SPDE Limits of the Random Walk Metropolis Algorithm in High Dimensions

Jonathan C. Mattingly, Natesh S. Pillai, Andrew M. Stuart
May 4, 2009

Department of Mathematics
Duke University, Durham
NC, 27708-0251
e-mail: jonm@math.duke.edu

CRiSM, Department of Statistics
Warwick University
CV4 7AL, UK
e-mail: n.pillai@warwick.ac.uk

Mathematics Institute
Warwick University
CV4 7AL, UK
e-mail: a.m.stuart@warwick.ac.uk

Abstract: Diffusion limits of MCMC methods in high dimensions provide a useful theoretical tool for studying efficiency. In particular they facilitate precise estimates of the number of steps required to explore the target measure, in stationarity, as a function of the dimension of the state space. However, to date such results have only been proved for target measures with a product structure, severely limiting their applicability to real applications. The purpose of this paper is to study diffusion limits for a class of naturally occuring high dimensional measures, found from the approximation of measures on a Hilbert space which are absolutely continuous with respect to a Gaussian reference measure. The diffusion limit to an infinite dimensional Hilbert space valued SDE (or SPDE) is proved.

1. Introduction

Metropolis-Hastings methods (N. Metropolis & Teller, 1953; Hastings, 1970) form a widely used class of MCMC methods (Liu, 2008; Robert & Casella, 2004) 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. Quantifying computational complexity of an MCMC method is most naturally undertaken by studying the behaviour of the method on a family of probability distributions indexed by a parameter, and studying the cost of the algorithm as a function of that parameter. In this paper we will study the cost as a function of dimension for algorithms applied to a family of probability distributions found from finite dimensional approximation of a measure on an infinite dimensional space. Our interest is focused on Metropolis-Hastings MCMC methods (Robert & Casella, 2004).

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

We study the simplest of these, the random walk Metropolis algorithm (RWM). Let π be a target distribution on $\mathbb{R}^N$. To sample from π, the RWM algorithm creates a π- reversible markov chain which moves from a current state x^0 to a new state x^1 via proposing a candidate y, using a symmetric Markov tranistion kernel such as a random walk, and accepting y with probability $\alpha(x^0, y)$, where $\alpha(x,y) = 1 \wedge \frac{\pi(y)}{\pi(x)}$. Although the proposal is somewhat naive, within the class of all Metropolis-Hastings algorithms, the RWM is still used in many applications because of its simplicity. The only computational cost involved in calculating the acceptance probabilities are the relative ratio of densities $\frac{\pi(y)}{\pi(x)}$, as compared to, say, the Langevin algorithm where one needs to evaluate the gradient of $\log \pi$.

A pioneering paper in the analysis of complexity for MCMC methods in high dimensions is Roberts *et al.* (1997). This paper studied the behaviour of random walk Metropolis methods when applied to target distributions with density

$$\pi^N(x) = \Pi_{i=1}^{N} f_i(x_i). \tag{1.1}$$

where $f_i(x)$ are one dimensional probability density functions. The authors considered a proposal of the form:

$$y = x + \sqrt{\delta}\rho$$
$$\rho \sim \mathrm{No}(0, \mathrm{I}_N)$$

The objective was to study the complexity of the algorithm as a function of the dimension N of the state space. It was shown that choosing the proposal variance δ to scale as $\delta = \ell^2 N^{-1}$ (we discuss the choice of $\ell > 0$ later) 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 a stochastic differential equation (SDE). If x^k denotes the k^{th} iterate of the Markov chain, started in stationarity, and $z^N(t) := x_i^{\lfloor Nt \rfloor}$ denotes a continuous time-interpolant of the i^{th} component of the Markov chain, then $z^N \Rightarrow z$ as $N \to \infty$ in $C([0,T]; \mathbb{R})$ where z solves the SDE

$$\frac{dz}{dt} = -\frac{1}{2} h(\ell) f'(z) + \sqrt{h(\ell)} \frac{dW}{dt}. \tag{1.2}$$

Note that the invariant measure of the sde (1.2) has the density f with respect to the lebesgue measure. This weak convergence result leads to the interpretation that, started in stationarity and applied to target measures of the form (1.1), the RWM algorithm will take on the order of N steps to explore the invariant measure. Furthermore it may 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 Metrpolis algorithms applied to targets (1.1), of approximately 0.234.

These ideas have been generalized to other proposals, such as those based on discretization of the Langevin equation as described in Roberts & Rosenthal (1998). For Langevin proposals the scaling of δ which achieves order 1 acceptance probabilities is $\delta = \ell^2 N^{-\frac{1}{3}}$ and the choice of ℓ which maximizes the speed of the limiting SDE results from an acceptance probability of approximately 0.574. The intuition behind obtaining $\mathcal{O}(1)$ acceptance probabilities is related to the 'optimality' of the algorithm and is explained a bit later in this section.

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

The work by Roberts and co-workers was amongst the first to develop a mathematical theory of Metropolis-Hastings methods in high dimension, and does so in a fashion which leads to clear criteria which practitioners can use to optimize algorithmic performance, for instance by tuning the acceptance probabilities to 0.234 or 0.574. Yet it is open to the criticism that, from a practioner's perspective, target measures of the form (1.1) are too limited a class of probability distributions to be useful and, in any case, can be tackled by sampling a single one-dimensional target because of the product structure. There have been papers which generalize this work to target measures which retain the product structure inherent in (1.1), but are no longer i.i.d (see Bédard (2007); Roberts & Rosenthal (2001)):

$$\pi_0^N(x) = \Pi_{i=1}^{N} \lambda_i^{-1} f(\lambda_i^{-1} x_i). \tag{1.3}$$

However, the same criticism may be applied to this scenario as well.

However, despite the apparent simplicity of target measures of the form (1.1), the intuition obtained from the study of Metropolis-Hastings methods applied to these 'toy models' is in fact extremely valuable. Inspired by these results, in this paper we seek diffusion limits for the RWM algorithm when applied to more complicated target measures π, which arise in practical applications. To this end, we adopt the framework used in Beskos *et al.* (2008), where the authors consider a target distribution π which lives in an infinite dimensional, real separable Hilbert space $\mathcal{H}$. Furthermore, π is absolutely continuous with respect to a Gaussian measure π_0 on $\mathcal{H}$ which has mean zero and covariance operator C (see section 2 for details). The Radon-Nikodym derivative $\frac{d\pi}{d\pi_0}$ has the form:

$$\frac{d\pi}{d\pi_0} = M_\Psi \exp(-\Psi(x)) \tag{1.4}$$

for a real valued functional $\Psi\colon\ \mathcal{H} \mapsto \mathcal{H}$ which is densely defined and M_Ψ is a normalizing constant. In applications of interest Ψ is strongly nonlinear. Thus care is required in choosing C to make sure that π is a legitimate probability measure on $\mathcal{H}$. We discuss this issue below. This infinite dimensional framework for the target measures, besides being able to capture a huge number of useful models arising in practice, has a number of advantages. We highlight two of these, particularly relevant to the context of this paper.

Firstly, the theory of Gaussian measures naturally generalizes from $\mathbb{R}^N$ to infinite dimensional Hilbert spaces. The covariance operator $C : \mathcal{H} \mapsto \mathcal{H}$ is a selfadjoint, positive, and trace class operator on $\mathcal{H}$ with a complete orthonormal eigenbasis $\{\lambda_k^2, \phi_k\}$:

$$C\phi_k = \lambda_k^2 \phi_k.$$

By the Karhunen-Loeve (Da Prato & Zabczyk (1992)) expansion, a realization from the Gaussian measure $x \sim \pi_0$ can be expressed as:

$$x \sim \sum_{k=1}^{\infty} \lambda_k \, \rho_k \, \phi_k, \;\; \rho_k \overset{\text{iid}}{\sim} \text{No}(0,1)$$

Thus the 'reference measure' π_0 has a product structure as in (1.1). More importantly absolute continuity of π with respect to π_0 means that a typical draw from the target measure π

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

must behave like a typical draw from π_0 in the large $k-$coordinates dictated by an exapnsion in the Karhunen-Loeve basis. This offers hope that the ideas the from the product case are applicable here as well; however the nonlinear nature of Ψ poses new genuine challenges in carrying over the techniques used by the previous authors. The fact that individual components of the markov chain converge to a scalar SDE, as proved in Roberts *et al.* (1997), is a direct consequence of the product structure inherent in (1.1). For target measures of the form (1.6) this structure is not present, and individual components of the Markov chain cannot be expected to converge to a scalar SDE. However it is natural to expect convergence of the entire Markov chain to an infinite dimensional continuous time stochastic process and the purpose of this paper is to carry out such a program.

Secondly, as proved in a series of recent papers (Hairer *et al.* (2005, 2007)), the target meausre π is invariant for $\mathcal{H}-$valued SDEs (or stochastic PDES – SPDEs) with the form

$$\frac{dz}{ds} = -\alpha\Big(z + C\nabla\Psi(z)\Big) + \sqrt{2\alpha\, C}\,\frac{dW}{ds}, \quad z(0) = z_0 \tag{1.5}$$

where W is a cylindrical Brownian motion (Da Prato & Zabczyk, 1992) in $\mathcal{H}$ and $\alpha > 0$. Thus the above result from SPDE theory gives us a natural candidate for the infinite dimensional limit of an MCMC method. Moreover the invariant measure is preserved for any positive value of α and we will see that the choice of $\alpha = 0.234$ leads to the optimal acceptance probability.

In many applications the measures π and π_0 might be thought of as posterior and prior distributions in the Baysian formulation for an inverse problem on $\mathcal{H}$. One example to keep in mind is the following. Let $\mathcal{H} \equiv L^2[0,1]$. Consider the heat equation for $v(x,t) : [0,1]\times[0,T] \to \mathbb{R}$:

$$\begin{aligned} \partial_t v &= \partial^2_{xx} v, \quad (x,t) \in (0,1)\times(0,T] \\ v &= 0, \quad (x,t) \in \{0,1\}\times(0,T] \\ v &= u, \quad (x,t) \in [0,1]\times\{0\}. \end{aligned}$$

We assume that we observe v noisily at discrete points in space time and so that we are given

$$y_{j,k} = v(x_j, t_k) + \eta_{j,k}$$

where the $\eta_{j,k}$ are centred Gaussian random variables. We also assume that all the t_k are positive. The goal is to estimate the initial condition u from y. In a Bayesian formulation, a natural prior distribution for u is a Gaussian prior π_0. Let $v(t,\cdot) \equiv \Omega(u,t,\cdot)$ denote the solution map of the heat equation. Then $\frac{d\pi}{d\pi_0}$ is given by the likelihood:

$$\frac{d\pi}{d\pi_0}(u) \propto \exp\Big(-\frac{1}{2}\sum_{j,k}\Big(v(x_j,t_k) - \Omega(u,x_j,t_k)\Big)^2\Big),$$

thus defining Ψ as a quadratic form, since$\Omega(\,\cdot,t,x)$ is linear. In this particular case the posterior measure is itself Gaussian and can be computed exactly, but nonlinear generalizations of this problem are not Gaussian and sampling methods are required to probe the posterior.

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

To sample from π numerically, one needs a finite dimensional approximation and this leads to a target measure π^N with the form

$$\frac{d\pi^N}{d\pi_0^N}(x) \propto \exp\Big(-\Psi^N(x)\Big) \tag{1.6}$$

whereΨ^N is an approximation of Ψ in an appropriate sense. For example, one might use a finite dimensional approximation based on the Karhunen-Loeve expansion. In this finite dimensional approximation, the proposal distribution for the RWM can be expressed as:

$$y = x + \sqrt{\frac{2\ell^2}{N^\beta}} C^{1/2}\xi$$

$$\xi = \sum_{i=1}^{N} \rho_k \phi_k, \; \rho_k \overset{\text{iid}}{\sim} \text{No}(0,1)$$

and $\beta > 0$. Thus the proposal variance $N^{-\beta}$ is large for smaller values of β. Identifying the optimal choice for β is a delicate excercise. Larger values of β corresponds to 'local' moves, and therefore for the algorithm to explore the state space rapidly, β needs to be as small as possible. However, there is another side to this coin: if we set β to be arbitrarily small, then the acceptance probability decreases to *zero* very rapidly as a function of N. In fact it was shown in Beskos & Stuart (2007); Beskos *et al.* (2008) that, for a variety of Metropolis-Hastings proposals, there is $\beta_c > 0$ such that choice of $\beta < \beta_c$ leads to average acceptance probabilities which are smaller than any inverse power of N. Thus in higher dimensions, smaller values of β lead to very poor mixing because of the negligible acceptance probability. However, it turns out that at the critical value β_c, the acceptance probability is $\mathcal{O}(1)$ as a function of N. The value of β_c was identified to be 1 and 1/3 for the RWM and Langevin proposals respectively. Finally, when using the scalings leading to $\mathcal{O}(1)$ acceptance probabilities, it was shown that the mean square distance moved is maximized by choosing the acceptance probabilities to be 0.234 or 0.574 as in the i.i.d product case (1.1). Guided by this intuition, we will set $\beta = \beta_c = 1$ for our proposal variance which leads to $\mathcal{O}(1)$ acceptance probabilities.

Summarizing the discussion so far, our goal is to obtain an invariance principle for the stationary RWM Markov chain applied to target measures of the form (1.4) with limiting diffusion given by the spde (1.5). This will show that, in stationarity and properly scaled to achieve $\mathcal{O}(1)$ acceptance probabilities, the random walk Metropolis algorithm takes $\mathcal{O}(N)$ steps to explore the target distribution. From a practical point of view the take home message of this work is that standard RWM algorithms applied to approximations of target measures with the form (1.4), can be tuned to behave optimally by adjusting the acceptance probability to be approximately 0.234. This will lead to $\mathcal{O}(N)$ steps to explore the target measure in stationarity. Although we only analyse the RWM algorithm, we believe that our techniques can be applied to a larger class of Metropolis-Hastings methods, including the Langevin algorithm.

We analyse the RWM algorithm started at stationarity, and thus do not attempt to answer the question of 'burn-in time': the number of steps required to reach stationarity and how

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

the proposal scaling affects the rate of convergence. These are important questions which we hope to answer in a future paper. Furthermore practitioners wishing to sample from probability measures on function space with the form (1.4) should be aware that for some examples, new generalizations of random walk Metropolis algorithms, defined on function space, can be more efficient than the standard random walk methods analyzed in this paper (Beskos *et al.* , 2008; Beskos & Stuart, 2007).

There exist several methods in the literature to prove invariance principles. For instance, because of the reversibility of the RWM Markov chain, utilizing the abstract but powerful theory of Dirichlet forms (Ma & Röckner (1992)) is appealing. Another promising alternative is to show the convergence of generators of the associated Markov processes (Ethier & Kurtz (1986)). However, we chose a more 'hands on' approach using simple probabilistic tools, thus gaining more intuition about the RWM algorithm in higher dimensions. We show that with the correct choice of scaling, the one step transition for the RWM markov chain behaves nearly like an Euler scheme applied to (1.5). This fact coupled with a martingale central limit theorem (Berger (1986)) leads to a direct proof of our main result, using preservation of weak convergence under a continuous mapping. Our arguments are very much in spirit to that of Walk (1977), where the author obtains an invariance principle for the Robbins-Monroe procedure.

In section 2 we set-up the notation that we use throughout the remainder of the paper. In section 3 we investigate the mathematical structure of the RWM algorithm when applied to target measures of the form (1.6). Before presenting details, a heurisitic but detailed outline of the proof strategy is given for communicating the main ideas. In section 4 we state our assumptions on Ψ and the covariance operator C, study the finite dimensional approximations of Ψ and derive some apriori bounds. Due to the technical nature of the drift and diffusion coefficient calculations, as done earlier we first present a heuristic proof in section 5, with the emphasis on the key aspects of the proof strategy. The rate of convergence to central limit theorem is derived using the Stein's method and we discuss this framework briefly in section 6. In section 7, the heuristic arguments given in the previous sections are made rigorous by peforming careful estimates to establish the mean drift and diffusion using the tools developed in section 6. In section 8 the main task of obtaining the invariance principle for the Markov chain is accompolished. Proofs of various technical results are contained in an Appendix.

2. Preliminaries

Let $\mathcal{H}$ be a separable Hilbert space of real valued funtions and C be a trace class postive operator on $\mathcal{H}$. Let $\{\phi_i, \lambda_i^2\}$ be the eigenvectors and eigenvalues of C respectively, so that

$$C\phi_i = \lambda_i^2 \, \phi_i, \; i \in \mathbb{N}.$$

We assume a normalization under which $\{\phi_i\}$ forms a complete orthonormal basis in $\mathcal{H}$. For every $h \in \mathcal{H}$, we write

$$x = \sum_{i=1}^{\infty} x_i \phi_i, \; x_i \equiv \langle x, \phi_i \rangle. \tag{2.1}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Using the above expansion, we define Sobolev spaces $\mathcal{H}^s \subseteq \mathcal{H}$, $s \geq 0$, with the norms defined by

$$\|x\|_s^2 \equiv \sum_{i=1}^{\infty} i^{2s} x_i^2. \tag{2.2}$$

Notice that $\mathcal{H}^0 = \mathcal{H}$ and we denote the $\mathcal{H}$ inner-product and norm by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. Let $\otimes$ denote the outer product operator in $\mathcal{H}$, *i.e.*,

$$(x \otimes y)z = \langle y, z \rangle x, \ \forall x, y, z \in \mathcal{H}. \tag{2.3}$$

For an operator $L : \mathcal{H} \mapsto \mathcal{H}$, we denote the operator norm on $\mathcal{H}$ by $\| \cdot \|_{\mathcal{L}(\mathcal{H},\mathcal{H})}$ defined by

$$\|L\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} = \sup_{\|x\|=1} \|Lx\|.$$

For self-adjoint L this is, of course, the spectral radius of L. For a positive operator $B : \mathcal{H} \mapsto \mathcal{H}$, define its trace:

$$\mathrm{tr}(B) = \sum_{k=1}^{\infty} \langle \phi_k, B\phi_k \rangle \tag{2.4}$$

The sum in (2.4) is invariant under the choice of the orthonormal basis $\{\phi_k\}$ (see Da Prato & Zabczyk (1992)). The operator B is said to be trace class if $\mathrm{tr}(B) < \infty$.

