Spectral Gaps for a Metropolis-Hastings Algorithm in Infinite Dimensions

M. Hairer, A. M. Stuart and S. J. Vollmer

14th February 2012

Abstract

We study the problem of sampling high and infinite dimensional target measures arising in applications such as conditioned diffusions and inverse problems. We focus on those that arise from approximating measures on Hilbert spaces defined via a density with respect to a Gaussian reference measure. We consider the Metropolis-Hastings algorithm that adds an accept-reject mechanism to a Markov chain proposal in order to have the target measure as an ergodic invariant measure. We focus on cases where the proposal is either a Gaussian random walk (RWM) with covariance equal to that of the reference measure or an Ornstein-Uhlenbeck proposal (pCN) for which the reference measure is invariant.

Previous results in terms of scaling and diffusion limits suggested that the pCN has a convergence rate that is independent of the dimension while the RWM method has undesirable dimension-dependent behaviour. We confirm this claim by showing dimension-independent Wasserstein spectral gap for pCN algorithm for a large class of target measures. In our setting this Wasserstein spectral gap implies an L^2-spectral gap. We use both spectral gaps to show that the ergodic average satisfies a strong law of large numbers, the central limit theorem and non-asymptotic bounds on the mean square error, all dimension independent. In contrast we show that the RWM algorithm applied to the reference measures degenerates as the dimension tends to infinity.

1 Introduction

The aim of this article is to study the complexity of certain sampling algorithms in high dimensions. Creating samples from a high dimensional probability distribution is important for Bayesian Inverse Problems [49] and Bayesian Statistics [32]. In Bayesian nonparametrics [19], which have recently become more and more important for applications, these are the main tools for extracting information from the posterior. Last but not least our results are applicable to a certain class of conditioned diffusions [23].

The most widely used method for general target measures are Markov chain Monte Carlo (MCMC) algorithms which run an ergodic Markov chain with the target measure as the invariant measure. Under certain conditions the empirical average of a function f (observable) applied to the steps of the Markov chain converges to the expectation of this function with respect to the target measure. The computational cost of such an algorithm is the product of the cost of one step and the number of steps necessary for a certain level of accuracy. While in most applications the cost of one step grows with dimensionality, a major result of this article is to show that under certain conditions an upper bound on the number of steps which are necessary is independent of the dimension.

For ease of presentation we work on a separable Hilbert space $(\mathcal{H}, \|\cdot\|)$ equipped with a mean-zero Gaussian reference measure γ with covariance operator $\mathcal{C}$. Let $\{\varphi_n\}_{n\in\mathbb{N}}$ be an orthonormal base of eigenvectors of $\mathcal{C}$ corresponding to eigenvalues $\{\lambda_n^2\}_{n\in\mathbb{N}}$. Thus γ can be written as its Karhunen-Loeve Expansion (c.f. [1])

$$\gamma = \mathcal{L}(\sum_{i=1}^{\infty} \lambda_i e_i \xi_i), \qquad \text{where } \xi_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$$

where $\mathcal{L}(\cdot)$ denotes the law of a random variable. The target measure μ is assumed to have a density with respect to γ of the form

$$\mu = M \exp(-\Phi(x))\gamma. \tag{1.1}$$

Gaussian measures have the property that there are always many Hilbert spaces which satisfy $\gamma(H) = 1$. We will assume that $\Phi : \mathcal{H} \to R$ is Lipschitz and that the reference measure γ has the property that $\gamma(H) = 1$. For Bayesian problems this amounts to the choice of prior; for conditioned diffusions it restricts the class of admissible target measures. With P_m the projection onto the first m basis elements we consider the following m-dimensional approximations to γ and μ

$$\begin{aligned} \gamma_m &= \mathcal{L}(\sum_{i=1}^{m} \lambda_i e_i \xi_i) \\ \mu_m &= M_m \exp(-\Phi(P_m x))\gamma_m. \end{aligned} \tag{1.2}$$

The approximation error, namely the difference between μ and μ_m, is already well studied ([15, 10] for example) an can be estimated in terms of the closeness between $\Phi \circ P_m$ and Φ.

In this article we consider Metropolis-Hastings MCMC methods ([36] and [24]). For an overview of other MCMC methods, which have been developed and analyzed, consult [43, 33]. The idea of the Metropolis-Hastings algorithm is to add an independent accept-reject mechanism to a Markov chain proposal in order to have the target measure as an ergodic invariant measure. We denote by $Q(x, dy)$ the transition kernel of the underlying Markov chain and with $\alpha(x, y)$ the acceptance probability for a proposed move from x to y. The transition kernel of the Metropolis-Hastings algorithm reads

$$\mathcal{P}(x, dz) = Q(x, dz)\alpha(x, z) + \delta_x(dz) \int (1 - \alpha(x, u))Q(x, du) \tag{1.3}$$

where $\alpha(x,y)$ is chosen such that $\mathcal{P}(x,dy)$ is reversible with respect to μ. According to [50], one considers $\nu = \mu(dx)Q(x,dy)$ and $\nu^T = \mu(dy)Q(y,dx)$ on a subset where they are mutually absolutely continuous and there one takes $\alpha(x,y) = 1 \wedge r(x,y)$ with $r = \frac{d\nu^T}{d\nu}$; on the complement of this subset $\alpha(x,y) = 0$. A common proposal kernel corresponds to the random walk

$$Q(x,dy) = \mathcal{L}(x + \sqrt{2\delta}\xi)$$

with $\xi \sim \gamma_m$ which leads to the acceptance probability

$$\alpha(x,y) = 1 \wedge \left(\Phi(x) - \Phi(y) + \frac{1}{2}\langle x, \mathcal{C}x\rangle - \frac{1}{2}\langle y, \mathcal{C}y\rangle \right). \tag{1.4}$$

Notice that the quadratic forms $\frac{1}{2}\langle x, \mathcal{C}x\rangle$ and $\frac{1}{2}\langle y, \mathcal{C}y\rangle$ are almost surely infinite in $\mathcal{H}$ since they correspond to the Cameron-Martin norm of x and y respectively. For this reason the RWM algorithm is not defined on the infinite dimensional Hilbert space $\mathcal{H}$ (see [11] for a discussion) and we will study it only on $m-$dimensional approximating spaces. Furthermore, it is intuitive that the algorithms we study will degenerate in some way as the dimension m increases. In this article we will demonstrate that the RWM can be considerably improved upon by using the preconditioned Crank-Nicolson (pCN), which is a well-defined algorithm on $\mathcal{H}$, and corresponds to

$$Q(x,dy) = \mathcal{L}((1-2\delta)^{\frac{1}{2}}x + \sqrt{2\delta}\xi) \tag{1.5}$$

$$\alpha(x,y) = 1 \wedge \exp(\Phi(x) - \Phi(y)) \tag{1.6}$$

with $\xi \sim \gamma$. The pCN was introduced in [5]. Numerical experiments in [11] demonstrate its favorable properties in comparison with the RWM algorithm. Notice that, in contrast to RWM, the acceptance probability is well-defined on Hilbert space and this fact gives an intuitive explanation for the theoretical results we derive in this paper in which we develop a theory which explains the superiority of pCN over RWM. Our main positive results about pCN can be summarised as (rigorous statement in Theorems 2.14, 2.15, 4.2 and 4.4):

Claim. Suppose Φ and its local Lipschitz constant both satisfy a growth assumption at infinity. Then the pCN algorithm applied to $\mu(\mu_m)$

1. has a unique invariant measure μ (μ_m);
2. has a Wasserstein spectral gap uniformly in m;
   (a) has an L^2-spectral gap $1-\beta$ uniform in m;

The corresponding sample average $S_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$

4 satisfies a strong law of large numbers and a central limit theorem (CLT) for a class of locally Lipschitz functionals for every initial condition;

5 For $f \in L^2_\mu$ $\left(L^2_{\mu_m}\right)$ S_n satisfies a CLT for μ (μ_m)-almost every initial condition with asymptotic variance uniformly bounded in m;

6 There is an explicit bound on the mean square error (MSE) between $S_n(f)$ and $\mu(f)$ for certain initial distributions ν.

These positive results about pCN clearly apply for $\Phi = 0$, which corresponds to the target measure γ and γ_m respectively; in this case the acceptance probability of pCN is always one, and the theorems mentioned are simply statements about a discretely sampled Ornstien-Uhlenbeck (OU) process on $\mathcal{H}$ in this case. On the other hand the RWM algorithm applied to the target measure γ_m has an L^2_μ spectral that converges to 0 as $m \to \infty$ as fast as any negative power of m see Theorem 2.17.

While it is a major contribution of this article to verify 1,2,4 and the negative result for the RWM, 3,5 and 6 are consequences of verifying conditions of known results.

In addition to the significance of the results themselves for the understanding of MCMC methods, we would also like to highlight the techniques of proof that we use. We use recently developed tools for the study of Markov chains on infinite dimensional spaces [22] that, for many problems, improve significantly on the machinery that has been used for the study of MCMC methods to date. The weak Harris theorem makes a Wasserstein spectral verifiable in practice and for reversible Markov processes it even implies an L^2-spectral gap. Previous results have been formulated in terms of the following three main types of convergence:

1. For a metric d on the space of measures the convergence rate is given as the decay rate of $d(\nu\mathcal{P}^n, \mu)$, where ν is the initial distribution of the Markov chain. The most prominent examples here are convergence in a (weighted) total variation and in a Wasserstein distance.

2. For the Markov operator $\mathcal{P}$ the convergence rate is given as the operator norm of $\mathcal{P}$ on a space of functions from $\mathcal{H}$ to $\mathbb{R}$ modulo constants. The most prominent example here is the L^2-spectral gap.

3. The (asymptotic) convergence rate of $S_n(f) = \sum_{i=1}^n f(X_i)$ to $\mu(f)$ for a class of functions f in form of a CLT or a MSE bound.

Between these notions of convergence, there are many fruitful relations, see e.g. [46]. All these convergence types have been used to study MCMC algorithms.

The first systematic approach to prove L^2-spectral gaps for Markov chains was developed in [31] using the conductance concept due to Cheeger ([9]). These results were extended and applied to the Metropolis-Hastings algorithm with uniform proposal and a log-concave target distribution on a bounded convex subset of $\mathbb{R}^n$ in [34]. The consequences of a spectral gap for the ergodic average in terms of a CLT and the MSE have been investigated in [26, 12] and [46] respectively and were first brought up in the MCMC literature in [18, 8].

For finite state Markov chains the spectral gap can be bounded in terms of quantities associated with its graph [16] and this idea has also been applied to the Metropolis-Algorithm in [48] and [17].

A different approach using the now called splitting chain technique was independently developed in [38] and [2] to bound the total variation distance between the n-step kernel and the invariant measure. Small and petite sets are used in order to split the trajectory of a Markov chain into independent blocks. This theory was fully developed in [37] and again adapted and applied to the Metropolis-Hastings algorithm in [44] resulting in a criterion for geometric ergodicity

$$\|\mathcal{P}(x,\cdot)^n - \mu\|_{\mathrm{TV}} \le C(x)c^n \qquad \text{for some } c < 1.$$

Moreover, they also established a criterion for a CLT. Extending this method, it was also possible to derive rigorous confidence intervals in [29].

In most infinite dimensional settings the splitting chain method cannot be applied since measures tend to be mutually singular. The method is hence not well-adapted to the high-dimensional setting. Even Gaussian measures with the same covariance operator are only equivalent if the difference between their means lies in the Cameron-Martin space. As a consequence, the discrete time Ornstein-Uhlenbeck process on a function space is not irreducible in the sense of [37], i.e. there is no non trivial measure φ such that $\varphi(A) > 0$ implies $\mathcal{P}(x, A) > 0$ for all x. By inspecting the Metropolis-Hastings transition kernel (1.3) the pCN algorithm is not irreducible, since if $x - y$ is not an element of the Cameron-Martin space, each measure in the decomposition for $\mathcal{P}(x, \cdot)$ is mutually singular to each measure in the same decomposition for $\mathcal{P}(y, \cdot)$. This may also be shown to be true for the n-step kernel by expressing it as a sum of densities times Gaussian measures and applying the Feldman-Hajek Theorem [13].

For these reasons, existing theoretical results concerning RWM and pCN in high dimensions have been confined to scaling results and derivation of diffusion limits. In [4] the RWM with a target that is absolutely continuous with respect to a product measure has been analyzed for its dependence on the dimension. The proposal distribution is a centered normal random variable with covariance matrix $\sigma_n I_n$. The main result there is that δ has to be chosen as a constant times a particular negative power of n to prevent the expected acceptance probability to go to one or zero. In a similar setup it was recently shown [35] that there is a μ-reversible SPDE limit if the product law is a truncated Karhunen-Loeve expansion. This SPDE limit suggests that the number of steps necessary for a certain level of accuracy grows like $\mathcal{O}(m)$, because in order to approximate the SPDE limit on $[0, T]$ $\mathcal{O}(m)$ steps are necessary. A similar result in [41] suggests that the pCN algorithm only needs $\mathcal{O}(1)$ steps.

Uniform contraction in a Wasserstein distance was first applied to MCMC in [25] in order to get bound on the variance and bias of the sample average of Lipschitz functionals. We use the weak Harris theorem to verify this contraction and using the results from [46] non-asymptotic bounds on the sample average of L^2_μ functionals.

