Computational Complexity of Metropolis-Hastings Methods in High Dimensions

Alexandros Beskos and Andrew Stuart

Abstract This article contains an overview of the literature concerning the computational complexity of Metropolis-Hastings based MCMC methods for sampling probability measures on $\mathbb{R}^d$, when the dimension d is large. The material is structured in three parts addressing, in turn, the following questions: (i) what are sensible assumptions to make on the family of probability measures indexed by d?; (ii) what is known concerning computational complexity for Metropolis-Hastings methods applied to these families?; (iii) what remains open in this area?

1 Introduction

Metropolis-Hastings methods [19, 15] form a widely used class of MCMC methods [17, 21] for sampling from complex probability distributions. It is therefore of considerable interest to develop mathematical analyses which explain the structure inherent in these algorithms, especially structure which is pertinent to understanding the computational complexity of the algorithm. In this short article we overview the literature concerning the computational complexity of Metropolis-Hastings based MCMC methods for sampling probability measures on $\mathbb{R}^d$, when the dimension d is large. The presentation will be discursive: theorems will not be given, rather we will outline the essential ideas and give pointers to the relevant literature where the theorems are stated and proved.

The article is organized around three sections. In section 2 we address the question of how to make sensible assumptions on the family of probability measures

Alexandros Beskos
Department of Statistical Science, University College, London, WC1E 6BT, UK,
http://www.homepages.ucl.ac.uk/~ucakabe/

Andrew Stuart
Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK,
http://www.maths.warwick.ac.uk/~stuart

indexed by dimension d. In section 3 we overview what is known concerning computational complexity for Metropolis-Hastings methods applied to these families. Section 4 highlights open questions.

2 Structure of the Target

2.1 Product Target

A pioneering paper in the study of Metropolis methods in high dimensions is [10]; it studied the behaviour of random walk Metropolis methods when applied to target distributions with density

$$\pi_0^d(x) = \Pi_{i=1}^d f(x_i). \tag{1}$$

A similar study was undertaken in [22] for Langevin based Metropolis methods. Whilst these were amongst the first papers to pursue a rigorous study of Metropolis methods in high dimensions, a natural objection to this work is that families of target measures of the form (1) are restrictive from an applied perspective and, in any case, can be tackled by sampling a single one-dimensional target, because of the product structure. Partly in response to this objection, there have been several papers which generalize this work to target measures which retain the product structure inherent in (1), but are no longer i.i.d.. To be precise, we introduce standard deviations $\{\lambda_{i,d}\}_{i=1}^d$ so that

$$\pi_0^d(x) = \Pi_{i=1}^d \lambda_{i,d}^{-1} f(\lambda_{i,d}^{-1} x_i). \tag{2}$$

The papers [2, 23] consider this form of measure in the case where $\lambda_{i,d} = \lambda_i$ only, when the standard deviations do not change with dimension. Similar objections maybe raised concerning applicability of this work, namely that the product structure renders the problem far from most applications.

2.2 Beyond the Product Structure

In [5, 3] a different approach was taken, motivated by an infinite dimensional perspective arising in many applications. The target measure π is defined on a function space and is absolutely continuous with respect to some simpler reference measure π_0:

$$\frac{d\pi}{d\pi_0}(x) \propto \exp\bigl(-\Psi(x)\bigr). \tag{3}$$

For example π and π_0 might be the posterior and prior distributions respectively in the Bayesian formulation for an inverse problem on function space [7], or might arise from a (possibly conditioned on observations, end-point constraints etc.) SDE

via the Girsanov formula [12]. Often π_0 has a product structure when written in an appropriate basis.

Perhaps the simplest context in which to see such a product structure is to consider the case where π_0 is a Gaussian distribution $\mathcal{N}(0,\mathscr{C})$ on a Hilbert space $\mathscr{H}$. The eigenvalue problem

$$\mathscr{C}\phi_i = \lambda_i^2 \phi_i \tag{4}$$

provides a basis $\{\phi_i\}_{i=1}^{\infty}$ in which the operator $\mathscr{C}$ is diagonal and hence may be used to create a coordinate system in which there is a product structure. For the Gaussian measure to be well defined, $\mathscr{C}$ must be a trace-class operator which in turn implies that the λ_i's are square summable [9]. Any function $x \in \mathscr{H}$ may be written as

