible, with $n \equiv 1$. Fix $x \in \mathbb{X}$ and $A \in \mathcal{B}(\mathbb{X})$ such that $\pi(A) > 0$. Let $\beta > 0$ be sufficiently small so that the set $B := \{y \in A : \pi(y) \geq \beta\} \subset A$ has a positive Lebesgue measure. Now,

$$\begin{aligned} M_q(x, A) &\geq \int_A \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} q(y - x)\mathrm{d}y \\ &\geq \min\left\{1, \frac{\beta}{\pi(x)}\right\} \int_B q(y - x)\mathrm{d}y > 0. \end{aligned}$$

Corollary 2 of Tierney [35] implies that the chain is Harris recurrent, Theorem 17.0.1(i) of Meyn and Tweedie [24] yields the strong law of large numbers (1). □

Theorem 1.2 serves as an example how minimal assumptions ensure the ergodicity of the random walk Metropolis sampler. See, for example, Nummelin [27] for other practically motivated assumptions ensuring the ergodicity of various Metropolis-Hastings chains.

1.4. **Geometric Ergodicity of Random Walk Metropolis.** The strong law of large numbers (1) holds very generally for a random walk Metropolis algorithm, as exemplified in Theorem 1.2. In practice, one is also interested on the properties of the sample average $I_n := n^{-1} \sum_{k=1}^n f(X_k)$, with some finite $n \geq 1$. For example, one could ask whether a central limit theorem holds, that is, $\sqrt{n} I_n$ converges in distribution to a Gaussian limit.

A common condition implying the central limit theorem, that can also be verified in practical situations, is the geometric ergodicity. This section summarises some necessary and sufficient conditions for the geometric ergodicity of random walk Metropolis chain. These results play also a central role in the analysis of the adaptive MCMC algorithms, as discussed in Section 3.

Definition 1.3. A Markov chain with transition kernel P on $\mathbb{X}$ is said to be

(i) *ergodic*, if $\|P^n(x, \cdot) - \pi(\cdot)\| \to 0$ as $n \to \infty$ for all $x \in \mathbb{X}$.

(ii) *geometrically ergodic* if there is a function $V : \mathbb{X} \to [1, \infty)$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\|_V \leq RV(x)\rho^n$$

for all $n \geq 1$, where $R < \infty$ and $\rho \in (0, 1)$ are constants.

(iii) *uniformly ergodic* if it is geometrically ergodic with $V \equiv 1$.

Uniform ergodicity is the strongest form of ergodicity. An ergodic chain evolving on a finite state space $\mathbb{X}$ is always uniformly ergodic. It is easy to see that a random walk Metropolis chain can be uniformly ergodic only if the support of π (the space $\mathbb{X}$) is bounded. When the support $\mathbb{X}$ is unbounded, the chain cannot be uniformly ergodic, but it can be geometrically ergodic.

In recent years, different conditions ensuring the geometric ergodicity of random walk Metropolis chains has been proposed [18, 22, 31]. In particular, the geometric ergodicity of the chain in $\mathbb{R}^d$ seems to be tightly related on the decay rate of the tails of the target distribution and the regularity of the tail contours [18].

Assumption 1.4. The target density π is supported on $\mathbb{R}^d$ and is continuously differentiable. The tails of π are super-exponentially decaying and have regular contours, that is,

$$\lim_{|x|\to\infty} \frac{x}{|x|} \cdot \nabla \log \pi(x) = -\infty \qquad \text{and} \tag{4}$$

$$\limsup_{|x|\to\infty} \frac{x}{|x|} \cdot \frac{\nabla \pi(x)}{|\nabla \pi(x)|} < 0, \tag{5}$$

respectively.

Theorem 1.5. *Suppose the proposal density q is bounded away from zero in some neighbourhood of the origin, that is, there exist $\delta_q > 0$ and $\epsilon_q > 0$ such that $q(z) \geq \epsilon_q$ for all $|x| \leq \delta_q$. If the target distribution satisfies Assumption 1.4, then the random walk Metropolis chain with the transition kernel M_q is geometrically ergodic.*

Proof. Theorem 4.3 of Jarner and Hansen [18]. □

Theorem 1.5 is based on establishing a so called geometric drift of the function $V := [\sup_z \pi(z)]^\gamma \pi^{-\gamma}(x)$ toward a compact small set D. Precisely, it is shown that there exist constants $\lambda, \delta \in (0,1)$ and $b < \infty$, and a probability measure ν concentrated on D such that

$$\begin{aligned} M_q V(x) &\leq \lambda V(x) + b\mathbb{1}_D(x) \qquad \text{and} \\ M_q(x, A) &\geq \mathbb{1}_D(x)\delta\nu(A) \end{aligned}$$

for all $x \in \mathbb{R}^d$ and all Borel sets $A \subset \mathbb{R}^d$. This condition implies that the chain is geometrically ergodic with the same function V and some constants $M < \infty$ and $\rho \in (0,1)$ that depend only (and explicitly) on the constants λ, b and δ [8, 25].

The conditions of Theorem 1.5 are quite close to optimal. Particularly, if the probability of rejection is not bounded away from one, or if the tails of π are heavier than exponential, the chain cannot be geometrically ergodic. More precisely,

Theorem 1.6. *Suppose that M_q is π-irreducible Metropolis kernel on the space $\mathbb{X} = \mathrm{supp}(\pi)$ and that*

$$\operatorname*{ess\,sup}_{x \in \mathbb{X}} M_q(x, \{x\}) = 1.$$

Then, M_q is not geometrically ergodic.

Proof. Theorem 5.1 of Roberts and Tweedie [31]. □

Theorem 1.7. *Suppose the proposal density is spherically symmetric, $q(z) = q(|z|)$, satisfying $\int_{\mathbb{R}^d} |z| q(z) \mathrm{d}z < \infty$. Then, if M_q is geometrically ergodic, there exists a $s > 0$ such that*

$$\int_{\mathbb{R}^d} e^{s|x|} \pi(x) \mathrm{d}x < \infty.$$

Proof. Corollary 3.4 of Jarner and Hansen [18], □

As Theorems 1.6 and 1.7 show, certain Metropolis chains cannot be geometrically ergodic. Recent advances establish bounds

$$\|P^n(x, \cdot) - \pi(\cdot)\|_V \leq r(x, n)$$

with a rate function $r(x, n)$ decaying at some sub-geometric rate as $n \to \infty$ [11, 19, 20]. The current results have, however, somewhat limited applicability, due to conditions that are either restrictive or hard to verify in practice.

2. Adaptive Markov Chain Monte Carlo Algorithms

As already mentioned in the introduction, 'adaptivity' has appeared in different meanings in the context of MCMC [13, 36]. Following the terminology of Tierney and Mira [36], the 'adaptive MCMC' algorithms considered here are continuous and infinite-horizon. That is, the adaptation takes place continuously during the simulation, and the whole simulated history is used for adaptation.

Section 2.1 starts by formulating a general framework for such adaptive MCMC algorithms, inspired by Robbins-Monro stochastic approximation [28]. Sections 2.2 and 2.3 introduce the two algorithms, the Adaptive Metropolis (AM) and the Adaptive Scaling Metropolis (ASM), that are analysed in detail in Sections 3.4 and 3.5, respectively. Section 2.4 outlines some other proposed algorithms that are closely related to the AM and the ASM algorithms.

2.1. General Adaptive MCMC Framework.

The adaptive MCMC framework introduced here is similar to the Robbins-Monro stochastic approximation proposed by Andrieu and Robert [2]. In what follows, the adaptation space $\mathbb{S}$ is a subset of $\mathbb{R}^{d_\mathbb{S}}$ for some integer $d_\mathbb{S} \geq 1$, and $\{P_s\}_{s \in \mathbb{S}}$ is a family of ergodic Markov transition kernels on $\mathbb{X} := \mathrm{supp}(\pi) \subset \mathbb{R}^d$ with the unique invariant density π.