Let π_0 be denote a mean zero Gaussian measure on $\mathcal{H}$ with covariance operator C, *i.e.,* $\pi_0 \equiv \mathrm{No}(0, C)$. If $x \sim \pi_0$, then by the Karuhnen-Loeve expansion,

$$x = \sum_{i=1}^{\infty} \lambda_i \, \rho_i \, \phi_i, \ \ \rho_i \overset{\text{iid}}{\sim} \mathrm{No}(0, 1). \tag{2.5}$$

Our goal is to sample from a measure π on $\mathcal{H}$, given by (1.4) with π_0 as constructed above. In order to sample from π we first approximate by a finite dimensional measure. For $N \in \mathbb{N}$, let $P^N : \mathcal{H} \mapsto X^N \subset \mathcal{H}$ be the projection operator in $\mathcal{H}$ onto $\mathrm{span}\{\phi_1, \phi_2, \cdots, \phi_N\}$, *i.e.*,

$$P^N x \equiv \sum_{i=1}^{N} x_i \phi_i, \ \ x_i = \langle x, \phi_i \rangle.$$

Notice that X^N is isomorphic to $\mathbb{R}^N$ via (2.1). Next, we approximate Ψ: $\mathcal{H} \mapsto \mathbb{R}$ by $\Psi^N : X^N \mapsto \mathbb{R}$ and attempt to sample from the following approximation to π; namely

$$\frac{d\pi^N}{d\pi_0}(x) \equiv M_{\Psi^N} \exp(-\Psi^N(P^N x))$$

and as done earlier, the constant M_{Ψ^N} is chosen so that $\pi^N(\mathcal{H}) = 1$. Notice that on X^N, π^N has Lebesgue density [1]

$$\pi^N(x) = M_{\Psi^N} \exp\Big(- \Psi^N\big(P^N x\big) - \frac{1}{2}\Big\langle P^N x, C^{-1}(P^N x)\Big\rangle\Big) \tag{2.6}$$

On $\mathcal{H} \backslash X^N$ we have that $\pi^N = \pi_0$. Later we will impose natural assumptions on Ψ , Ψ^N, which are motivated by applications.

[1]For ease of notation we do not distinguish measure and its density

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

3. Random Walk Metropolis Algorithm

Recall that our goal is to sample from (2.6) with $x \in X^N$. As explained in the introduction, we set the proposal variance $\delta = \frac{\ell^2}{N}$ and use a RWM proposal:

$$y = x + \sqrt{\frac{2\ell^2}{N} C}\, \xi, \ell \quad \in \mathbb{R}_+ \tag{3.1}$$
$$\xi = \sum_{i=1}^{N} \xi_i\, \phi_i, \;\; \xi_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$$

Hence even though the Markov chain evolves in $\mathcal{H}$, x and y in (3.1) differ only in the first N coordinates. Notice that the noise ξ is also independent of x. The acceptance probability is

$$\alpha(x, y) = 1 \wedge \exp(Q(x,\xi\,)) \tag{3.2}$$
$$Q(x,\xi\,) \equiv \left\{ \frac{1}{2} \left\| C^{-1/2}(P^N x) \right\|^2 - \frac{1}{2} \left\| C^{-1/2}(P^N y) \right\|^2 + \Psi^N(P^N x) - \Psi^N(P^N y) \right\} \tag{3.3}$$

The successive accepted draws x^k, $k \in \mathbb{N}$ are thus given by

$$\begin{aligned} x^{k+1} &= \gamma^k y^k + (1 - \gamma^k)\, x^k \\ y^k &= x^k + \sqrt{\frac{2\ell^2}{N}} C^{1/2}\, \xi^k \\ \xi^k &= \sum_{i=1}^{N} \xi_i^k \phi_i, \;\; \xi_i^k \overset{\text{iid}}{\sim} \mathrm{No}(0,1) \\ \gamma^k \equiv \gamma(x^k, \xi^k) &\sim \mathrm{Bern}\Big(\alpha(x^k, \xi^k) \Big) \end{aligned} \tag{3.4}$$

where $\alpha(x^k, \xi^k)$ is the acceptance probability defined in equation (3.2).

Recall that the target measure π in (1.4) is the invariant measure of the SPDE (1.5). Our goal is to obtain an invariance principle for the Markov chain $\{x^k\}$ started in stationarity, *i.e.*, to show weak convergence of the continuous interpolant of the markov chain x^k, with step size $\delta = \ell^2/N$ to the SPDE (1.5), as the dimension of the noise $N \to \infty$.

In the rest of the section, we will give a heuristic outline of our main argument. The emphasis will be on the proof strategy and main ideas. So, we will not pay careful attention to the error bounds, and use the symbol "$\approx$" to indicate so. Once the main skeleton is outlined, we retrace our arguments and make them rigorous.

3.1. Convergence to SPDE: Outline of the proof strategy

Let $\mathcal{F}_k$ denote the sigma algebra generated by $\{\gamma^n, \xi^n, n \le k\}$. For notational convenience, we denote the conditional expectations $\mathbb{E}(\cdot|\mathcal{F}_k)$ by $\mathbb{E}_k(\cdot)$. We first compute the one-step expected drift of the Markov chain $\{x^k\}$. Let $x_i^k, i \le N$ denote the i^{th} coordinate of x^k. Also for notational convenience let $x^0 = x$ and $\xi^0 = \xi$. Define the constant β:

$$\beta = 2\Phi(-\frac{\ell}{\sqrt{2}}) \tag{3.5}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

where Φ denotes the CDF of the standard normal distribution. Then, under the assumptions onΨ , Ψ^N given in section 4, we prove the following Theorem in section 7.

THEOREM 7.1. *Let $\{x^k\}$ be the RWM Markov chain with $x^0 = x \sim \pi^N$. Then*

$$N\,\mathbb{E}_0(x_i^1 - x_i) = -\ell^2\beta\Big(P^N x + C\nabla\Psi^N(P^N x)\Big)_i + r^N(i)$$
$$\lim_{N\to\infty} \mathbb{E}^\pi \|r^N\|^2 = 0.$$

Thus the discrete time Markov chain $\{x^k\}$ obtained by the successive accepted samples of the RWM algorithm has approximately the expected drift as that of the SPDE (1.5). Similarly, in section 7 (Theorem 7.9) we show that :

THEOREM 7.9. *Let $\{x^k\}$ be the RWM Markov chain with $x^0 = x \sim \pi^N$. Then*

$$N\,\mathbb{E}_0\Big[(x^1 - x)\otimes(x^1 - x)\Big] = 2\ell^2\beta\, C^N + E^N$$
$$\lim_{N\to\infty} \mathbb{E}^{\pi^N} \|E^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} = 0$$

where C^N is the covariance operator C restricted to X^N.

Thus the quantity $\ell^2\beta$ can be considered as the expected acceptance probability for very large N. Optimising $\ell^2\beta = 2\ell^2\Phi(-\frac{\ell}{\sqrt{2}})$ over the parameter ℓ leads to the *optimal* acceptance probability 0.234 which was obtained in the case of iid targets by Roberts *et al.* (1997). It is also crucial to our subsequent arguments that the error terms r^N and E^N converge to 0 in the Hilbert space norm and the operator norm respectively.

Once we have that the expected drift and diffusion terms of the discrete time Markov chain converge to the corresponding terms of the SPDE (1.5), all that is to be done now is to establish an invariance principle on $C([0,T],\mathcal{H})$ for the process $\{x^k\}$; if $x^0 \sim \pi^N$ *i.e.*, the initial condition is drawn from the stationary distribution, then the Markov chain $\{x^k\}$ converges weakly to the SPDE given in equation (1.5) with $O(1/N)$ time scale. However, as described below, our setup is slightly different from the traditional setup of discrete time Markov chains converging to diffusion processes.

Define

$$m(\cdot) \equiv \Big(P^N\cdot + C\nabla\Psi(P^N\cdot)\Big) \tag{3.6}$$

$$\Gamma^{k,N} \equiv \sqrt{\frac{N}{2\ell^2\beta}}\Big(x^{k+1} - x^k - \mathbb{E}_k(x^{k+1} - x^k)\Big) \tag{3.7}$$

Thus,

$$x^{k+1} = x^k + \mathbb{E}_k(x^{k+1} - x^k) + \sqrt{\frac{2\ell^2\beta}{N}}\,\Gamma^{k,N} \tag{3.8}$$

From Theorem 7.1, for large enough N

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

$$x^{k+1} \approx x^k - \frac{\ell^2\beta}{N}\Big(P^N x^k + C\nabla\Psi(P^N x^k)\Big) + \sqrt{\frac{2\ell^2\beta}{N}}\,\Gamma^{k,N}$$
$$= x^k - \frac{\ell^2\beta}{N} m(x^k) + \sqrt{\frac{2\ell^2\beta}{N}}\,\Gamma^{k,N} \qquad (3.9)$$

From the definition of$\Gamma^{k,N}$ in (3.7), and from Theorem 7.9,

$$\mathbb{E}_k(\Gamma^{k,N}) = 0$$
$$\mathbb{E}_k(\Gamma^{k,N} \otimes \Gamma^{k,N}) \approx C^N$$

Therefore large enough N, equation (3.9) 'resembles' the Euler scheme for simulating the finite dimensional approximation of the SPDE (1.5) on $\mathbb{R}^N$, with drift function $m(\cdot)$ and covariance operator C^N:

$$x^{k+1} \approx x^k - \ell^2\beta\, m(x^k)\Delta t + \sqrt{2\ell^2\beta\Delta t}\,\Gamma^{k,N}, \ \Delta t \equiv \frac{T}{N}, T > 0.$$

Notice that the appearence of $\ell^2\beta$ above does not change the invariant measure and $\ell^2\beta$ is related proportional to the speed measure of the limiting diffusion. Also as mentioned earlier $\ell^2\beta$ is the limiting value of the expected acceptance probability of RWM algorithm. The central idea in Roberts *et al.* (1997) was to choose ℓ to maximize $\ell^2\beta$ (the maximum value being 0.234).

Note that there is an important difference in analysing the weak convergence from the traditional Euler scheme. In our case for any fixed $N \in \mathbb{N}$, $\Gamma^{k,N} \in X^N$ is finite dimensional, but clearly the dimension of$\Gamma^{k,N}$ grows with N. Also the distribution of the initial condition $x(0) \sim \pi^N$ changes with N, unlike the case of the traditional Euler scheme where the distribution of $x(0)$ does not change with N. Moreover, for any fixed N, the "noise" process $\{\Gamma^{k,N}\}$ are not independent random variables. However they are identically distributed (a stationary sequence) because the Metropolis algorithm preserves stationarity. To obtain an invariance principle, we first use a version of the martingale central limit theorem (Theorem 8.1) to show that the noise process $\{\Gamma^{k,N}\}$ when rescaled and summed converges weakly to a Brownian motion on $C([0,T],\mathcal{H})$ with covariance operator C, with $T = \mathcal{O}(1)$.

Before we proceed, we introduce some notation. Define the constants, for any $T > 0$,

$$\Delta t \equiv \frac{T}{N},$$
$$t^k \equiv k\Delta t, \qquad (3.10)$$
$$\eta^{k,N} \equiv \sqrt{\Delta t}\sum_{l=0}^{k-1}\Gamma^{l,N}$$

Define

$$W^N(t) \equiv \eta^{\lfloor Nt\rfloor,N} + \frac{Nt - \lfloor Nt\rfloor}{\sqrt{N}}\,\Gamma^{\lfloor Nt\rfloor+1,N}, \ t \in [0,T] \qquad (3.11)$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Let $W(t), t \in [0,T]$ be a $\mathcal{H}$ valued brownian motion with covariance operator C. Using the martingale central limit theorem, we show in section 8 (Lemma 8.4) that

Lemma 8.4.

$$W^N(t) \Rightarrow W(t), \ \text{in } C[0,T,\mathcal{H}].$$

Once we have the invariance principle for the noise process, because the noise process is additive (the diffusion coefficient is constant), the invariance principle for the markov chain follows from a standard continuous mapping argument. Let us define:

$$\bar{z}^N(t) = x^k, \ t \in [t^k, t^{k+1})$$

We can use $\bar{z}^N$ to construct a continuous piecewise linear interpolant of x^k by defining

$$z^N(t) = z_0 + \ell^2\beta \int_0^t m(\bar{z}^N(s))\, ds + \sqrt{2\ell^2\beta}\, W^N(t) \tag{3.12}$$

The SPDE (1.5) (with $\alpha = \ell^2\beta$) may be written as the integral equation:

$$z(t) = z_0 - \ell^2\beta \int_0^t m(Z(s))\ ds + \sqrt{2\ell^2\beta}\, W(t). \tag{3.13}$$

Using the invariance principle for the noise process and an argument based on the continuous mapping theorem, in section 8 (Thereom 8.7), we prove our main result :

Theorem 8.7. *The stochastic process $z^N(t)$ from* (3.12)*, which is a piecewise linear, continuous interpolant of the RWM algorithm under the assumptions given in section 4, converges weakly in $C([0,T],\mathcal{H})$ to the diffusion process $z(t)$ given by equation* (1.5)*.*

4. Assumptions onΨ, C

In this section we state our assumptions on the covariance operator C and the functionalΨ. To avoid technicalities we assume that$\Psi(\ x)$ is quadratically bounded, with first derivative linearly bounded and second derivative globally bounded. Weaker assumptions could be dealt with stopping time arguments.

Assumptions 4.1. *The operator C and functional Ψ satisfy the following:*

1. ***Decay of Eigenvalues of C:*** *There exists $M_-, M_+ > 0$ and $k > \frac{1}{2}$ such that*

$$M_- \le i^k \lambda_i \le M_+, \ \forall i \in \mathbb{Z}_+ \tag{4.1}$$

2. ***Assumptions on Ψ:*** *There exist constants $M_i \in \mathbb{R}, i \le 4$ and $s \in [0, k-1/2)$ such that*

$$M_1 \le \Psi(x) \le M_2\Big(1 + \|x\|_s^2\Big) \qquad \forall x \in \mathcal{H}^s \tag{4.2}$$

$$\|\nabla\Psi(x)\| \le M_3\Big(1 + \|x\|_s\Big) \qquad \forall x \in \mathcal{H}^s \tag{4.3}$$

$$\|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} \le M_4 \qquad \forall x \in \mathcal{H}^s. \tag{4.4}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Remark 4.2. *In Equation* (4.1)*, the condition* $k > \frac{1}{2}$ *ensures that* C *is a trace class operator. Also, the* $\mathcal{H}^s$ *norm of* $x \sim No(0, C)$ *is almost surely finite for* $s < k - \frac{1}{2}$ *(Da Prato & Zabczyk, 1992). Notice also that the above assumptions on* Ψ *imply that for all* $x, y \in \mathcal{H}^s$*,*

$$|\Psi(x) - \Psi(y)| \leq M_5\Big(1 + \|x\|_s + \|y\|_s\Big)\|x - y\|_s \tag{4.5a}$$

$$\Psi(y) = \Psi(x) + \langle \nabla\Psi(x), y - x\rangle + \mathrm{rem}(x, y) \tag{4.5b}$$

$$\mathrm{rem}(x, y) \leq M_6\|x - y\|^2 \tag{4.5c}$$

for some constants $M_5, M_6 \in \mathbb{R}_+$*.*

It is natural to consider approximations to Ψ , denoted by Ψ^N, which converge to Ψ in such a fashion as to ensure that the Ψ^N also satisfy Assumptions 4.1. This is the content of the following:

Assumptions 4.3. ***Assumptions on*** Ψ^N*:*

1. Ψ^N *satisfies the same conditions imposed on* Ψ *given by equations* (4.2),(4.3) *&* (4.4) *with the same constants uniformly in* N*.*
2. *There exists a constant* M_7 *and real numbers* $\theta(N)$ *such that*

$$\begin{aligned} |\Psi^N(x) - \Psi(x)| &\leq M_7\,\theta(N)\,\|x\|_s^2,\ x \in \mathcal{H}_s \\ \|\nabla\Psi^N(x) - \nabla\Psi(x)\| &\leq M_8\,\theta(N)\,\|x\|_s^2,\ x \in \mathcal{H}_s \\ \lim_{N\to\infty} \theta(N) &= 0. \end{aligned} \tag{4.6}$$

Thus Ψ^N also satisfies (4.5) with constants independent of N. As is shown in Lemma 4.5, the above assumptions on Ψ^N imply that the sequence $\{\pi^N\}$ converges to π in the KL topology and therefore are good candidates for *finite dimensional* approximations of π.

We now show that the normalizing constants M_{Ψ^N} are uniformly bounded and use this to obtain uniform bounds on moments of functionals in $\mathcal{H}$ under π^N:

Lemma 4.4. *Under the assumptions on* Ψ^N*,*

1. $$\sup_{N\in\mathbb{N}} M_{\Psi^N} < \infty$$

2. *For any measurable functional* $f : \mathcal{H} \mapsto \mathbb{R}$*, and any* $p \geq 1$*,*

$$\sup_{N\in\mathbb{N}} \mathbb{E}^{\pi^N}|f(x)|^p \leq M\mathbb{E}^{\pi_0}|f(x)| \tag{4.7}$$

The estimate given in (4.7) will be used repeatedly in the sequel. The following lemma shows that π^N converges to π in the Kullback-Leibler topology .