$$x = \sum_{i=1}^{\infty} x_i \phi_i. \tag{5}$$

If $x \sim \mathcal{N}(0,\mathscr{C})$ then the $\{x_i\}$ form a sequence of independent Gaussian random variables on $\mathbb{R}$ with $x_i \sim \mathcal{N}(0,\lambda_i^2)$. Thus we may write

$$x = \sum_{i=1}^{\infty} \xi_i \lambda_i \phi_i \tag{6}$$

where the ξ_i are an i.i.d. sequence of standard unit Gaussians. This shows that any Gaussian measure can be identified with a product measure on $\mathbb{R}^{\infty}$, an important idea which underlies the connections between simple product measures and quite complex measures π given by (3). The representation (6) is known as the *Karhunen-Loéve* expansion.

In the applications cited in [7, 12], the exponent Ψ is shown to satisfy useful properties which can be exploited in the design and analysis of sampling methods. In particular, Ψ can be shown to be bounded from below and above polynomially, and to be Lipschitz, on some Banach space X of full measure under π_0.

Consideration of some finite dimensional approximation of (3) will lead to a target measure π^d of the form

$$\frac{d\pi^d}{d\pi_0^d}(x) \propto \exp\bigl(-\Psi^d(x)\bigr) \tag{7}$$

where π_0^d is given by (2). Such measures are no longer of product form. However, the fact that they arise as approximations of measures on function space which are absolutely continuous with respect to a product measure leads to certain properties of Ψ^d being uniform in d. Furthermore, absolute continuity of π with respect to π_0 means, in rough terms, that if we expand a sample from π and one from π_0 in an orthonormal basis for $\mathscr{H}$, then the expansion coefficients are asymptotically (in the parameter indexing the expansion) identical: indeed absolute continuity sets strict conditions on the rate at which this asymptotic behaviour must occur (the Feldman-Hajek theorem [9]). Intuitively this allows for insight gleaned from the case of product measures

to be transferred to this more applicable context and explains the importance of the initial work in [10, 22, 23] concerning product measures.

These insights enable proof that, in some contexts, Ψ^d is bounded from below and above polynomially and is Lipschitz, with constants uniform in dimension d, provided appropriate norms are chosen to reflect the underlying infinite dimensional norm on X; see [3].

To give a little more detail on the nature of finite dimensional approximations of (3) we continue with the case where the reference measure is symmetric Gaussian and the Karhunen-Loéve expansion (5). If we denote by P^d the orthogonal projection of $\mathscr{H}$ onto the linear span

$$P^d\mathscr{H} := \text{span}\{\phi_1, \ldots, \phi_d\}$$

then we may define the measure π_{KL} on $\mathscr{H}$ by

$$\frac{d\pi_{\text{KL}}}{d\pi_0}(x) \propto \exp\bigl(-\Psi(P^d x)\bigr). \tag{8}$$

This measure is identical to π_0 on $\mathscr{H}\backslash P^d\mathscr{H}$, i.e. on the orthogonal complement of $P^d\mathscr{H}$.

On $P^d\mathscr{H}$, it provides a measure with the structure (7) and with the reference measure π_0^d given by (2) for $\lambda_{i,d} = \lambda_i$ given by (4); see [3] for details. Further approximation may be necessary, or desirable, as it may not be possible to evaluate Ψ, even on $P^d\mathscr{H}$. In the case of SDE (possibly conditioned on observations, end-point constraints etc.), and finite difference approximations (Euler-Maruyama method) one again obtains a measure of the form (7) with the reference measure π_0^d given by (2), but now the $\lambda_{i,d}$ depend on d and satisfy $\lambda_{i,d} \to \lambda_i$ as $d \to \infty$, for each fixed i; see [3] for details.

In summary, the early foundations of the study of the computational complexity of Metropolis methods in high dimension are based in the study of families of product measures (2); see [23] for an overview. More recently, this has given way to the study of wider classes of problems arising in applications with target measure of the form (3); see [5] for an overview. Whilst product measures might seem unduly restrictive, it turns out that a great deal of intuition can be transferred from this situation to the more applied problems, whenever the underlying reference measure in (3) has a product structure, a situation arising frequently in practice. With this in mind we now turn to the study of complexity.