To unify notations, let us consider the following new formal definition. First of all, let $d' \geq 0$ be an integer and define the extended state space $\tilde{\mathbb{X}} := \mathbb{X} \times \mathbb{R}^{d'}$, with the convention that $\tilde{\mathbb{X}} = \mathbb{X}$ if $d' = 0$. Consider the collection of transition kernels $\{\tilde{P}_s\}_{s \in \mathbb{S}}$ on $\tilde{\mathbb{X}}$ such that each $\tilde{P}_s$ is an extension of P_s in the following sense.

Definition 2.1. The transition kernel $\tilde{P}$ on $\mathbb{X} \times \mathbb{R}^{d'}$ is an *extension* of the transition kernel P on $\mathbb{X}$, if

$$\tilde{P}((x,z), A \times B) = \tilde{P}(x, A \times B) \qquad \text{and} \tag{6}$$

$$\tilde{P}(x, A \times \mathbb{R}^{d'}) = P(x, A) \tag{7}$$

for all $x \in \mathbb{X}$, $z \in \mathbb{R}^{d'}$, $s \in \mathbb{S}$, $A \in \mathcal{B}(\mathbb{X})$ and $B \in \mathcal{B}(\mathbb{R}^{d'})$.

Observe that each one of the extended transition kernels $\tilde{P}_s$ is ergodic with $\pi_s := \int_{\mathbb{R}^d} \pi(x)\tilde{P}_s(x, \cdot)\mathrm{d}x$ as the unique invariant measure. Moreover, the marginal of π_s is always π, that is, $\pi_s(A \times \mathbb{R}^{d'}) = \int_{\mathbb{R}^d} \pi(x)P_s(x, A)\mathrm{d}x = \pi(A)$ for all $s \in \mathbb{S}$.

Having defined the extended transition kernels, the adaptation is assumed to have the following form, which is slightly more general than the definitions given in the included articles. The process starts at some fixed points $X_1 \equiv x_1 \in \mathbb{X}$ and $S_1 \equiv s_1 \in \mathbb{S}$, and the variables $(X_n, Z_n, S_n)_{n \geq 2}$ are defined recursively through

$$(X_{n+1}, Z_{n+1}) \sim \tilde{P}_{S_n}(X_n, \cdot) \qquad \text{and} \tag{8}$$

$$S_{n+1} = S_n + \eta_{n+1} H(S_n, X_{n+1}, Z_{n+1}) \tag{9}$$

where $(\eta_n)_{n \geq 2}$ is a sequence of constant non-negative step sizes decaying to zero and $H : \mathbb{S} \times \mathbb{X} \times \mathbb{R}^{d'} \to \mathbb{R}^{d_{\mathbb{S}}}$ is a measurable adaptation function.

In an ideal situation, stochastic approximation algorithms implementing recursion (9) seek s^*, the unique root of the *mean field* $h : \mathbb{S} \to \mathbb{R}^{d_{\mathbb{S}}}$, defined as

$$h(s) := \int_{\mathbb{R}^d} H(s, x, z)\pi_s(\mathrm{d}x \times \mathrm{d}z).$$

In the context of adaptive MCMC, however, the mean field h may be complicated and may even have multiple roots. Many practical algorithms can be formulated in this framework; examples of adaptation functions H and the corresponding mean field functions h are given below.

Remark 2.2. Observe that the above defined sequence $(X_n, Z_n, S_n)_{n \geq 2}$ forms, in fact, an inhomogeneous Markov chain. For this reason, some authors call this type of adaptation *Markovian* [29], which is a possible source of confusion. The term 'non-Markovian' employed here refers to the chain $(X_n)_{n \geq 1}$, which is not Markov.

Remark 2.3. There are various results on the *convergence* of several 'stochastic approximation' algorithms, with different assumptions; see for example

the monographs [10, 21]. For example, if (X_{n+1}, Z_{n+1}) in (9) were independent draws from π_s, the parameters S_{n+1} would indeed converge to s^* with some natural assumptions. Unfortunately, these classical results are mostly inappropriate for adaptive MCMC, as $(X_n)_{n\geq 2}$ and $(Z_n)_{n\geq 2}$ typically have a complicated dependence on $(S_n)_{n\geq 1}$. In this work, the main concern is that the averages involving $(X_n)_{n\geq 1}$ converge. For this, it is primarily important to establish the *stability* of S_n, not necessary the convergence. The theoretical aspects of the adaptation are discussed in more detail in Section 3.

2.2. **Adaptive Metropolis.** The Adaptive Metropolis (AM) algorithm due to Haario, Saksman and Tamminen [14] was the first adaptive MCMC algorithm of this kind. Suppose $X_1 \equiv x_1 \in \mathbb{X}$ and let $\overline{C}_1 \equiv c_1 \in \mathbb{R}^{d\times d}$ be positive definite. The variables $(X_n)_{n\geq 2}$ are then defined recursively through

$$X_{n+1} \sim M_{q_{\overline{C}_n}}(X_n, \cdot)$$

where M_{q_c} is the Metropolis kernel (3) with a zero-mean Gaussian proposal density q_c with covariance matrix $\theta^2 c + \epsilon I$, where the constant $\theta > 0$ is a scaling factor and $\epsilon > 0$ is a small constant multiplier of the identity matrix $I \in \mathbb{R}^{d\times d}$. The original algorithm was based on the unbiased covariance estimate $\overline{C}_n$ of the history of the chain X_n defined for $n \geq 2$ through[3]

$$\overline{C}_n := \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)(X_k - \overline{X}_n)^T$$

where $\overline{X}_n$ stands for the average of $X_1, \ldots, X_n$.

One can check that $(\overline{C}_n)_{n\geq 2}$ and $(\overline{X}_n)_{n\geq 2}$ can be computed recursively by

$$\overline{X}_{n+1} = \frac{n}{n+1}\overline{X}_n + \frac{1}{n+1}X_{n+1} \qquad \text{and} \tag{10}$$

$$\overline{C}_{n+1} = \frac{n-1}{n}\overline{C}_n + \frac{1}{n+1}(X_{n+1} - \overline{X}_n)(X_{n+1} - \overline{X}_n)^T. \tag{11}$$

This algorithm does not, strictly speaking, fit into the stochastic approximation framework of Section 2.1, but would require a sequence of adaptation functions $(H_n)_{n\geq 1}$. In this work, the following slight modification proposed by Andrieu and Robert [2] is considered.

Define $\mathrm{S} = \mathbb{R}^d \times \mathcal{C}^d$, where $\mathcal{C}^d \subset \mathbb{R}^{d\times d}$ stands for the symmetric and positive definite matrices, and let $S_n := (M_n, C_n)$ with $M_1 \equiv x_1$ and $C_1 \equiv c_1$ where $c_1 \in \mathcal{C}^d$. Let $q_s = q_{(m,c)} = q_c$ stand for a zero-mean Gaussian density with the covariance matrix $\theta^2 c + \epsilon I$ for some $\epsilon \geq 0$. Define $\tilde{P}_s =$

[3] In the original setting, the initial covariance s_1 was employed during the whole *burn-in* period. The burn-in is not considered here, as it does not affect the asymptotical behaviour of the algorithm.

$P_s = M_{q_s}$ as the Metropolis kernel (3) with the proposal density q_s. Then, set

$$M_{n+1} := M_n + \eta_{n+1}(X_{n+1} - M_n) \qquad \text{and} \tag{12}$$

$$C_{n+1} := C_n + \eta_{n+1}\big((X_{n+1} - M_n)(X_{n+1} - M_n)^T - C_n\big). \tag{13}$$

This corresponds to the stochastic approximation scheme with the adaptation function

$$H_{\mathrm{AM}}(s,x) = H_{\mathrm{AM}}\big((m,c),x\big) := \begin{bmatrix} x-m \\ (x-m)(x-m)^T - c \end{bmatrix}.$$

If the weights are defined to be $\eta_n := n^{-1}$ and the same $(X_n)_{n\geq 1}$ are used in recursions (10)–(13), $M_n = \overline{X}_n$ for all $n \geq 1$, and the value of C_{n+1} obtained from recursion (13) compared with the original (11) differs by an order of $n^{-2}|C_n|$.