In this paper we substantiate these ideas, by using spectral gaps derived by applying the weak Harris theory of [22]. Section 2 contains statement of our main results, namely Theorems 2.9, 2.11 and 2.13 concerning the desirable dimension-independence properties of the pCN method, and Theorem 2.16 concerning the undesirable dimension dependence of the RWM method. Section 2 starts by specifying the RWM and pCN algorithms as Markov chains, statement of the weak Harris theorem, and a discussion of the relationship between exponential convergence in a Wasserstein distance and L^2_μ spectral gaps. Proofs of the theorems from Section 2 are given in Section 3. In Section 4 we exploit the Wasserstein and L^2_μ spectral gaps in order to derive a law of large numbers (LLN), central limit theorems (CLTs) and mean square error (MSE) bounds for sample-path ergodic averages of the pCN method, again emphasizing dimension-independence of the results. We draw some overall conclusions in Section 5.

Acknowledgement. We thank Feng-Yu Wang for pointing out the connection between Wasserstein and L^2 spectral gap. Financial support was kindly provided to MH by EPSRC grant EP/D071593/1, by the Royal Society through a Wolfson Research Merit Award, and by the Leverhulme Trust through a Philip Leverhulme Prize. AMS is grateful to EPSRC and ERC for financial support. SJV gratefully acknowledges support from an ERC stipend.

2 Main Results

In Section 2.1 we specify the RWM and pCN algorithms and in Section 2.2 we summarize the weak Harris theorem and present how a Wasserstein spectral gap implies an L^2_μ-spectral gap. In Section 2.3 we give necessary conditions on the target measure for the pCN algorithm to have a dimension independent spectral gap in a Wasserstein distance. In Section 2.4 we highlight the downside of the RWM by giving an example that satisfies our assumption for the pCN algorithm for which the spectral gap of the RWM algorithm converges to zero as fast as any negative power of m for $m \to \infty$.

2.1 Algorithms

We focus on convergence results for the pCN algorithm (Algorithm 1) that generates a Markov chain $\{X^n\}_{n\in\mathbb{N}}$ with $X^n \in H$ and $\{X^n_m\}_{n\in\mathbb{N}}$ when applied to a measure μ and μ_m respectively. The corresponding transition Markov kernels are called $\mathcal{P}$ and $\mathcal{P}_m$ respectively. We use the same notation for the Markov chain generated by the RWM (Algorithm 2). This should not cause confusion as statements concerning the pCN and RWM algorithms occur in separate subsections.

Algorithm 1 Preconditioned Crank-Nicolson

Initialize X_0.
For $n \geq$ do:

1. Generate $\xi \sim \gamma$ and set $p_{X_n}(\xi) = X_n + \sqrt{2\delta}\,\xi$.
2. Set

$$X_{n+1} = \begin{cases} p_{X_n} & \text{with probability } a(X_n, p_{X_n}) = 1 \wedge \exp(\Phi(x) - \Phi(y) \\ X_n & \text{otherwise} \end{cases}$$

Algorithm 2 Random Walk Metropolis

Initialize X_0.
For $n \geq$ do:

1. Generate $\xi \sim \gamma_m$ and set $p_{X_n}(\xi) = X_n + \sqrt{2\delta}\,\xi$.
2. Set

$$X_{n+1} = \begin{cases} p_{X_n} & \text{with probability } a(X_n, p_{X_n}) = \\ & 1 \wedge \exp(\Phi(x) - \Phi(y) + \frac{1}{2}\langle x, \mathcal{C}x\rangle - \frac{1}{2}\langle y, \mathcal{C}y\rangle) \\ X_n & \text{otherwise} \end{cases}$$

2.2 Preliminaries

Here we introduce Lyapunov functions, Wasserstein distances, d-small sets and d-contracting Markov kernels in order to state a weak Harris theorem recently proved in [22]. We use this theorem to prove our main results. By weakening the notion of a small set, this theorem gives a sufficient condition for exponential convergence in a Wasserstein distance. We explain how this in turn implies an L^2-spectral gap which is a major reason for the importance of the weak Harris theorem.

2.2.1 Weak Harris Theorem

Definition 2.1. Given a Polish space $\mathbf{E}$, a function $d : \mathbf{E} \times \mathbf{E} \to \mathbb{R}_+$ is a *distance-like* function if it is symmetric, lower semi-continuous and $d(x, y) = 0$ is equivalent to $x = y$.

This induces the 1-Wasserstein “distance” associated with d for measures ν_1, ν_2

$$d(\nu_1, \nu_2) = \inf_{\pi \in \Gamma(\nu_1,\nu_2)} \int_{\mathbf{E}\times\mathbf{E}} d(x, y)\pi(dx, dy) \tag{2.1}$$

where $\Gamma(\nu_1, \nu_2)$ is the set of couplings of ν_1 and ν_2 (all measures on $\mathbf{E} \times \mathbf{E}$ with marginals ν_1 and ν_2). If d is a metric the Monge-Kantorovich duality states

$$d(\nu_1, \nu_2) = \sup_{\|f\|_{Lip(d)}=1} \int f d\nu_1 - \int f d\nu_2.$$

We use the same notation for the distance and the associated Wasserstein distance; we hope that this does not lead to any confusion.

Definition 2.2. A Markov kernel $\mathcal{P}$ is *d-contracting* if there is $0 < c < 1$ such that $d(x, y) < 1$ implies

$$d(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) \le c \cdot d(x, y).$$

Definition 2.3. Let $\mathcal{P}$ be a Markov operator over a Polish space $\mathbf{E}$ endowed with a distance-like function $d : \mathbf{E}\times\mathbf{E} \to [0, 1]$. A set $S \subset \mathbf{E}$ is said to be *d-small* if there exists $0 < s < 1$ such that for every $x, y \in S$

$$d(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) \le s.$$

The d-Wasserstein distance associated with $d(x, y) = \chi_{\{x\neq y\}}(x, y)$ coincides with the total variation distance. If S is a small set (c.f. [37]) there is a probability measure ν such that $\mathcal{P}$ can be decomposed into

$$\mathcal{P}(x, dz) = s\tilde{\mathcal{P}}(x, dz) + (1 - s)\nu(dz) \qquad \text{for } x \in S,$$

which implies $d_{TV}(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) \le s$ hence S is d-small, too.

Definition 2.4. A Markov kernel $\mathcal{P}$ has a Wasserstein spectral gap if there is a $\lambda > 0$ and a $C > 0$ such that

$$d(\nu_1\mathcal{P}^n, \nu_2\mathcal{P}^n) \leq C\exp(-\lambda n)d(\nu_1, \nu_2) \text{ for all } n \in \mathbb{N}.$$

Definition 2.5. V is a *Lyapunov* function for the Markov operator $\mathcal{P}$ if there exist $K > 0$ and $0 \leq l < 1$ such that

$$\mathcal{P}^n V(x) \leq l^n V(x) + K \text{ for all } x \in \mathbf{E} \text{ and all } n \in \mathbb{N}. \tag{2.2}$$

Remark. Sometimes referred to as a drift condition because it implies that $\mathbb{E}(V(X_{n+1}))$ is smaller than $V(X_n)$ if $V(X_n) \geq \frac{1}{1-l}K$.

Proposition 2.6. *(**Weak Harris Theorem** [22]) Let $\mathcal{P}$ be a Markov kernel over a Polish space $\mathbf{E}$. Assume that:*

1. *$\mathcal{P}$ has a Lyapunov function V such that (2.2) holds;*
2. *$\mathcal{P}$ is d-contracting for a distance-like function $d : \mathbf{E} \times \mathbf{E} \to [0, 1]$;*
3. *the set $S = \{x \in \mathbf{E} : V(x) \leq 4K\}$ is d-small.*

Then there exists $\tilde{n}$ such that for any ν_1, ν_2 be probability measures on $\mathbf{E}$ we have

$$\tilde{d}(\nu_1\mathcal{P}^{\tilde{n}}, \nu_2\mathcal{P}^{\tilde{n}}) \leq \frac{1}{2}\tilde{d}(\nu_1, \nu_2)$$

where $\tilde{d}(x, y) = \sqrt{d(x, y)(1 + V(x) + V(y))}$ and $\tilde{n}(l, K, c, s)$ is increasing in l, K,c and s. Moreover, if there exists a complete metric d_0 on $\mathbf{E}$ such that $d_0 \leq \sqrt{d}$ and such that $\mathcal{P}_t$ is Feller on $\mathbf{E}$, then there is a unique invariant measure μ for $\mathcal{P}_t$.

Remark. For $\nu_2 = \mu$ we obtain the convergence rate to the invariant measure.

2.2.2 Wasserstein implies L^2-spectral Gap

In this section we explain why a Wasserstein spectral gap under mild assumption implies an L^2_μ- spectral gap.

Definition 2.7. (L^2_μ-spectral gap) A Markov operator $\mathcal{P}$ with invariant measure μ has an L^2_μ-spectral gap $1 - \beta$ if for $L^2_0 = \{f \in L^2_\mu \mid \mu(f) = 0\}$

$$\beta = \|\mathcal{P}\|_{L^2_0 \to L^2_0} = \sup \frac{\|\mathcal{P}^n f - \mu(f)\|_2}{\|f - \mu(f)\|_2} < 1.$$

The following proposition is due to F.-Y. Wang and is a discrete-time version of Theorem 2.1(2) [51]. It was also rediscovered in [39]. The proof given below is from private communication with F.-Y. Wang and is presented because of its beauty and the tremendous consequences in combination with weak Harris theorem.

Proposition 2.8. *([45] Private Communication) Let $\mathcal{P}$ be a Markov transition operator that is reversible with respect to μ and suppose $Lip(\tilde{\delta}) \cap L^\infty_\mu \cap L^2_\mu$ is dense in L^2_μ for some C, then*

$$\tilde{\delta}((P_t f)\mu, \mu) \leq C \exp(-\lambda n) \tilde{\delta}(f\mu, \mu)$$

implies the L^2_μ-spectral gap

$$\|\mathcal{P}^n f - \mu(f)\|_2^2 \leq \|f - \mu(f)\|_2^2 \exp(-\lambda n). \tag{2.3}$$

Proof. Let $0 \leq f \in Lip \cap L^\infty(\mu)$ with $\mu(f) = 1$ and π be the optimal coupling between $(\mathcal{P}^{2n} f)\mu$ and μ for the Wasserstein distance associated with d. Reversibility implies $\int (\mathcal{P}^n f)^2 d\mu = \int (\mathcal{P}^{2n} f) f d\mu$ which leads to

$$\begin{aligned}
\|\mathcal{P}^n f - \mu(f)\|_2^2 &= \mu\left((\mathcal{P}^n f)^2\right) - 1 = \int (f(x) - f(y)) d\pi \\
&\leq Lip(f) \int \tilde{\delta}(x, y) d\pi \leq Lip(f) \tilde{\delta}(\mathcal{P}^{2n} f\mu, \mu) \\
&= Lip(f) \tilde{\delta}((f\mu)\mathcal{P}^{2n}, \mu) \leq C Lip(f) \exp(-2\lambda n).
\end{aligned}$$

Since the above extends to $a \cdot f + b$ for general $f \in L^\infty \cap Lip(\tilde{\delta})$, we note that

$$\|P_t f - \mu(f)\|_2^2 \leq 2 \left\|P_t f^+ - \mu(f^+)\right\|_2^2 + 2 \left\|P_t f^- - \mu(f^-)\right\|_2^2.$$

By Lemma 2.9, the bound (2.3) holds for functions in $Lip \cap L^\infty(\mu)$, hence the result follows by taking limits of such functions. □

Lemma 2.9. *Let $\mathcal{P}$ be a Markov transition operator that is reversible with respect to μ. If* for some $f \in L^2(\mu)$ and constants $C(f)$ and $\lambda > 0$

$$\|\mathcal{P}^n f - \mu(f)\|_2^2 \leq C(f) \exp(-\lambda n),$$

then for all $n \in N$

$$\|\mathcal{P}^n f - \mu(f)\|_2^2 \leq \|f - \mu(f)\|_2^2 \exp(-\lambda n).$$

Proof. Without loss of generality we assume $\mu(\hat{f}^2) = 1$ where $\hat{f} = f - \mu(f)$. Applying the spectral theorem to $\mathcal{P}$ yields the existence of a unitary map $U : L^2(\mu) \mapsto L^2(X, \nu)$ such that UPU^{-1} is a multiplication operator by m. Moreover, $\mu(\hat{f}^2) = 1$ implies that $(U\hat{f})^2\nu$ is a probability measure such that for $k \in \mathbb{N}$

$$\begin{aligned}
\int (\mathcal{P}^n \hat{f}(x))^2 d\mu &= \int m(x)^{2n} (U\hat{f})^2(x) d\nu = \int m(x)^{(2n+k)\frac{2n}{2n+k}} d(U\hat{f})^2\nu \\
&\leq \left(\int m(x)^{2n+k} d(U\hat{f})^2\nu\right)^{\frac{2n}{2n+k}} \leq C^{\frac{2n}{2n+k}} \exp(-\lambda 2n),
\end{aligned}$$

letting $k \to \infty$ yields the required claim. □

2.3 Dimension-Independent Spectral Gaps for RWM

Using the weak Harris theorem we give necessary conditions on μ (see (1.1)) in terms of regularity and growth of Φ to have a uniform spectral gap in a Wasserstein distance for X^n and X^n_m. We need Φ to be at least locally Lipschitz; the case where it is globally Lipschitz is more straightforward and is presented first. Using the notation $\rho = 1-(1-2\delta)^{\frac{1}{2}}$ we can express the proposal of the pCN algorithm as

$$p_{X^n}(\xi) = (1-\rho)X^n + \sqrt{2\delta}\,\xi.$$

The mean of the proposal $(1-\rho)X^n$ suggests that we can prove that $f(\|\cdot\|)$ is a Lyapunov function for certain f and that $\mathcal{P}$ is d-contracting (for a suitable metric) if we have a lower bound on the probability of X_{n+1} being in a ball around the mean. In fact, our assumptions are stronger since we assume a uniform lower bound on $\mathbb{P}(p_x \text{ is accepted}|p_x = z)$ for z in $B_{r(\|x\|)}((1-\rho)x)$.