Lemma 4.5. *Under the above assumptions on* Ψ *and* Ψ^N*,*

$$\lim_{N\to\infty}\Big(D_{KL}(\pi||\pi^N) + D_{KL}(\pi^N||\pi)\Big) = 0$$

where $D_{KL}(\pi^N||\pi)$ *denotes the KL divergence between* π^N *and* π*.*

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Since the total variation distance is bounded above by the KL distance, the above lemma implies that $\lim_{N\to 0}\|\pi-\pi^N\|_{TV}=0$. Thus sampling the approximate measure π^N will lead, for large N, to approximate samples from π.

5. Expected onestep drift and diffusion: heuristic argument

The calculations of the expected onestep drift and diffusion calculations are long and technical. So, as to enhance the readibility, here we outline the sketch of our proof strategy emphasising the key calculations.

Recall the set-up from section 3. After some algebra, we obtain:

$$Q(x,\xi) = -\sqrt{\frac{2\ell^2}{N}}\langle\zeta,\xi\rangle - \frac{\ell^2}{N}\|\xi\|^2 - r(x,\xi) \tag{5.1}$$

$$\zeta \equiv C^{-1/2}(P^N x) + C^{1/2}\ \nabla\Psi^N(P^N x) \tag{5.2}$$

$$r(x,\xi) \equiv \Psi^N(P^N y) - \Psi^N(P^N x) - \langle\nabla\Psi(P^N x), P^N y - P^N x\rangle$$

The reason behind doing the Taylor expansion here is to obtain the gradient term $\nabla\Psi$ which appears in the drift term in the SPDE (1.5). By (4.5) and Assumptions 4.1, 4.3 on $\Psi$and Ψ^N, we have a global bound on the remainder term:

$$|r(x,\xi)| \le M_6\, 2\frac{\ell^2}{N}\,\|C^{1/2}\xi\|^2 \tag{5.3}$$

Remark 5.1. *If $x\sim\pi_0$ in $\mathcal{H}$, then the random variable $C^{-1/2}x$ is not well defined because $C^{-1/2}$ is not a trace class operator. However equation* (5.2) *is still well defined because the operator $C^{-1/2}$ acts only in X^N for any* fixed *N.*

Before we proceed, we need the following simple lemma:

Lemma 5.2. *Let $Z_\ell \sim No(-\ell^2, 2\ell^2)$. Then,*

$$\mathbb{P}(Z_\ell > 0) = \mathbb{E}\Big(e^{Z_\ell}1_{Z_\ell<0}\Big) = \Phi\Big(-\frac{\ell}{\sqrt{2}}\Big).$$

Hence from the above lemma, if $Z_l \sim \mathrm{No}(-\ell^2, 2\ell^2)$,

$$\mathbb{E}(1\wedge e^{Z_\ell}) = 2\Phi\Big(-\frac{\ell}{\sqrt{2}}\Big) = \beta \tag{5.4}$$

Lemma 5.2 gives us the relation between the expected drift and diffusion coefficients which ensures π invariance, as will be seen later in this section.

5.1. Expected Drift

In this section, we will give heuristic arguments which underly Theorem 7.1. Let $x_i^k, i\le N$ denote the i^{th} coordinate of x^k. Recall that $\mathcal{F}_k$ denotes the sigma algebra generated by $\{\gamma^n,\xi^n, n\le k\}$ and the conditional expectations $\mathbb{E}(\cdot|\mathcal{F}_k)$ are denoted by $\mathbb{E}_k(\cdot)$. Thus $\mathbb{E}_0(\cdot)$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

denotes the expecation with respect to ξ^0 and γ^0 with x^0 fixed. Also for notational convinience set $x^0 = x$ and $\xi^0 = \xi$. It follows that,

$$\begin{aligned} N\,\mathbb{E}_0(x_i^1 - x_i^0) &= N\,\mathbb{E}_0\Big(\gamma^0(y_i^0 - x_i)\Big) \\ &= N\,\mathbb{E}_0^{\xi}\Big(\alpha(x,\xi\,)\sqrt{\frac{2\ell^2}{N}}(C^{1/2}\xi)_i\Big) \end{aligned}$$

where $\mathbb{E}_0^{\xi}$ denotes the expectation with respect to ξ

$$\begin{aligned} &= \lambda_i\,\sqrt{2\ell^2 N}\,\mathbb{E}_0^{\xi}\Big(\alpha(x,\xi\,)\,\xi_i\Big) \\ &= \lambda_i\,\sqrt{2\ell^2 N}\,\mathbb{E}_0^{\xi}\Big((1\wedge e^{Q(x,\xi)})\xi_i\Big) \end{aligned} \tag{5.5}$$

It can be seen equation (5.1) and the law of large numbers that for large N,

$$Q(x,\xi\,) \approx \mathrm{No}\Big(-\ell^2, 2\ell^2\frac{\|\zeta\|^2}{N}\Big).$$

It turns out that (see lemma 6.2) $\lim_{N\to\infty}\frac{\|\zeta\|^2}{N} = 1$, π almost surely. Therefore for large enough N,

$$Q(x,\xi\,) \Rightarrow \mathrm{No}\Big(-\ell^2, 2\ell^2\Big),\ \ \pi \text{ a.s.} \tag{5.6}$$

To evaluate (5.5), it is easier to first factorize $Q(x,\xi\,)$ into components involving ξ_i and orthogonal to them. To this end we introduce the following terms:

$$R(x,\xi\,) \equiv Q(x,\xi\,) + r(x,\xi\,) = -\sqrt{\frac{2\ell^2}{N}}\sum_{j=1}^{N}\zeta_j\xi_j - \frac{\ell^2}{N}\sum_{j=1}^{N}\xi_j^2 \tag{5.7}$$

$$R_i(x,\xi\,) \equiv R(x,\xi\,) + \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_i + \frac{\ell^2}{N}\xi_i^2 = -\sqrt{2\frac{\ell^2}{N}}\sum_{j=1,j\neq i}^{N}\zeta_j\xi_j - \frac{\ell^2}{N}\sum_{j=1,j\neq i}^{N}\xi_j^2 \tag{5.8}$$

The utility of rewriting $R(x,\xi\,)$ in terms of $R_i(x,\xi\,)$ is due to the following formula:

Lemma 5.3. *Let $z \sim \mathrm{No}(0,1)$. Then*

$$\mathbb{E}\Big[z\Big(1\wedge\exp(az+b)\Big)\Big] = a\exp(a^2/2+b)\,\Phi\Big(-\frac{b}{|a|} - |a|\Big)$$

Now for large enough N,

$$Q(x,\xi_{\,i}) \approx R_i(x,\xi\,) - \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_i \tag{5.9}$$

The important observation here is that conditional on x, the random variable $R_i(x,\xi\,)$ is independent of ξ_i. Hence the expectation $\mathbb{E}_0^{\xi}\Big((1\wedge e^{Q(x,\xi)})\xi_i\Big)$ can be computed by first computing it over ξ_i and then over $\xi\setminus\xi_i$. Let $\mathbb{E}^{\xi_i^-}, \mathbb{E}^{\xi_i}$ denote the expectation with respect to

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

$\xi \setminus \xi_i, \xi_i$ respectively. Using the relation (5.9), and applying lemma 5.3 with $a = -\sqrt{\frac{2\ell^2}{N}}\zeta_i$, $z = \xi_i$ and $b = R_i(x,\xi)$, we obtain

$$\mathbb{E}_0^{\xi}\Big((1 \wedge e^{Q(x,\xi)})\xi_i\Big) \approx -\sqrt{\frac{2\ell^2}{N}}\zeta_i\, \mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi)+\frac{\ell^2}{N}\zeta_i^2}\, \Phi\Big(\frac{-R_i(x,\xi)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|} - \sqrt{\frac{2\ell^2}{N}}|\zeta_i|\Big)$$
$$\approx -\sqrt{\frac{2\ell^2}{N}}\zeta_i\, \mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi)}\, \Phi\Big(\frac{-R_i(x,\xi)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|}\Big) \tag{5.10}$$

Now from the relation (5.9) and the weak convergence of $Q(x,\xi)$ deduced in (5.6), it follows that

$$R_i(x,\xi) \Rightarrow \mathrm{No}(-\ell^2, 2\ell^2), \;\; \pi \text{ a.s.} \tag{5.11}$$

Since for large enough N, $\Phi\Big(\frac{-R_i(x,\xi)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|}\Big) \approx 1_{R_i(x,\xi)<0}$, and therefore by (5.10),

$$\mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi)+\frac{\ell^2}{N}\zeta_i^2}\, \Phi\Big(\frac{-R_i(x,\xi)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|} - \sqrt{\frac{2\ell^2}{N}}|\zeta_i|\Big) \approx \mathbb{E}_0^{\xi_i^-}\; \Big(e^{R_i(x,\xi)}1_{R_i(x,\xi)<0}\Big) \to \mathbb{E}e^{Z_\ell}1_{Z_\ell<0} = \beta/2. \tag{5.12}$$

Hence from (5.5), (5.10) and (5.12), we gather that for large N,

$$N\,\mathbb{E}_0(x_i^1 - x_i^0) \approx -\ell^2\beta\;\lambda_i\zeta_i$$

The following lemma identifies the drift term:

Lemma 5.4.

$$\lambda_i\zeta_i = \Big(P^N x + C\nabla\Psi^N(P^N x)\Big)_i$$

Hence for large enough N, we deduce that (atleast heuristically), that the expected drift in the i^{th} coordinate after one step of the Markov chain $\{x^k\}$ is well approximated by:

$$N\,\mathbb{E}_0(x_i^1 - x_i^0) \approx -\ell^2\beta\Big(P^N x + C\nabla\Psi^N(P^N x)\Big)_i \approx -\ell^2\beta\Big(P^N x + C\nabla\Psi(P^N x)\Big)_i$$

which is exactly the drift term that appears in the SPDE (1.5)! Therefore the above heuristic arguments show how the Metropolis algorithm achieves the 'change of measure' by mapping π_0 to π. Now, the approximations made above will be identities had $Q(x,\xi)$ been a Gaussian random variable! So we write,

$$N\,\mathbb{E}_0(x_i^1 - x_i^0) = \ell^2\beta\Big(P^N x + C\nabla\Psi^N(P^N x)\Big)_i + r^N(i)$$

It is easy to show by qualitative arguments (for instance, by the dominated convergence theorem), that for each fixed i, the error term $r^N(i) \to 0$, π almost surely. However, we are interested in an invariance principle in infinite dimensions and therefore need explicit control

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

of the error $r^N(i)$ (and uniformly over i). To obtain such control we need careful bookkeeping of the error terms and this is the primary technical difficulty to be overcome. We show in Lemma 7.8 that

$$\lim_{N\to\infty} \mathbb{E}^\pi \|r^N\|^2 = \lim_{N\to\infty} \mathbb{E}^\pi \sum_{i=1}^N |r^N(i)|^2 = 0.$$

We take advantage of the fact that $Q(x,\xi)$ converges weakly to a Gaussian random variable. So to obtain control over the error terms, the explicit rate of converge $Q(x,\xi)$ to the Gaussian random variable Z_ℓ is needed (such as the Berry-Essen theorem). We use Stein's method (see section 6.1), a natural tool for obtaining convergence rates, to obtain the required bounds.

5.2. Expected diffusion coefficient

As done in the expected drift calculations, we now give the heuristic arguments for the expected diffusion coefficient, after one step of the Markov chain $\{x^k\}$. The arguments used here are much simpler than the drift calculations. Recall that $X^N \subset \mathcal{H}$ is the subspace spanned by $\{\phi_1, \phi_2, \cdots, \phi_N\}$. Let C^N denote the covariance operator on X^N.

The strategy is the same as in the drift case. We look at the covariance between two coordinates x_i^1, x_j^1. For $1 \le i, j \le N$,

$$\begin{aligned} N\,\mathbb{E}_0\Big[(x_i^1 - x_i^0)(x_j^1 - x_j^0)\Big] &= N\,\mathbb{E}_0^\xi\Big[(y_i^0 - x_i)(y_j^0 - x_j)\,\alpha(x,\xi)\Big] \\ &= N\,\mathbb{E}_0^\xi\Big[(y_i^0 - x_i)(y_j^0 - x_j)\,\Big(1 \wedge \exp Q(x,\xi)\Big)\Big] \\ &= 2\ell^2\,\mathbb{E}_0^\xi\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\,\Big(1 \wedge \exp Q(x,\xi)\Big)\Big] \end{aligned} \tag{5.13}$$

Now notice that

$$\mathbb{E}_0^\xi\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\Big] = \lambda_i\lambda_j\delta_{ij}$$

Similarly to the calculations used when evaluating the expected drift we express $Q(x,\xi)$ in terms of:

$$\begin{aligned} R_{ij}(x,\xi) &= R(x,\xi) + \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_i + \frac{\ell^2}{N}\xi_i^2 + \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_j + \frac{\ell^2}{N}\xi_j^2 \\ &= -\sqrt{\frac{2\ell^2}{N}}\sum_{k=1,k\neq i,j}^N \zeta_k\xi_k - \frac{\ell^2}{N}\sum_{k=1,k\neq i,j}^N \xi_k^2 \end{aligned} \tag{5.14}$$

Therefore we replace $Q(x,\xi)$ in equation (7.8) by $R_{ij}(x,\xi)$ and take advantage of the fact that $R_{ij}(x,\xi)$ is conditionlly independent of ξ_i, ξ_j. However the additional error term introduced is easy to estimate because the function $f(x) \equiv (1 \wedge e^x)$ is 1-Lipschitz. So, for large enough N,

$$\mathbb{E}_0^\xi\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\Big(1 \wedge \exp R(x,\xi)\Big)\Big] \approx \mathbb{E}_0^\xi\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\,\Big(1 \wedge \exp R_{ij}(x,\xi)\Big)\Big]$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

$$= \lambda_i \lambda_j \delta_{ij} \, \mathbb{E}_0^{\xi_{ij}^-} \Big[\Big(1 \wedge \exp R_{ij}(x,\xi\,) \Big) \Big] \tag{5.15}$$

Remark 5.5. *This trick of "decoupling" the coordinates i, j from the rest using a simple application of Cauchy-Schwartz inequality does not work in the drift calculations, since the expected drift term carries information about the change of measure. This is the main difficulty one has to over come while analyzing random walk metropolis algorithms. However a similar decoupling occurs in the drift for Langevin algorithms (see (Roberts & Rosenthal, 1998)).*

Once again we have the weak limit:

$$R_{ij}(x,\xi\,) \Rightarrow \mathrm{No}(-\ell^2, 2\ell^2), \;\; \pi \text{ a.s.}$$

So by dominated convergence theorem

$$\lim_{N\to\infty} \mathbb{E}^{\xi_{ij}^-} \Big[\Big(1 \wedge \exp R_{ij}(x,\xi\,) \Big) \Big] = \beta.$$

Therefore for large N,

$$\begin{aligned} N\,\mathbb{E}_0 \Big[(x_i^1 - x_i^0)(x_j^1 - x_j^0) \Big] &\approx 2\ell^2 \beta \lambda_i \lambda_j \delta_{ij} \\ &= 2\ell^2 \beta \langle \phi_i, C\phi_j \rangle \end{aligned}$$

Notice that we have recovered the correct covariance operator C, with diffusion term $2\ell^2\beta$ which is twice the drift coefficient $\ell^2\beta$. The approximations made above can be turned into identities, once we have control over the error estimates. So, we write:

$$N\,\mathbb{E}_0 \Big[(x_i^1 - x_i^0)(x_j^1 - x_j^0) \Big] = 2\ell^2 \beta \lambda_i \lambda_j \delta_{ij} + u_{ij}^N$$

and once again use Stein's method to obtain the convergence rate of error of u_{ij}^N in the operator norm. To summarize, we show in theorem 7.9 that,

$$\begin{aligned} N\,\mathbb{E}_0 \Big[(x^1 - x^0) \otimes (x^1 - x^0) \Big] &= 2\ell^2 \beta\, C^N + u^N \\ \lim_{N\to\infty} \mathbb{E}^{\pi^N} \|u^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} &= 0 \end{aligned}$$

6. Gaussian approximation

In this section we discuss the framework which is used later to derive the rates of convergence of the error terms. We start with a simple estimate. Recall from equations (5.7), (5.8) & (5.14) that:

$$R(x,\xi\,) = -\sqrt{\frac{2\ell^2}{N}} \sum_{j=1}^{N} \zeta_j \xi_j - \frac{\ell^2}{N} \sum_{j=1}^{N} \xi_j^2$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

$$R_i(x,\xi) = R(x,\xi) + \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_i + \frac{\ell^2}{N}\xi_i^2$$
$$R_{ij}(x,\xi) = R(x,\xi) + \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_i + \frac{\ell^2}{N}\xi_i^2 + \sqrt{\frac{2\ell^2}{N}}\zeta_j\xi_j + \frac{\ell^2}{N}\xi_j^2$$

These quantities were introduced so that the term in the exponential of the acceptance probability $Q(x,\xi)$ could be replaced with $R_i(x,\xi)$ and $R_{ij}(x,\xi)$ to take advantage of the fact that, conditional on x, $R_i(x,\xi)$ is independent of ξ_i and $R_{ij}(x,\xi)$ independent of ξ_i, ξ_j. In the next lemma, we estimate the additional error due to this replacement of $Q(x,\xi)$:

Lemma 6.1.

$$\mathbb{E}_x^{\xi}|Q(x,\xi) - R_i(x,\xi)|^2 \leq M\frac{1}{N}(1+|\zeta_i|^2) \tag{6.1}$$
$$\mathbb{E}_x^{\xi}\Big(Q(x,\xi) - R_{ij}(x,\xi)\Big)^2 \leq M\frac{1}{N}(1+|\zeta_i|^2+|\zeta_j|^2) \tag{6.2}$$

The random variables $R(x,\xi)$,$R_i(x,\xi)$ and $R_{ij}(x,\xi)$ are approximately Gaussian random variables. Indeed it can be readily seen that,

$$R(x,\xi) \approx \mathrm{No}(-\ell^2, 2\frac{\ell^2}{N}\|\zeta\|^2).$$

The next lemma contains a crucial observation. We show that the sequence of random variables $\{\frac{\|\zeta\|^2}{N}\}$ converges to 1 almost surely under both π_0 and π. Thus $R(x,\xi)$ converges almost surely to $Z_\ell \equiv \mathrm{No}(-\ell^2, 2\ell^2)$ random variable and thus the expected acceptance probability $\mathbb{E}\alpha(x,\xi) = 1 \wedge e^{Q(x,\xi)}$ converges to $\beta = \mathbb{E}(1 \wedge e^{Z_\ell})$.