Suppose that the target density π has finite second moments. One can compute the mean field

$$h_{\mathrm{AM}}(s) = h_{\mathrm{AM}}(m,c) = \begin{bmatrix} m_\pi - m \\ c_\pi - c + (m_\pi - m)(m_\pi - m)^T \end{bmatrix}$$

having a unique root at $[m_\pi, c_\pi]^T$, the mean and covariance of π, respectively. Therefore, the AM algorithm seeks the true covariance c_π of the target distribution π. Section 3.4 gives conditions ensuring the ergodicity of the AM algorithm, and implying that C_n indeed converges to c_π.

2.3. **Adaptive Scaling Metropolis.** It is well-known for practitioners that the mean acceptance probability of a random walk Metropolis algorithms should not be too low or too high in general. Indeed, in the case of a spherically symmetric multivariate Gaussian target, Gelman, Gilks and Roberts observed that in certain sense optimal acceptance probability is approximately 0.44 in dimension one and declines to 0.234 as the dimension increases [12, 32]. This '0.234 rule' has then been verified to hold also with some other target distributions [33], but it may not always be optimal [9].

Gilks, Roberts and Sahu [13] proposed over a decade ago an adaptive MCMC scheme that tries to find a scale admitting certain mean acceptance probability. Their approach was based on adaptation upon certain regeneration times, which may occur rarely and may be difficult to identify in practice. Andrieu and Robert [2] proposed the same approach using continuous MCMC adaptation, and different variations have then been proposed by several authors [3, 5, 6, 30].

Let us formulate the adaptive scaling Metropolis (ASM) algorithm considered here, following Atchadé and Fort [5] and Andrieu and Thoms [3]. Let $S_1 \equiv s_1 \in \mathbb{S} = \mathbb{R}$ and $X_1 \equiv x_1 \in \mathbb{X}$, and recursively for $n = 1, 2, \ldots$

(S1) simulate $Y_{n+1} = X_n + e^{S_n} W_{n+1}$ with $W_{n+1} \sim q$,

(S2) with probability $\alpha_{n+1} := \alpha(X_n, Y_{n+1})$ the proposal is accepted and $X_{n+1} = Y_{n+1}$; otherwise $X_{n+1} = X_n$, and

(S3) set $S_{n+1} = S_n + \eta_{n+1}(\alpha_{n+1} - \alpha^*)$.

where q is a symmetric probability density in $\mathbb{R}^d$ and the constant $\alpha^* \in (0,1)$ is the desired mean acceptance probability, for example $\alpha^* = 0.234$. The steps (S1) and (S2) above can be written using the Metropolis kernel as $X_{n+1} \sim P_{S_n}(X_n, \cdot\,)$ by defining the proposal densities $\{q_s\}_{s\in\mathbb{R}}$ through $q_s(z) := e^{-ds} q(e^{-s}z)$. For example, if q is a zero-mean Gaussian density with identity covariance, then q_s has the covariance $e^{2s} I$.

This algorithm fits the stochastic approximation framework by extending the state space to $\tilde{\mathbb{X}} := \mathbb{X} \times \mathbb{R}$ and considering the following extension of M_{q_s}:

$$\tilde{P}_s(x, A \times B) := \int_A \mathbb{1}_B\big(\alpha(x,y)\big)\alpha(x,y)q_s(x,y)\mathrm{d}y + \mathbb{1}_A(x)\int_{\mathbb{R}^d} \mathbb{1}_B\big(\alpha(x,y)\big)\big(1-\alpha(x,y)\big)q_s(x,y)\mathrm{d}y.$$

That is, $Z_n = \alpha_n(X_{n-1}, Y_n)$ for all $n \geq 2$. In this case, one may write $H_{\mathrm{ASM}}\big(s,(x,\tilde{\alpha})\big) = \tilde{\alpha} - \alpha^*$, and the mean field equals $h(s) = A(s) - \alpha^*$, with

$$A(s) = \int_{\mathbb{R}} \tilde{\alpha}\pi_s(\mathbb{R}^d \times \mathrm{d}\tilde{\alpha}) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \alpha(x,y)q_s(x,y)\mathrm{d}y\pi(x)\mathrm{d}x$$

since one can write $\pi_s(\mathbb{R}^d \times B) = \int_{\mathbb{R}^d} \pi(x)\big[\int_{\mathbb{R}^d} \mathbb{1}_B\big(\alpha(x,y)\big)q_s(x,y)\mathrm{d}y\big]\mathrm{d}x$ for any $B \in \mathcal{B}(\mathbb{R})$.

One can show that the mean acceptance rate $A(s) \to 0$ if $s \to \infty$ and $A(s) \to 1$ if $s \to -\infty$. Therefore, for any $\alpha^* \in (0,1)$, there is a 'negative drift' when S_n is large and 'positive drift' when S_n is small, and one can expect a stable behaviour on $(S_n)_{n\geq 1}$. Section 3.5 shows that the ASM algorithm has the correct ergodic properties when $\alpha^* \in (0, 1/2)$ and the target distribution satisfies certain smoothness and tail decay conditions.

Remark 2.4. There are results that show the convergence of S_n toward s^* such that $h(s^*) = 0$, that is, $A(s^*) = \alpha^*$, in the case of a (modified) ASM algorithm [6]. The mean field h may not, however, have a unique root, and it may be difficult in general to check the uniqueness in practice [16, Section 4.4].

2.4. **Some Related Algorithms.** The two algorithms described above, the AM and the ASM, can be naturally combined [3, 5, 6]. Define $S_n := (M_n, C_n, T_n)$ where M_n and C_n are the AM mean and covariance as defined as in Section 2.2 and T_n corresponds to the ASM scaling S_n defined in Section 2.3. That is, define $\{\tilde{P}_s\}_{s\in\mathbb{S}}$ as in Section 2.3 and the adaptation function

$$H_{\mathrm{AM+ASM}}\big((m,c,t),(x,\tilde{\alpha})\big) = \begin{bmatrix} x - m \\ (x-m)(x-m)^T - c \\ \tilde{\alpha} - \alpha^* \end{bmatrix}.$$

Then, define the proposal distribution $q_s = q_{(m,c,t)}$ as a zero-mean Gaussian with covariance $e^{2t}c$. The stability and ergodicity results on the ASM algorithm described in Section 3.5 apply also for the analysis of this algorithm.

Andrieu and Thoms [3] have proposed a 'Rao-Blackwellised' AM algorithm. In the general framework of Section 2.1, the definition $S_n = (M_n, C_n)$, $Z_{n+1} = (X_n, Y_{n+1})$ (with a suitable definition of $\tilde{P}_s$) and the adaptation function

$$H_{\mathrm{RBAM}}\big((m,c),(x,x_-,y)\big) = \begin{bmatrix} \alpha(x_-,y)y + [1-\alpha(x_-,y)]x_- - m \\ \alpha(x_-,y)(y-m)(y-m)^T + [1-\alpha(x_-,y)](x_- - m)(x_- - m)^T - c \end{bmatrix}$$

yields this algorithm. In words, one uses a combination of the proposed value Y_{n+1} and the previous state X_n in the update, weighted by the acceptance probability, instead of the current state X_{n+1}. This interesting modification corresponds to a one step Rao-Blackwellisation, but it is unknown whether this algorithm has benefits, in general, compared to the original AM algorithm [3].

3. Ergodicity of Adaptive MCMC

This chapter summarises the main contributions of this work. Section 3.1 starts by discussing a so called simultaneous geometric ergodicity condition. It is very commonly used when verifying the ergodicity of adaptive MCMC in practice. Section 3.2 continues by a short overview of the previous ergodicity results in the literature, especially regarding the AM and ASM algorithms.