3 Computational Complexity

We study Metropolis methods applied to the target measure π^d given by (7), and based on approximating (3). We assume that there are constants $0 < C^- \le C^+ < \infty$ and $\kappa \ge 0$ such that, for all indices i and dimensions d,

$$C^- \leq i^\kappa \lambda_{i,d} \leq C^+$$

giving bounds on the standard deviations.

Note that this setup subsumes the simpler cases (1) and (2) - by choosing $\Psi \equiv 0$ for both cases, and $\lambda_{i,d} \equiv 1$ for the first. In real applications a wide range of $\kappa > 0$ are encountered. For SDEs, possibly conditioned on observations, we have $\kappa = 1$. For Gaussian random field priors based on covariance operators which are fractional powers of the Laplacian in spatial dimension 2 (not to be confused with the dimension d of the approximating space) we require $\kappa > \frac{1}{2}$ to obtain almost surely continuous fields; more generally, increasing κ will correspond to random fields with increasing regularity, almost surely.

3.1 The Algorithms

The Metropolis methods we will consider are based on proposals on $\mathbb{R}^d$ with kernel $Q^d(x,dy)$ derived from the following expression in which the parameter $\delta > 0$ and the square root is applied to a positive-definite matrix:

$$\frac{y-x}{\delta} = \alpha \mathscr{A} \nabla \log \pi_0^d(x) + \sqrt{\frac{2\mathscr{A}}{\delta}}\,\xi, \quad \xi \sim \mathscr{N}(0,I). \tag{9}$$

In the case $\alpha = 0$ we refer to *random walk methods* and for $\alpha = 1$ to *Langevin methods*. We will take $\mathscr{A} = I$ or $\mathscr{A} = \mathscr{C}_d$ where we define the diagonal matrix $\mathscr{C}_d = \mathrm{diag}\{\lambda_{1,d}^2, \cdots, \lambda_{d,d}^2\}$.

In the case where π_0^d is Gaussian we will also be interested in proposals of the form, for $\theta \in [0,1]$,

$$\frac{y-x}{\delta} = \theta \mathscr{A} \nabla \log \pi_0^d(y) + (1-\theta)\mathscr{A} \nabla \log \pi_0^d(x) + \sqrt{\frac{2\mathscr{A}}{\delta}}\,\xi \tag{10}$$

for $\xi \sim \mathscr{N}(0,I)$. For both classes of proposal we will refer to δ as the *proposal variance*. All these proposals can be viewed as being derived from Euler-Maruyama-like discretizations of stochastic differential equations (SDEs) which are either π^d-invariant or π_0^d-invariant. Note, for instance, that proposals (9) for $\alpha = 1$ and (10) could be conceived as approximations (the first an explicit, the second an *implicit* one, see [16] for background on numerical approximations of SDEs) of the π_0-invariant SDE:

$$\frac{dx}{dt} = \mathscr{A} \nabla \log \pi_0^d(x) + \sqrt{2\mathscr{A}}\,\frac{dW}{dt}$$

driven by d-dimensional Brownian motion W. See [1, 14, 13, 11, 24] for more details on this interpretation. In this setting δ is the time-step in the Euler-Maruyama discretization.

The Metropolis-Hastings MCMC method [19, 15] creates a π^d invariant Markov chain $\{x^n\}$ as follows. Let $a(x,y)$ denote the acceptance probability, that is:

$$a(x,y) = 1 \wedge \frac{\pi^d(y)\,Q^d(y,x)}{\pi^d(x)\,Q^d(x,y)}.$$

Given x^n we make a proposal $y^n \sim Q^d(x^n,\cdot)$. With (independent) probability $a(x^n,y^n)$ we set $x^{n+1} = y^n$; otherwise we set $x^{n+1} = x^n$.

3.2 Complexity

Application of the Metropolis-Hastings accept-reject rule to proposals generated by the kernels Q described above gives rise to a π^d-invariant Markov chain $\{x^n\}_{n=0}^{\infty}$ on $\mathbb{R}^d$; we are interested in the computational complexity of running this chain to explore π^d. Let y^n denote the proposed state at step n, calculated from setting $x = x^n$ and $y = y^n$ in (9) or (10). The cost of each update is usually straightforward to compute, as a function of dimension, and thus the question of computational complexity boils down to understanding the number of steps required to explore π^d. Complexity of Metropolis methods on $\mathbb{R}^d$, for d large, is a difficult subject and the work we are about to overview does not provide the kind of complete analysis that is currently available for MCMC methods applied to some combinatorial problems. We will overview results related to optimizing choices of $\mathscr{A}$, α and θ (and δ, as a function of the dimension d) according to four (inter-twined) criteria, which we now describe.