Lemma 6.2.

$$\frac{1}{N}\|\zeta\|^2 = 1, \;\; \pi_0 \;\; a.s. \tag{6.3}$$
$$\frac{1}{N}\|\zeta\|^2 = 1, \;\; \pi \;\; a.s. \tag{6.4}$$
$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\big(\big|1 - \frac{1}{N}\|\zeta\|^2\big|^2\big) = 0 \tag{6.5}$$
$$\limsup_{N\to\infty} \mathbb{E}^{\pi^N} e^{c\frac{1}{N}\|\zeta\|^2} < \infty, \;\; \textit{for any } c > 0. \tag{6.6}$$

The proof of Lemma 6.2 proceeds by showing the conclusions first in the case when $x \sim \pi_0$; this is easier because the finite dimensional distributions are Gaussian and by Fernique's theorem x has exponential moments. Next we notice that the almost sure properties are preserved under the change of measure π. To show the convergence of moments, we use our hypothesis that the Radon-Nikodym derivative $\frac{d\pi^N}{d\pi_0}$ is bounded above independent of N as shown in Lemma 4.4, equation (4.7).

From lemma 6.1 it follows that $R_i(x,\xi)$ and $R_{ij}(x,\xi)$ also are approximately Gaussian. Therefore the conclusion of Lemma 6.2 leads to the reasoning that, for any fixed realization of $x \sim \pi$, the random variables $R(x,\xi)$, $R_i(x,\xi)$ and $R_{ij}(x,\xi)$ all converge to the same weak limit

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

$\mathrm{No}(-\ell^2, 2\ell^2)$ as the dimension of the noise ξ goes to ∞. Of course this last deduction would be rigorous had $R(x,\xi)$ been a Gaussian random variable itself! In the rest of this subsection, we rigorize this argument by deriving a Berry-Essen bound for the weak convergence of $R(x,\xi)$ to Z_ℓ, using Stein's method.

6.1. Stein's method

Stein's method gives quantative bounds for functionals of the form $\mathbb{E}(g(W) - g(Z))$, where W is a random variable "close" in distribution to a standard normal random variable Z and g is either a bounded function or Lipscitz continuous. See the monograph Diaconis & Holmes (2004) for a clear exposition. Consider the following ODE (called Stein equation):

$$f'(w) - wf(w) = h(w) - \mathbb{E}(h(Z)) \tag{6.7}$$

If h is bounded, by Stein's lemma (Diaconis & Holmes (2004), Chapter 1), there exists an absolutely continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ solving (6.7) such that,

$$\|f\|_\infty \le M\|h - \mathbb{E}h\|_\infty, \|f'\|_\infty \le \|h - \mathbb{E}h\|_\infty \tag{6.8}$$

Furthermore if h is Lipschitz and hence differentiable a.e.

$$\|f''\|_\infty \le M\|h'\|_\infty \tag{6.9}$$

The fact that the solution f of the Stein equation may not have a second derivative when h is not Lipschitz will turn out to be crucial for our calculations.

Our goal is to obtain quantitative bounds for the weak convergence of $R(x,\xi)$ to $Z_\ell \sim \mathrm{No}(-\ell^2, 2\ell^2)$. For our purposes, it is natural and convenient to obtain these bounds in the Wasserstein metric. Recall that the Wasserstein distance between two random variables $\mathrm{Wass}(X, Y)$ is defined by

$$\mathrm{Wass}(X, Y) \equiv \sup_{f \in \mathcal{D}} \mathbb{E}(f(X) - f(Y))$$

where $\mathcal{D}$ is the class of 1-Lipschitz functions. Define

$$W \equiv \frac{1}{\sqrt{2\ell^2}}(R(x,\xi) + \ell^2) \tag{6.10}$$

$$W_i \equiv \frac{1}{\sqrt{2\ell^2}}(R_i(x,\xi) + \ell^2) \tag{6.11}$$

This rescaling is done since it is convinient to apply Stein's method to obtain Berry-Essen bounds for the weak convergence to a standard normal random variable. The following lemma, proved using Stein's method, gives a bound for the Wasserstein distance between W and Z, where $Z \sim \mathrm{No}(0, 1)$.

Lemma 6.3.

$$\mathrm{Wass}(W, Z) \le M\Big(\frac{1}{\sqrt{N}} + \frac{1}{N^{3/2}}\sum_{j=1}^{N} |\zeta_j|^3 + \Big|1 - \frac{1}{N}\|\zeta\|^2\Big|\Big) \tag{6.12}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Also a simple estimate will yield that

$$\text{Wass}(W, W_i) \le M \frac{1}{\sqrt{N}}(|\zeta_i| + 1) \tag{6.13}$$

Hence from equations (6.13) and (6.12), noticing that the function $f(x) = 2x+1$ has Lipschitz constant $\sqrt{2}$, we obtain

$$\text{Wass}(R_i(x,\xi\,), Z_\ell) \le M\Big(\frac{1}{\sqrt{N}}(|\zeta_i| + 1) + \frac{1}{N^{3/2}}\sum_{j=1}^{N}|\zeta_j|^3 + \Big|1 - \frac{1}{N}\|\zeta\|^2\Big|\Big) \tag{6.14}$$

We conclude this section with the following observation which will be used later. Recall the Kolmogorov-Smirnov (KS) distance between two random variables (X, Y):

$$\text{KS}(X, Y) \equiv \sup_{t\in\mathbb{R}} |\mathbb{P}(X \le t) - \mathbb{P}(Y \le t)| \tag{6.15}$$

Moreover if the random variable Y has a density with respect to the Lebesgue measure, bounded by a constant M, then

$$\text{KS}(X, Y) \le \sqrt{4M\text{Wass}(X, Y)} \tag{6.16}$$

7. Expected onestep drift and diffusion

Now we are ready to rigorously derive the one step expected drift and diffusion terms.

7.1. *Expected Drift*

Theorem 7.1. *Let $\{x^k\}$ be the RWM Markov chain with $x^0 = x \sim \pi^N$. Then*

$$N\,\mathbb{E}_0(x_i^1 - x_i^0) = -\ell^2\beta\Big(P^N x + C\nabla\Psi^N(P^N x)\Big)_i + r^N(i)$$
$$\lim_{N\to\infty} \mathbb{E}^\pi\|r^N\|^2 = 0.$$

Proof of Theorem 7.1: We prove the theorem via a series of lemmas.

Lemma 7.2.

$$N\,\mathbb{E}_0(x_i^1 - x_i) = \lambda_i\,\sqrt{2\ell^2 N}\,\mathbb{E}_0^\xi\Big((1 \wedge e^{R_i(x,\xi) - \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_i})\xi_i\Big) + \omega_0(i)$$
$$|\omega_0(i)| \le \frac{M}{\sqrt{N}}\lambda_i\sqrt{1 + |\zeta_i|}$$

Applying Lemma 5.3 with $a = -\sqrt{\frac{2\ell^2}{N}}\zeta_i$, $z = \xi_i$ and $b = R_i(x,\xi\,)$, we obtain

$$\mathbb{E}_0^\xi\Big((1 \wedge e^{R_i(x,\xi) - \sqrt{\frac{2\ell^2}{N}}\zeta_i\xi_i})\xi_i\Big) = -\sqrt{\frac{2\ell^2}{N}}\zeta_i\,\mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi) + \frac{\ell^2\zeta_i{}^2}{N}}\,\Phi\Big(\frac{R_i(x,\xi\,)}{\sqrt{2\delta}|\zeta_i|} - \sqrt{2\delta}|\zeta_i|\Big) \tag{7.1}$$

Using the Lipschitzness ofΦ , we obtain:

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Lemma 7.3.

$$\mathbb{E}_0^{\xi_i^-}\, e^{R_i(x,\xi)+\frac{\ell^2 {\zeta_i}^2}{N}}\,\Phi\Big(\frac{R_i(x,\xi\,)}{\sqrt{2\delta}|\zeta_i|}-\sqrt{2\delta}|\zeta_i|\Big)=\mathbb{E}_0^{\xi_i^-}\, e^{R_i(x,\xi)+\frac{\ell^2 {\zeta_i}^2}{N}}\,\Phi\Big(\frac{-R_i(x,\xi\,)}{\sqrt{2\delta}|\zeta_i|}\Big)+\omega_1(i) \quad (7.2)$$

$$|\omega_1(i)|\le M|\zeta_i|\frac{1}{\sqrt{N}}e^{\frac{\ell^2}{N}\|\zeta\|^2} \quad (7.3)$$

The next two lemmas are the key technical components the proof of Theorem 7.1.

Lemma 7.4.

$$\mathbb{E}_0^{\xi_i^-}\, e^{R_i(x,\xi)+\frac{\ell^2 {\zeta_i}^2}{N}}\,\Phi\Big(\frac{-R_i(x,\xi\,)}{\sqrt{2\delta}|\zeta_i|}\Big)=\mathbb{E}_0^{\xi_i^-}\, e^{R_i(x,\xi)+\frac{\ell^2 {\zeta_i}^2}{N}}1_{R_i(x,\xi)<0}+\omega_2(i)$$

$$|\omega_2(i)|\le e^{\frac{\ell^2}{N}\|\zeta\|^2}(|\zeta_i|+1)^2\Big[\mathbb{E}_0^{\xi}\frac{1}{(1+|R(x,\xi\,)|\sqrt{N})^2}\Big]^{1/2} \quad (7.4)$$

Notice that in the above error estimate in ω_2 has $R(x,\xi\,)$ instead of $R_i(x,\xi\,)$. This lemma formalizes the intuition that, since $R_i(x,\xi\,)$ for all $i\in\mathbb{N}$ has the same weak limit as $R(x,\xi\,)$, the additional error term due to the replacement of $R_i(x,\xi\,)$ by $R(x,\xi\,)$ in the expression $\mathbb{E}_0^{\xi_i^-}\, e^{R_i(x,\xi)+\frac{\ell^2 {\zeta_i}^2}{N}}\,\Phi\Big(\frac{-R_i(x,\xi)}{\sqrt{2\delta}|\zeta_i|}\Big)$ can be controlled uniformly over i.

Lemma 7.5.

$$\mathbb{E}_0^{\xi_i^-}\, e^{R_i(x,\xi)+\frac{\ell^2 {\zeta_i}^2}{N}}1_{R_i(x,\xi)<0}=\frac{\beta}{2}+\omega_3(i)$$

$$|\omega_3(i)|\le M\frac{\zeta_i^2}{N}e^{\ell^2\frac{\|\zeta\|^2}{N}}+M\sqrt{\Big(\frac{1+|\zeta_i|}{\sqrt{N}}+\frac{1}{N^{3/2}}\sum_{j=1}^{N}|\zeta_j|^3+|(1-\frac{1}{N}\|\zeta\|^2)|\Big)}$$

Remark 7.6. *For deriving the error bounds in Lemma 7.5, we cannot directly apply the Wasserstein bounds obtained in equation* (6.14)*, because the function* $b(x)\equiv e^x 1_{x<0}$ *is not Lipschitz on* $\mathbb{R}$*. However as noted in* (6.16)*, the KS distance between* $R_i(x,\xi\,)$ *and* Z_ℓ *is bounded above the Wasserstein distance, and this is enough to get the required error bounds. Also notice that the function* $g(x)\equiv\Phi(x\sqrt{N})$ *has a Lipschitz constant* $\sqrt{N}$*, and the Berry-Essen bounds from applying Stein's method to the function* $g(x)\equiv e^x\Phi(x\sqrt{N})$ *will not be bounded in* N*. Hence we need Lemma 7.4 to overcome this difficulty.*

The following lemma identifies the drift term:

Lemma 7.7.

$$\lambda_i\zeta_i=\Big(P^N x+C\nabla\Psi^N(P^N x)\Big)_i$$

Lemma 7.7 has the drift term $P^N x+C\nabla\Psi^N(P^N x)$ instead of $P^N x+C\nabla\Psi(P^N x)$. Set

$$\omega_4(i)\equiv(C\nabla\Psi^N(P^N x))_i-(C\nabla\Psi(P^N x))_i=\langle C\Big(\nabla\Psi^N(P^N x))-\nabla\Psi(P^n x)\Big),\phi_i\rangle \quad (7.5)$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Putting together all the above estimates, we obtain

$$N\mathbb{E}_0^{\xi}[x_i^1 - x_i] = -\ell^2\beta\,(P^N x + C\nabla\Psi(P^N x))_i + r_i^N \tag{7.6}$$

$$|r_i^N| \leq |\,\omega_0(i)| + M\lambda_i|\zeta_i|\Big(|\omega_1(i)| + |\omega_2(i)| + |\omega_3(i)| + |\omega_4(i)|\Big) \tag{7.7}$$

Now we estimate the norm of the error term r^N:

Lemma 7.8.

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\|r^N\|^2 = \lim_{N\to\infty} \mathbb{E}^{\pi^N}\sum_{i=1}^N |r_i^N|^2 = 0.$$

Hence by equations (7.6), (7.7) and lemma 7.8,

$$N\mathbb{E}_0^{\xi}[x^1 - x] = -\ell^2\beta\left(P^N x + C\nabla\Psi^N(P^N x)\right) + r^N,$$

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\|r^N\|^2 = 0$$

and we have proved the theorem. □

7.2. Expected diffusion coefficient

Recall that $X^N \subset \mathcal{H}$ is the subspace spanned by $\{\phi_1, \phi_2, \cdots, \phi_N\}$. Let C^N denote the covariance operator on X^N. Now we are ready to derive the expected diffusion term:

Theorem 7.9. *Let $\{x^k\}$ be the RWM Markov chain with $x^0 = x \sim \pi^N$. Then*

$$N\,\mathbb{E}_k\Big[(x^1 - x^0)\otimes(x^1 - x^0)\Big] = 2\ell^2\beta\,C^N + E^N$$

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\|E^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} = 0$$

Moreover,

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N} N\mathrm{tr}\Big[\mathbb{E}_0(x^1 - x^0)\otimes(x^1 - x^0)\Big] = 2\ell^2\beta\,\mathrm{tr}(C)$$

Proof. The proof strategy is the same as in the drift case. We look at the covariance between two coordinates x_i^k, x_j^k. For $1 \leq i, j \leq N$,

$$\begin{aligned} N\,\mathbb{E}_0\Big[(x_i^1 - x_i^0)(x_j^1 - x_j^0)\Big] &= N\,\mathbb{E}_0^{\xi}\Big[(y_i^0 - x_i)(y_j^0 - x_j)\,\alpha(x,\xi\,)\Big] \\ &= N\,\mathbb{E}_0^{\xi}\Big[(y_i^0 - x_i)(y_j^0 - x_j)\,\Big(1\wedge\exp Q(x,\xi\,)\Big)\Big] \\ &= 2\ell^2\,\mathbb{E}_0^{\xi}\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\,\Big(1\wedge\exp Q(x,\xi\,)\Big)\Big] \end{aligned} \tag{7.8}$$

Now notice that

$$\mathbb{E}_k^{\xi}\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\Big] = \lambda_i\lambda_j\delta_{ij}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Lemma 7.10.

$$\mathbb{E}_0^{\xi}\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\,\Big(1\wedge\exp Q(x,\xi\,)\Big)\Big]=\mathbb{E}^{\xi}\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\,\Big(1\wedge\exp R_{ij}(x,\xi\,)\Big)\Big]+\kappa_{ij}$$
$$|\kappa_{ij}|\le M\lambda_i\lambda_j(1+|\zeta_i|^2+|\zeta_j|^2)^{1/2}\frac{1}{\sqrt{N}}$$

Now due to the conditional independence of $R_{ij}(x,\xi\,)$ and ξ_i^k,ξ_j^k, it follows that:

$$\mathbb{E}^{\xi}\Big[(C^{1/2}\xi)_i(C^{1/2}\xi)_j\,\Big(1\wedge\exp R_{ij}(x,\xi\,)\Big)\Big]=\lambda_i\lambda_j\delta_{ij}\,\mathbb{E}^{\xi_{ij}^-}\Big[\Big(1\wedge\exp R_{ij}(x,\xi\,)\Big)\Big]$$

Since $R_{ij}(x,\xi\,)$ has a weak limit $\mathrm{No}(-\ell^2,2\ell^2)$, by Lemma 5.2 and dominated convergence theorem

$$\lim_{N\to\infty}\mathbb{E}^{\xi_{ij}^-}\Big[\Big(1\wedge\exp R_{ij}(x,\xi\,)\Big)\Big]=\beta$$