The first main contribution in the article [A] is a general sequential truncation approach within adaptive MCMC, described in Section 3.3. This sequential truncation approach is employed in all the included articles [A]–[C] to establish ergodicity results for the unconstrained AM and ASM algorithms. These results are summarised in Sections 3.4 and 3.5.

3.1. Simultaneous Geometric Ergodicity.

Most of the current practically verifiable ergodicity results on adaptive MCMC rely on 'uniform' properties of the Metropolis-Hastings kernels P_s with respect to the adaptation parameter $s \in \mathrm{S}$ [1, 5, 6, 29]. In the case of the AM and ASM algorithms, one commonly assumes the following simultaneous geometric drift and minorisation conditions.

Assumption 3.1. There exist a set $D \in \mathcal{B}(\mathbb{R}^d)$, a function $V : \mathbb{R}^d \to [1,\infty)$, constants $\delta, \lambda \in (0,1)$ and $b < \infty$, and a probability measure ν concentrated on D such that

$$P_s V(x) \leq \lambda V(x) + \mathbb{1}_D(x) b \qquad \text{and} \tag{14}$$

$$P_s(x,A) \geq \mathbb{1}_D(x)\delta\nu(A) \tag{15}$$

for all $x \in \mathbb{R}^d$, $s \in \mathbb{S}$ and $A \in \mathcal{B}(\mathbb{R}^d)$.

As already mentioned in Section 1.4, under Assumption 3.1, there are constants $\rho \in (0,1)$ and $R < \infty$ depending only on δ, λ and b such that

$$\|P_s^n(x, \cdot) - \pi(\cdot)\|_V \leq RV(x)\rho^n$$

holds for all $x \in \mathbb{R}^d$ and $s \in \mathbb{S}$.

Assumption 3.1 can be shown to hold under practically verifiable conditions. For example, Andrieu and Moulines [1] show, modifying the proof of Jarner and Hansen [18] the following.

Proposition 3.2. *Suppose the target distribution π satisfies Assumption 1.4 and that the proposal distributions q_s are zero-mean Gaussian with covariance s, and let $0 < a \leq b < \infty$ be constants. Let $K_{a,b} \subset \mathbb{R}^{d\times d}$ stand for the positive definite matrices s with all eigenvalues $\lambda_1(s), \ldots, \lambda_d(s) \in (a,b)$. Then, Assumption 3.1 holds for the Metropolis kernels $\{M_{q_s}\}_{s\in K_{a,b}}$.*

In fact, the result of Andrieu and Moulines is more general allowing one to employ also non-Gaussian proposal distributions. As discussed in Section 1.4, the conditions on the target distribution are quite close to optimal.

It may be also worth mentioning that the bound $[a,b]$ on the eigenvalues of s in Proposition 3.2 is necessary.

Proposition 3.3. *Let q_s and $K_{a,b}$ be defined as in Proposition 3.2, and assume that $a = 0$ or $b = \infty$. Then, Assumption 3.1 cannot hold for $\{M_{q_s}\}_{s\in K_{a,b}}$.*

Proof. Denote $P_s = M_{q_{sI}}$ and suppose Assumption 3.1 holds for all $s \in \mathbb{S} = \mathbb{R}_+$, the positive real numbers. The set D must have a positive π-measure, since (14) implies that for all $x \in \mathbb{R}^d$ there is a positive integer $n \geq 1$ such that $P_s^n(x, D) > 0$ [e.g. 24, Theorem 11.3.4].

Suppose for a moment that D is bounded and let $R < \infty$ stand for the diameter of D. Observe next that the probability measure ν satisfying (15) must be absolutely continuous with respect to the Lebesgue measure. For if not, then there is a Lebesgue null set $A \in \mathcal{B}(\mathbb{X})$ such that $\nu(A) > 0$. But then, for every $x \in D$, it must hold $P_s(x, A) = P_s(x, \{x\} \cap A)$ since the remainder of P_s is absolutely continuous. Therefore, $P_s(x, \{x\} \cap A) \geq \delta\nu(A) > 0$, implying that $A = D$, which is a contradiction. Therefore, for all $x \in D$ there is a $r > 0$ such that

$$\begin{aligned}\nu(D) - \frac{1}{2} &\leq \nu\big(D \setminus B(x,r)\big) \leq \delta^{-1} P_s\big(x, D \setminus B(x,r)\big) \\ &\leq \delta^{-1} \int_{r\leq|y|\leq R} q_{sI}(y)\mathrm{d}y \to 0\end{aligned}$$

if $s \to \infty$ or $s \to 0$. Therefore, $\nu(D) \leq 1/2$, which is a contradiction.

Finally, it is easy to see that an unbounded D cannot satisfy (15) for any $s > 0$ and $\delta > 0$. □

3.2. **Overview of Previous Ergodicity Results.** The seminal article [14] established the correct ergodicity properties of the AM algorithm, when the target density π is bounded and compactly supported. A strong law of large numbers was shown to hold for bounded functions (on the support of π). The assumptions on the target distribution implied the uniform ergodicity of the Metropolis samplers, and also implied that the covariance estimator was naturally bounded. The proof was based on deterministic estimates and so called *mixingales*; see, for example, [15, Theorem 2.21].

Atchadé and Rosenthal [6] extended the original mixingale proof [14] and applied it to show a strong law of large numbers for (a version of) the ASM algorithm, assuming simultaneous geometric ergodicity such as described above in Section 3.1. Their result yields the ergodicity of a truncated version of the ASM algorithm. For example, suppose $-\infty < a \leq b < \infty$ are constants and replace (9) with

$$S_{n+1} = \max\{a, \min\{b, S_n + \eta_{n+1} H_{\text{ASM}}(S_n, X_{n+1}, Z_{n+1})\}\}. \tag{16}$$

It is obvious that the constants a and b above need to be chosen with care, so that S_n can get such values that admit good mean acceptance rates.

Andrieu, Moulines and Priouret [1, 4] considered a Robbins-Monro stochastic approximation adaptation similar the one formulated in Section 2.1. They applied a martingale approximation based on the Poisson equation and showed that, under certain conditions, the stochastic approximation process converges. Moreover, in the case of convergence, they showed that a central limit theorem holds. They applied the technique to prove the convergence and the correct ergodicity of the AM algorithm in the sequential reprojection framework. For example, suppose $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ are positive sequences decaying to zero and increasing to infinity, respectively. Start to run the AM algorithm of Section 2.2 with the proposal q_c having a covariance $\theta^2 c$. Continue until the adaptation 'skips outside' $[a_1, b_1]$. Then, essentially restart the algorithm and change the truncation set to $[a_2, b_2]$. More precisely, introduce two counters initialised to $\kappa = \xi = 1$. Instead of (8) and (9), compute

$$\tilde{X}_{n+1} \sim P_{S_n}(X_n, \cdot) \quad \text{and} \tag{17}$$

$$(\tilde{M}_{n+1}, \tilde{C}_{n+1}) = \tilde{S}_{n+1} = S_n + \eta_{\kappa+\xi} H_{\text{AM}}(S_n, X_{n+1}). \tag{18}$$

Then, if $|\tilde{M}_{n+1}| \leq b_\kappa$ and all eigenvalues $\lambda_1(\tilde{C}_{n+1}), \ldots, \lambda_d(\tilde{C}_{n+1}) \in [a_\kappa, b_\kappa]$, set $S_{n+1} = \tilde{S}_{n+1}$, $X_{n+1} = \tilde{X}_{n+1}$ and increment the counter ξ by one. Otherwise, reinitialise $S_{n+1} = S_1$, $X_{n+1} = X_1$ and $\xi = 1$, and increment the truncation set counter κ by one.