Assume that $x^0 \sim \pi^d$ and that $y \mid x$ is given by one of the proposals (9), (10) above. The four criteria are:

1. choose proposal parameters to maximize the mean square jump

$$\mathbb{E}\,\|x^{n+1} - x^n\|^2;$$

2. choose proposal parameters to maximize the mean time-step

$$\delta \times \mathbb{E}\,[a(x^n,y^n)];$$

3. choose proposal parameters to maximize the proposal variance subject to the constraint that the average acceptance probability is bounded away from zero, uniformly in dimension:

$$\liminf_{d\to\infty} \mathbb{E}\,[a(x^n,y^n)] > 0;$$

4. choose proposal parameters to maximize the proposal variance for which there exists a π-invariant diffusion limit for $z^d(t) := x^{\lfloor \delta t \rfloor}$, as $d \to \infty$.

In all four cases we use the rule of thumb that the number of steps $M(d)$ required to sample the invariant measure is given by the expression

$$M(d) \propto \delta^{-1}, \tag{11}$$

where the constant of proportionality is independent of dimension d, but the proposal variance δ depends on d. Later in this section we discuss the theory which justifies this decision. In the final section we will discuss the relations among these criteria and ideal criteria for convergence of Markov chains. For now we proceed on the assumption that the four criteria listed are useful in practice.

In [3] it is shown that, for proposals of the form (9), the optimality criteria 1., 2. and 3. all lead to the same conclusion (in an asymptotic sense, as $d \to \infty$) about optimal scaling of the proposal variance, hence to the same expression for $M(d)$. We summarise the specification of $M(d)$ for the different choices of α and $\mathscr{A}$ in Table 1.

Briefly, for $\alpha = 0$ and $\mathscr{A} = I$ we find that $M(d) = d^{2\kappa+1}$; for $\alpha = 0$ and $\mathscr{A} = \mathscr{C}_d$ we remove the κ-dependence, at the expense of inverting a covariance operator, and find that $M(d) = d$. Similar considerations apply for the case when $\alpha = 1$, only now the corresponding values are $d^{2\kappa+1/3}$ and $d^{1/3}$.

Proposal (9), with $\alpha = 0$ and $\mathscr{A} = I$	$M(d) = d^{2\kappa+1}$
Proposal (9), with $\alpha = 0$ and $\mathscr{A} = \mathscr{C}_d$	$M(d) = d$
Proposal (9), with $\alpha = 1$ and $\mathscr{A} = I$	$M(d) = d^{2\kappa+1/3}$
Proposal (9), with $\alpha = 1$ and $\mathscr{A} = \mathscr{C}_d$	$M(d) = d^{1/3}$
Proposal (10), with $\theta = 1/2$ and π_0 Gaussian	$M(d) = \mathscr{O}(1)$

Table 1 Number of steps $M(d)$ to sample the invariant measure for each of the various MCMC algorithms derived via proposals (9) and (10).

In [4] we show that, by choosing $\theta = \frac{1}{2}$ in proposal (10), it is possible to achieve $M(d) = \mathscr{O}(1)$ when the reference measure is Gaussian. In [4] numerical illustration is given only in the case of SDEs conditioned to start and end at specified points (diffusion bridges); however, [8] shows application of the same algorithmic idea to the problem of data assimilation for the Navier-Stokes equation.

3.3 A Special Result: Diffusion Limit

We now turn to the subject of diffusion limits. This will enable us to connect criterion 4. with criteria 1., 2. and 3., providing substantiation for the use of the heuristic (11) to measure the number of steps required to explore the target distribution in stationarity.