Once again we use Stein's method to derive the convergence rate:

Lemma 7.11.

$$\mathbb{E}^{\xi_{ij}^-}\Big[\Big(1\wedge\exp R_{ij}(x,\xi\,)\Big)\Big]=\beta+\rho_{ij}$$
$$|\rho_{ij}|\le M\Big(\frac{1}{\sqrt{N}}(1+|\zeta_i|+|\zeta_j|)+\frac{1}{N^{3/2}}\sum_{s=1}^{N}|\zeta_s|^3+\Big|1-\frac{1}{N}\|\zeta\|^2\Big|\Big)$$

Putting together all the estimates,

$$\begin{aligned}N\,\mathbb{E}_0\Big[(x_i^1-x_i^0)(x_j^1-x_j^0)\Big]&=2\ell^2\beta\lambda_i\lambda_j\delta_{ij}+E_{ij}^N\\ |E_{ij}^N|&\le|\,\kappa_{ij}|+\lambda_i\lambda_j\delta_{ij}|\rho_{ij}|\end{aligned}\tag{7.9}$$

Finally we estimate the error of u_{ij}^N in the operator norm:

Lemma 7.12.

$$\lim_{N\to\infty}\mathbb{E}^{\pi^N}\|E^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})}=0$$

Therefore we have shown

$$N\,\mathbb{E}_0\Big[(x_i^1-x_i^0)(x_j^1-x_j^0)\Big]=2\ell^2\beta\langle\phi_i,C\phi_j\rangle+E^N$$
$$\lim_{N\to\infty}\mathbb{E}^{\pi^N}\|E^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})}=0$$

The convergence of the trace also easily follows from the previous estimates:

Lemma 7.13.

$$\lim_{N\to\infty}\mathbb{E}^{\pi^N}N\,\mathrm{tr}\Big[\mathbb{E}_0(x^1-x^0)\otimes(x^1-x^0)\Big]=2\ell^2\beta\,\mathrm{tr}(C)$$

and hence the theorem is proved. □

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

8. Convergence to the SPDE Limit

Let $x^k \sim \pi^N$ and $x(0) \equiv x^0$. We first write the identity (see (3.7))

$$x^{k+1} = x^k + \mathbb{E}_k(x^{k+1} - x^k) + \sqrt{\frac{2\ell^2\beta}{N}}\,\Gamma^{k,N} \tag{8.1}$$

From Theorem 7.1

$$\begin{aligned} x^{k+1} &= x^k - \frac{\ell^2\beta}{N}\Big(P^N x^k + C\nabla\Psi(P^N x^k)\Big) + \frac{r^{k,N}}{N} + \sqrt{\frac{2\ell^2\beta}{N}}\,\Gamma^{k,N} \\ &= x^k - \frac{\ell^2\beta}{N} m(x^k) + \frac{r^{k,N}}{N} + \sqrt{\frac{2\ell^2\beta}{N}}\,\Gamma^{k,N} \end{aligned} \tag{8.2}$$

where, by (3.6)

$$m(x^k) = \Big(P^N x^k + C\nabla\Psi(P^N x^k)\Big)$$

and from (8.1) we have,

$$\begin{aligned} \mathbb{E}_k(\Gamma^{k,N}) &= 0 \\ \mathbb{E}_k(\Gamma^{k,N}\otimes\Gamma^{k,N}) &= \frac{N}{2\ell^2\beta}\Big[\mathbb{E}_k((x^{k+1}-x^k)\otimes(x^{k+1}-x^k)) - \mathbb{E}_k(x^{k+1}-x^k)\otimes\mathbb{E}_k(x^{k+1}-x^k)\Big] \\ &= C^N + \frac{1}{2\ell^2\beta}\,E^{k,N} - \frac{N}{2\ell^2\beta}\Big[\mathbb{E}_k(x^{k+1}-x^k)\otimes\mathbb{E}_k(x^{k+1}-x^k)\Big] \end{aligned} \tag{8.3}$$

8.1. Weak convergence of the noise process

Fix $T > 0$, and recall the constants from (3.10). Define

$$W^N(t) \equiv \eta^{\lfloor Nt\rfloor,N} + \sqrt{\Delta t}\,(Nt - \lfloor Nt\rfloor)\,\Gamma^{\lfloor Nt\rfloor+1,N},\ \ t\in[0,T] \tag{8.4}$$

Our goal in this subsection is to show the weak convergence of the stochastic process $W^N(t)$ to $W(t)$ in $C([0,T],\mathcal{H})$.

We need the following functional central limit theorem for Hilbert space valued martingale difference arrays. Let $k_N, N\in\mathbb{N}$ be a sequence of nondecreasing, right-continuous functions $k_N : [0,1] \to \mathbb{Z}_+$ such that $k_N(0) = 0$ and $k_N(1) \geq 1$. Let $\{M^{i,N}, \mathcal{F}^{i,N}\}_{1\leq i\leq k_N(1)}$ be an $\mathcal{H}$ valued martingale array, *i.e.*,

$$\begin{aligned} \mathbb{E}(M^{i,N}|\mathcal{F}^{i-1,N}) &= 0 \\ \mathbb{E}(\|M^{i,N}\|^2|\mathcal{F}^{i-1,N}) &< \infty,\ \text{a.s.} \\ \mathcal{F}^{j,N} &\subset \mathcal{F}^{j+1,N} \end{aligned}$$

Theorem 8.1 (Berger (1986), Theorem 5.1)**.** *Let $S : \mathcal{H} \to \mathcal{H}$ be a self-adjoint, positive definite, operator with finite trace. If for all $x\in\mathcal{H}, \epsilon > 0, t\in[0,1]$:*

1.

$$\lim_{N\to\infty}\sum_{i=1}^{k_N(1)} \mathbb{E}(\|M^{i,N}\|^2|\mathcal{F}^{i-1,N}) = \text{trace}(S),\ \ i.p. \tag{8.5}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

2.

$$\lim_{N\to\infty}\sum_{i=1}^{k_N(t)}\mathbb{E}(\langle M^{i,N},x\rangle^2|\mathcal{F}^{i-1,N})=t\langle Sx,x\rangle,\ \ i.p. \tag{8.6}$$

3.

$$\lim_{N\to\infty}\sum_{i=1}^{k_N(1)}\mathbb{E}(\langle M^{i,N},x\rangle^2 1_{|\langle M^{i,N},x\rangle|\geq\epsilon}|\mathcal{F}^{i-1,N})=0,\ \ i.p. \tag{8.7}$$

then the sequence of random variables $W^N\in C([0,1],\mathcal{H})$ defined by $W^N(t)=\sum_{i=1}^{k_N(t)}M^{i,N}$ if $k_N(t)>k_N(t-)$ and by linear interpolation otherwise, converges in distribution to a Brownian motion W in $\mathcal{H}$, with $W(0)=0$, $\mathbb{E}(W(1))=0$, and with covariance operator S.

Remark 8.2. *The first two hypotheses of the above theorem ensure the weak convergence of finite dimensional distributions of $W^N(t)$ using the martingale central limit theorem in $\mathbb{R}^N$; the last hypothesis is needed to verify the tightness of the family $\{W^N(\cdot)\}$.*

Remark 8.3. *As noted in Chen & White (1998), hypothesis 2 (equation (8.6)) of Theorem 8.1 is implied by*

$$\lim_{N\to\infty}\sum_{i=1}^{k_N(t)}\mathbb{E}(\langle M^{i,N},e_n\rangle\langle M^{i,N},e_m\rangle|\mathcal{F}^{i-1,N})=t\langle Se_n,e_m\rangle,\ \ i.p. \tag{8.8}$$

where $\{e_n\}$ is any orthonormal set of $\mathcal{H}$. Hypothesis 3 (equation (8.6)) of Theorem 8.1 is implied by the Lindberg type condition:

$$\lim_{N\to\infty}\sum_{i=1}^{k_N(1)}\mathbb{E}(\|M^{i,N}\|^2 1_{\|M^{i,N}\|\geq\epsilon}|\mathcal{F}^{i-1,N})=0,\ \ i.p. \tag{8.9}$$

Now we are ready prove the functional CLT:

Lemma 8.4. *The process $W^N(t)$ defined in equation (8.4) converges weakly to W in $C([0,T],\mathcal{H})$ as N tends to ∞, where W is a Brownian motion in time with covariance operator C in $\mathcal{H}$.*

8.2. Weak convergence to SPDE

In order to extend the Markov chain to continuous time, we first define piecewise constant interpolants of the Markov chain, and the mean drift error term, by defining the cádlág process, $t\in[t_k,t_{k+1})$,

$$\bar{z}^N(t)=x^k,$$
$$r_1^N(t)\equiv r^{k,N}.$$

where $r^{k,N}$ is the error term which satisfies the estimate obtained in theorem 7.1. We can use $\bar{z}^N$ to construct a continuous piecewise linear interpolant of x^k by defining

$$z^N(t)=z_0+\ell^2\beta\int_0^t m(\bar{z}^N(s))\,ds+\int_0^t r_1^N(s)\,ds+\sqrt{2\ell^2\beta}\,W^N(t), \tag{8.10}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

The SPDE (1.5) (with $\alpha = \ell^2\beta$) has the integral form:

$$z(t) = z_0 - \ell^2\beta \int_0^t \Big(z(s) + C\nabla\Psi(z(s))\Big)\, ds + \sqrt{2\ell^2\beta}\, W(t). \tag{8.11}$$

In order to facilitate proof of convergence of z^N to z we rewrite (8.10) as

$$z^N(t) = z_0 - \ell^2\beta \int_0^t \Big(z^N(s) + C\nabla\Psi(z^N(s))\Big)\, ds + \int_0^t \Big(r_1^N(s) + r_2^N(s)\Big) ds + \sqrt{2\ell^2\beta}\, W^N(t)$$

where

$$r_2^N(s) \equiv \ell^2\beta\Big(z^N(s) - \bar{z}^N(s) + C\nabla\Psi(z^N(s)) - C\nabla\Psi(\bar{z}^N(s))\Big)$$

Lemma 8.5. *Define* $e^N(t) \equiv \int_0^t \Big(r_1^N(s) + r_2^N(s)\Big)\, ds$. *Then*

$$\lim_{N\to 0} \mathbb{E}^{\pi^N}\big(\sup_{t\in[0,T]} \|e^N(t)\|^2\big) = 0$$

Lemma 8.6. *Fix any* $T > 0$. *For every* $W \in C([0,T], \mathcal{H})$ *the integral equation* (8.11) *has a unique solution* $z \in C([0,T], \mathcal{H})$. *Furthemore, let* $\Theta : C([0,T], \mathcal{H}) \mapsto C([0,T], \mathcal{H})$ *be the following map:*

$$\Theta(W) \equiv z \tag{8.12}$$

where z *solves the integral equation* (8.11). *Then* Θ *is continuous.*

Proof. The proof is a standard contraction mapping argument, using (4.3)–(4.4). We indicate the continuity argument which also underpins the contraction argument. Let z_i solve (8.11) with $W = W_i, i = 1, 2$. Subtracting the two equations and using the fact that $z \mapsto z + C\nabla\Psi(z)$ is globally Lipschitz on $\mathcal{H}$ gives

$$\|z_1(t) - z_2(t)\| \le M \int_0^t \|z_1(s) - z_s(s)\| ds + \sqrt{2\ell^2\beta}\, \|W_1(t) - W_2(t)\|.$$

Thus

$$\sup_{0\le t\le T} \|z_1(t) - z_2(t)\| \le M \int_0^T \sup_{0\le\tau\le s} \|z_1(\tau) - z_s(\tau)\| ds + \sqrt{2\ell^2\beta} \sup_{0\le t\le T} \|W_1(t) - W_2(t)\|.$$

The Gronwall lemma gives continuity in the desired spaces. □

Now we are ready to prove the main result of the paper:

Theorem 8.7. *The stochastic process* $z^N(t)$ *from* (8.10), *which is a piecewise linear, continuous interpolant of the RWM algorithm, converges weakly in* $C([0,T], \mathcal{H})$ *to the diffusion process* $z(t)$ *given by equation* (1.5).

Proof. The proof is a straightforward consequence of the continuous mapping theorem. Let $\widehat{W}^N = W^N + e^N$. Let Ω denote the probability space generating the Markov chain in stationarity. Then by Lemma 8.5, $e^N \to 0$ in $L^2\Big(C([0,T], \mathcal{H}); \Omega\Big)$ and by Lemma 8.4, W^N converges weakly to W in $C([0,T], \mathcal{H})$. Thus $\widehat{W}^N$ converges weakly to W^N in $C([0,T], \mathcal{H})$. Notice that $z^N = \Theta(\widehat{W}^N)$, where Θ is defined as in Lemma 8.6. Since Θ is a continuous map by Lemma 8.6, we deduce from the continuous mapping theorem that the process z^N converges weakly in $C([0,T], \mathcal{H})$ to z with law given by $\Theta(W)$. This is precisely the law of the SPDE given by (1.5). □

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

References

Bédard, Mylène. 2007. Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Probab.*, **17**(4), 1222–1244.

Berger, Erich. 1986. Asymptotic behaviour of a class of stochastic approximation procedures. *Probab. Theory Relat. Fields*, **71**(4), 517–552.

Beskos, A., & Stuart, A.M. 2007. MCMC Methods for Sampling Function Space. *ICIAM Plenary Lectures*. To Appear.

Beskos, A., Roberts, G.O., Stuart, A.M., & Voss, J. 2008. An MCMC Method for diffusion bridges. *Stochastics and Dynamics*, **8**(3), 319–350.

Chen, Xiaohong, & White, Halbert. 1998. Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory*, **14**(2), 260–284.

Da Prato, Giuseppe, & Zabczyk, Jerzy. 1992. *Stochastic equations in infinite dimensions*. Encyclopedia of Mathematics and its Applications, vol. 44. Cambridge: Cambridge University Press.

Diaconis, Persi, & Holmes, Susan (eds). 2004. *Stein's method: expository lectures and applications*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 46. Beachwood, OH: Institute of Mathematical Statistics. Papers from the Workshop on Stein's Method held at Stanford University, Stanford, CA, 1998.

Ethier, Stewart N., & Kurtz, Thomas G. 1986. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc. Characterization and convergence.

Hairer, M., A.M.Stuart, Voss, J., & Wiberg, P. 2005. Analysis of SPDEs Arising in Path Sampling. Part 1: The Gaussian Case. *Comm. Math. Sci.*, 587–603.

Hairer, M., Stuart, A. M., & Voss, J. 2007. Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.*, **17**(5-6), 1657–1706.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 97–109.

Liu, Jun S. 2008. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. New York: Springer.

Ma, Zhi Ming, & Röckner, Michael. 1992. *Introduction to the theory of (nonsymmetric) Dirichlet forms*. Universitext. Berlin: Springer-Verlag.

N. Metropolis, A.W. Rosenbluth, M.N. Teller, & Teller, E. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.

Robert, Christian P., & Casella, George. 2004. *Monte Carlo statistical methods*. Second edn. Springer Texts in Statistics. New York: Springer-Verlag.

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

Roberts, Gareth O., & Rosenthal, Jeffrey S. 1998. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60**(1), 255–268.

Roberts, Gareth O., & Rosenthal, Jeffrey S. 2001. Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, **16**(4), 351–367.

Walk, H. 1977. An invariance principle for the Robbins-Monro process in a Hilbert space.

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, **39**(2), 135–150.

Appendix

In this section we give the proofs of all technical lemmas.