The reprojection approach has some benefits compared to the truncation to a fixed set such as (16). It is only required that within each truncation set, which are determined above by a_n and b_n, the Metropolis kernels satisfy certain uniform assumptions, including the geometric ergodicity discussed in Section 3.1. Moreover, the AM covariance C_n may ultimately

have any positive definite value; indeed, with certain assumptions, C_n converges to the true covariance of π [1, 4]. On the other hand, each time the adaptation 'escapes' from the current truncation set, the adaptation is essentially reinitialised to its starting values. This procedure can make the adaptation quite inefficient, if the reprojections occur frequently. Therefore, the reprojection sets (the sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ above) need to be chosen with care.

Roberts and Rosenthal [29] introduced another technique for establishing the ergodicity of adaptive MCMC algorithms. Their framework, based on coupling constructions, introduces quite weak assumptions. The technique, using mostly elementary arguments, also allows quite simple and elegant proofs. As an example, they applied their technique and showed the correct ergodicity and a weak law of large numbers for the AM algorithm with the same assumptions as in the original work [14]. The general assumptions, while weak, are quite implicit and therefore often difficult to verify in practice. In the recent preprint [7], Bai, Roberts and Rosenthal establish an ergodicity result on certain type of unconstrained AM algorithm, which will be discussed in Section 3.4.

The recent advances in the theoretical side of adaptive MCMC include the work of Atchadé and Fort [5], who consider some sub-geometric ergodicity assumptions, which are weaker than the ones discussed in Section 3.1. They use a different martingale approximation technique than Andrieu and Moulines [1], and apply also coupling arguments. The authors apply their result on an algorithm combining the AM and the ASM algorithms as described in Section 2.4, but involving truncations both for the scaling parameters as in (16) and for the AM mean and covariance. As already mentioned in Section 1.4, the current techniques allow the subgeometric ergodicity to be verified in practice only in rare cases.

3.3. **Sequential Truncation.** The article [A] formulates a so called *sequential truncation* approach within adaptive MCMC. It is similar to the reprojection method of Andrieu and Moulines [1], but does not involve reinitialisations. It is shown that under certain conditions, a strong law of large numbers and a central limit theorem hold for the sequential truncation adaptation. These general results have interest in their own right. In this work, they have a central role in analysing the unconstrained AM and ASM algorithms, as discussed in Sections 3.4 and 3.5.

Let $K_1 \subset K_2 \subset \cdots \subset K_n \in \mathcal{B}(\mathbb{S})$ be an increasing sequence of measurable subsets of the adaptation parameter space $\mathbb{S}$, assume $\tilde{S}_1 \equiv s_1 \in K_1$ and let $\tilde{X}_1 \equiv x_1 \in \mathbb{X}$. The sequentially truncated process $(\tilde{X}_n, \tilde{Z}_n, \tilde{S}_n)_{n\geq 2}$ is defined recursively through

$$(\tilde{X}_{n+1}, \tilde{Z}_{n+1}) \sim \tilde{P}_{\tilde{S}_n}(\tilde{X}_n, \cdot) \qquad \text{and} \tag{19}$$

$$\tilde{S}_{n+1} = \sigma_{n+1}\big(\tilde{S}_n,\, \eta_{n+1} H(\tilde{S}_n, \tilde{X}_{n+1}, \tilde{Z}_{n+1})\big) \tag{20}$$

where $\sigma_n : \mathbb{S} \times \mathbb{R}^{d_{\mathbb{S}}} \to \mathbb{S}$ are defined as

$$\sigma_n(s, s') = \begin{cases} s + s', & \text{if } s + s' \in K_n \text{ and} \\ s, & \text{otherwise.} \end{cases}$$

The truncation functions σ_n only ensure that $\tilde{S}_n \in K_n$ for all $n \geq 1$. Otherwise, the above defined process $(\tilde{X}_n, \tilde{Z}_n, \tilde{S}_n)_{n \geq 2}$ coincides with the process $(X_n, Z_n, S_n)_{n \geq 2}$ following the general framework of Section 2.1. In fact, the results in Sections 3.4 and 3.5 below are based on independent estimates that guarantee that the unconstrained and the sequentially truncated adaptive MCMC processes coincide with a large probability.

Suppose that the adaptation algorithm has the above described form, and the following assumptions are satisfied for some constants $c \geq 1$ and $\epsilon \geq 0$.

(A1) For each $n \geq 1$, the simultaneous geometric drift and minorisation conditions hold. In particular, there is a drift function $V : \mathbb{X} \to [1, \infty)$ such that for all $n \geq 1$ Assumption 3.1 holds for $s \in K_n$, with the minorising set D_n and the minorisation measure ν_n. Furthermore, the drift constants $\lambda_n \in (0, 1)$ and $b_n \in (0, \infty)$ are increasing, and the minorisation constants $\delta_n \in (0, 1]$ are decreasing with respect to n, and they are polynomially bounded so that

$$(1 - \lambda_n)^{-1} \vee \delta_n^{-1} \vee b_n \leq c n^{\epsilon}.$$

(A2) For all $n \geq 1$ and any $r \in (0, 1]$, there is $c' = c'(r) \geq 1$ such that for all $s, s' \in K_n$,

$$\|P_s f - P_{s'} f\|_{V^r} \leq c' n^{\epsilon} \|f\|_{V^r} |s - s'|.$$

(A3) There is a $\beta \in [0, 1/2]$ such that for all $n \geq 1$, $s \in K_n$, $x \in \mathbb{X}$ and $z \in \mathbb{R}^{d'}$

$$|H(s, x, z)| \leq c n^{\epsilon} V^{\beta}(x).$$

Theorem 3.4. *Assume (A1)–(A3) hold and let f be a function with $\|f\|_{V^\alpha} < \infty$ for some $\alpha \in (0, 1 - \beta)$. Assume $\epsilon < \kappa_*^{-1} [(1/2) \wedge (1 - \alpha - \beta)]$, where $\kappa_* \geq 1$ is an independent constant, and that $\sum_{k=1}^{\infty} k^{\kappa_* \epsilon - 1} \eta_k < \infty$. Then,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(\tilde{X}_k) = \int f(x) \pi(x) \mathrm{d}x \qquad \textit{almost surely.}$$

Proof. Theorem 3.4 is a slight modification of Theorem 1 in [A]. In particular, (A3) is modified to include H with the additional argument z, but as the bound in (A3) is independent of z, the proof of Theorem 1 in [A] applies essentially without changes. □

In practice, the simultaneous geometric ergodicity in (A1) must be established with specific constants. For this purpose, the article [A] introduces the following decay condition, which is slightly more stringent than in Assumption 1.4.

Assumption 3.5. Suppose that the target distribution π satisfies Assumption 1.4. Moreover, for for some constant $\rho > 1$

$$\lim_{r\to\infty} \sup_{|x|\geq r} \frac{x}{|x|^\rho} \cdot \nabla \log \pi(x) = -\infty. \tag{21}$$

Having Assumption 3.5, it is possible to establish the following 'refinement' of Proposition 3.2, which only assumes a lower bound on the covariance matrices.

Proposition 3.6. *Suppose the target density π satisfies Assumption 3.5 and let $P_s = M_{q_s}$ stand for the Metropolis kernel* (3) *with q_s, the zero-mean Gaussian density with covariance matrix s. Let $a > 0$ be a real number, and let $\mathcal{P}_a \subset \mathcal{C}^d$ stand for the symmetric and positive definite matrices with eigenvalues greater than a.*

There exist a compact set $D \subset \mathbb{R}^d$, a probability measure ν on D and a constant $b \in [0, \infty)$ such that for all $s \in \mathcal{P}_a$, $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$

$$\begin{aligned} P_s V(x) &\leq \lambda_s V(x) + b\mathbb{1}_D(x) \qquad \text{and} \\ P_s(x, A) &\geq \delta_s \mathbb{1}_D(x)\nu(A) \end{aligned}$$

where $V(x) := [\sup_z \pi(z)]^{1/2}\pi^{-1/2}(x) \geq 1$ and the constants $\lambda_s, \delta_s \in (0,1)$ satisfy the bound

$$(1-\lambda_s)^{-1} \vee \delta_s^{-1} \leq c|\det(s)|^{-1}$$

for some constant $c \geq 1$.