Assumption 2.10. *There is $R > 0$ and a function $r : \mathbb{R}^+ \mapsto \mathbb{R}^+$ satisfying $r(s) \leq \frac{\rho}{2}s$ for all $|s| \geq R$ such that for $x \in B_R(0)^c$*

$$\inf_{z\in B_{r(\|x\|)}((1-\rho)x)} -\Phi(z) + \Phi(x) > \alpha_l. \tag{2.4}$$

Assumption 2.11. *Let Φ in (1.1) have global Lipschitz constant L and assume that $\exp(-\Phi)$ is γ-integrable.*

Theorem 2.12. *Let Assumption 2.10 and 2.11 be satisfied with either*

1. *$r(\|x\|) = r\|x\|^a$ where $r \in \mathbb{R}^+$ for any $a \in (\frac{1}{2}, 1)$ then we consider $V = \|x\|^i$ with $i \in \mathbb{N}$ or $V = \exp(v\|x\|)$, or*
2. *$r(\|x\|) = r \in R$ for $r \in \mathbb{R}^+$ then we take $V = \|x\|^i$ with $i \in \mathbb{N}$.*

Then μ (μ_m) is the unique invariant measure for the Markov chain associated with the pCN algorithm applied to μ (μ_m). Moreover, define

$$\begin{aligned}\tilde{d}(x,y) &= \sqrt{d(x,y)(1+V(x)+V(y))} \text{ with}\\ d(x,y) &= 1 \wedge \frac{\|x-y\|}{\epsilon}.\end{aligned}$$

Then for ϵ small enough there is an $\tilde{n}$ such that for all ν_1, ν_2 probability measures on $\mathcal{H}$ and on $P_m H$ respectively and for all $m \in \mathbb{N}$

$$\begin{aligned}\tilde{d}(\nu_1\mathcal{P}^{\tilde{n}}, \nu_2\mathcal{P}^{\tilde{n}}) &\leq \frac{1}{2}\tilde{d}(\nu_1,\nu_2),\\ \tilde{d}(\nu_1\mathcal{P}^{\tilde{n}}_m, \nu_2\mathcal{P}^{\tilde{n}}_m) &\leq \frac{1}{2}\tilde{d}(\nu_1,\nu_2).\end{aligned}$$

Proof. The conditions of weak Harris theorem (Proposition 2.6) are satisfied by Lemmas 3.3, 3.4 and 3.5 and the uniqueness follows by Proposition 3.9. □

A key step in the proof is to verify the d-contraction. In order to get an upper bound on $d(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot))$ (see (2.1)) we choose a particular coupling between the algorithm started at x and y and distinguish between the cases when both proposals are accepted, both are rejected and only one is accepted. The case when only one of them accepts is the most difficult to tackle. By choosing $d = 1 \wedge \frac{\|x-y\|}{\epsilon}$ with ϵ small, it turns out that enough the Lipschitz constant of $\alpha(x,y)$ can be brought under control.

By changing the distance function d we can also handle the locally Lipschitz case provided that the local Lipschitz constant does not grow too fast.

Assumption 2.13. *Let* $\exp(-\Phi)$ *be integrable with respect to* γ *and assume that for any* $\kappa > 0$ *there is an* M_κ *such that*

$$\phi(r) = \sup_{x\neq y\in B_r(0)} \frac{|\Phi(x)-\Phi(y)|}{\|x-y\|} \leq M_\kappa e^{\kappa r}.$$

Theorem 2.14. *Let Assumption 2.10 and 2.13 be satisfied with* $r(\|x\|) = r\|x\|^a$ *with* $r\in\mathbb{R}$, $a\in(\frac{1}{2},1)$ and either $V = \|x\|^i$ with $i\in\mathbb{N}$ or $V = \exp(v\|x\|)$.

Then μ (μ_m) *is the unique invariant measure for the Markov chain associated with the pCN algorithm applied to* μ (μ_m).

For $\mathsf{A}(T,x,y) := \{\psi \in C^1([0,T],\mathcal{H}), \psi(0) = x, \psi(T) = y, \|\dot{\psi}\| = 1\}$, $\tilde{d}$ *as above with*

$$d(x,y) \quad = \quad 1 \wedge \inf_{T,\psi\in\mathsf{A}(T,x,y)} \frac{1}{\epsilon}\int_0^T \exp(\eta\|\psi\|)dt$$

and η *and*ϵ *small enough there is an* $\tilde{n}$ *such that for all* ν_1,ν_2 *probability measures on* $\mathcal{H}$ *and on* $P_m H$ *respectively and* $m\in\mathbb{N}$

$$\begin{aligned}
\tilde{d}(\nu_1\mathcal{P}^{\tilde{n}},\nu_2\mathcal{P}^{\tilde{n}}) &\leq \frac{1}{2}\tilde{d}(\nu_1,\nu_2)\\
\tilde{d}(\nu_1\mathcal{P}_m^{\tilde{n}},\nu_2\mathcal{P}_m^{\tilde{n}}) &\leq \frac{1}{2}\tilde{d}(\nu_1,\nu_2).
\end{aligned}$$

Remark. A Wasserstein spectral gap for the $\tilde{n}$-step transition kernel and an estimate of the form

$$d(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot)) \leq Cd(x,y) \tag{2.5}$$

implies a spectral gap for the one-step Kernel. Using that V is a Lyapunov function and $\mathcal{P}$ is $\tilde{d}$ contracting a straightforward calculation shows (2.5).

Proof. This time Lemmas 3.3, 3.7 and 3.8 verify the conditions of the weak Harris theorem (Proposition 2.6) and Proposition 3.9 yields again the uniqueness. □

Remark. Our arguments work for $\delta\in(0,\frac{1}{2}]$; for $\delta = \frac{1}{2}$ Assumption 2.10 degenerates to

$$\sup_{h\in\mathcal{H}}\Phi(h) - \inf_{h\in\mathcal{H}}\Phi(h) < \infty.$$

In this case $\mathcal{P}(x,\cdot)$ and $\mathcal{P}(y,\cdot)$ are not mutually singular any more and the theory of Meyn and Tweedie [37] applies.

In order to get the same lower bound for the L^2_μ-spectral gap we just have to verify that $Lip(\tilde{\delta}) \cap L^\infty_\mu \cap L^2_\mu$ is dense in L^2_μ.

Theorem 2.15. *If the conditions of Theorem 2.12 or 2.14 are satisfied, then we have the same lower bound on the L^2_μ-spectral gap of $\mathcal{P}$ and $\mathcal{P}_m$ uniformly in m.*

Proof. By Proposition 2.8 we only have to show that $Lip(\tilde{d}) \cap L^\infty(\mu)$ is dense in $L^2(H, \mathcal{B}, \mu)$.

By Lemma 4.1 and 4.3 $Lip(\|\cdot\|) \subseteq Lip(\tilde{d})$ hence it is enough to show that $Lip(\|\cdot\|) \cap L^\infty(\mu)$ is dense in $L^2(H, \mathcal{B}, \mu)$. Suppose not then there is $0 \neq g \in L^2(\mu)$ such that

$$\int fgd\mu = 0 \quad \text{for all } f \in Lip \cap L^\infty(\mu).$$

Since all measures on a separable Banach space equipped with the Borel σ-algreba are characterised by their characteristic functional (Bochner's theorem e.g. [7]), in particular they are characterised by bounded Lipschitz functions with respect to to $\|\cdot\|$. Hence $gd\mu$ is the zero measure so that $g \equiv 0$ in L^2_μ. $\square$

2.4 Dimension-Dependent Spectral Gaps for RWM

In order to prove negative results on the spectral gap is suffices to consider a particular case, and the analysis is made relatively straightforward by considering the case $\Phi = 0$ so that our target measure is μ_m, and by choosing a particular covariance operator. In this setting Theorem 2.15 shows that pCN has an $m-$independent L^2_μ spectral gap; in contrast we will now show that the spectral gap for RWM degenerates as m grows, on a specific example. We consider the family of measures μ_m on the scale of Hilbert spaces and then into (2.6). So far we have shown convergence results for the pCN, so subsequently we present an example where these results apply but the spectral gap of the RWM goes to 0 as m tends to infinity. We consider the target measures μ on

$$\mathcal{H}^\sigma_m := \left\{ x | \, \|x\|_\sigma = \sum_{i=1}^m i^{2\sigma} x_i^2 < \infty \right\}$$

with $0 < \sigma < \frac{1}{2}$ given by

$$\mu_m = \gamma_m = \mathcal{L}\left(\sum_{i=1}^m \frac{1}{i} \xi_i e_i \right) \qquad \xi \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1). \tag{2.6}$$

In the setting of (1.1) this corresponds to $\Phi = 0$. Hence the assumptions of Theorem 2.14 are satisfied and we obtain a uniform lower bound on the L^2_μ-spectral gap for the pCN. For the RWM algorithm we show that the spectral gap converges to zero faster than any negative power of m if we scale $\delta = s\,m^{-a}$ for any $a \in [0, 1)$.

Using the notion of conductance

$$\mathsf{C} = \inf_{\mu(A)\le\frac{1}{2}} \frac{\int_A \mathcal{P}(x, A^c)d\mu(x)}{\mu(A)}, \tag{2.7}$$

we obtain an upper bound on the spectral gap by Cheeger's inequality (c.f. [31, 48])

$$\frac{\mathsf{C}^2}{2} \le 1-\beta \le 2\mathsf{C}. \tag{2.8}$$

For the Metropolis-Hastings algorithm we can use $\alpha(x) = \int \alpha(x, y)Q(x, dy)$ to bound C.

Proposition 2.16. *Let $\mathcal{P}$ be a Metropolis-Hastings transition kernel for a target measure μ with acceptance probability $\alpha(x, y)$. For any set B with $\mu(B) \le \frac{1}{2}$, the spectral gap can be bounded by*

$$1-\beta \le 2 \sup_{x\in B} \alpha(x).$$

Proof. The algorithm can only move from B to B^c if it accepts the move. Hence

$$\mathcal{P}(x, B^c) \le \alpha(x).$$

Since this yields the bound

$$\mathsf{C} \;=\; \inf_{\mu(A)\le\frac{1}{2}} \frac{\int_A \mathcal{P}(x, A^c)d\mu(x)}{\mu(A)} \le \frac{\int_B \alpha(x)d\mu(x)}{\mu(B)} \le \sup_{x\in B} \alpha(x),$$

the claim follows from Cheeger's inequality. ☐

Theorem 2.17. *Let $\mathcal{P}_m$ be the Markov kernel and α be the acceptance probability associated with the RWM algorithm applied to μ_m as in (2.6).*

1. *For $\delta_m \sim m^{-a}$, $a \in [0, 1)$ and any p there is a $K(p, a)$ such that the spectral gap of $\mathcal{P}_m$ satisfies*

$$1-\beta_m \le K(p, a)m^{-p}.$$

2. *For $\delta_m \sim m^{-a}$, $a \in [1, \infty)$ there is a $K(a)$ such that the spectral gap of $\mathcal{P}_m$ satisfies*

$$1-\beta_m \le K(a)m^{-\frac{a}{2}}.$$

Proof. For the first part we work on the space H_σ with $\sigma \in [0, \frac{1}{2})$ and σ is determined later. We choose $B_r(0)$ such that $\mu\left(B_r(0)\right) \le \frac{1}{4}$ and by (3.1) we know that $\mu_m\left(B_r^m(0)\right)$ is decreasing towards $\mu\left(B_r(0)\right)$. Hence for all m larger than some M we know that $\mu\left(B_r^m(0)\right) \le \frac{1}{2}$. In order to apply Proposition 2.16, we have to get an upper bound on $\alpha(x)$ on $B_r^m(0)$. Thus we use $u \wedge v \le u^\lambda v^{1-\lambda}$ to bound

$$\alpha(x,y) = 1 \wedge \exp\left(-\frac{1}{2}\sum_{i=1}^m i^2(y_i^2 - x_i^2)\right) \le \exp\left(-\frac{1}{2}\sum_{i=1}^m i^2(y_i^2 - x_i^2)\lambda\right).$$

Using this inequality, we can find an upper bound on the acceptance probability

$$\alpha(x) = \int \alpha(x,y)Q(x,dy) \le \int \frac{m!}{(4\delta\pi)^{\frac{m}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^m i^2\left[(y_i^2 - x_i^2)\lambda + \frac{(x_i - y_i)^2}{2\delta}\right]\right) dy.$$