Proof of Lemma 4.4: By definition,

$$\begin{aligned} M_{\Psi^N}^{-1} &= \int_{\mathcal{H}} \exp\{-\Psi^N(x)\}\pi_0(dx) \\ &\geq \int_{\mathcal{H}} \exp\{-M(1+\|x\|_s^2)\}\pi_0(dx) \geq e^{-2M}\mathbb{P}(\|x\|_s \leq 1) \end{aligned}$$

and therefore

$$\inf_{N\in\mathbb{N}} M_{\Psi^N}^{-1} > 0 \Rightarrow \sup_{N\in\mathbb{N}} M_{\Psi^N} < \infty.$$

For any $f : \mathcal{H} \mapsto \mathbb{R}$

$$\begin{aligned} \sup_{N\in\mathbb{N}} \mathbb{E}^{\pi^N}|f(x)| &\leq \sup_{N\in\mathbb{N}} M_{\Psi^N}\mathbb{E}^{\pi_0}(e^{-\Psi^N(x)}|f(x)|) \\ &\leq M\mathbb{E}^{\pi_0}|f(x)| \end{aligned}$$

proving the lemma. □

Proof of Lemma 4.5: Notice that

$$D_{KL}(\pi||\pi^N) = \mathbb{E}^{\pi}\left|\Psi^N(P^N x) - \Psi(x) + \log\frac{M_\Psi}{M_{\Psi^N}}\right| \tag{8.13}$$

We estimate the two terms of the right hand side of equation (8.13) separately.

$$\begin{aligned} \mathbb{E}^{\pi}|\Psi^N(P^N x) - \Psi(x)| &\leq e^{-M_1}\mathbb{E}^{\pi_0}|\Psi^N(P^N x) - \Psi(x)| + \log\frac{M_\Psi}{M_{\Psi^N}} \\ &\leq e^{-M_1}\mathbb{E}^{\pi_0}|\Psi^N(P^N x) - \Psi(P^N x)| + e^{-M_1}\mathbb{E}^{\pi_0}|\Psi(P^N x) - \Psi(Px)| \\ &\leq e^{-M_1}\, M_7\,\theta(N)\,\mathbb{E}^{\pi_0}\|P^N x\|_s^2 + \\ &\qquad e^{-M_1}\, M_5\,\mathbb{E}^{\pi_0}\Big(1 + \|x\|_s + \|P^N x\|_s\Big)\Big(\|x - P^N x\|\Big) \\ &\leq e^{-M_1}\, M_7\,\theta(N)\,\mathbb{E}^{\pi_0}\|x\|_s^2 + \\ &\qquad e^{-M_1}\, M_5\left[\mathbb{E}^{\pi_0}\Big(\|x - P^N x\|_s^2\Big)^2\right]^{1/2}\left[\mathbb{E}^{\pi_0}(1 + 2\|x\|_s)^2\right]^{1/2} \end{aligned}$$

The first term above goes to 0 because $\theta(N) \to 0$ and $\mathbb{E}^{\pi_0}(\|x\|_s^2) < \infty$ as $s < k - 1/2$. The second term above goes to 0 because $\mathbb{E}^{\pi_0}\Big(\|x - P^N x\|_s^2\Big)^2 \to 0$ and $\mathbb{E}^{\pi_0}(\|x\|_s^2) < \infty$. Now we estimate the second term of equation (8.13).

$$\begin{aligned} |M_\Psi^{-1} - M_{\Psi^N}^{-1}| &= |\int_{\mathcal{H}} \exp\left(-\Psi(x)\right) - \exp\left(-\Psi^N(x)\right)\pi_0(dx)| \\ &\leq \int_{\mathcal{H}} |\exp\left(-\Psi(x)\right) - \exp\left(-\Psi^N(x)\right)|\pi_0(dx) \end{aligned}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

$$\leq M \int_{\mathcal{H}} |\Psi(x) - \Psi^N(x)| \pi_0(dx) \leq M\theta(N)\mathbb{E}^{\pi_0}(\|x\|_s^2)$$

where the inequality follows from the assumption in equation (4.6). Since $\mathbb{E}^{\pi_0}(\|x\|_s^2) < \infty$, it follows that

$$\lim_{N\to\infty} |M_\Psi^{-1} - M_{\Psi^N}^{-1}| = 0$$

Since $M_\Psi < \infty$, it follows that

$$\lim_{N\to\infty} |\frac{M_\Psi}{M_{\Psi^N}} - 1| = 0 \Rightarrow \lim_{N\to\infty} \log \frac{M_\Psi}{M_{\Psi^N}} = 0$$

Hence we have shown that $\lim_{N\to\infty} D_{KL}(\pi || \pi^N) = 0$. Similar calculations yield that $D_{KL}(\pi^N || \pi) \to 0$. Hence the lemma is proved. $\square$

Proof of Lemma 5.2: For $Z \sim \mathrm{No}(\mu,\sigma^2)$, it follows that

$$\mathbb{P}(Z > 0) = 1 - \Phi(-\mu/\sigma).$$

Also,

$$\mathbb{E}\Big(e^Z 1_{Z<0}\Big) = \Phi\Big(-\frac{\mu+\sigma^2}{\sigma}\Big)e^{\mu+\frac{1}{2}\sigma^2}$$

Substituting $\mu = -\ell^2$ and $\sigma^2 = 2\ell^2$ yields

$$\mathbb{P}(R > 0) = \mathbb{E}\Big(e^R 1_{R<0}\Big) = \Phi(-\ell/\sqrt{2})$$

and the lemma is proved. $\square$

Proof of Lemma 5.3:

$$\mathbb{E}\Big[z\Big(1 \wedge \exp(az+b)\Big)\Big] = \int_{\{az+b\geq 0\}} \frac{1}{\sqrt{2\pi}} z\, e^{-z^2/2}\, dz + \int_{\{az+b\leq 0\}} \frac{1}{\sqrt{2\pi}} z\, e^{-z^2/2+az+b}\, dz$$

$$= \int_{\{az+b\geq 0\}} \frac{1}{\sqrt{2\pi}} z\, e^{-z^2/2}\, dz + \int_{\{az+b\leq 0\}} \frac{1}{\sqrt{2\pi}} z\, e^{\left(-\frac{1}{2}(z-a)^2+b+a^2/2\right)}\, dz$$

Replacing $z \mapsto z + a$ in the second term above

$$= \int_{\{az+b\geq 0\}} -\frac{1}{\sqrt{2\pi}} \frac{d}{dz}\Big(e^{-z^2/2}\Big)\, dz + e^{(b+a^2/2)} \int_{\{a^2+az+b\leq 0\}} -\frac{1}{\sqrt{2\pi}} z\, \frac{d}{dz}\Big(e^{-z^2/2}\Big)\, dz$$
$$+ ae^{(b+a^2/2)} \int_{\{a^2+az+b\leq 0\}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

If $a > 0$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{b^2}{2a^2}} - e^{b+a^2/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(a+\frac{b}{a})^2} + ae^{b+a^2/2}\Phi(-b/a - a)$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

The first two terms cancel and we obtain

$$= ae^{b+a^2/2}\Phi(-b/a-a)$$

Similarly, if $a < 0$

$$\mathbb{E}\Big[z\Big(1\wedge\exp(az+b)\Big)\Big] = ae^{b+a^2/2}\Phi(b/a+a)$$

Hence the lemma is proved. □

Proof of Lemma 6.1: Since $C^{1/2}\xi$ is Gaussian and is finite a.s. in $\mathcal{H}$

$$\mathbb{E}\|C^{1/2}\xi\|^4 \le 3(\mathbb{E}\|C^{1/2}\xi\|^2)^2 \le 3(\sum_{j=1}^{\infty}\lambda_j^2)^2 < \infty. \tag{8.14}$$

From (5.7), it follows that

$$\mathbb{E}^{\xi}|Q(x,\xi\,) - R_i(x,\xi\,)|^2 \le M\Big(\mathbb{E}|r(x,\xi\,)|^2 + \frac{1}{N}\zeta_i^2 + \frac{1}{N^2}\mathbb{E}\xi_i^2\Big)$$

using the estimates (5.3) and (8.14),

$$\begin{aligned}
&\le M\Big(\frac{1}{N^2}\mathbb{E}\|C^{1/2}\xi\|^4 + \frac{1}{N}\zeta_i^2 + \frac{1}{N^2}\mathbb{E}\xi_i^2\Big)\\
&\le M\frac{1}{N}\Big(1+\zeta_i^2\Big)
\end{aligned}$$

verifying the first part of the lemma. An identical argument for the second part finishes the proof. □

Proof of Lemma 6.2: First let $x \sim \pi_0$. Recall that $\zeta = C^{-\frac{1}{2}}(P^N x) + C^{\frac{1}{2}}\nabla\Psi^N(P^N x)$ and

$$\|\nabla\Psi^N(P^N x)\| \le M_3(1+\|x\|_s)$$

and therefore

$$\|C^{1/2}\nabla\Psi^N(P^N x)\| \le M_3\lambda_1(1+\|x\|_s) \tag{8.15}$$

uniformly in N. Also since under π_0, x is Gaussian, from equation (2.5), $C^{-1/2}(P^N x) = \sum_{k=1}^N \rho_k\phi_k$, where ρ_k are iid No$(0,1)$. Note that

$$\begin{aligned}
\frac{1}{N}\|\zeta\|^2 &= \frac{1}{N}\|C^{-\frac{1}{2}}(P^N x) + C^{\frac{1}{2}}\nabla\Psi^N(P^N x)\|^2\\
&= \frac{1}{N}\Big(\|C^{-\frac{1}{2}}(P^N x)\|^2 + 2\langle C^{-\frac{1}{2}}(P^N x), C^{\frac{1}{2}}\nabla\Psi^N(P^N x)\rangle + \|C^{\frac{1}{2}}\nabla\Psi^N(P^N x)\|^2\Big)\\
&= \frac{1}{N}\Big(\|C^{-\frac{1}{2}}(P^N x)\|^2 + 2\langle P^N x, \nabla\Psi^N(P^N x)\rangle + \|C^{\frac{1}{2}}\nabla\Psi^N(P^N x)\|^2\Big)\\
&= \frac{1}{N}\sum_{k=1}^N \rho_k^2 + \gamma
\end{aligned} \tag{8.16}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

where

$$|\gamma| \leq \frac{1}{N}\Big(2\|x\|\|\nabla\Psi^N(P^N x)\| + \|C^{\frac{1}{2}}\nabla\Psi^N(P^N x)\|^2\Big)$$
$$\leq \frac{M}{N}\Big(2\|x\|(1+\|x\|_s) + (1+\|x\|_s)^2\Big) \tag{8.17}$$

Under π_0, we have $\|x\| < \infty$ a.s., and hence by equation (8.17), under π_0

$$\lim_{N\to\infty} |\gamma| = 0, \text{ a.s.}$$

Now, by the Strong law of large numbers, $\frac{1}{N}\sum_{k=1}^N \rho_k^2 \to 1$ almost surely. Hence from equation (8.16) we obtain that under π_0, $\lim_{N\to\infty} \frac{1}{N}\|\zeta\|^2 = 1$ almost surely, proving equation (6.3). Now equation (6.4) follows by noting that almost sure limits are preserved under a (absolutely continuous) change of measure. Next notice that by equation (8.16) and Cauchy-Schwartz inequality, for any $c > 0$:

$$(\mathbb{E}^{\pi_0} e^{c\frac{1}{N}\|\zeta\|^2})^2 \leq \Big(\mathbb{E}^{\pi_0} e^{2\frac{c}{N}\sum\rho_k^2}\Big)\Big(\mathbb{E}^{\pi_0} e^{2c\gamma}\Big)$$
$$= \Big(\mathbb{E}^{\pi_0} e^{2\frac{c}{N}\sum\rho_k^2}\Big)\Big(\mathbb{E}^{\pi_0} e^{\frac{M}{N}\|x\|_s^2}\Big)$$

for sufficiently large N

$$\leq M e^{-\frac{N}{2}\log(1-\frac{4c}{N})}\Big(\mathbb{E}^{\pi_0} e^{\frac{M}{N}\|x\|_s^2}\Big) \leq M$$

where the last inequality follows from Fernique's theorem since

$$\Big(\mathbb{E}^{\pi_0} e^{\frac{M}{N}\|x\|^2)}\Big) < \infty$$

for sufficiently large N. Hence it follows that $\limsup_{N\to\infty} \mathbb{E}^{\pi^0} e^{c\frac{1}{N}\|\zeta\|^2} < \infty$ and therefore

$$\limsup_{N\to\infty} \mathbb{E}^{\pi^N} e^{c\frac{1}{N}\|\zeta\|^2} < \infty$$

by applying Lemma 4.4, equation (4.7). Hence we have verified (6.6). Similarly, a straightfoward calculation yields that

$$\mathbb{E}^{\pi^0}(\Big|1 - \frac{1}{N}\|\zeta\|^2\Big|^2) \leq M\frac{1}{N}$$

hence again by Lemma 4.4

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}(\Big|1 - \frac{1}{N}\|\zeta\|^2\Big|^2) = 0$$

verifying (6.5). Hence we have proved the lemma. □

Proof of Lemma 6.3: Our argument follows the standard route for obtaining Berry-Essen bounds in Wasserstein distance. Let $h(x) : \mathbb{R} \mapsto \mathbb{R}$ be a Lipschitz function. Then by Stein's

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

lemma (as explained in (6.8),(6.9)), there exists a function f with $\|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty \le M$, such that

$$\mathbb{E}(f'(W) - Wf(W)) = \mathbb{E}(h(W) - \mathbb{E}h(Z)) \tag{8.18}$$

Now

$$\begin{aligned}\mathbb{E}^\xi(Wf(W)) &= \mathbb{E}^\xi\Big[\frac{1}{\sqrt{2\ell^2}}(R(x,\xi\,) + \ell^2)f(W)\Big] \\ &= \mathbb{E}^\xi\Big[-\frac{1}{\sqrt{N}}\sum_{j=1}^N \zeta_j\xi_j f(W)\Big] + \gamma_1 \qquad (8.19)\\ |\gamma_1| &\le M\|f\|_\infty \mathbb{E}^\xi\Big[|(1-\frac{1}{N}\sum_{j=1}^N {\xi_j}^2)|\Big] \le \frac{M}{\sqrt{N}}. \qquad (8.20)\end{aligned}$$

The distribution of W remains unchanged if replace ξ_j by $-\xi_j$, but simplifies our calculation and we make this change. Let us introduce $V_j \equiv W - \frac{1}{\sqrt{N}}\zeta_j\xi_j$. Continuing from equation (8.19) and observing that V_j is independent of ξ_j and $\xi_j \sim \mathrm{No}(0,1)$,

$$\begin{aligned}\mathbb{E}^\xi[\frac{1}{\sqrt{N}}\zeta_j\xi_j f(W)] &= \mathbb{E}^\xi\Big[\frac{1}{\sqrt{N}}\zeta_j\xi_j\Big(f(W) - f(V_j)\Big)\Big] \\ &= \frac{1}{\sqrt{N}}\mathbb{E}^\xi\Big[\zeta_j\xi_j\Big(f(W) - f(V_j) - f'(V_j)(W-V_j)\Big)\Big] + \frac{1}{N}\mathbb{E}^\xi{\zeta_j}^2{\xi_j}^2 f'(V_j)\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}^\xi[Wf(W)] &= \frac{1}{N}\sum_{j=1}^N \mathbb{E}^\xi{\zeta_j}^2 f'(V_j) + \gamma_2 + \gamma_1 \\ |\gamma_2| &\le \Big|\sum_{j=1}^N \frac{1}{\sqrt{N}}\mathbb{E}^\xi\Big[\zeta_j\xi_j\Big(f(W) - f(V_j) - f'(V_j)(W-V_j)\Big)\Big]\Big| \\ &\le \|f''\|_\infty \frac{1}{N^{3/2}}\sum_{j=1}^N |\zeta_j|^3 \qquad (8.21)\end{aligned}$$

Now,

$$\begin{aligned}\Big|\mathbb{E}^\xi\Big(f'(W) - Wf(W)\Big)\Big| &= \Big|\frac{1}{N}\sum_{j=1}^N \mathbb{E}^\xi\,{\zeta_j}^2\Big(f'(W) - f'(V_j)\Big) - \\ &\quad \mathbb{E}^\xi\Big(Wf(W) - \frac{1}{N}\sum_{j=1}^N {\zeta_j}^2\,f'(V_j)\Big) + \mathbb{E}^\xi f'(W)(1-\frac{\|\zeta\|^2}{N})\Big| \\ &\le \Big|\frac{1}{N}\sum_{j=1}^N \mathbb{E}^\xi\Big(f'(W) - f'(V_j)\Big)\Big| + \Big|\frac{1}{N}\sum_{j=1}^N \mathbb{E}^\xi\Big(Wf(W) - f'(V_j)\Big)\Big| + M\Big|1 - \frac{1}{N}\|\zeta^k\|^2\Big| \\ &\le \frac{1}{N^{3/2}}\sum_{j=1}^N |\zeta|^3 + M\Big|1-\frac{1}{N}\|\zeta\|^2\Big| + M\frac{1}{\sqrt{N}} \qquad (8.22)\end{aligned}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

where the estimate for bounding the second term in the right hand side comes from (8.21) and we are done. □

Proof of Lemma 7.2 :

$$\begin{aligned}
N\,\mathbb{E}_0(x_i^1 - x_i^0) &= N\,\mathbb{E}_0\Big(\gamma^0(y_i^0 - x_i)\Big) \\
&= N\,\mathbb{E}_0^{\xi}\Big(\alpha(x,\xi\,)\sqrt{\frac{2\ell^2}{N}}(C^{1/2}\xi)_i\Big) \\
&= \lambda_i\,\sqrt{2\ell^2 N}\,\mathbb{E}_0^{\xi}\Big(\alpha(x,\xi\,)\,\xi_i\Big) \\
&= \lambda_i\,\sqrt{2\ell^2 N}\,\mathbb{E}_0^{\xi}\Big((1\wedge e^{Q(x,\xi)})\xi_i\Big)
\end{aligned}$$

Now we write

$$\mathbb{E}_0^{\xi}\Big((1\wedge e^{Q(x,\xi)})\xi_i\Big) = \mathbb{E}_0^{\xi}\Big((1\wedge e^{R_i(x,\xi)-\sqrt{\frac{2\ell^2}{N}}\xi_i\zeta_i})\xi_i\Big) + \omega_0(i)$$

By Lemma 6.1 and noticing that $1\wedge e^x$ is Lipshcitz,

$$\begin{aligned}
|\omega_0(i)| &\le M\lambda_i\mathbb{E}_0^{\xi}\Big|\Big((1\wedge e^{Q(x,\xi)}) - (1\wedge e^{R_i(x,\xi)-\sqrt{\frac{2\ell^2}{N}}\xi_i\zeta_i})\Big)\,\xi_i\Big| \\
&\le M\lambda_i\Big[\mathbb{E}_0^{\xi}\Big|Q(x,\xi\,) - R_i(x,\xi\,)\Big|^2 + \frac{1}{N^2}\Big]^{1/2}\Big[\mathbb{E}_0(\xi_i)^2\Big]^{1/2} \le \frac{M}{\sqrt{N}}\lambda_i\sqrt{1+|\zeta_i|}
\end{aligned}$$

proving the lemma. □

Proof of lemma 7.3: We have,

$$\begin{aligned}
\mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi)+\frac{\ell^2\zeta_i{}^2}{N}}\,\Phi\Big(\frac{R_i(x,\xi\,)}{\sqrt{2\delta}|\zeta_i|} - \sqrt{2\delta}|\zeta_i|\Big) &= \mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi)+\frac{\ell^2\zeta_i{}^2}{N}}\,\Phi\Big(\frac{-R_i(x,\xi\,)}{\sqrt{2\delta}|\zeta_i|}\Big) + \omega_1(i) \\
|\omega_1(i)| &\le M|\zeta_i|\frac{1}{\sqrt{N}}\mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi)+\frac{\ell^2\zeta_i{}^2}{N}} \\
&\le M|\zeta_i|\frac{1}{\sqrt{N}}e^{\frac{\ell^2}{N}\|\zeta\|^2} \qquad (8.23)
\end{aligned}$$