Proof. Proposition 18 in [A]. □

Remark 3.7. Proposition 3.6 is extended for non-Gaussian proposal densities in Appendix B of the article [B]. In particular, Proposition 3.6 is verified to hold with heavy-tailed multivariate Student distributions.

In order to establish a central limit theorem, one more condition is required to hold, with the same constants $c \geq 1$ and $\epsilon \geq 0$ as in (A1)–(A3).

(A4) There is a $\beta \in [0, 1/2]$ such that (A3) holds, and for all $n \geq 1$, $x \in \mathbb{X}$, $z \in \mathbb{R}^{d'}$ and $s, s' \in K_n$,

$$|H(s, x, z) - H(s', x, z)| \leq cn^\epsilon |s - s'| V^\beta(x).$$

Theorem 3.8. *Assume (A1)–(A4) hold. Let f be a function with $\|f\|_{V^\alpha} < \infty$ for some $\alpha \in (0, (1-\beta)/2)$. Assume $\epsilon < \kappa_{**}^{-1}[1/2 \wedge (1 - 2\alpha - \beta)]$ and $\sum_{k=1}^\infty k^{\kappa_{**}\epsilon - 1/2}\eta_k < \infty$, where $\kappa_{**} \geq 1$ is an independent constant. Furthermore, assume that $\tilde{S}_k$ converges a.s. to some constant limit s_∞ in the interior of K_N for some index $N < \infty$. Then,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \left[f(\tilde{X}_k) - \pi(f) \right] \xrightarrow{n\to\infty} N(0, \sigma^2)$$

in distribution, where $N(0, \sigma^2)$ stands for the zero-mean Gaussian distribution with the variance $\sigma^2 = \sigma^2(f, s_\infty) < \infty$.

Proof. Theorem 3.8 is a simplified version of Theorem 7 in [A]. □

The assumption in Theorem 3.8 requiring that the adaptation parameter $\tilde{S}_n$ converges may be difficult to check in practice. In the case of the AM algorithm it can be verified to hold; see Section 3.4 below.

3.4. **Ergodicity of the Unconstrained AM Algorithm.** Consider the AM algorithm as defined in Section 2.2. Throughout this section, it is assumed that the tail of the adaptation weight sequence is defined as $\eta_n := \mu n^{\gamma}$ for some constants $\mu \in (0, 1]$ and $\gamma \in (1/2, 1]$.

There are some special cases that are considered separately. The article [A] considers the AM algorithm as described in Section 2.2, having a covariance lower bound induced by the factor ϵI.

Theorem 3.9. *Let the proposal densities q_c be zero-mean Gaussian with covariance $\theta^2 c + \epsilon I$, where $\theta > 0$ and $\epsilon > 0$ are constants. Suppose the target density π satisfies Assumption 3.5. Then, for any measurable function $f : \mathbb{R}^d \to \mathbb{R}$ with $\sup_x |f(x)|^{\xi} \pi^{1/2}(x) < \infty$ for some constant $\xi > 1$,*

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{n \to \infty} \pi(f) \tag{22}$$

almost surely. If, in addition, $\sup_x |f(x)|^{\xi} \pi^{1/4} < \infty$ and $\eta_n := n^{-1}$, then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} [f(X_k) - \pi(f)] \xrightarrow{n \to \infty} N(0, \sigma^2) \tag{23}$$

in distribution, where $\sigma^2 = \sigma^2(f) < \infty$ is a constant.

Proof. Theorem 13 in [A]. □

Remark 3.10. If the conditions of Theorem 3.9 are satisfied, $\pi(x)$ decays faster than any exponential, and hence (22) and (23) hold for exponential moments. In particular, they hold for power moments, that is, for $f(x) = |x|^p$ for any $p \geq 0$. Therefore, if $\eta_n := n^{-1}$, the adaptation parameter $S_n \to (m_\pi, c_\pi)$ where m_π and c_π are the mean and covariance of π.

The article [C] considers, in general, the case $\epsilon = 0$. That is, the proposal q_c is a Gaussian density with the covariance $\theta^2 c$, with eigenvalues having no explicit lower bound. The first result involves an additional fixed proposal component ensuring the stability.

Theorem 3.11. *Let $\beta \in (0, 1]$ and $\theta > 0$ be constants and let q_{fix} be a probability density bounded away from zero in some neighbourhood of the origin. Assume the proposal densities q_c are defined as a mixture*

$$q_c(z) := \beta q_{fix}(z) + (1 - \beta) \tilde{q}_c(z)$$

where $\tilde{q}_c$ denotes a zero-mean Gaussian density with covariance $\theta^2 c$. Suppose also that the target density π satisfies Assumption 3.5. Then, the strong law of large number (22) *and the central limit theorem* (23) *hold as stated in Theorem 3.9.*

Proof. Theorem 31 in [C]. The proof is essentially based on Theorem 3.9 and an independent martingale argument implying that the fixed component βq_{fix} guarantees that the eigenvalues of C_n are bounded away from zero. □

While omitting the parameter $\epsilon > 0$, Theorem 3.11 includes two additional parameters: the mixing probability $\beta \in (0, 1)$ and the fixed symmetric proposal distribution q_{fix}. It has the advantage that the 'worst case scenario' having ill-defined q_{fix} only 'wastes' the fixed proportion β of samples, while S_n can take any positive definite value on adaptation.

An analogous result as in Theorem 3.11 was obtained by Bai, Roberts and Rosenthal [7]. In particular, the authors show that if the target density π satisfies Assumption 1.4 and the fixed component q_{fix} is a uniform density on a centred ball $B(0, R)$ having a large enough radius $R > 0$, then the AM algorithm has the correct ergodic properties. Theorem 3.11 is, however, obtained using a different technique, allowing less stringent assumptions on q_{fix} and ensuring a strong law of large numbers for possibly unbounded functionals.

The article [C] contains also some results without this fixed component. They are, however, significantly weaker, applying only for a one-dimensional case.

Theorem 3.12. *Assume $d = 1$, the proposal density q_c is defined as a zero-mean Gaussian with variance $\theta^2 c$ and $\log \pi$ is uniformly continuous. Then, there is a constant $b > 0$ such that $\liminf_{n\to\infty} C_n \geq b$.*

Proof. Theorem 21 in [C]. □

In the article [C], Theorem 3.12 is also shown to imply a strong law of large numbers for a Laplace target $\pi(x) := (2b)^{-1} e^{-b|x|}$. This result has little direct practical relevance, but it is the first ergodicity result for such a fully adaptive and unconstrained version of the AM algorithm.

In addition to results ensuring the correct ergodicity, the article [C] includes some analysis on the behaviour of the AM covariance, when the algorithm is applied to an improper uniform target $\pi \equiv \text{constant} > 0$. In this case, every proposal is accepted, and the algorithm forms an 'adaptive random walk'

$$X_{n+1} = X_n + \theta C_n^{1/2} W_{n+1} \tag{24}$$

where $(W_n)_{n\geq 2}$ are independent standard Gaussian $N(0, I)$ random variables. Fix a unit vector $u \in \mathbb{R}^d$ and consider the expectations

$$\begin{aligned} a_n &:= \mathbb{E}\left[|u^T(X_n - M_{n-1})|^2\right] \qquad \text{and} \\ b_n &:= \mathbb{E}\left[u^T C_n u\right]. \end{aligned}$$

for $n \geq 1$, with the convention that $M_0 \equiv X_1$. It is straightforward to show that, in case of the 'adaptive random walk' (24), a_n and b_n can be computed

recursively through

$$\begin{aligned} a_{n+1} &= (1-\eta_n)^2 a_n + \theta^2 b_n \qquad \text{and} \\ b_{n+1} &= (1-\eta_{n+1}) b_n + \eta_{n+1} a_{n+1}. \end{aligned}$$

If $\theta < 1$, numerical computations show that these sequences may decay quite rapidly for a long time before starting to grow. Asymptotically, however, their behaviour can be characterised as given below.