Completing the square and using the normalisation constant yields

$$\begin{aligned}\alpha(x) &\le \int \frac{m!}{(4\delta\pi)^{\frac{m}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^m i^2\left[(\lambda + \frac{1}{2\delta})\left(y_i - \frac{x_i}{2\delta\lambda + 1}\right)^2 - \frac{2\delta\lambda^2 x_i^2}{(2\delta\lambda+1)}\right]\right) dy \\ &\le (1+2\lambda\delta)^{-\frac{m}{2}} \exp\left(\sum_{i=1}^m \frac{\delta\lambda^2}{(2\delta\lambda+1)} i^2 x_i^2\right).\end{aligned}$$

For $x \in B_r^m(0)$ in $\mathcal{H}_\sigma$, using $\delta = m^{-a}$ and setting $\lambda = m^{-b}$

$$\alpha(x) \le (1 + 2m^{-(a+b)})^{-\frac{m}{2}} \exp\left(\frac{rm^{2-2\sigma-a-2b}}{3}\right).$$

In order to get decay from the first factor we need $a + b < 1$ and to prevent growth from the second $a + 2b > 2 - \sigma$ which corresponds to $a + 2b > 1$ for σ sufficiently close to $\frac{1}{2}$. This can be satisfied with $b = \frac{2(1-a)}{3}$ and $\sigma = \frac{2+a}{6} < \frac{1}{2}$. In this case the first factor decays faster than any negative power of m since

$$(1 + 2m^{-(a+b)})^{-\frac{m}{2}} = \exp\left(-\frac{m}{2}\log(1 + 2m^{-(a+b)})\right) \le \exp(-Cm^{1-(a+b)}).$$

For the second part of the poof we use $\alpha(x,y) \le 1$ and $A = \{x \in \mathbb{R}^n | x_1 \ge 0\}$, which by symmetry satisfies $\gamma_m(A) = \frac{1}{2}$, to bound the conductance

$$\begin{aligned}\frac{\mathsf{C}}{2} &\le \int_A P(x, A^c) d\mu \le \int_A\int_{A^c} \frac{\alpha(x,y) n!^2}{(2\pi)^n (2\delta)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^m i^2(x_i^2 + (x_i - y_i)^2/(2\delta)\right) dxdy \\ &\le \int_0^\infty \int_{-\infty}^0 \frac{\exp(-\frac{1}{2}\frac{(y_1-x_1)^2}{2\delta})}{2\pi\sqrt{2\delta}} dy_1 \exp\left(-\frac{1}{2}x_1^2\right) dx_1 \\ &= \int_0^\infty \int_{-\infty}^{-\frac{x_1}{\sqrt{2\delta}}} \frac{\exp(-\frac{1}{2}z^2)}{2\pi} dy_1 \exp\left(-\frac{1}{2}x_1^2\right) dx_1.\end{aligned}$$

Combining Fernique's theorem and Markov's inequality (Lemma A.2) yields

$$\mathsf{C} \le K \int_0^\infty \exp(-\frac{1}{2}(\frac{\delta+1}{\delta})x_1^2)dx \le K\sqrt{2\pi\frac{\delta}{\delta+1}} \le \tilde{K} m^{-\frac{a}{2}},$$

so that the claim follows again from Cheeger's inequality. □

3 Spectral Gap: Proofs

We check the three conditions of the weak Harris theorem (Proposition 2.6) for globally and locally Lipschitz Φ (see (1.1)) in Sections 3.1 and 3.2 respectively. For each condition we use the following lemma for the dependence of constants l, K, c and s in the weak Harris theorem on m. This allows us to conclude that there is $\tilde{n}(m) \leq \tilde{n}$ such that

$$\begin{aligned}\tilde{d}(\nu_1\mathcal{P}^{\tilde{n}}, \nu_2\mathcal{P}^{\tilde{n}}) &\leq \frac{1}{2}\tilde{d}(\nu_1, \nu_2)\\ \tilde{d}(\nu_1\mathcal{P}_m^{\tilde{n}(m)}, \nu_2\mathcal{P}_m^{\tilde{n}(m)}) &\leq \frac{1}{2}\tilde{d}(\nu_1, \nu_2)\end{aligned}$$

for all measures ν_1, ν_2 measures on H and P_mH respectively.

Replacing $r(s) \wedge \frac{\rho}{2}s$ only weakens the condition (2.4) so we can and will assume that $r(s) \leq \rho s/2$.

Lemma 3.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be monotone increasing, then*

$$\int f(\|\xi\|)d\gamma_m(\xi) \leq \int f(\|\xi\|)d\gamma(\xi)$$

and in particular

$$\gamma_m(B_R(0)) \geq \gamma(B_R(0)). \tag{3.1}$$

Proof. The truncated Karhunen-Loeve expansion relates γ_m and γ and yields

$$\sum_{i=1}^{m} \lambda_i\xi_i^2 \leq \sum_{i=1}^{\infty} \lambda_i\xi_i^2.$$

Hence the result follows by monotonicity of the integral and f

$$\int f(\|\xi\|)d\gamma_m(\xi) = \mathbb{E}(\sqrt{f(\sum_{i=1}^{m}\lambda_i\xi_i^2)}) \leq \mathbb{E}(\sqrt{f(\sum_{i=1}^{\infty}\lambda_i\xi_i^2)}) = \int f(\|\xi\|)d\gamma(\xi).$$

This yields (3.1) by inserting $f = \chi_{B_R(0)^c}$. □

We conclude this section by showing μ respectively μ_m are the unique invariant measure for $\mathcal{P}$ respectively $\mathcal{P}_m$.

3.1 Global log-Lipschitz density

In this section we will prove Theorem 2.12 by checking the three conditions of the weak Harris theorem for

$$d(x, y) = 1 \wedge \frac{\|x - y\|}{\epsilon}. \tag{3.2}$$

3.1.1 Lyapunov Functions

Under Assumption 2.10 we show the existence of a Lyapunov function V. This relies on the decay of V on $B_{r(\|x\|)}((1-\rho)x)$ and the fact that probability of the next step of the algorithm lying in that ball can be bounded below by the Fernique's theorem which we recall here

Proposition 3.2. *(Fernique's theorem see e.g. [6, 13, 20]) Let $\gamma = \mathcal{N}(m, \mathcal{C})$ be a Gaussian measure on a Banach space, then for β small enough*

$$\int_X \exp(\beta \|u\|^2) d\gamma(u) = F_\beta < \infty.$$

Moreover, to deal with a proposals outside $B_{r(\|x\|)}((1-\rho)x)$ we use

Proposition. *A.1 (Appendix) For small enough β and $\alpha \in \mathbb{R}$ there is a constant $C_{\alpha.\beta}$ such that*

$$\int_{\{\|u\| \geq K\}} \exp(\alpha \|u\|) d\gamma(u) \leq C_{\alpha,\beta} e^{-\beta K^2 + \alpha K}.$$

Lemma 3.3. *Suppose Assumption 2.10 is satisfied with either:*

1. *$r(\|x\|) = r \in \mathbb{R}$; or*
2. *$r(\|x\|) = r\|x\|^a)$, $\kappa > 0$ and $a \in (\frac{1}{2}, 1)$.*

Then the function $V(x) = \|x\|^i$ with $i \in \mathbb{N}$ in the first case and additionally $V(x) = \exp\big(\ell\|x\|\big)$ in the second case, are Lyapunov functions for both $\mathcal{P}$ and $\mathcal{P}_m$, with constants l and K uniform in m.

Proof. In both cases we choose R as in Assumption 2.10 set

$$\sup_{x \in B_R(0)} \mathcal{P}V(x) \leq \sup_{x \in B_R(0)} \int \Big(\|x\| + \sqrt{2\delta}\,\|\xi\|\Big)^i d\gamma(\xi) \leq R^i + C =: K_1 < \infty$$

by Fernique's theorem. Now let $x \in B_R(0)^c$, then there is $0 < \tilde{l} < 1$ such that

$$\sup_{y \in B_{r(\|x\|)}((1-\rho)x)} V(y) \quad \leq \tilde{l} V(x). \tag{3.3}$$

We denote by $A = \{\omega | \sqrt{2\delta}\,\|\xi\| \leq r(\|x\|)\}$ the event that the proposal lies in a ball with a lower bound acceptance probability due to Assumption 2.10 to bound

$$\begin{aligned}
\mathcal{P}V &\leq \mathbb{P}(A)\Big[\mathbb{P}(accept|A)\tilde{l}V(x) + \mathbb{P}(reject|A)V(x)\Big] + \mathbb{E}(V(p_x) \vee V(x); A^c) \\
&\leq \mathbb{P}(A)\Big[(1 - \mathbb{P}(accept|A)(1-\tilde{l})\Big] V(x) + \mathbb{E}(V(p_x) \vee V(x); A^c) \\
&\leq \theta \mathbb{P}(A) V(x) + \mathbb{E}(V(p_x) \vee V(x); A^c)
\end{aligned}$$

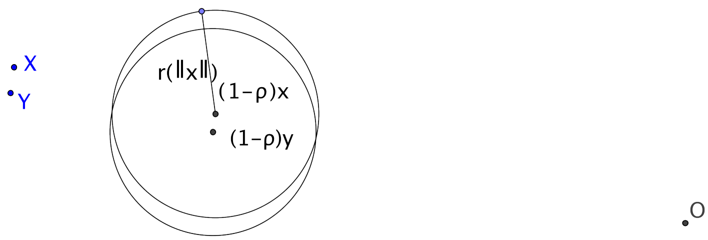

Figure 1: Contraction

for some $\theta < 1$. It remains to consider $\mathbb{E}(V(p_x)\vee V(x);A^c)$ where the differences will arise between cases 1 and 2. For the first case we have by Fernique's theorem

$$\begin{aligned}\mathbb{E}(V(p_x)\vee V(x);A^c) &\leq \int_{\sqrt{2\delta}\|\xi\|\geq c} \|x\|^i \vee \left((1-\rho)\|x\| + \sqrt{2\delta}\,\|\xi\|\right)^i d\gamma(\xi) \\ &\leq \int_{\|\xi\|\geq \frac{c}{\sqrt{2\delta}}} \left(\|x\|^i + K\|\xi\|^p\right) d\gamma(\xi) \leq \mathbb{P}(A^c)V(x) + K_2\end{aligned}$$

Since a ball around the mean of a Gaussian always has positive mass (Theorem 3.6.1 in [6]) we note

$$\mathcal{P}V \leq V(x)(\mathbb{P}(A)\theta + \mathbb{P}(A^c)) + K_2 \leq lV + K_2.$$

For the second case we estimate

$$\mathbb{E}(V(p_x)\vee V(x);A^c) \quad\leq\quad M_v \int_{\|\eta\|>r\|x\|^a} e^{v(\|x\|+\sqrt{2\delta}\|\xi\|)} d\gamma(\xi).$$

The right hand side above is uniformly bounded in $x \in B_R(0)^c$ by some K_2 due to Proposition A.1. Hence in both cases there is an $l < 1$ such that

$$\mathcal{P}V(x) \leq lV(x) + \max(K_1, K_2) \quad \forall x.$$

For the m-dimensional approximation the probability of the event A is larger by Lemma 3.1 and $\mathbb{P}(accept|A)$ has the same lower bound and therefore $l(m)$ is smaller than l. Similarly $K_i(m)$ is smaller then K_i by Lemma 3.1. □

3.1.2 The d-Contraction

In this section we show that $\mathcal{P}$ is d-contracting for $d(x,y) = 1 \wedge \frac{\|x-y\|}{\epsilon}$ by bounding $d(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot))$ (see (2.1)) with a particular coupling. For x and y we choose the same noise ξ giving rise to the proposals $p_x(\xi)$ and $p_y(\xi)$ and the same uniform random variable for acceptance. Subsequently we will refer to this coupling as the basic coupling and bound the expectation of d under this coupling by inspecting the following cases:

1. The proposals for the algorithm started at x and y are both accepted.

2. Both proposals are rejected.

3. One of the proposals is accepted and the other rejected.

Lemma 3.4. *If Φ in (1.1) satisfies Assumption 2.10 and 2.11, then $\mathcal{P}$ and $\mathcal{P}_m$ are d-contracting for d as in (3.2) with a contraction constant uniform in m.*

Proof. By Definition 2.2 we only need to consider x and y such that $d(x,y) < 1$, which implies $\|x-y\| < \epsilon$. Later we will choose $\epsilon \ll 1$ hence if $\|x-y\| < \epsilon$ then either $x, y \in B_R(0)$ or $x, y \in B^c_{\tilde{R}}(0)$ with $\tilde{R} = R-1$, and we will treat the two cases separately. We assume without loss of generality $\|y\| \geq \|x\|$.