The estimate leading to (8.23) follows from the fact that Φ is globally Lipschitz and:

$$\begin{aligned}
\mathbb{E}_0^{\xi_i^-}\; e^{R_i(x,\xi)+\frac{\ell^2\zeta_i{}^2}{N}} &= \mathbb{E}_0^{\xi_i^-}\;(e^{-\sqrt{\frac{2\ell^2}{N}}\sum_{j=1,j\neq i}^N\zeta_j\xi_j - \frac{1}{N}\sum_{j=1,j\neq i}^N{\xi_j}^2+\frac{\ell^2}{N}{\zeta_i}^2}) \\
&\le \mathbb{E}_0^{\xi_i^-}\;(e^{-\sqrt{\frac{2\ell^2}{N}}\sum_{j=1,j\neq i}^N\zeta_j\xi_j+\frac{\ell^2}{N}{\zeta_i}^2}) = e^{\frac{\ell^2}{N}\|\zeta\|^2} \qquad (8.24)
\end{aligned}$$

proving the lemma. □

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Lemma 8.8. *Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and CDF of the standard normal distribution respectively. For any $x > 0$ and $\epsilon \geq 0$,*

$$1 - \Phi(x) \leq \frac{1+\epsilon}{x+\epsilon}$$

Proof.

$$\begin{aligned} 1 - \Phi(x) &= \int_x^\infty \phi(u)du \\ &\leq \int_x^\epsilon \frac{u+\epsilon}{x+\epsilon}\,\phi(u)\,du \\ &\leq \frac{\phi(x)+\epsilon}{x+\epsilon} \leq \frac{1+\epsilon}{x+\epsilon} \end{aligned}$$

proving the lemma. □

Proof of Lemma 7.4: We have

$$|\omega_2(i)| \leq \mathbb{E}_0^{\xi_i^-}\left[e^{R_i(x,\xi)+\frac{\ell^2}{N}\zeta_i{}^2}\left|1_{R_i(x,\xi)<0} - \Phi\Big(\frac{-R_i(x,\xi\,)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|}\Big)\right|\right] \tag{8.25}$$

Notice that

$$\begin{aligned} \left|1_{R_i(x,\xi)<0} - \Phi\Big(\frac{-R_i(x,\xi\,)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|}\Big)\right| &= 1 - \Phi\left(\frac{|R_i(x,\xi\,)|}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|}\right) \\ &= 1 - \Phi\left(\frac{|R_i(x,\xi\,)|\sqrt{N}}{\sqrt{2}\ell|\zeta_i|}\right) \end{aligned}$$

By applying Lemma 8.8 with $\epsilon = \frac{1}{\sqrt{2}\ell|\zeta_i|}$,

$$\leq (1+\sqrt{2}\ell|\zeta_i|)\frac{1}{1+|R_i(x,\xi\,)|\sqrt{N}} \tag{8.26}$$

Remark 8.9. *Notice that the usual Mill's ratio for Gaussian random variables imply the above bound when we apply lemma 8.8 with $\epsilon = 0$. However in that case the bound in* (8.26) *will be of the form*

$$\left|1_{R_i(x,\xi)<0} - \Phi\Big(\frac{-R_i(x,\xi\,)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|}\Big)\right| \leq M\frac{1}{|R_i(x,\xi\,)|\sqrt{N}}$$

and since $R_i(x,\xi\,)$ is approximately Gaussian, the right hand side of the above bound will not have moments. However, the generalized version of the Mill's ratio derived in lemma 8.8 gets around this difficulty.

The right hand side of the estimate (8.26) depends on i but we need estimates which are independent of i. The main idea here is that the above estimate holds true if $R_i(x,\xi\,)$ is

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

replaced by $R(x,\xi)$ with the extra error term (due to the replacement of $R_i(x,\xi)$ by $R(x,\xi)$) having a bound independent of i. Indeed,

$$\begin{aligned}
\frac{1}{1+|R_i(x,\xi)|\sqrt{N}} &= \frac{1}{1+|R(x,\xi)|\sqrt{N}} + \gamma \\
\mathbb{E}_0^{\xi}|\gamma| &\leq \left|\frac{1}{1+|R_i(x,\xi)|\sqrt{N}} - \frac{1}{1+|R(x,\xi)|\sqrt{N}}\right| \\
&\leq \mathbb{E}_0^{\xi}\frac{\sqrt{N}|R(x,\xi)-R_i(x,\xi)|}{(1+|R_i(x,\xi)|\sqrt{N})(1+|R(x,\xi)|\sqrt{N})} \\
&\leq \mathbb{E}_0^{\xi}\frac{\sqrt{2}\ell|\zeta_i||\xi_i| + \frac{1}{\sqrt{N}}{\xi_i}^2}{(1+|R_i(x,\xi)|\sqrt{N})(1+|R(x,\xi)|\sqrt{N})} \\
&\leq \mathbb{E}_0^{\xi}\frac{\sqrt{2}\ell|\zeta_i||\xi_i| + \frac{1}{\sqrt{N}}{\xi_i}^2}{(1+|R(x,\xi)|\sqrt{N})} \\
&\leq M({\zeta_i}^2 + \frac{1}{N})^{1/2}\Big[\mathbb{E}_0^{\xi}\frac{1}{(1+|R(x,\xi)|\sqrt{N})^2}\Big]^{1/2} \\
&\leq M(|\zeta_i|+1)\Big[\mathbb{E}_0^{\xi}\frac{1}{(1+|R(x,\xi)|\sqrt{N})^2}\Big]^{1/2} \qquad (8.27)
\end{aligned}$$

where in the last inequality the constant M is possibly bigger than the one appearing in the penultimate one. Now from equation (8.25), by Cauchy Schwartz we obtain that,

$$\begin{aligned}
|\omega_2(i)| &\leq \mathbb{E}_0^{\xi_i^-}\left[e^{R_i(x,\xi)+\frac{\ell^2}{N}{\zeta_i}^2}\left|1_{R_i(x,\xi)<0} - \Phi\Big(\frac{-R_i(x,\xi)}{\sqrt{\frac{2\ell^2}{N}}|\zeta_i|}\Big)\right|\right] \\
&\leq \mathbb{E}_0^{\xi_i^-}\left[e^{2R_i(x,\xi)+2\frac{\ell^2}{N}{\zeta_i}^2}\right]^{1/2}\left[\mathbb{E}_0^{\xi_i^-}\left|1_{R_i(x,\xi)<0} - \Phi\Big(\frac{-R_i(x,\xi)}{\sqrt{2\delta}|\zeta_i|}\Big)\right|\right]^{1/2}
\end{aligned}$$

Now by applying the estimates obtained in (8.24), (8.26) and (8.27), we obtain

$$\leq Me^{\frac{\ell^2}{N}\|\zeta\|^2}(|\zeta_i|+1)^2\Big[\mathbb{E}_0^{\xi}\frac{1}{(1+|R(x,\xi)|\sqrt{N})^2}\Big]^{1/2}$$

proving the lemma. □

Proof of Lemma 7.5: This is the lemma where the estimates obtained from using Stein's method are going to be useful. Let W_i be as defined in equation (6.11). Set $g(x) \equiv e^{\sqrt{2}x-1}1_{x\leq 1}$ (we set $\ell = 1$ in this proof as it doesn't affect the bounds). we first need to estimate the following :

$$|\mathbb{E}(g(W_i) - g(Z))|,\ Z \sim \mathrm{No}(0,1)$$

Notice that the function $g(\cdot)$ is not Lipscitz and therefore the Wasserstein bounds obtained earlier cannot be used directly. However we use the fact that the Normal distribution has a density which is bounded above. So by Lemma 6.3, (6.14) and (6.16),

$$\mathrm{KS}(W_i,Z) \leq 2\sqrt{\mathrm{Wass}(W_i,Z)} \leq M\sqrt{\frac{1+|\zeta_i|}{\sqrt{N}} + \frac{1}{N^{3/2}}\sum_{j=1}^{N}|\zeta_j|^3 + |(1-\frac{1}{N}\|\zeta\|^2)|}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Since g is positive on $(-\infty, 1]$, for a real valued random continuous variable X,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{1} g'(t)\Big(\mathbb{P}(X > t)\Big)dt + g(1)\mathbb{P}(X \geq 1)$$

Hence,

$$\begin{aligned}|\mathbb{E}(g(W_i) - g(Z))| &\leq \Big|\int_{-\infty}^{1} g'(t)\Big(\mathbb{P}(W_i > t) - \mathbb{P}(Z > t)\Big)dt\Big| + g(1)|\mathbb{P}(W_i > 1) - \mathbb{P}(Z > 1)| \\ &\leq \mathrm{KS}(W_i, Z)(\int_{-\infty}^{1} g'(t)dt + g(1)) \leq M\mathrm{KS}(W_i, Z)\end{aligned}$$

Hence putting the above calculations together and noticing that $\mathbb{E}(e^{\sqrt{2}Z-1}1_{Z<\frac{1}{\sqrt{2}}}) = \beta/2$, we have just shown that

$$\Big|\mathbb{E}_0^{\xi}(e^{R_i(x,\xi)}1_{R_i(x,\xi)<0}) - \frac{\beta}{2}\Big| \leq M\sqrt{\frac{1+|\zeta_i|}{\sqrt{N}} + \frac{1}{N^{3/2}}\sum_{j=1}^{N}|\zeta_j|^3 + |(1 - \frac{1}{N}\|\zeta\|^2)|}$$

Notice that

$$\begin{aligned}|\omega_3(i)| &\leq | e^{\ell^2\frac{\zeta_i^2}{N}}\mathbb{E}_0^{\xi}(e^{R_i(x,\xi)}1_{R_i(x,\xi)<0}) - \beta/2| \\ &\leq |e^{\ell^2\frac{\zeta_i^2}{N}} - 1| + |\mathbb{E}_0^{\xi}(e^{R_i(x,\xi)}1_{R_i(x,\xi)<0}) - \beta/2| \\ &\leq M\frac{\zeta_i^2}{N}e^{\ell^2\frac{\|\zeta\|^2}{N}} + |\mathbb{E}_0^{\xi}(e^{R_i(x,\xi)}1_{R_i(x,\xi)<0}) - \beta/2|\end{aligned}$$

finishing the proof. □

Proof of Lemma 7.7: Indeed,

$$\begin{aligned}\lambda_i\zeta_i &= \lambda_i\Big\langle C^{-1/2}(P^N x) + C^{1/2}\nabla\Psi^N(P^N x), \phi_i\Big\rangle \\ &= \lambda_i\Big\langle C^{-1/2}(P^N x) + C^{-1/2}C\nabla\Psi^N(P^N x), \phi_i\Big\rangle\end{aligned}$$

Since $C^{-1/2}$ is self adjoint and $i \leq N$, we have $\lambda_i C^{-1/2}\phi_i = \phi_i$ and thus

$$= \Big\langle P^N x + C\nabla\Psi^N(P^N x), \phi_i\Big\rangle$$

and we are done. □

Proof of Lemma 7.8: By (7.7),

$$|r_i^N| \leq | \omega_0(i)| + M\lambda_i|\zeta_i|(|\omega_1(i)| + |\omega_2(i)| + |\omega_3(i)| + |\omega_4(i)|)$$

Therefore

$$\mathbb{E}^{\pi_0}\sum_{i=1}^{N}|r_i^N|^2 \leq M\mathbb{E}^{\pi_0}\sum_{i=1}^{N}\Big(|\omega_0(i)|^2 + \lambda_i^2{\zeta_i}^2(|\omega_1(i)|^2 + |\omega_2(i)|^2 + |\omega_3(i)|^2 + |\omega_4(i)|^2)\Big)$$

Before proceeding, let us first record the following lemma which will be used repeatedly.

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Lemma 8.10. *For any $m \in \mathbb{N}$, $\alpha \geq 2$ and for any $c \geq 0$,*

$$\limsup_{N\to\infty} \mathbb{E}^{\pi^N} \sum_{i=1}^{N} \lambda_i^{\alpha} \, |\zeta_i|^m \, e^{c\frac{1}{N}\|\zeta\|^2} < \infty.$$

Proof. If $x \sim \pi_0$, then as in (2.5) we have,

$$x = \sum_{i=1}^{\infty} \lambda_i \rho_i \phi_i, \;\; \rho_i \overset{\text{iid}}{\sim} \text{No}(0,1)$$

From (5.2) and (4.3), we have the bound $|\zeta_i| \leq \lambda_i^{-1}|x_i| + (1+\|x\|_s)\lambda_i \leq |\rho_i| + (1+\|x\|_s)\lambda_i$. Thus for any for any m, and for any $N \in \mathbb{N}$,

$$\begin{aligned}
\sum_{i=1}^{N} \mathbb{E}^{\pi_0} \lambda_i^{\alpha} |\zeta_i|^m e^{c\frac{1}{N}\|\zeta\|^2} &\leq \sum_{i=1}^{N} \Big(\mathbb{E}^{\pi_0} \lambda_i^{2\alpha} |\zeta_i|^{2m}\Big)^{1/2} \Big(\mathbb{E}^{\pi_0} e^{2c\frac{1}{N}\|\zeta\|^2}\Big)^{1/2} \\
&\leq M \Big(\mathbb{E}^{\pi_0} e^{2c\frac{1}{N}\|\zeta\|^2}\Big)^{1/2} \sum_{i=1}^{N} \Big(\lambda_i^{\alpha} (\mathbb{E}^{\pi_0}|\rho_i|^{2m})^{1/2} + \lambda_i^{\alpha+m} (\mathbb{E}^{\pi_0}(1+\|x\|_s)^{2m})^{1/2}\Big) \\
&\leq M (\sum_{i=1}^{\infty} \lambda_i^{\alpha} + \sum_{i=1}^{\infty} \lambda_i^{\alpha+m}) \leq M
\end{aligned}$$

where in the penultimate step we have used the estimate (6.6) and the last step follows from the fact that $\sum_{i=1}^{\infty} \lambda_i^{\alpha} < \infty$ for $\alpha \geq 2$, since the covariance operator C is trace class. Thus we have

$$\limsup_{N\to\infty} \mathbb{E}^{\pi^N} \sum_{i=1}^{N} \lambda_i^{\alpha} |\zeta_i|^m e^{c\frac{1}{N}\|\zeta\|^2} \leq M \limsup_{N\to\infty} \mathbb{E}^{\pi_0} \sum_{i=1}^{N} \lambda_i^{\alpha} |\zeta_i|^m e^{c\frac{1}{N}\|\zeta\|^2} < \infty$$

and we are done. □

Now we proceed to the proof of Lemma 7.8. We have,

$$\begin{aligned}
\sum_{i=1}^{N} \mathbb{E}^{\pi^N} |\omega_0(i)|^2 &\leq M \frac{1}{N} \mathbb{E}^{\pi^N} \sum_{i=1}^{N} \lambda_i^2 (1+|\zeta_i|) \qquad (8.28) \\
&\leq M \frac{1}{N} (\sum_{i=1}^{\infty} \lambda_i^2 + \mathbb{E}^{\pi^N} \lambda_i^2 |\zeta_i|) \to 0
\end{aligned}$$

where the last estimate follows from Lemma 8.10.

We now show that $\lim_{N\to\infty} \mathbb{E}^{\pi^N} \sum_{i=1}^{N} \lambda_i^2 \zeta_i^2 |\omega_j(i)|^2 = 0$, for $j = 1,2,3,4$. By equation (7.3) and Lemma 8.10,

$$\mathbb{E}^{\pi^N} \sum_{i=1}^{N} \lambda_i^2 \zeta_i^4 |\omega_1(i)|^2 \leq M \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}^{\pi^N} \lambda_i^2 |\zeta_i|^4 e^{\frac{2\ell^2}{N}\|\zeta\|^2} \to 0.$$

From equation (7.4) and Cauchy-Schwartz, we obtain

$$\sum_{i=1}^{N} \mathbb{E}^{\pi^N} \lambda_i^2 |\zeta_i|^2 |\omega_2(i)|^2 \leq M \Big(\mathbb{E}^{\pi^N}\Big[\mathbb{E}_0^{\xi} \frac{1}{(1+|R(x,\xi)|\sqrt{N})^2}\Big]^2\Big)^{1/2} \sum_{i=1}^{N} \Big(\mathbb{E}^{\pi^N} e^{\frac{4\ell^2}{N}\|\zeta\|^2} \lambda_i^4 (|\zeta_i|^8 + |\zeta_i|^4)\Big)^{1/2}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Proceeding similarly as done in Lemma 8.10 it follows that

$$\sum_{i=1}^{N}\Big(\mathbb{E}^{\pi^N} e^{\frac{4\ell^2}{N}\|\zeta\|^2}\lambda_i^4(|\zeta_i|^8+|\zeta_i|^4)\Big)^{1/2}<\infty. \tag{8.29}$$

Since, with $x\sim\pi_0$, $R(x,\xi)$ converges weakly to Z_ℓ as $N\to\infty$, by the bounded convergence theorem we obtain

$$\lim_{N\to\infty}\mathbb{E}^{\pi_0}\Big[\mathbb{E}_0^{\xi}\frac{1}{(1+|R(x,\xi)|\sqrt{N})^2}\Big]^2=0$$

and therefore

$$\lim_{N\to\infty}\sum_{i=1}^{N}\mathbb{E}^{\pi^N}|\zeta_i|^2\lambda_i^2|\omega_2(i)|^2=0.$$

After some algebra we obtain,

$$\begin{aligned}\mathbb{E}^{\pi^N}\sum_{i=1}^{N}\lambda_i^2|\zeta_i|^2|\omega_3(i)|^2 &\le M\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}^{\pi^N}\lambda_i^2|\zeta_i|^6e^{2\ell^2\frac{\|\zeta\|^2}{N}}+M\frac{1}{\sqrt{N}}\mathbb{E}^{\pi^N}\sum_{i=1}^{N}\lambda_i^2\zeta_i^2\Big(1+|\zeta_i|\Big)\\ &+M\Big[\Big(\mathbb{E}^{\pi^N}(\frac{1}{N^{3/2}}\sum_{j=1}^{N}|\zeta_j|^3)^2\Big)^{1/2}+\Big(\mathbb{E}^{\pi^N}|(1-\frac{1}{N}\|\zeta\|^2)|^2\Big)^{1/2}\Big]\sum_{i=1}^{N}\Big(\mathbb{E}^{\pi^N}\lambda_i^4\zeta_i^4\Big)^{1/2}\end{aligned}$$

Similar to the previous calculations using lemma 8.10, it can be shown that the first two terms above converge to 0. By Lemma 6.2, equation (6.5), we have $\mathbb{E}^{\pi^N}|(1-\frac{1}{N}\|\zeta\|^2)|^2\to 0$. Also using similar arguments as that of Lemma 8.10, it can be shown that $\mathbb{E}^{\pi^N}(\frac{1}{N^{3/2}}\sum_{j=1}^{N}|\zeta_j|^3)^2\to 0$. Thus we obtain, $\lim_{N\to\infty}\sum_{i=1}^{N}\mathbb{E}^{\pi^N}\lambda_i^2|\zeta_i|^2|\omega_3(i)|^2=0$.