First we consider the simplest case where the target measure has the form (1). In [10] it was shown that, using (9) with $\alpha = 0$ and $\mathscr{A} = I$, and choosing the proposal variance δ to scale as $\delta = \ell^2 d^{-1}$, for some constant $\ell > 0$, leads to an average acceptance probability of order 1. Furthermore, with this choice of scaling, individual components of the resulting Markov chain converge to the solution of an SDE. Analytically, if the Markov chain is started in stationarity, and

$$z^d(t) := x_i^{\lfloor d\cdot t\rfloor}$$

denotes a continuous-time interpolant of the i^{th} component of the Markov chain, then $z^d \Rightarrow z$ as $d \to \infty$ in $C([0,T];\mathbb{R})$, where z solves the SDE

$$\frac{dz}{dt} = h(\ell)\,(\log f)'(z) + \sqrt{2\,h(\ell)}\,\frac{dW}{dt}. \tag{12}$$

Here $h(\ell)$ is often termed the *speed measure* and simply sets a time-scale for the SDE; it is identified explicitly in [10].

The diffusion limit leads to the interpretation that, started in stationarity, and applied to target measures of the form (1), the random walk Metropolis algorithm will require an order of δ^{-1} steps to explore the invariant measure; it also provides the justification for (11). Furthermore, the existence of a diffusion limit in this case shows that optimality criteria 1., 2., 3. and 4. all coincide. But the diffusion limit contains further information: it can be shown that the value of ℓ which maximizes $h(\ell)$, and therefore maximizes the speed of convergence of the limiting diffusion, leads to a universal acceptance probability, for random walk Metropolis algorithms applied to targets (1), of approximately 0.234. This means that, for the stated class of target distributions and algorithms, optimality can be obtained simply by tuning the algorithm to attain this desired acceptance probability.

These ideas have been generalized to other proposals, such as those based on (9) with $\alpha = 1$ and $\mathscr{A} = I$ in [22]. In this case, the choice $\delta = \ell^2 d^{-1/3}$ leads to a diffusion limit for

$$z^d(t) := x_i^{\lfloor d^{1/3}t\rfloor}\ ,$$

again implying that optimality criteria 1., 2., 3. and 4. all coincide. This leads to the interpretation that the algorithm will take time of order $d^{1/3}$ to explore the invariant measure. Furthermore, the choice of ℓ which maximizes the speed of the limiting SDE can be identified and results from an acceptance probability of approximately 0.574.

These papers of Roberts and coworkers concerning i.i.d. product measures are extended to non-i.i.d. products in [2, 23]. The impact of this work has been very high, in part because of the simple criteria for optimality when expressed in terms of the average acceptance probabilities 0.234 and 0.574, and in part because the existence of a diffusion limit provides an important conceptual understanding of the behaviour of MCMC methods in high dimensions. It is therefore natural to wonder if these optimal average acceptance probabilities arise also in the nonproduct case and if diffusion limits can then be found. We finish this section by discussing these two issues.

As mentioned above, [5, 3] study the question of optimal scaling of the proposal variance according to criteria 1., 2. and 3., for proposals (9), with $\alpha \in \{0,1\}$ and $\mathscr{A} \in \{I, \mathscr{C}_d\}$, for non-product target measures of the form (7). There, it is shown that the mean square jumping distance (criterion 1.) and the mean time-step (criterion 2.) are both maximized by choosing the acceptance probabilities to be 0.234 (for

$\alpha = 0$) or 0.574 (for $\alpha = 1$) as in the i.i.d. product case (1). It is also shown that such a choice corresponds to optimizing with respect to criterion 3.

For target measures of the form (7), individual components of the Metropolis Markov chain cannot be expected to converge to a scalar SDE as happens for (1). However, it is natural to expect convergence of the entire Markov chain to an infinite dimensional continuous time stochastic process. In [13, 14] it is shown that the target measure π given by (3) is invariant for $\mathscr{H}$-valued SDEs (or stochastic PDEs, labelled SPDEs) with the form

$$\frac{dz}{ds} = -z - \mathscr{C}\nabla\Psi(z) + \sqrt{2\mathscr{C}}\frac{dW}{ds}, \tag{13}$$

where W is cylindrical Brownian motion (see [9] for a definition) in $\mathscr{H}$. In [18], we show that for proposal (9) with $\alpha = 0$ and $\mathscr{A} = \mathscr{C}_d$, started in stationarity, and $z^d(t) := x^{\lfloor dt \rfloor}$, $z^d \Rightarrow z$ as $d \to \infty$ in $C([0,T];\mathscr{H})$. This generalizes the work in [10, 22, 23] to the non-product set-up and shows that, in stationarity, the random walk Metropolis algorithm requires $\mathscr{O}(d)$ steps to explore the target distribution.