For $x, y \in B_R(0)$ and $A = \{\omega | \sqrt{2\delta}\,\|\xi\| \leq R\}$ the basic coupling yields

$$\begin{aligned} d(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot)) \quad \leq \quad & \mathbb{P}(A)\left[\mathbb{P}(\text{both accept}|A)(1-\rho)d(x,y) + \right. \\ & \left.\mathbb{P}\left(\text{both reject}|\text{A}\right)d(x,y))\right] + \mathbb{P}(A^c)d(x,y) + \\ & \int_{\mathcal{H}} |\alpha(x,p_x)(\xi) - \alpha(y,p_y)(\xi)|\, d\gamma(\xi) \end{aligned} \tag{3.4}$$

where the last term bounds the case that only one of the proposals is accepted. Using the bound $\mathbb{P}(\text{both reject}|\text{A}) \leq 1 - \mathbb{P}(\text{both accept}|A)$ yields a non-trivial convex combination of d and $(1-\rho)d$, since the probability $\mathbb{P}(\text{both accept}|A)$ is bounded below by $\exp(-\sup\{\Phi(z)|\,\|z\| \leq 2R\} + \inf\{\Phi(z)|\,\|z\| \leq 2R\})$ due to (1.5). The first two summands in (3.4) form again a non-trivial convex combination, since $\mathbb{P}(A) > 0$, so that there is $\tilde{c} < 1$ with

$$d(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot)) \leq \tilde{c}d(x,y) + \int_{\mathcal{H}} |\alpha(x,p_x)(\xi) - \alpha(y,p_y)(\xi)|\, d\gamma(\xi).$$

Note that $\tilde{c}$ is independent of ϵ. For the last term we use that $1 \wedge \exp(\cdot)$ has Lipschitz constant 1

$$\begin{aligned} \int_X |\alpha(x,p_x)(\xi) - \alpha(y,p_y)(\xi)|\, d\gamma(\xi) \quad & \leq \quad \int |\Phi(p_x) - \Phi(p_y)| + |\Phi(x) - \Phi(y)|\, d\gamma(\xi) \\ & \leq \quad 2L\,|x-y| \leq 2L\epsilon d(x,y) \end{aligned}$$

which yields an overall contraction for ϵ small enough.

Similarly we get for $x, y \in B_{\tilde{R}}(0)^c$ and $B = \{\omega | \sqrt{2\delta}\,\|\zeta\| \leq r(\|x\| \wedge \|y\|)\}$

$$\begin{aligned} d(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot)) \quad \leq \quad & \mathbb{P}(B)(\mathbb{P}(\text{both accept}|B)(1-\rho) + \mathbb{P}(\text{both reject}|\text{B}))d(x,y) \\ & \mathbb{P}(B^c)d(x,y) + \int_{\mathcal{H}} |\alpha(x,p_x)(\xi) - \alpha(y,p_y)(\xi)|\, d\gamma(\xi). \end{aligned}$$

The lower bound for $\mathbb{P}(\text{both accept}|B)$ is this time due to Assumption 2.10.

All occurring ball probabilities are larger in the m-dimensional approximation due to Lemma 3.1 and the acceptance probability is larger since inf and sup are applied to smaller sets, thus the contraction constant is uniform in m. ☐

3.1.3 The d-Smallness

The d-smallness of the level sets of V is achieved by replacing the Markov kernel by the n-step one. This preserves the d-contraction and the Lyapunov function. The variable n is chosen large enough so that if the algorithms started at x and y both accept n times in a row d drops below $\frac{1}{2}$, hence

$$d(\mathcal{P}^n(x,\cdot),\mathcal{P}^n(y,\cdot)) \leq 1-\frac{1}{2}\mathbb{P}(\text{accept n-times}).$$

Remark. It is necessary to replace the one step Markov kernel with the n-step which can be seen by considering the Wiener measure on $(\mathcal{C}([0,1]),\|\cdot\|_\infty)$ and $\Phi=0$ (our theory also applies to Banach spaces see Section (5)). For the constant zero path ψ and

$$\phi_n(x)=\begin{cases} nx & x\leq 1/n \\ 1 & x\geq 1/n \end{cases}$$

$\|\psi-\phi_n\|_\infty=1$ but the transition to a common ϵ neighborhood using the proposal (1.5) converges to zero as $n\to\infty$.

Lemma 3.5. *If S is bounded, then there is an n and $0<s<1$ such that for all $x,y\in S$, $m\in\mathbb{N}$ and for d as in (3.2)*

$$d\left(\mathcal{P}_m^n(x,\cdot),\mathcal{P}_m^n(y,\cdot)\right)\leq s \quad \textit{and} \quad d\left(\mathcal{P}^n(x,\cdot),\mathcal{P}^n(y,\cdot)\right)\leq s\ .$$

Proof. In order to get an upper bound for $d\left(\mathcal{P}^n(x,\cdot),\mathcal{P}^n(y,\cdot)\right)$ we choose the basic coupling (see Section 3.1.2) as before. Let R_S be such that $S\subset B_{R_S}(0)$ and B be the event, that both instances of the algorithm accept n times in a row. In the event of B we have using (3.2)

$$d(X_n,Y_n)\leq\frac{1}{\epsilon}\|X_n-Y_n\|\leq\frac{1}{\epsilon}(1-\rho)^n\|X_0-Y_0\|\leq\frac{1}{\epsilon}(1-\rho)^n\operatorname{diam}S\leq\frac{1}{2}$$

which implies that if X_0 and Y_0 are in S then $d(X_n,Y_n)\leq\frac{1}{2}$, hence

$$d(\mathcal{P}^n(x,\cdot),\mathcal{P}^n(y,\cdot))\leq\mathbb{P}(B)\frac{1}{2}+(1-\mathbb{P}(B))\cdot 1<1$$

We write ξ^i for the noise in the i-th step and bound

$$\begin{aligned}\mathbb{P}(B) &\geq \mathbb{P}\left(\left\|\sqrt{2\delta}\xi^i\right\|\leq\frac{R}{n}\, i=1\dots n\right)\mathbb{P}\left(\text{both accept } n \text{ times} \mid \left\|\xi^i\right\|\leq\frac{R}{n}\right)\\ &\geq \mathbb{P}(\|\zeta\|\leq\frac{R}{n})^n\exp\left(-\sup_{z\in B_{2R}(0)}\Phi(z)+\inf_{z\in B_{2R}(0)}\Phi(z)\right)^n>0,\end{aligned}$$

uniformly for all $X_0,Y_0\in B_R(0)$. For the m-dimensional approximation the lower bound exceeds that in the infinite dimensional case due to Lemma 3.1 and the fact that

$$-\sup_{z\in B_{2R}(0)}\Phi(z)+\inf_{z\in B_{2R}(0)}\Phi(z)\leq-\sup_{z\in B_{2R}(0)}\Phi(P_nz)+\inf_{z\in B_{2R}(0)}\Phi(P_nz)$$

so that the claim follows. $\square$

3.2 Local log-Lipschitz density

Now we allow the local Lipschitz constant

$$\phi(r) = \sup_{x \neq y \in B_r(0)} \frac{|\Phi(x) - \Phi(y)|}{\|x - y\|}$$

to grow in r. In order to deal with the situation where only one proposal is accepted, in proving $\mathcal{P}$ is d-contracting, we choose d in a way such that two points far out have to be closer in $\|\cdot\|_{\mathcal{H}}$ in order to be considered "close" i.e. $d(x,y) < 1$. This is inspired by constructions in [21, 22]. Setting

$$\mathsf{A}(T,x,y) := \{\psi \in C^1([0,T],\mathcal{H}), \psi(0) = x, \psi(T) = y, \|\dot{\psi}\| = 1\},$$

we define metrics d and $\bar{d}$ by

$$d(x,y) = 1 \wedge \bar{d}(x,y) \qquad \bar{d}(x,y) = \inf_{T,\psi \in \mathsf{A}(T,x,y)} \tfrac{1}{\epsilon} \int_0^T \exp(\eta \|\psi\|) dt, \tag{3.5}$$

where ϵ and η i s chosen along the way depending on Φ and γ. The situation is different from before because even in the case "both accept" the distance can increase because of the weight. In order to control this we note

Lemma 3.6. *Let ψ be a path connecting x, y then for $\bar{d}$ as in (3.5)*

1. $\frac{1}{\epsilon}\int_0^T \exp(\eta \|\psi\|)dt < 1$ *implies* $T \leq J := \epsilon \exp\left(-\eta(\|x\| \vee \|y\| - \epsilon) \vee 0\right) \leq \epsilon$.
2. $\bar{d}(x,y) \leq \frac{\|x-y\|}{\epsilon} \exp\left(\eta(\|x\| \vee \|y\|)\right)$ *and for* $\bar{d} < 1$
$$\frac{\|x-y\|}{\epsilon} \exp\left(\eta(\|x\| \vee \|y\| - J) \vee 0\right) \leq \bar{d}(x,y)$$
3. *For* $\bar{d} < 1$ *we have*
$$\frac{\bar{d}(p_x, p_y)}{\bar{d}(x,y)} \leq (1 - 2\delta)^{\frac{1}{2}} e^{-\eta\rho(\|x\| \vee \|y\| + \eta(\left\|\sqrt{2\delta}\xi\right\| + J)}.$$

Proof. For the first statement, observe that

$$\epsilon \geq \int_0^T e^{\eta |\|x\| \vee \|y\| - t|} dt \geq T e^{\eta(\|x\| \vee \|y\| - T) \vee 0)} \geq T e^{\eta(\|x\| \vee \|y\| - \epsilon) \vee 0)}.$$

For the second part we set ψ to be the line connecting x and y to get the upper bound and for the lower bound we use $\|\psi\| \geq (\|x\| \vee \|y\| - J) \vee 0$ from the first part combined with the fact that $T \leq \epsilon$. Using 2. we get

$$
\begin{aligned}
\bar{d}(p_x, p_y) &\leq \frac{1}{\epsilon}(1-2\delta)^{\frac{1}{2}} \|x-y\| \, e^{\eta\left[(\|x\|\vee\|y\|)-\rho(\|x\|\vee\|y\|)+\sqrt{2\delta}\|\xi\|\right]} \\
&\leq (1-2\delta)^{\frac{1}{2}} e^{\eta\left[-\rho(\|x\|\vee\|y\|)+\sqrt{2\delta}\|\xi\|+J\right]} \frac{1}{\epsilon} \|x-y\| \, e^{\eta(\|x\|\vee\|y\|-J)} \\
&\leq (1-2\delta)^{\frac{1}{2}} e^{\eta\left[-\rho(\|x\|\vee\|y\|)+\sqrt{2\delta}\|\xi\|+J\right]} \bar{d}(x,y),
\end{aligned}
$$

which is precisely the required bound. $\square$

3.2.1 Lyapunov Functions

This condition neither depends on the distance function d nor on the Lipschitz properties of Φ hence Lemma 3.3 applies.

3.2.2 The d-Contraction

Lemma 3.7. *If Φ satisfies Assumption 2.10 and 2.13 then $\mathcal{P}$ and $\mathcal{P}_m$ are d-contracting for d as in (3.5) with a contraction constant uniform in m.*

Proof. First suppose $x, y \in B_R(0)$ with $d(x,y) < 1$ and denote the event $A = \left\{\omega | \, \|\xi\| \leq \frac{2R}{\sqrt{2\delta}}\right\}$. We will first choose R large then η small and at last ϵ small. We have

$$
\begin{aligned}
d(\mathcal{P}(x,\cdot), \mathcal{P}(y,\cdot)) \leq \; & \mathbb{P}(A) \left[\mathbb{P}(\text{both accept}|A)(1-\tilde{\rho}) d(x,y)\right. && (3.6) \\
& + \left[\mathbb{P}(\text{both reject}|A) d(x,y)\right] \\
& + \mathbb{E}((\alpha(x,p_x) \wedge \alpha(y,p_y)) d(p_x,p_y); A^c) \\
& + \mathbb{E}((1-\alpha(x,p_x) \vee \alpha(y,p_y)) d(x,y); A^c) && (3.7) \\
& + \mathbb{P}(\text{only one accepts}) \cdot 1
\end{aligned}
$$

where the first two lines deal with both accept and both reject in the case of A, the third and fourth line considers the same case in the event of A^c. The last line takes care of only one accepts. For the first two lines of (3.6) we argue that

$$
\mathbb{P}(\text{both accept}|A) \geq \inf_{x,z \in B_{3R}(0)} \mathbb{P}(\text{accepts}|p_x = z) = \exp(-\Phi^+(3R) + \Phi^-(3R)).
$$