Now for the last term, using (4.6) we have

$$\mathbb{E}^{\pi^N}\sum_{i=1}^{N}\lambda_i^2|\zeta_i|^2|\omega_4(i)|^2\le M_8\theta(N)\Big(\mathbb{E}^{\pi^N}\|x\|_s^4\Big)^{1/2}\sum_{i=1}^{N}\Big(\mathbb{E}^{\pi^N}\lambda_i^4|\zeta_i|^4\Big)^{1/2}\to 0$$

since $\theta(N)\to 0$, and we are done. □

Proof of Lemma 7.10: We have

$$\begin{aligned}|\kappa_{ij}| &\le \mathbb{E}_0^{\xi}\Big[\Big|(C^{1/2}\xi)_i(C^{1/2}\xi)_j\Big(\{1\wedge e^{Q(x,\xi)}\}-\{1\wedge e^{R_{ij}(x,\xi)}\}\Big)\Big|\Big]\\ &\le M\lambda_i\lambda_j\,\mathbb{E}_0^{\xi}\Big[\Big|\xi_i\xi_j\Big(\{1\wedge e^{Q(x,\xi)}\}-\{1\wedge e^{R_{ij}(x,\xi)}\}\Big)\Big|\Big]\end{aligned}$$

By Cauchy Schwartz inequality

$$\le M\lambda_i\lambda_j\Big(\mathbb{E}_0^{\xi}|(1\wedge e^{Q(x,\xi)})-(1\wedge e^{R_{ij}(x,\xi)})\Big)^{1/2}$$

$$\le M\lambda_i\lambda_j\Big(\mathbb{E}_0^{\xi}|Q(x,\xi)-R_{ij}(x,\xi)|^2\Big)^{1/2}$$

using the estimate obtained in equation (6.2),

$$\le M\lambda_i\lambda_j(1+|\zeta_i|^2+|\zeta_j|^2)^{1/2}\frac{1}{\sqrt{N}}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

proving the lemma. $\square$

Proof of Lemma 7.11: As in the proof of Lemma 7.5, set $W_{ij} = \frac{1}{\sqrt{2}\ell}(R_{ij}(x,\xi) + \ell^2)$. Here also we set $\ell^2 = 1$ without loss of generality. We need to bound:

$$\mathbb{E}(g(W_{ij}) - g(Z)),\ Z \sim \mathrm{No}(0,1)$$

where $g(x) \equiv 1 \wedge e^{\sqrt{2x}-1}$. Notice that $\mathbb{E}(g(Z)) = 2\beta$. Since $g(\cdot)$ is lipschitz,

$$|\mathbb{E}(g(W_{ij}) - g(Z))| \leq M\mathrm{Wass}(W_{ij}, Z) \tag{8.30}$$

Recall that $W \equiv \frac{1}{\sqrt{2}}(R(x,\xi) + 1)$. A simple calculation will yield that

$$\mathrm{Wass}(W_{ij}, W) \leq M(|\zeta_i| + |\zeta_j| + 1)\frac{1}{\sqrt{N}}$$

Therefore by the triangle inequality and Lemma 6.3,

$$\mathrm{Wass}(W_{ij}, W) \leq M\Big(\frac{1}{\sqrt{N}}(1 + |\zeta_i| + |\zeta_j|) + \frac{1}{N^{3/2}}\sum_{s=1}^{N}|\zeta_s|^3 + \Big|1 - \frac{1}{N}\|\zeta\|^2\Big|\Big)$$

Hence the result follows from the observation made in equation (8.30). $\square$

Proof of Lemma 7.12: By definion:

$$\|E^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} = \Big\{\sup_a \Big(\sum_{i=1}^{N}\Big|\sum_{j=1}^{N} E^N_{ij} a_j\Big|^2\Big)^{1/2},\ \sum_{j=1}^{N} a_j^2 = 1\Big\}$$

By Cauchy-Schwartz

$$\|E^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} \leq (\sum_{i=1}^{N}\sum_{j=1}^{N}|u^N_{ij}|^2)^{1/2}$$

and noting that $|E^N_{ij}|^2 \leq 2(|\kappa^2_{ij} + \lambda_i^2\lambda_j^2\,\delta_{ij}\,\rho^2_{ij}|)$

$$\sum_{i=1}^{N}\sum_{j=1}^{N}|u^N_{ij}|^2 \leq 2\sum_{i=1}^{N}\sum_{j=1}^{N}|\kappa_{ij}|^2 + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i^2\lambda_j^2\,\delta_{ij}\,\rho_{ij}|^2$$

$$\begin{aligned}
\mathbb{E}^{\pi^N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\kappa_{ij}|^2 &\leq M\mathbb{E}^{\pi_0}\sum_{i=1}^{N}\sum_{j=1}^{N}|\kappa_{ij}|^2 \\
&\leq M\mathbb{E}^{\pi_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i^2\lambda_j^2(1 + |\zeta_i|^4 + |\zeta_j|^4)\frac{1}{N} \\
&\leq M\frac{1}{N}(1 + \mathbb{E}^{\pi_0}\sum_{i=1}^{N}\lambda_i^2|\zeta_i|^4) \\
&\to 0
\end{aligned}$$

as done in the calculations in lemma 7.8. Now the second term of E^N:

$$\mathbb{E}^{\pi^N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\lambda_i^2\lambda_j^2\,\delta_{ij}\,\rho_{ij}|^2 \leq M\mathbb{E}^{\pi_0}\sum_{i=1}^{N}\lambda_i^2\Big(\frac{1}{N}(1 + |\zeta_i|^2) + \Big(\frac{1}{N^{3/2}}\sum_{s=1}^{N}|\zeta_s|^3)^2 + \Big|1 - \frac{1}{N}\|\zeta\|^2\Big|^2\Big)$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

The first term above goes to zero just by the arguments done for κ_{ij}, the second term goes to zero by the same arguments for the error term $\omega_3(i)$ in Lemma 7.8, and the last term goes to zero by equation (6.6). Therefore we have shown that:

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N} \|E^N\|_{\mathcal{L}(\mathcal{H},\mathcal{H})} = 0$$

and hence the lemma is proved. □

Proof of Lemma 7.13: First notice that for any $x \in \mathcal{H}$, the outer product $x \otimes x$ is a positive operator since for any $u \in \mathcal{H}$, $\langle u, x \otimes x\, u\rangle = \langle x, u\rangle^2 \geq 0$ and thus $\mathrm{tr}(x \otimes x) = \sum_{k=1}^{\infty} \langle \phi_k, x \otimes x\, \phi_k\rangle$. Since $(x^1 - x)$ is N dimensional,

$$\sum_{i=1}^{N} \mathbb{E}_0 |x_i^1 - x_i^0|^2 = \mathrm{tr}(\mathbb{E}_0 (x^1 - x^0) \otimes (x^1 - x^0)) = \sum_{i=1}^{N} \langle \phi_i, \mathbb{E}_0 (x^1 - x^0) \otimes (x^1 - x^0) \phi_i\rangle$$

From (7.9), we obtain

$$\sum_{i=1}^{N} \langle \phi_i, \mathbb{E}_0 (x^1 - x^0) \otimes (x^1 - x^0) \phi_i\rangle = 2\ell^2 \beta \sum_{i=1}^{N} \lambda_i^2 + \sum_{i=1}^{N} |E_{ii}| \tag{8.31}$$

Proceeding similarly as that of Lemma 7.12, it can be shown that $\sum_{i=1}^{N} \mathbb{E}^{\pi^N} |E_{ii}| \to 0$, which together with (8.31) with imply that $\sum_{i=1}^{N} \mathbb{E}^{\pi^N} \langle \phi_i, \mathbb{E}_0 (x^1 - x^0) \otimes (x^1 - x^0) \phi_i\rangle \to 2\ell^2 \beta \mathrm{tr}(C)$ and the lemma is proved. □

Proof of Lemma 8.4: We apply Theorem 8.1 with $k^N(t) \equiv \lfloor Nt \rfloor$ and $X^{i,N} \equiv \frac{1}{\sqrt{N}} \Gamma^{k,N}$. Since the chain is stationary, the noise process $\{\Gamma^{k,N}, 0 \leq k < N\}$ are identically distributed, and so are the errors $r^{k,N}$ and $E^{k,N}$ from Theorems 7.1 and 7.9. We first verify condition 8.5: By stationarity

$$\begin{aligned} \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E}^{\pi^N} (\|\Gamma^{k,N}\|^2) &= \mathbb{E}^{\pi^N} (\|\Gamma^{1,N}\|^2) = \sum_{n=1}^{N} \mathbb{E}^{\pi^N} \langle \Gamma^{1,N}, \phi_n\rangle^2 \\ &= \sum_{n=1}^{N} \mathbb{E}^{\pi^N} \langle \phi_n, \Gamma^{0,N} \otimes \Gamma^{0,N} \phi_n\rangle = \mathbb{E}^{\pi^N} \mathrm{tr}(\Gamma^{0,N} \otimes \Gamma^{0,N}) \end{aligned} \tag{8.32}$$

From equation (8.3) we have

$$\begin{aligned} \sum_{n=1}^{N} \mathbb{E}^{\pi^N} \langle \phi_k, \mathbb{E}_0 (\Gamma^{0,N} \otimes \Gamma^{0,N}) \phi_k\rangle &= \frac{N}{2\ell^2\beta} \sum_{k=1}^{N} \mathbb{E}^{\pi^N} \langle \phi_k, \mathbb{E}_0 (x^1 - x^0) \otimes (x^1 - x^0), \phi_k\rangle - \frac{N}{2\ell^2\beta} \mathbb{E}^{\pi^N} \|\mathbb{E}_0 (x^1 - x^0)\|^2 \\ &= \frac{N}{2\ell^2\beta} \mathbb{E}^{\pi^N} \mathrm{tr}\, \mathbb{E}_0 (x^1 - x^0) \otimes (x^1 - x^0) - \frac{N}{2\ell^2\beta} \mathbb{E}^{\pi^N} \|\mathbb{E}_0 (x^1 - x^0)\|^2 \end{aligned} \tag{8.33}$$

The first term above converges to $\mathrm{tr}(C)$ from Theorem 7.9. For the second term notice that, from Theorem 7.1 we have

$$\begin{aligned} \mathbb{E}^{\pi^N} \frac{N}{2\ell^2\beta} \|\mathbb{E}_0 (x^1 - x^0)\|^2 &\leq MN \Big(\frac{1}{N^2} \mathbb{E}^{\pi^N} \|m(x^0)\|^2 + \frac{1}{N^2} \|r^{1,N}\|^2\Big) \\ &\leq M \frac{1}{N} \Big(\mathbb{E}^{\pi^N} (1 + \|x\|_s)^2 + \mathbb{E}^{\pi^N} \|r^{k,N}\|^2 \Big) \to 0 \end{aligned} \tag{8.34}$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

Hence by (8.32), (8.33) and Markov's inequality, it follows that $\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} \mathbb{E}_k(\|\Gamma^{k,N}\|^2) = \text{trace}(C)$ in probability verifying 8.5.

To verify (8.6), by remark 8.3 it is enough to verify (8.8). Since C is diagonal, $\langle C\phi_n, \phi_m\rangle = \lambda_n^2\delta_{nm}$. To show (8.8), by stationarity it is enough to show that

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}(\langle\Gamma^{0,N}, \phi_n\rangle\langle\Gamma^{0,N}, \phi_m\rangle) = \lambda_n^2\delta_{nm} \tag{8.35}$$

Indeed, $\langle\Gamma^{0,N}, \phi_n\rangle\langle\Gamma^{0,N}, \phi_m\rangle = \langle\phi_n, \Gamma^{0,N}\otimes\Gamma^{0,N}\phi_m\rangle$ and from (8.2), Theorem 7.1 and (7.9) we obtain

$$\mathbb{E}^{\pi^N}\langle\phi_n, \Gamma^{0,N}\otimes\Gamma^{0,N}\phi_m\rangle = \lambda_{nm}^2\delta_{nm} + \mathbb{E}^{\pi^N}E_{ij}^{0,N} - \frac{N}{2\ell^2\beta}\mathbb{E}^{\pi^N}\mathbb{E}_0(x_n^1 - x_n^0)\mathbb{E}_0(x_m^1 - x_m^0) \tag{8.36}$$

As done in the proof of Lemma 7.12, it can be shown that $\lim_{N\to\infty}\mathbb{E}^{\pi^N}E_{ij}^{0,N} = 0$. Also notice that $\mathbb{E}^{\pi^N}|\mathbb{E}_0(x_n^1 - x_n^0)\mathbb{E}_0(x_m^1 - x_m^0)| \le \mathbb{E}^{\pi^N}\|\mathbb{E}_0(x_n^1 - x_n^0)\|^2 \to 0$ by the calculation done in (8.34). Therefore equation (8.35) follows and thus (8.6) follows from Markov's inequality.

To verify equation (8.9): fix $\epsilon > 0$,

$$\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}^{\pi^N}(\|\Gamma^{j,N}\|^2 1_{\{\|\Gamma^{j,N}\|^2\ge\epsilon N\}}) = \mathbb{E}^{\pi^N}(\|\Gamma^{0,N}\|^2 1_{\{\|\Gamma^{0,N}\|^2\ge\epsilon N\}}) \to 0$$

by the dominated convergence theorem since $\lim_{N\to\infty}\mathbb{E}^{\pi^N}\|\Gamma^{0,N}\|^2 = \text{tr}(C) < \infty$.

Thus we have verified all the three hypothesis of Theorem 8.1 and the theorem is proved. □

Proof of Lemma 8.5: First notice that, for a constant C independent of N, but possibly changing from instance to instance,

$$\sup_{t\in[0,T]}\|e^N(t)\|^2 \le C\Big(\sup_{t\in[0,T]}\int_0^t\|r_1^N(s)\|^2 ds + \sup_{t\in[0,T]}\int_0^t\|r_2^N(s)\|^2 ds\Big).$$

Also

$$\begin{aligned}\mathbb{E}^{\pi^N}\sup_{t\in[0,T]}\int_0^t\|r_1^N(s)\|^2 ds &\le \mathbb{E}^{\pi^N}\int_0^T\|r_1^N(s)\|^2 ds\\ &\le C\frac{1}{N}\mathbb{E}^{\pi^N}\sum_{k=1}^{N}\|r^{k,N}\|^2\\ &= C\mathbb{E}^{\pi^N}\|r^{1,N}\|^2 \to 0\end{aligned}$$

where we used stationarity of $r^{k,N}$ and Theorem 7.1 in the last step. We now estimate the second term similarly to complete the proof. Note that the function $z \mapsto z + C\nabla\Psi(z)$ is Lipschitz on $\mathcal{H}$ by (4.4). Thus $\|r_2^N(s)\|^2 \le M\|z^N(s) - \bar{z}^N(s)\|^2$. But for any $s \in [t_k, t_{k+1})$, we have

$$\|z^N(s) - \bar{z}^N(s)\| \le \|x^{k+1} - x^k\|$$

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism

because

$$\bar{z}^N(s) = x^k, \quad z^N(s) = \frac{1}{\Delta t}\Big((s - t_k)x^{k+1} + (t_{k+1} - s)x^k\Big).$$

Thus

$$\|r_2^N(s)\|^2 \le M\|x^{k+1} - x^k\|^2.$$

By Theorem 7.9, we deduce that

$$\mathbb{E}^{\pi^N}\|r_2^N(s)\|^2 \to 0$$

as $N \to \infty$ and hence that

$$\begin{aligned}\mathbb{E}^{\pi^N} \sup_{t\in[0,T]} \int_0^t \|r_2^N(s)\|^2 ds &\le \mathbb{E}^{\pi^N} \int_0^T \|r_2^N(s)\|^2 ds \\ &\le \int_0^T \mathbb{E}^{\pi^N}\|r_2^N(s)\|^2 ds \to 0\end{aligned}$$

and we have proved the lemma. □

CRiSM Paper No. 09-19, www.warwick.ac.uk/go/crism