4 Open Questions

There are, of course, many open questions in the broad area of analyzing and constructing efficient MCMC methods in infinite dimensions. We mention a few interesting avenues in this general area, reflecting our personal tastes.

- *Rigorous complexity estimates.* Perhaps the primary open question concerning the work described herein is whether it can be used as the basis of the proof of a spectral gap for the Markov chain $\{x^n\}_{n=0}^{\infty}$, and determination of how the spectral gap scales with dimension d. A natural approach to this problem would be to use the theory highlighted in [20]. This theory provides a methodology for establishing convergence results of the following form: there are constants $C > 0, \lambda < 1$ and function $V : \mathbb{R}^d \mapsto [1,\infty)$ such that, for every $x^0 \in \mathbb{R}^d$, and every function g with $|g| \le V$,

$$|\mathbb{E}[g(x^n)] - \pi^d(g)| \le CV(x^0)\lambda^n.$$

  The distance of the constant λ from 1 can be used to estimate the spectral gap of the Markov chain. In typical proofs, the value of λ reflects both the mixing rate of the Markov chain in the center of the state space (in a *small set*) and the rate of return to the center of the state space. Since the results outlined in the previous section are concerned with behaviour in stationarity, it is likely that they reflect behaviour in the center of the state space. Thus, they do not contain information about travel times from outside the center of the state space; indeed this may lead to optimal scalings of the proposal variance which differ from those in the center of the state space, as shown in [6]. This relates to the burn-in time of the

algorithm, whereas the work described in section 3 is primarily concerned with behaviour in stationarity.

- *[Number of steps]/[work per step] trade-off.* We have indicated above that the work in [4] demonstrates that, for measures of the form (7) with π_0^d Gaussian, it is possible to construct algorithms which explore the state space in a number of steps independent of dimension. These algorithms use proposals given by (10) with $\theta = \frac{1}{2}$. However, the algorithms require, at each step, either drawing a sample from the Gaussian reference measure $\mathcal{N}(0, \mathscr{C}_d)$ (in the case $\mathscr{A} = \mathscr{C}_d$), or inversion of the operator $I + \frac{\delta}{2}\mathscr{C}_d^{-1}$ (in the case $\mathscr{A} = I$). In contrast, proposal (9) with $\mathscr{A} = I$ is potentially considerably cheaper per step, but does require $\mathcal{O}(d^{2\kappa+1})$ steps to explore the invariant measure. There is, therefore, a trade-off between cost per step, and number of steps, for proposals based on (9) and (10). For probability measures arising from SDEs (possibly conditioned by observations, end-point constraints etc.) the linear algebra associated with proposals of the form (10) is (asymptotically in d) no more expensive than the cost of an update under (9) with $\mathscr{A} = I$, so it is clear that methods based on (10) with $\theta = \frac{1}{2}$ have a significant advantage; this advantage is illustrated numerically in [4]. However, for other classes of problems the trade-off remains to be studied. This poses an interesting avenue for study.
- *Non-Gaussian reference measures.* At the end of subsection 3.2 we highlighted the fact that certain probability measures can be explored in number of steps *independent of the dimension* d, when started in stationarity. However this relies heavily on the assumption that the reference measure π_0^d in (7) is Gaussian. It remains an open question whether similar ideas to those in [4, 8] can be developed in the case of non-Gaussian reference measures. This is intimately related to the development of π-invariant SPDEs for measures of the form (3) [1, 13].
- *Other proposals.* The proposals we have discussed have been based on the discretization of π- or π_0-reversible SPDEs, leading to the Metropolis and Metropolis-Hastings variants of MCMC methods. However, there are many other proposals known to be effective in practice. In particular, Hybrid Monte Carlo (HMC) methods are widely used by practitioners. These methods double the size of the state space, from d to $2d$, by adding a momentum variable; they then use randomized Hamiltonian mechanics to explore the probability measure. Practical experience indicates that these methods can be very effective and theoretical studies of these proposals, of the type described in this review, would be of interest. More generally, there may be other proposals which yield improved complexity and this area is likely to be fruitful for further development.

Acknowledgements This article describes work overviewed in the second author's plenary lecture at MCQMC08, held in Montreal in July 2008. The authors are grateful to the organizers of this conference for the opportunity to present this material in lectured and written form. The authors are also indebted to all coauthors on papers listed in the bibliography, as well as to Jeff Rosenthal, for helpful discussions relating to the material in this paper. They are grateful to the European Research Council, the Engineering and Physical Sciences Research Council (UK) and the Office of Naval Research (USA) for funding all of their coauthored research referenced in the bibliography.