If both are accepted we know from Lemma 3.6 that

$$
\begin{aligned}
\frac{\bar{d}(p_x,p_y)}{\bar{d}(x,y)} &\leq (1-2\delta)^{\frac{1}{2}} \exp\left(-\eta\rho\left(\|x\| \vee \|y\|\right) + \eta(\left\|\sqrt{2\delta}\xi\right\| + J)\right) \\
&\leq (1-2\delta)^{\frac{1}{2}} e^{\eta(3R+J)} \leq (1-\tilde{\rho})
\end{aligned}
$$

where the last step follows for small enough η. Using the complementary probability we can estimate

$$\mathbb{P}(\text{both reject}|A) \leq 1 - \mathbb{P}(\text{both accept}|A)$$

Combining both estimates we get $\mathbb{P}(A)\left(1 - \mathbb{P}(\text{both accept}|A)(1-\tilde{\rho})\right)$ as coefficient in front of $d(x,y)$. In order to show contraction we have to show that the expression in the third and fourth line of (3.6) is close to $\mathbb{P}(A^c) \cdot d(x,y)$. We note that

$$\begin{aligned}
&\mathbb{E}\left((1-\alpha(x,p_x)\vee\alpha(y,p_y))d(x,y);A^c\right) + \mathbb{E}\left((\alpha(x,p_x)\wedge\alpha(y,p_y))d(p_x,p_y);A^c\right)\\
&\leq \mathbb{E}\left(d(p_x,p_y)\vee d(x,y);A^c\right) \leq \bar{d}(x,y)\mathbb{E}\frac{\bar{d}(p_x,p_y)}{\bar{d}(x,y)} \vee 1\\
&\leq d(x,y) \int\limits_{\sqrt{2\delta}\|\xi\|>2R} 1 \vee e^{\eta(\sqrt{2\delta}\|\xi\|+J)} d\gamma(\xi)
\end{aligned}$$

where the last step followed by Lemma 3.6. For small η the above is arbitrarily close to $\mathbb{P}(A^c) \cdot d(x,y)$ by the dominated convergence theorem. By writing the integrand as $\chi_{\sqrt{2\delta}\|\xi\|>2R}\left(1 \vee \exp(\eta(\sqrt{2\delta}\,\|\xi\| + J)\right)$ and applying Lemma 3.1 we conclude that this holds uniformly in m. Combing the first four lines, the coefficient in front of $d(x,y)$ is less than 1 independently of ϵ. Only $\mathbb{P}(\text{only one accepts}) \cdot 1$ is left to bound in terms of $d(x,y)$:

$$\begin{aligned}
\mathbb{P}(\text{only one accepts}) &= \int |\alpha(x,p_x) - \alpha(y,p_y)|\, d\gamma(\xi)\\
&\leq \int (|\Phi(p_x) - \Phi(p_y)| + |\Phi(x) - \Phi(y)|) d\gamma(\xi)\\
&\leq \epsilon d(x,y) \int (\phi((1-\rho)R + \sqrt{2\delta}\,\|\xi\|) + \phi(R)) d\gamma(\xi)
\end{aligned}$$

The integral above is bounded by Fernique's theorem, hence for ϵ small enough combining with the result above we get an overall contraction.

Now let $x, y \in B^c_{\tilde{R}}(0)$ with $d(x,y) < 1$ and without loss of generality $\|y\| \geq \|x\|$. Analogous to the above with $A = \{\omega | \|\sqrt{2\delta}\zeta\| \leq r(\|x\|)\}$ we have

$$\begin{aligned}
d(\mathcal{P}(x,\cdot),\mathcal{P}(y,\cdot)) \leq\ & \mathbb{P}(A)\left[\mathbb{P}(\text{both accept}|A)(1-\rho)d(x,y) + \right.\\
& \left.\mathbb{P}(\text{both reject}|A)d(x,y)\right] + \mathbb{E}\left(d(x,y) \vee d(p_x,p_y);A^c\right)\\
& + \mathbb{P}(\text{only one accepts}) \cdot 1
\end{aligned}$$

If "both accept" in the event of A the contraction constant is smaller than $(1-\rho)$

since $r(\|x\|) \leq \frac{\rho}{2}\|x\|$ and using Lemma 3.6. For the next term it yields

$$\begin{aligned}
\mathbb{E}\left(d(p_x,p_y)\vee d(x,y);A^c\right) &\leq \bar{d}(x,y)\mathbb{E}\frac{\bar{d}(p_x,p_y)}{\bar{d}(x,y)}\vee 1 \\
&\leq \bar{d}(x,y)\int_{A^c} 1\vee e^{-\rho\eta(\|y\|)+\eta(\|\sqrt{2\delta}\xi\|+J)}d\gamma(\xi).
\end{aligned}$$

We denote the integral above by I, its integrand by $f(\zeta)$ and $F>0$ then

$$I \leq I_1+I_2 = \int\limits_{\rho(\|y\|-J)+F\geq\|\sqrt{2\delta}\xi\|\geq r(\|x\|\wedge\|y\|)} f(\xi)d\gamma(\xi) + \int\limits_{\|\sqrt{2\delta}\xi\|\geq\rho(\|y\|-J)+F} f(\xi)d\gamma(\xi)$$

for the first part we have the upper bound $\mathbb{P}(A^c)e^{\sqrt{2\delta}\eta F}$. For the second part we take $g\in X^\star$ with $\|g\|=1$ and note that $\{x|g(x)>R\}\subseteq B_R(0)^c$ which yields

$$\gamma(B_R(0)^c)\geq\gamma(\{x|g(x)>R\})\geq\exp(-\tilde{\beta}R^2+\zeta)$$

using the one dimensional lower bound. For the uniformity in m we choose $g=e_1^\star$. We incorporate all occurring constants into ζ and use Proposition A.1 to bound

$$\begin{aligned}
I_2 \leq\ & \mathbb{P}(A^c)\exp\left(\tilde{\beta}\frac{r(\|x\|)^2}{2\delta}-\rho\eta(\|y\|-J)\right. \\
& \left.\eta\sqrt{2\delta}(\rho(\|y\|-J)+F)-\beta\sqrt{2\delta}(\rho(\|y\|-J)+F)^2+\zeta\right).
\end{aligned}$$

For any $\tau>0$ first we choose F large enough and then η small enough so that $I\leq(1+\tau)\mathbb{P}(A^c)d(x,y)$. Again the estimates above are independent of ϵ which we choose small in order to bound $\mathbb{P}(\text{only one accepts}|A^c)$ in terms of $d(x,y)$. We calculate as above

$$\begin{aligned}
& \int|\alpha(x,p_x)-\alpha(y,p_y)|\,d\gamma(\xi) \\
\leq\ & \int|\Phi(x)-\Phi(y)|+|\Phi(p_x)-\Phi(p_y)|\,d\gamma(\xi) \\
\leq\ & \int(\phi(\|y\|)+\phi(\|p_x\|\vee\|p_y\|)\,d\gamma(\xi)\,\|x-y\| \\
\leq\ & \left(M_\kappa e^{\kappa\|y\|}+\int\phi((1-\rho)\|y\|+\sqrt{2\delta}\|\xi\|)d\gamma(\xi)\right)\|x-y\| \\
\leq\ & CM_\kappa\epsilon e^{-\eta(\|x\|\vee\|y\|-\epsilon)\vee 0+\kappa\|y\|}\bar{d}(x,y)
\end{aligned}$$

where the last step follows using the upper bound for $\|x-y\|$ from Lemma 3.6. Choosing $\kappa=\frac{\eta}{2}$ and ϵ small enough, we can guarantee a uniform contraction. Checking line by line, the same is true for the m-dimensional approximation. $\square$

3.2.3 The d-Smallness

Similarly to the globally Lipschitz case we have

Lemma 3.8. *If S is bounded, then $\exists n \in \mathbb{N}$ and $0 < s < 1$ such that for all $x, y \in S$, $m \in \mathbb{N}$ and for d as in (3.5)*

$$d(\mathcal{P}_m^n(x,\cdot), \mathcal{P}_m^n(y,\cdot)) \leq s \quad \textit{and} \quad d(\mathcal{P}^n(x,\cdot), \mathcal{P}^n(y,\cdot)) \leq s \ .$$

Proof. By Lemma 3.5 d and $\|\cdot\|$ are comparable on bounded sets. If $X_0, Y_0 \in B_R(0)$ and both algorithms accept n proposals in a row that all lay in $B_{2R}(0)$,

$$d(X_n, Y_n) \leq \frac{\exp(\eta(2R+J))}{\epsilon} \operatorname{diam}(S)(1-2\delta)^{n/2} \leq \frac{1}{2}.$$

Hence the result follows analog to Lemma 3.5. □

3.3 Uniqueness of the Invariant Measure

Proposition 3.9. *If the conditions of one of Theorem 2.12 or 2.14 are satisfied, then μ and μ_m are the unique invariant measures for $\mathcal{P}$ and $\mathcal{P}_m$ respectively.*

Proof. The space $(\mathcal{H}, d_0)$ with $d_0 = 1 \wedge \|x-y\| \leq d$ is complete because $(\mathcal{H}, \|\cdot\|)$ is complete and convergence in both spaces is equivalent. Using the dominated convergence theorem for

$$\mathcal{P}\phi(x) = \int \alpha_{x.p_x} \phi(p_x) d\gamma(\xi) + \phi(x) \int (1-\alpha_{x.p_x}) d\gamma(\xi),$$

the Markov kernel $\mathcal{P}$ is Feller. The result is now a direct consequence of the second part of the weak Harris theorem. □

4 Results Concerning the Sample-Path Average

In this section we focus on sample path properties of the pCN algorithm. We prove a strong law of large numbers, a CLT and bound on the MSE. This allows us to quantify the approximation of $\mu(f)$ by

$$S_{n,n_0}(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_{i+n_0})$$

. We present the results that are consequences of the Wasserstein and the L^2_μ-spectral gap in Section 4.1 and 4.2 respectively.

4.1 Consequences of the Wasserstein Spectral Gap

In this section we show the consequences of the Wasserstein spectral gap on the sample average. n Compared to the results from the L^2-spectral gap since they apply to a smaller class of observables, but they hold for the algorithm started at any deterministic point. Moreover, similar results also apply to non-reversible Markov processes that have a Wasserstein spectral gap.

4.1.1 Proper Metric and Lipschitz Functionals

For the CLT below we need a Wasserstein spectral gap with respect to a metric, since the Monge-Kantorovich duality is used for its proof [27]. The distance

$$\tilde{d} = \sqrt{(1 + \|x\|^i + \|y\|^i) \wedge \inf_{T,\psi \in \mathsf{A}(T,x,y)} \frac{1}{\epsilon} \int_0^T \exp(\eta \|\psi\|) dt (1 + \|x\|^i + \|y\|^i)}$$

does not necessarily satisfy the triangle inequality. Therefore we introduce

$$d' = \sqrt{(1 + \|x\|^i + \|y\|^i) \wedge \inf_{T,\psi \in \mathsf{A}(T,x,y)} \frac{1}{\epsilon} \int_0^T \exp(\eta \|\psi\|)(1 + \|\psi\|^i) dt}. \tag{4.1}$$

and show that $\tilde{d} \leq d' \leq C\tilde{d}$, thus exponential convergence transfers from $\tilde{d}$ to d'.

Lemma 4.1. *For the distance-like function $\tilde{d}$ and metric d' as above there is a C such that*

$$d' \leq \tilde{d} \leq Cd'.$$

Proof. Subsequently we assume without loss of generality that $\|y\| \geq \|x\|$. For any path $\psi \in \mathsf{A}$ we denote

$$F(\psi) = \frac{1}{\epsilon} \int_0^T \exp(\eta \|\psi\|)(1 + \|\psi\|^i) dt$$

by reflecting all points $\psi(t)$ in $B_{\|y\|}(0)^c$ at $\partial B_{\|y\|}(0)$ we make $F(\psi)$ smaller, hence we only have to consider ψ that satisfy

$$\|\psi(t)\| \leq \|y\|, \quad t \in [0, T] \tag{4.2}$$

. The first part follows due to $1 + \|\psi\|^i \leq 1 + \|x\|^i + \|y\|^i$.

For the second part we will use that only have to consider x and y such that

$$\inf_{T,\psi \in \mathsf{A}(T,x,y)} \frac{1}{\epsilon} \int_0^T \exp(\eta \|\psi\|) dt \leq (1 + \|x\|^i + \|y\|^i). \tag{4.3}$$

since the minimum expression in $\tilde{d}$ and d' have $(1 + \|x\|^i + \|y\|^i)$ in common.