Theorem 3.13. *For all $\lambda > 1$ there is an index $n_0 \geq m$ such that for all $n \geq n_0$ and $k \geq 1$, the following bounds hold:*

$$\frac{1}{\lambda}\left(\theta \sum_{j=n+1}^{n+k} \sqrt{\eta_j}\right) \leq \log\left(\frac{a_{n+k}}{a_n}\right) \leq \lambda\left(\theta \sum_{j=n+1}^{n+k} \sqrt{\eta_j}\right).$$

Proof. Theorem 1 in [C]. □

Remark 3.14. Theorem 3.13 applied to $\eta_n := \mu n^{-\gamma}$ implies the following asymptotical growth rate, when $(X_n)_{n\geq 2}$ follows the 'adaptive random walk' recursion (24):

$$\mathbb{E}\left[u^T C_n u\right] \simeq \exp\left(\frac{\theta\sqrt{\mu}}{1-\frac{\gamma}{2}} n^{1-\frac{\gamma}{2}}\right).$$

Particularly, in the original setting $\eta_n := n^{-1}$, one has $\mathbb{E}\left[u^T C_n u\right] \simeq e^{2\theta\sqrt{n}}$.

Intuitively speaking, this shows how the AM algorithm behaves, when it has a very small covariance parameter C_n compared to the true scale of a sufficiently smooth target distribution. In that case, most of the proposals are accepted, and it is expected that the covariance parameter C_n grows at the above mentioned speed, until it reaches the scale of the target.

3.5. **Ergodicity of the Unconstrained ASM Algorithm.** The article [B] establishes a strong law of large numbers for the ASM algorithm of Section 2.3, without any constraints or modifications. It relies on independent martingale arguments implying that the paths $\log S_n$ are bounded away from zero, and have a controlled polynomial growth.

The additional assumption required by the lower bound is that the tail contours of the target distribution must be uniformly smooth.

Definition 3.15. Suppose that $\{A_i\}_{i\in I}$ is a collection of sets $A_i \subset \mathbb{R}^d$ each consisting of finitely many disjoint components that are closures of C_1-domains. Let $n_i(x)$ stand for the outer-pointing normal at x in the boundary ∂A_i. Then, $\{A_i\}_{i\in I}$ have *uniformly continuous normals* if for all $\epsilon > 0$ there is a $\delta > 0$ such that for any $i \in I$ it holds that $|n_i(x) - n_i(y)| \leq \epsilon$ for all $x, y \in \partial A_i$ such that $|x - y| \leq \delta$.

This definition essentially states that the boundaries ∂A_i must be regular enough to ensure that if one looks at ∂A_i at a small enough scale, it will look locally almost like a plane.

The two main results of [B] consider compactly supported targets and targets having an unbounded support separately.

Theorem 3.16. *Assume π has a compact support $\mathbb{X} \subset \mathbb{R}^d$ and π is continuous, bounded and bounded away from zero on $\mathbb{X}$. Moreover, assume that the set $\mathbb{X}$ has a uniformly continuous normal in the sense of Definition 3.15. Then, for any $0 < \alpha^* < 1/2$ and a bounded measurable function f, the strong law of large numbers* (22) *holds.*

Proof. Theorem 2 in [B]. □

Theorem 3.17. *Suppose π fulfils Assumption 3.5 and there is a $t_0 > 0$ such that the contour sets $\{L_t\}_{0<t\leq t_0}$ where $L_t := \{x \in \mathbb{R}^d : \pi(x) \geq t\}$ have uniformly continuous normals in the sense of Definition 3.15. Then, for any $0 < \alpha^* < 1/2$ and any measurable function $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $\sup_x |f(x)|^\xi \pi^{1/2}(x) < \infty$ for some constant $\xi > 1$, the strong law of large numbers* (22) *holds.*

Proof. Theorem 4 in [B]. □

Remark 3.18. For many practical target densities satisfying Assumption 3.5 the tail contours are (essentially) scaled copies of each other, in which case they have automatically uniformly continuous normals. This indicates that Theorem 3.17 is practically a counterpart of Theorem 3.9 verifying the ergodicity of the Adaptive Metropolis algorithm.

Remark 3.19. The 'safe' values for the desired acceptance rate stipulated by Theorems 3.16 and 3.17 are $\alpha^* \in (0, 1/2)$ including probably the most commonly used values for a random walk Metropolis algorithms $\alpha^* = 0.234$ and $\alpha^* = 0.44$ as discussed in Section 2.3.

Remark 3.20. The results of article [B] extend also for the analysis of the algorithm combining the AM and the ASM algorithms, as described in Section 2.4. It is, however, required that the AM covariance parameter C_n is constrained so that ratios of the eigenvalues of C_n are bounded; See Remark 19 in the article [B].

4. Discussion

Most theoretical results on adaptive MCMC in the literature prior to this work were based on assumptions that typically require, either explicitly or implicitly, one to modify a natural adaptation scheme by some additional constraints. The constraint parameters may be difficult to choose in practice, and the algorithms are generally sensitive to these parameters. In the worst case, poor choices can render the algorithms useless.

This work addressed the stability and ergodicity of two commonly applied adaptive MCMC algorithms without such constraints. The practical implications of the results are twofold. First, the algorithms are shown to be intrinsically stable under certain conditions, implying that they are fairly 'safe' to apply in practice. Second, the unconstrained algorithms

are more universal and 'fully adaptive,' applying for target distributions with different scales and shapes, with less parameters to adjust before the algorithm can be applied to some practical problem. In addition, the techniques developed for the analysis may be applied also to other adaptive MCMC algorithms.

Some of the present results have a limited applicability due to technical assumptions. Many of the results are in a sense 'preliminary' as it is expected that many of the assumptions can be relaxed or made substantially weaker in the future. One important issue is provided by the heavy-tailed target distributions for which the Metropolis algorithms fail to be geometrically ergodic. New tools are necessary to allow easy practical verification of a suitable sub-geometric ergodicity condition. It is also important to develop more general conditions to ensure the stability of other type of adaptation schemes.

5. Summary of Included Articles

Article [A]: ***On the ergodicity of the adaptive Metropolis algorithm on unbounded domains.*** Sufficient conditions are considered to ensure the correct ergodicity of the Adaptive Metropolis (AM) algorithm for target distributions with a non-compact support. The conditions ensuring a strong law of large numbers and a central limit theorem require that the tails of the target density decay super-exponentially and have regular contours. The result is based on the ergodicity of an auxiliary process that is sequentially truncated to feasible adaptation sets, and independent estimates of the growth rate of the AM chain and the corresponding geometric drift constants.

The proof of the central limit theorem was omitted in the article that was accepted for publication in the Annals of Applied Probability. The extended preprint version containing the central limit theorem is included in the dissertation.

Article [B]: ***On the stability and ergodicity of an adaptive scaling Metropolis algorithm.*** The stability and ergodicity properties of an adaptive random walk Metropolis algorithm are considered. Unlike the previously proposed forms of this algorithm, the adapted scaling parameter is not constrained within a predefined compact interval. This makes the algorithm more generally applicable and 'automatic,' with two parameters less to be adjusted. A strong law of large numbers is shown to hold when the target density is smooth enough and has either compact support or super-exponentially decaying tails.

Article [C]: ***Can the adaptive Metropolis algorithm collapse without the covariance lower bound?*** This article considers variants of the AM algorithm that do not explicitly bound the eigenvalues of S_n away from zero. The behaviour of S_n is studied in detail, indicating that the eigenvalues of

S_n do not tend to collapse to zero in general. In dimension one, it is shown that S_n is bounded away from zero if the logarithmic target density is uniformly continuous. For a modification of the AM algorithm including an additional fixed component in the proposal distribution, the eigenvalues of S_n are shown to stay away from zero with a practically non-restrictive condition. This result implies a strong law of large numbers for super-exponentially decaying target distributions with regular contours.