References

1. Apte, A., Hairer, M., Stuart, A., Voss, J.: Sampling the posterior: an approach to non-Gaussian data assimilation. Physica D **230**, 50–64 (2007)
2. Bédard, M.: Weak convergence of Metropolis algorithms for non-IID target distributions. Ann. Appl. Probab. **17**, 1222–1244 (2007)
3. Beskos, A., Roberts, G., Stuart, A.: Optimal scalings for local Metropolis-Hastings chains on non-product targets in high dimensions. Ann. Appl. Prob. (2009)
4. Beskos, A., Roberts, G., Stuart, A., Voss, J.: MCMC methods for diffusion bridges. Stochastics and Dynamics **8**(3), 319–350 (2008)
5. Beskos, A., Stuart, A.: MCMC methods for sampling function space. In: Proceedings of the International Congress of Industrial and Applied Mathematicians, (Zurich, 2007) (2009)
6. Christensen, O., Roberts, G., Rosenthal, J.: Optimal scaling for the transient phase of local Metropolis-Hastings algorithms. J. Roy. Stat. Soc. B **67**, 253–268 (2005)
7. Cotter, S., Dashti, M., Robinson, J., Stuart, A.: Data assimilation problems arising in fluid mechanics: Bayesian formulation in function space. Inverse Problems **Submitted** (2009)
8. Cotter, S., Dashti, M., Robinson, J., Stuart, A.: Data assimilation problems arising in fluid mechanics: MCMC methods in function space **Submitted** (2009)
9. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions, *Encyclopedia of Mathematics and its Applications*, vol. 44. Cambridge University Press (1992)
10. Gelman, A., Gilks, W., Roberts, G.: Weak convergence and optimal scaling of random walk metropolis algorithms. Ann. Appl. Prob. **7**, 110–120 (1997)
11. Hairer, M., Stuart, A., Voss, J.: Sampling conditioned diffusions. In: Von Weizsäcker Proceedings, p. To appear. Cambridge University Press (2009)
12. Hairer, M., Stuart, A., Voss, J.: Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods. In: Handbook of Nonlinear Filtering. Oxford University Press, Editors D. Crisan and B. Rozovsky (2010)
13. Hairer, M., Stuart, A.M., Voss, J.: Analysis of SPDEs arising in path sampling. II. The nonlinear case. Ann. Appl. Probab. **17**(5-6), 1657–1706 (2007)
14. Hairer, M., Stuart, A.M., Voss, J., Wiberg, P.: Analysis of SPDEs arising in path sampling. I. The Gaussian case. Commun. Math. Sci. **3**(4), 587–603 (2005)
15. Hastings, W.: Monte Carlo sampling methods using markov chains and their applications. Biometrika **57**, 97–109 (1970)
16. Kloeden, P.E., Platen, E.: Numerical solution of stochastic differential equations, *Applications of Mathematics (New York)*, vol. 23. Springer-Verlag, Berlin (1992)
17. Liu, J.: Monte Carlo Strategies in Scientific Computing. Springer Texts in Statistics. Springer-Verlag (2001)
18. Mattingly, J., Pillai, N., Stuart, A.: SPDE limits of the Random Walk Metropolis algorithm in high dimensions (2009)
19. Metropolis, N., Rosenbluth, R., Teller, M., Teller, E.: Equations of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)
20. Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Communications and Control Engineering Series. Springer-Verlag, London (1993)
21. Robert, C., Casella, G.: Monte Carlo Statistical Methods. Springer Texts in Statistics. Springer-Verlag (1999)
22. Roberts, G., Rosenthal, J.: Optimal scaling of discrete approximations to Langevin diffusions. J. Roy. Stat. Soc. B **60**, 255–268 (1998)
23. Roberts, G., Rosenthal, J.: Optimal scaling for various Metropolis-Hastings algorithms. Statistical Science **16**, 351–367 (2001)
24. Stuart, A., Voss, J., Wiberg, P.: Conditional path sampling of SDEs and the Langevin MCMC method. Commun. Math. Sci. **2**(4), 685–697 (2004)