We will first use this to show that x and y have to be close, if they are far out we will show that any path close to the infimum has to satisfy $\|y\| \geq \psi \geq \frac{\|y\|}{2}$, hence $1 + \|\psi\|^i$ and $(1 + \|x\|^i + \|y\|^i)$ are comparable. On fixed bounded sets d' and $\tilde{d}$ are comparable. In order to get a lower on $F(\psi)$ we distinguish between ψ intersects or does not intersect $B_R(0)$. If the path lies completely outside the ball we have

$$F(\psi) \geq \frac{1}{\epsilon} \|x - y\| \exp(\eta R)(1 + R^i)$$

if ψ and $B_R(0)$ have an intersection then ψ is longer than the shortest path to $B_R(0)$

$$\begin{aligned} F(\psi) &\geq \frac{1}{\epsilon}\int_0^{\|y\|-R} \exp(\eta(\|y\|-t))(1+(\|y\|-t)^i)dt \\ &\geq (\|y\|-R)\exp(\eta(\|y\|-R))(1+(\|y\|-R)^i)dt \end{aligned}$$

We choose $R = \frac{\|y\|}{2}$ and note that $\frac{\|y\|}{2} \geq \frac{\|x-y\|}{4}$, which yields in both cases

$$F(\psi) \geq \frac{1}{4\epsilon}\|x-y\|\exp(\eta\|y\|/2)(1+(\|y\|/2)^i).$$

By (4.3) this implies

$$\|x-y\| \leq 4\epsilon\exp(-\eta\frac{\|y\|}{2})\frac{1+\|x\|^i+\|y\|^i}{(1+(\frac{\|y\|}{2})^i)} \leq 4\epsilon\exp(-\eta\frac{\|y\|}{2})2^{i+1}. \tag{4.4}$$

For x and y in $B_Q(0)$ we have that $\tilde{d} \leq (2Q^i+1)^{\frac{1}{2}}d'$ because of (4.2). It is only left consider $x, y \in B_{\tilde{Q}}(0)^c$ for $\tilde{Q} = Q - 4\epsilon\exp(-\eta\frac{Q}{2})2^{i+1}$ since (4.4) holds. Subsequently we will show that for Q and hence $\tilde{Q}$ large enough it is sufficient to consider paths ψ that do not intersect $B_R(0)$ for $R = \frac{\|y\|}{2}$. Suppose the shortest the path would intersect $B_R(0)$ then the functional is larger than the shortest path to the boundary of the ball, hence

$$\begin{aligned} F(\psi) &\geq \frac{1}{\epsilon}\int_0^{\|y\|-R} e^{\eta(\|y\|-t)}(1+(\|y\|-t)^i)dt \\ &= \frac{1}{\epsilon}\Bigg[\exp(\eta\|y\|)(\eta^{-1}(1+\|y\|^i)+\sum_{j=1}^n \eta^{-1-j}\frac{i!}{(i-j)!}\|y\|^{i-j}) \\ &\qquad -\exp(\eta R)(\eta^{-1}(1+R^i)+\sum_{j=1}^n \eta^{-1-j}\frac{i!}{(i-j)!}R^{i-j})\Bigg] \end{aligned} \tag{4.5}$$

by $i+1$ integration by parts. Let l be the line connecting x and y, then using (4.5) yields

$$F(l) \leq \frac{1}{\epsilon}\|x-y\| e^{\eta\|y\|}(1+\|y\|^i) \leq 4\exp(\eta\frac{\|y\|}{2})2^{i+1}(1+\|y\|^i).$$

For Q and in turn $\tilde{Q}$ large enough we have $F(\psi) > F(l)$ by plugging $R = \frac{\|y\|}{2}$ into (4.5). Hence for all $t \in [0, L]$ $\|y\| \geq \psi \geq \|\mathrm{y}\|/2$ and therefore

$$2^{i+1}(1+\|\psi\|^i) \geq (1+\|x\|^i+\|y\|^i)$$

which yields that $\max(2L^i, 2^{i+1})d' \geq \tilde{d}$. $\square$

4.1.2 Strong Law of Large Numbers

In this section we will prove a strong law of large numbers for Lipschitz functions. Since μ (μ_m) is the unique invariant measures for $\mathcal{P}$ ($\mathcal{P}_m$), μ (μ_m) is ergodic and Birkhoff's ergodic theorem applies. Hence we only have to extend Birkhoff's theorem from almost every to every initial condition to get a strong law of large numbers.

Theorem 4.2. *In the setting of Theorem 2.12 or 2.14, suppose supp $\mu = \mathcal{H}$ and $h : \mathcal{H} \to \mathbb{R}$ has Lipschitz constant L with respect to $\tilde{d}$, then for arbitrary $X_0 \in \mathcal{H}$*

$$\left| \frac{1}{n} \sum_{i=1}^{n} h(X^i) - \mathbb{E}_\mu h \right| \overset{a.s}{\to} 0.$$

Proof. By Birkhoff's ergodic theorem we know that this is true for measurable h and a.e. initial condition. Because μ has full support for any $t > 0$ we can choose Y_0 such $\tilde{d}(X_0, Y_0) \leq t^2$ and Birkhoff's theorem applies to Y_0. Hence

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^{n} h(X^i) - \mathbb{E}_\mu h \right| &\leq \left| \frac{1}{n} \sum_{i=1}^{n} h(Y^i) - \mathbb{E}_\mu h \right| + \left| \frac{1}{n} \sum_{i=1}^{n} (h(X^i) - h(Y^i)) \right| \\ &\leq \left| \frac{1}{n} \sum_{i=1}^{n} h(Y^i) - \mathbb{E}_\mu h \right| + \frac{1}{n} \sum_{i=1}^{n} L\tilde{d}(X^i, Y^i). \end{aligned}$$

By the Wasserstein spectral gap we can couple X_n and Y_n such that

$$\mathbb{E}\tilde{d}(X^n, Y^n) \leq C r^n \tilde{d}(X^0, Y^0)$$

for some $0 < r < 1$. We then apply Markov's inequality to get

$$\mathbb{P}\left(\tilde{d}(X^n, Y^n) \geq c \right) \leq C \frac{r^n d(X^0, Y^0)}{c}.$$

Since Birkhoff's theorem applies to the Markov process started at Y_0 we have

$$\begin{aligned} \mathbb{P}\left(\limsup \left| \frac{1}{n} \sum_{i=1}^{n} h(X^i - \mathbb{E}_\mu h) \right| \geq c \right) &= \mathbb{P}\left(\limsup \frac{1}{n} \sum_{i=1}^{n} \left| h(X^i) - h(Y^i) \right| \geq c \right) \\ &\leq C \frac{L}{c(1-r)} d(X^0, Y^0). \end{aligned}$$

Setting $c = \frac{t}{L}$ yields

$$\mathbb{P}\left(\limsup \left| \frac{1}{n} \sum_{i=1}^{n} h(X^i - \mathbb{E}_\mu h) \right| \leq t \right) \geq 1 - t \frac{C}{1-r}.$$

Using the fact that for $A_1 \supseteq A_2 \dots \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ with $A = \bigcap_{i=1}^{\infty} A_i$ the result follows

$$\mathbb{P}\left(\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n} h(X^i) - \mathbb{E}_\mu h\right| = 0\right) = 1.$$

□

The above Theorem applies to a large class of functionals by the following sufficient criterion for $\tilde{d}$ Lipschitzness.

Lemma 4.3. *If $f : \mathcal{H} \to \mathbb{R}$ satisfies for all $R \in \mathbb{R}^+$*

$$\sup_{x,y\in B_R(0)} \frac{|f(x) - f(y)|}{\|x - y\|} \leq Ce^{\kappa R} \text{ for } \kappa < \eta \text{ and } \sup_{x\in B_R(0)} f(x) \leq C(1 + R^{\frac{i}{2}})$$

for all $R \in \mathbb{R}$, then f is Lipschitz with respect to $\tilde{d}$.

Proof. Subsequently we assume without loss of generality that $\|y\| \geq \|x\|$. From the arguments in Lemma 4.1 we know that for $\|x - y\| \geq 4\epsilon \exp(-\eta\frac{\|y\|}{2})2^{i+1}$ we have $\tilde{d} \geq \sqrt{1 + \|x\|^i + \|y\|^i}$, hence

$$|f(x) - f(y)| \leq |f(x)| + |f(y)| \leq Cd'$$

and we only have to consider x and y that are very close. Consider x and y such that $\|x - y\| \leq 4\epsilon \exp(-\eta\frac{\|y\|}{2})2^{i+1}$ then we have by arguments similar to those in the proof of Lemma 3.6:

$$\begin{aligned}
|f(x) - f(y)| &\leq \|f\|_{\mathrm{Lip}}\left(B_{\|x\|\vee\|y\|}(0)\right)\|x - y\| \\
&\leq \|f\|_{\mathrm{Lip}}\left(B_{\|x\|\vee\|y\|}(0)\right)\frac{\epsilon\exp(-\eta(\|x\|\vee\|y\| - \epsilon)\vee 0)}{1 + ((\|x\|\vee\|y\| - \epsilon)\vee 0)^i}\tilde{d}(x, y),
\end{aligned}$$

where the coefficient in front of $\tilde{d}$ is bounded by assumption. □

4.1.3 Central Limit Theorem

The result above does not give any rate of convergence. With a CLT on the other hand it is possible to derive (asymptotic) confidence intervals and so estimate the error for a finite n. We state a CLT that was proved by Komorowski and Walczuk in [27] and show that if the conditions of Theorem 2.12 or 2.14 is satisfied, then the result of Komorowski and Walczuk applies. This leads to:

Theorem 4.4. *If the conditions of Theorem 2.12 or 2.14 are satisfied, then there exists $\sigma \in [0, +\infty)$ such that*

$$\lim_{n\to+\infty}\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}\tilde{f}(X_s)\right)^2 = \sigma^2,$$

where $\tilde{f} := f - \mu(f)$ and f is Lipschitz with respect to d'. Moreover, we have

$$\lim_{T\to\infty} \mathbb{P}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{f}(X_s) < \xi) = \Phi_\sigma(\xi), \quad \forall \xi \in \mathbb{R}$$

where $\Phi_\sigma(\cdot)$ is the distribution function of $\mathcal{N}(0,\sigma^2)$ a zero mean normal law whose variance equals σ^2.

Let (E,ρ) be a Polish metric space and $\mathcal{P}^t$ be the transition probability semigroup for the E-valued Markov process X_t such that

Assumption 4.5.

1. The semigroup is Feller i.e $\mathcal{P}^t C_b(E) \subset C_b(E)$ and stochastic continuous i.e.
$$\lim_{t\to 0+} P^t f(x) = f(x), \quad \forall x \in E,\ f \in C_b(E).$$
2. we have $\mu P^t \in \mathcal{P}_i := \{\sigma | \sigma(E) = 1\,\&\,\int \rho_{x_0}(x)^i d\sigma(x) \leq \infty\}$ for any $\mu \in \mathcal{P}_i$ and $t \geq 0$,
3. for some $x_0 \in E$ there exist $\delta > 0$ such that for all $R < \infty$, and $T \geq 0$
$$\sup_{t\in[0,T]} \sup_{x\in B_R(x_0)} \int \rho_{x_0}^{2+\delta} P^t(x,dy) < \infty$$
4. there exist $x_0 \in E$ and $\delta > 0$ such that
$$A_* := \sup_{t\geq 0} \mathbb{E}\rho_{x_0}^{2+\delta}(X_t) < \infty$$
5. there exist $\hat{c}$, $\gamma > 0$ such that
$$d_1(\mu P^t, \nu P^t) \leq \hat{c}e^{-\gamma t} d_1(\mu,\nu) \in \mathcal{P}_1$$

Under this assumption their result reads

Proposition 4.6. *[27]Suppose that Assumption 4.5 is satisfied with $i = 1{,}2$ and μ_0 - the law of X_0 - belongs to $\mathcal{P}_1$. Then, for any observable $\psi \in Lip(E)$ the following are true:*

1. *(the weak law of large numbers) there exist $v_\star \in \mathbb{R}$ such that*
$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T \psi(X_s)ds = v_\star \quad \textit{in probability.}$$
2. *(asymptotic variance) For $\tilde{\psi} := \psi(x) - v_\star$ there is $\sigma \in [0,+\infty)$ such that*
$$\lim_{T\to+\infty} \frac{1}{T}\mathbb{E}(\int_0^T \tilde{\psi}(X_s)ds)^2 = \sigma^2.$$

3. *(the CLT) Let Φ_σ be the c.d.f. of a $\mathcal{N}(0,\sigma^2)$ random variable then*

$$\lim_{T\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{T}}\int_0^T \tilde{\psi}(X_s)ds < \xi\right) = \Phi_\sigma(\xi), \quad \forall \xi \in \mathbb{R}.$$

Proof of Theorem 4.4. Since convergence in $\|\cdot\|_{\mathcal{H}}$ is equivalent to convergence in d' $(\mathcal{H}, d')$ is complete and we will verify Assumption 4.5 for $\rho = d'$. The first part of the assumption is satisfied since $\mathcal{P}$ is Feller (c.f. section3.3) and stochastic continuity is not needed for time discrete processes. To verify the other assumptions we note that $\rho_{x_0}(x)^i$ is a Lyapunov function for $\mathcal{P}$ using the same argument as in Lemma 3.3. For the second part we note for every finite i

$$\mu\mathcal{P}^n(\rho_{x_0}^i) = \int l^n \rho_{x_0}^i + K d\mu(x) < \infty.$$

The fourth assumptions follows because $\rho_{x_0}^{2+\delta}$ is a Lyapunov function such that

$$\mathcal{P}^n \rho_{x_0}^{2+\delta} \leq l^n \rho_{x_0}^{2+\delta}(X_0) + K$$

and we can bound $A^\star \leq \rho_{x_0}^{2+\delta}(X_0) + K$. For third assumption we note

$$\sup_{i=0\ldots n} \sup_{x\in B_R(x_0)} \mathcal{P}\rho_{x_0}^{2+\delta}(x) \leq \sup_{x\in B_R(x_0)} \rho_{x_0}^{2+\delta}(x) + \frac{1}{1-l}K.$$

The last part is a consequence of Lemma 4.1 and Theorem 2.14.