REFERENCES

[1] C. Andrieu and É. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.*, 16(3):1462–1505, 2006.

[2] C. Andrieu and C. P. Robert. Controlled MCMC for optimal sampling. Technical Report Ceremade 0125, Université Paris Dauphine, 2001.

[3] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statist. Comput.*, 18(4):343–373, Dec. 2008.

[4] C. Andrieu, É. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44 (1):283–312, 2005.

[5] Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 2009. to appear.

[6] Y. F. Atchadé and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

[7] Y. Bai, G. O. Roberts, and J. S. Rosenthal. On the containment condition for adaptive Markov chain Monte Carlo algorithms. Preprint, July 2008. URL `http://probability.ca/jeff/research.html`.

[8] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1A): 700–738, 2005.

[9] M. Bédard. Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Process. Appl.*, 118(12):2198–2222, 2008.

[10] H.-F. Chen. *Stochastic Approximation and Its Applications*. Number 64 in Nonconvex Optimization and Its Applications. Kluwer Academic Publishers, 2002. ISBN 1-4020-0806-6.

[11] R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.

[12] A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. In *Bayesian Statistics 5*, pages 599–607. Oxford University Press, 1996.

[13] W. R. Gilks, G. O. Roberts, and S. K. Sahu. Adaptive Markov chain Monte Carlo through regeneration. *J. Amer. Statist. Assoc.*, 93(443): 1045–1054, 1998.

[14] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.
[15] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980. ISBN 0-12-319350-8.
[16] D. Hastie. *Toward Automatic Reversible Jump Markov Chain Monte Carlo*. PhD thesis, University of Bristol, Mar. 2005.
[17] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, Apr. 1970.
[18] S. F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85:341–361, 2000.
[19] S. F. Jarner and G. O. Roberts. Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.*, 12(1):224–247, 2002.
[20] S. F. Jarner and G. O. Roberts. Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.*, 34(4):781–815, Dec. 2007.
[21] H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Number 35 in Applications of Mathematics: Stochastic Modelling and Applied Probability. Springer-Verlag, second edition, 2003. ISBN 0-387-00894-2.
[22] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.
[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, June 1953.
[24] S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, second edition, 2009. ISBN 978-0-521-73182-9.
[25] S. P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 4(4):981–1011, 1994.
[26] E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators*. Number 83 in Cambridge Tracts in Mathematics. Cambridge University Press, 1984. ISBN 0-521-60494-X.
[27] E. Nummelin. MC's for MCMC'ists. *International Statistical Review*, 70 (2):215–240, 2002.
[28] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22:400–407, 1951.
[29] G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.*, 44(2):458–475, 2007.
[30] G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *J. Comput. Graph. Statist.*, 18(2):349–367, 2009.

[31] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

[32] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

[33] C. Sherlock and G. Roberts. Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli*, 15(3): 774–798, 2009.

[34] A. N. Shiryaev. *Probability*. Springer-Verlag, New York, second edition, 1996. ISBN 0-387-94549-0.

[35] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1728, Dec. 1994.

[36] L. Tierney and A. Mira. Some adaptive Markov chain Monte Carlo methods for Bayesian inference. *Statist. Med.*, 18:2507–2515, 1999.

87. LLORENTE, JOSÉ G., Discrete martingales and applications to analysis. (40 pp.) 2002
88. MITSIS, THEMIS, Topics in harmonic analysis. (52 pp.) 2003
89. KÄRKKÄINEN, SALME, Orientation analysis of stochastic fibre systems with an application to paper research. (53 pp.) 2003
90. HEINONEN, JUHA, Geometric embeddings of metric spaces. (44 pp.) 2003
91. RAJALA, KAI, Mappings of finite distortion: Removable singularities. (23 pp.) 2003
92. FUTURE TRENDS IN GEOMETRIC FUNCTION THEORY. RNC WORKSHOP JYVÄSKYLÄ 2003. Edited by D. Herron. (262 pp.) 2003
93. KÄENMÄKI, ANTTI, Iterated function systems: Natural measure and local structure. (14 pp.) 2003
94. TASKINEN, SARA, On nonparametric tests of independence and robust canonical correlation analysis. (44 pp.) 2003
95. KOKKI, ESA, Spatial small area analyses of disease risk around sources of environmental pollution: Modelling tools for a system using high resolution register data. (72 pp.) 2004
96. HITCZENKO, PAWEŁ, Probabillistic analysis of sorting algorithms. (71 pp.) 2004
97. NIEMINEN, TOMI, Growth of the quasihyperbolic metric and size of the boundary. (16 pp.) 2005
98. HAHLOMAA, IMMO, Menger curvature and Lipschitz parametrizations in metric spaces. (8 pp.) 2005
99. MOLTCHANOVA, ELENA, Application of Bayesian spatial methods in health and population studies using registry data. (55 pp.) 2005
100. HEINONEN, JUHA, Lectures on Lipschitz analysis. (77 pp.) 2005
101. HUJO, MIKA, On the approximation of stochastic integrals. (19 pp.) 2005
102. LINDQVIST, PETER, Notes on the p-Laplace equation. (80 pp.) 2006
103. HUKKANEN, TONI, Renormalized solutions on quasi open sets with nonhomogeneous boundary values. (41 pp.) 2006
104. HÄHKIÖNIEMI, MARKUS, The role of representations in learning the derivative. (101 pp.) 2006
105. HEIKKINEN, TONI, Self-improving properties of generalized Orlicz–Poincaré inequalities. (15 pp.) 2006
106. TOLONEN, TAPANI, On different ways of constructing relevant invariant measures. (13 pp.) 2007
107. HORPPU, ISMO, Analysis and evaluation of cell imputation. (248 pp.) 2008
108. SIRKIÄ, SEIJA, Spatial sign and rank based scatter matrices with applications. (29 pp.) 2007
109. LEIKAS, MIKA, Projected measures on manifolds and incomplete projection families. (16 pp.) 2007
110. TAKKINEN, JUHANI, Mappings of finite distortion: Formation of cusps. (10 pp.) 2007
111. TOLVANEN, ASKO, Latent growth mixture modeling: A simulation study. (201 pp.) 2007
112. VARPANEN, HARRI, Gradient estimates and a failure of the mean value principle for p-harmonic functions. (66 pp.) 2008
113. MÄKÄLÄINEN, TERO, Nonlinear potential theory on metric spaces. (16 pp.) 2008
114. LUIRO, HANNES, Regularity properties of maximal operators. (11 pp.) 2008
115. VIHOLAINEN, ANTTI, Prospective mathematics teachers' informal and formal reasoning about the concepts of derivative and differentiability. (86 pp.) 2008
116. LEHRBÄCK, JUHA, Weighted Hardy inequalities and the boundary size. (21 pp.) 2008
117. NISSINEN, KARI, Small area estimation with linear mixed models from unit-level panel and rotating panel data. (230 pp.) 2009
118. BOJARSKI, B.V., Generalized solutions of a system of differential equations of the first order and elliptic type with discontinuous coefficients. (64 pp.) 2009
119. RAJALA, TAPIO, Porosity and dimension of sets and measures. (22 pp.) 2009
120. MYLLYMÄKI, MARI, Statistical models and inference for spatial point patterns with intensity-dependent marks. (115 pp.) 2009
121. AVIKAINEN, RAINER, On generalized bounded variation and approximation of SDEs. (18 pp.) 2009
122. ZÜRCHER, THOMAS, Regularity of Sobolev–Lorentz mappings on null sets. (13 pp.) 2009
123. TOIVOLA, ANNI, On fractional smoothness and approximations of stochastic integrals. (19 pp.) 2009

ISBN 978-951-39-3809-3
ISSN 1457-8905