□

4.2 Consequences of L^2_μ-Spectral Gap

Under the assumption of Theorem (2.12) or (2.14) we have shown the existence of an L^2_μ-spectral gap in Section 2.2.2. Now we can use all existing consequences for the ergodic average with and without burn in ($n_0 = 0$):

$$S_{n,n_0}(f) = \frac{1}{n}\sum_{j=1}^n f(X_{j+n_0}) \qquad S_n = S_{n,0}.$$

First of all we recall a general form of the spectral theorem for self-adjoint bounded operators (e.g. [42])

Proposition 4.7. *Let P be a bounded self-adjoint operator on some Hilbert space H. Then exist λ, Λ such that $\sigma(P) \subseteq [\lambda, \Lambda]$ and a operator-valued spectral measure with support in $[\lambda, \Lambda]$ such that*

$$\left\langle P^k f, g\right\rangle = \int_\lambda^\Lambda \alpha^k \left\langle E(d\alpha)f, g\right\rangle, \quad f, g \in H \text{ and } k \in \mathbb{N}.$$

Let $F : [\lambda, \Lambda] \to \mathbb{R}$ be a continuous function. Then one has by the continuous functional calculus a self-adjoint operator $F(P)$ with

$$\langle F(P)f, g\rangle = \int_{\lambda}^{\Lambda} F(\alpha) \langle E(d\alpha)f, g\rangle \;\; f, g \in H,$$

and

$$\|F(P)\|_{H\to H} = \max_{\alpha\in\sigma(P)} |F(\alpha)| .$$

In the setting of Proposition 4.7 we have due to the L^2_μ spectral gap $[\lambda, \Lambda] \subset [-\beta, \beta]$. As a consequence, the following result of [26] yields a CLT.

Proposition 4.8. *([26] Statement adapted from [30]).*

Suppose we have a reversible and ergodic Markov chain and a function $f \in L^2$. If

$$\sigma^2_{f,\mathcal{P}} = \int\limits_{[-1,1]} \frac{1+x}{1-x} \langle E(dx)f, f\rangle < \infty,$$

then for $X_0 \sim \mu$ the expression $\sqrt{n}(S_n - \mu(f))$ converges weakly to $\mathcal{N}(0, \sigma^2_{f,P})$.

In our case $\sigma^2_{f,\mathcal{P}}$ is bounded by $\frac{2\mu(f^2)}{1-\beta}$ which yields a uniform lower bound on the asymptotic variance in m. The result above has been extended to μ almost every initial condition in [12] which also applies to our case.

A different approach due to [46] is to consider the MSE

$$e_\nu(S_{n,n_0}, f) = (\mathbb{E}_{\nu,K} \left|S_{n,n_0}(f) - \mu(f)^2\right|)^{1/2}.$$

Using Tschebyscheff inequality this results in a confidence interval for $S(f)$. We can bound it by using the following proposition from [46]:

Proposition 4.9. *Suppose that we have a Markov chain with Markov operator $\mathcal{P}$ which has an L^2_μ-spectral gap $1 - \beta$. For $p \in (2, \infty]$ let $n_0(p)$ be the smallest natural number which is greater or equal to*

$$\frac{1}{\log(\beta^{-1})} \begin{cases} \frac{p}{2(p-2)} \log(\frac{32p}{p-2}) \left\|\frac{d\nu}{d\mu} - 1\right\|_{\frac{p}{p-2}} & p \in (2, 4) \\ \log(64) \left\|\frac{d\nu}{d\mu} - 1\right\|_{\frac{p}{p-2}} & p \in [4, \infty]. \end{cases} \tag{4.6}$$

Then

$$\sup_{\|f\|_p\leq 1} e_\nu(S_{n,n_0}, f) \leq \frac{2}{n(1-\beta)} + \frac{2}{n^2(1-\beta)^2}.$$

In our setting $n_0(p)$ is finite for $\nu = \gamma$ under the additional assumption that for all $u_1 > 0$ there is a u_2 such that

$$\Phi(\|x\|) \leq u_1 \|x\|^2 + u_2.$$

Using Fernique's theorem this implies that that $\frac{d\gamma}{d\mu} - 1$ has moments of all orders.

5 Conclusion

From an applications perspective, the primary thrust of this paper is to develop an understanding of MCMC methods in high dimension. Our work has concentrated on identifying the (possibly lack of) dimension dependence of spectral gaps for the standard random walk method RWM, and a recently developed variant pCN adapted to measures defined via density w.r.t a Gaussian. There are also variants of the Metropolis-adjusted Langevin algorithm (MALA) [5], as well as Hybrid Monte-Carlo methods [3] adapted to the sampling of measures defined via density w.r.t a Gaussian, and it would be interesting to employ the weak Harris theory to study these algorithms. Other classes of target measure, such as those arising from Besov prior measures [28, 14], or the uniform measures in [47], would also provide interesting applications. More generally, we expect that the weak Harris theory will be well-suited to the study of many MCMC methods in high dimensions, because of its roots in the study of Markov processes in infinite dimensional spaces [22]. In contrast, the theory developed in [37] does not work well for the kind of high dimensional problems that are studied here.

From a methodological perspective, we have demonstrated a particular application of the theory developed in [22], demonstrating its versatility for the analysis of rates of convergence in Markov chains. We have also shown how that theory, whose cornerstone is a Wasserstein spectral gap, may usefully be extended to study L^2 spectral gaps, and resulting sample path properties. These observations will be useful in a variety of applications, not just those arising in the study of MCMC.

All our results were presented for separable Hilbert spaces, but in fact all our results hold on an arbitrary Banach space by using a Gaussian series (c.f. Section 3.5 in [6]) instead of the Karhunen-Loeve expansion and the m-independence is due to Theorem 3.3.6 in [6].

A Gaussian measures

In this section we will derive the estimates for Gaussian measure that we needed above. In the whole section γ is Gaussian measure on a Banach space with covariance operator C_γ. Many estimates for Gaussian measures exploit their quadratic-exponential moments (see 3.2). Fernique's Theorem is often used to bound integrals over the whole domain. We will use it to derive bounds on an integral over the complement of a large ball:

$$\int_{\{\|u\|\geq K\}} h(u)d\gamma(u).$$

We need this to show that $\mathcal{P}$ and $\mathcal{P}_m$ is d-contracting (see Section 3.2.2).

Proposition A.1. *(Tail estimates)*

1. *For $f : \mathbb{R} \to \mathbb{R}$ we have*

$$\int\limits_{\|x\| \geq K} f(\|x\|) d\gamma = f(K)\gamma(\|x\| \geq K) + \int_K^\infty \gamma(\|x\| \geq t) f' dt$$

2. *For β small enough and $\alpha \in \mathbb{R}^+$ there is a constant $C_{\alpha.\beta}$ such that for $K > \frac{\alpha}{2\beta}$*

$$\int\limits_{\{\|u\| \geq K\}} \exp(\alpha \|u\|) d\gamma(u) \leq C_{\alpha,\beta} e^{-\beta K^2 + \alpha K}.$$

Proof. Using integration by parts we get the first part

$$\int\limits_{\|x\| \geq K} f(\|x\|) d\gamma \quad = \quad f(K)\gamma(\|x\| \geq K) + \int_K^\infty \gamma(\|x\| \geq t) f' dt.$$

For the second part we set $f(x) = \exp(\alpha x)$ in the above and use Lemma A.2

$$\int\limits_{\|x\| \geq K} \exp(\alpha \|x\|) d\gamma \quad \leq \quad F_\beta \exp(-\beta K^2 + \alpha K) + F_\beta \alpha \int_K^\infty \exp(-\beta t^2 + \alpha t).$$

For the integral on the right hand side we use substitution an a result from [40]

$$\begin{aligned}
\int_K^\infty \exp(-\beta t^2 + \alpha t) &= \exp(\frac{\alpha^2}{2\beta}) \int_K^\infty \exp\left(-\beta(t - \frac{\alpha}{2\beta})^2\right) dt \\
&= \exp(\frac{\alpha^2}{2\beta}) \int_{\sqrt{\beta}(K - \frac{\alpha}{2\beta})}^\infty \exp(-s^2) ds \\
&\leq \frac{1}{\sqrt{\beta}(K - \frac{\alpha}{2\beta}) + \sqrt{\beta(K - \frac{\alpha}{2\beta})^2 + \frac{4}{\pi}}} \exp(-\beta(K - \frac{\alpha}{2\beta})^2 + \frac{\alpha^2}{2\beta})
\end{aligned}$$

□

Lemma A.2. *Let u be distributed according to $\gamma = \mathcal{N}(0, C)$, then we have for $0 < \beta < \frac{1}{2\|C_\gamma\|}$*

$$\mathbb{P}(\|u\| \geq K) \leq F_\beta e^{-\beta K^2}.$$

Proof. By Fernique's theorem 3.2 we know that $\mathbb{E}(e^{\beta \|u\|^2}) = F_\beta < \infty$. By Markov's inequality it follows that

$$\mathbb{P}(\|u\| \geq K) \leq \frac{\mathbb{E}(e^{\beta \|u\|^2})}{e^{\beta K^2}} = F_\beta e^{-\beta K^2}.$$

□

References

[1] Robert J. Adler. *An introduction to continuity, extrema, and related topics for general Gaussian processes.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 12. Institute of Mathematical Statistics, Hayward, CA, 1990.

[2] K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, 245:493–501, 1978.

[3] A. Beskos, F Pinski, J. M. Sanz-Serna, and A. M. Stuart. Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Processes and Applications*, 2011.

[4] A. Beskos, G. O. Roberts, and A. M. Stuart. Optimal scalings for local Metropolis-Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.*, 19(3):863–898, 2009.

[5] A. Beskos, G. O. Roberts, A. M. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8(3):319–350, 2008.

[6] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[7] Vladimir I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[8] K. S. Chan and C. J. Geyer. Discussion: Markov chains for exploring posterior distributions. *The Annals of Statistics*, pages 1747–1758, 1994.

[9] Jeff Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pages 195–199. Princeton Univ. Press, Princeton, N. J., 1970.

[10] S. L. Cotter, M. Dashti, and A. M. Stuart. Approximation of Bayesian inverse problems for PDEs. *SIAM J. Numer. Anal.*, 48(1):322–345, 2010.

[11] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: Modifying old algorithms to make them faster. *preprint*, 2011.

[12] C. Cuny and M. Lin. Pointwise ergodic theorems with rate and application to the CLT for Markov chains. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 45, pages 710–733. Institut Henri Poincaré, 2009.

[13] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1992.

[14] M. Dashti, S. Harris, and A. M. Stuart. Besov priors for Bayesian inverse problems. http://arxiv.org/abs/1105.0889, 2011.

[15] M. Dashti and A. M. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. submitted, 2011.

[16] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[17] A. Frigessi, P. Di Stefano, C. R. Hwang, and S. J. Sheu. Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 205–219, 1993.

[18] C. J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

[19] J. K. Ghosh and R. V. Ramamoorthi. *Bayesian nonparametrics.* Springer Series in Statistics. Springer-Verlag, New York, 2003.

[20] M. Hairer. *An Introduction to Stochastic PDEs.* Lecture Notes, 2010.

[21] M. Hairer and A. J. Majda. A simple framework to justify linear response theory. *Nonlinearity*, 23(4):909–922, 2010.

[22] M. Hairer, J. C. Mattingly, and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2):223–259, 2011.

[23] M. Hairer, A. M. Stuart, and J. Voss. Analysis of SPDEs Arising in Path Sampling. Part 2: The Nonlinear Case. *Annals of Applied Probability*, pages 1657–1706, 2007.

[24] W. K. Hastings. Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97, 1970.

[25] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.

[26] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19, 1986.

[27] T. Komorowski and A. Walczuk. Central limit theorem for Markov processes with spectral gap in Wasserstein metric. *Arxiv preprint arXiv:1102.18422*, 2011.

[28] M. Lassas, E. Saksman, and S. Siltanen. Discretization invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging*, pages 87–122, 2009.

[29] K. Latuszynski and W. Niemiro. Rigorous confidence bounds for MCMC under a geometric drift condition. *J. Complexity*, 27(1):23–38, 2011.

[30] K. Latuszynski and G.O. Roberts. CLTs and asymptotic variance of time-sampled Markov chains. *Methodology and Computing in Applied Probability*, pages 1–11, 2011.

[31] G. F. Lawler and A. D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *American Mathematical Society*, 309(2), 1988.

[32] Peter M. Lee. *Bayesian statistics.* Arnold, London, third edition, 2004. An introduction.

[33] Jun S. Liu. *Monte Carlo strategies in scientific computing.* Springer Verlag, 2008.

[34] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms*, 4(4):359–412, 1993.

[35] J. C. Mattingly, N. S. Pillai, and A. M. Stuart. Diffusion limits of the random walk metropolis algorithm in high dimensions. *to appear Ann. Appl. Prob*, 2010.

[36] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087, 1953.

[37] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[38] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Probability Theory and Related Fields*, 43(4):309–318, 1978.

[39] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009.

[40] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, editors. *NIST Handbook of Mathematical Functions.* Cambridge University Press, 2010.

[41] N. S. Pillai, A. M. Stuart, and A. H. Thiery. On the random walk metropolis algorithm for Gaussian random field priors and the gradient flow. submitted, 2011.

[42] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I. Functional analysis.* Academic Press, New York, 1972.

[43] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2005.

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

[45] M. Röckner and F.-Y. Wang. Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups. *J. Funct. Anal.*, 185(2):564–603, 2001.

[46] Daniel Rudolf. *Explicit error bounds for Markov chain Monte Carlo*. PhD thesis, to appear in Dissertationes Mathematicae, Friedrich-Schiller-Universität Jena, 2011.

[47] Ch. Schwab and A. M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. Submitted `http://arxiv.org/abs/1103.4522`, 2011.

[48] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989.

[49] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[50] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.

[51] F.-Y. Wang. Functional inequalities for the decay of sub-Markov semigroups. *Potential Anal.*, 18(1):1–23, 2003.