The Annals of Applied Probability
2014, Vol. 24, No. 6, 2455–2490
DOI: 10.1214/13-AAP982

SPECTRAL GAPS FOR A METROPOLIS–HASTINGS ALGORITHM IN INFINITE DIMENSIONS

BY MARTIN HAIRER[1], ANDREW M. STUART[2] AND SEBASTIAN J. VOLLMER[3]

University of Warwick

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 make the chain reversible with respect to the target 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 exhibiting a 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 nonasymptotic bounds on the mean square error, all dimension independent. In contrast we show that the spectral gap of 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 an essential tool in Bayesian inverse problems [Stuart (2010)], Bayesian statistics [Lee (2004)], Bayesian nonparametrics [Hjort et al. (2010)], and conditioned diffusions [Hairer, Stuart and Voss (2007)]. For example, in inverse problems, some input data such as initial conditions or parameters for a forward mathematical model have to be determined from observations of noisy output. In the Bayesian approach, assuming a prior on the unknown input, and conditioning on the data, results in the posterior distribution, a natural target

Received December 2011; revised February 2013.

[1]Supported by EPSRC, the Royal Society, and the Leverhulme Trust.

[2]Supported by EPSRC and ERC.

[3]Supported by ERC.

MSC2010 subject classifications. 65C40, 60B10, 60J05, 60J22.

Key words and phrases. Wasserstein spectral gaps, L^2-spectral gaps, Markov chain Monte Carlo in infinite dimensions, weak Harris theorem, random walk Metropolis.

for sampling algorithms. In fact these sampling algorithms are also used in optimisation in form of simulated annealing [Geyer and Thompson (1995), Pillai, Stuart and Thiéry (2011)].

The most widely used method for general target measures are Markov chain Monte Carlo (MCMC) algorithms which use a Markov chain that in stationarity yields dependent samples from the target. Moreover, under weak conditions, a law of large numbers holds for the empirical average of a function f (observable) applied to the steps of the Markov chain. We quantify the computational cost of such an algorithm as

$$\text{number of necessary steps} \times \text{cost of a step.}$$

While for most algorithms the cost of one step grows with the dimension, a major result of this article is to exhibit an algorithm which, when applied to measures defined via a finite-dimensional approximation of a measure defined by a density with respect to a Gaussian random field, requires a number of steps independent of the dimension in order to achieve a given level of accuracy.

For ease of presentation we work on a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ equipped with a mean-zero Gaussian reference measure γ with covariance operator $\mathcal{C}$. Let $\{\varphi_n\}_{n\in\mathbb{N}}$ be an orthonormal basis of eigenvectors of $\mathcal{C}$ corresponding to the eigenvalues $\{\lambda_n^2\}_{n\in\mathbb{N}}$. Thus γ can be written as its Karhunen–Loève expansion [Adler (1990)]

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

and 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

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

With P_m being the projection onto the first m basis elements, we consider the following m-dimensional approximations to γ and μ:

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

The approximation error, namely the difference between μ and μ_m, is already well studied [Dashti and Stuart (2011), Mattingly, Pillai and Stuart (2012)] and can be estimated in terms of the closeness between $\Phi \circ P_m$ and Φ.

In this article we consider Metropolis–Hastings MCMC methods [Hastings (1970), Metropolis et al. (1953)]. For an overview of other MCMC methods, which have been developed and analysed, we refer the reader to Liu (2008), Robert and Casella (2004). The idea of the Metropolis–Hastings algorithm is to add an accept–reject mechanism to a Markov chain proposal in order to make the resulting Markov chain reversible with respect to the target measure. We denote the transition kernel of the underlying Markov chain by $Q(x, dy)$ and the acceptance probability for a proposed move from x to y by $\alpha(x, y)$. The transition kernel of the Metropolis–Hastings algorithm reads

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

where $\alpha(x, y)$ is chosen such that $\mathcal{P}(x, dy)$ is reversible with respect to μ [Tierney (1998)]. For the random walk Metropolis algorithm (RWM) the proposal kernel corresponds to

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

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

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

Notice that the quadratic form $\frac{1}{2}\langle y, \mathcal{C}^{-1}y\rangle$ is almost surely infinite with respect to the proposal because it corresponds to the Cameron–Martin norm of y. For this reason the RWM algorithm is not defined on the infinite dimensional Hilbert space $\mathcal{H}$ [consult Cotter et al. (2011) for a discussion], and we will study it only on m-dimensional approximating spaces. In this article we will demonstrate that the RWM can be considerably improved by using the preconditioned Crank–Nicolson (pCN) algorithm which is defined via

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

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

with $\xi \sim \gamma$. The pCN was introduced in Beskos et al. (2008) as the PIA algorithm, in the case $\alpha = 0$. Numerical experiments in Cotter et al. (2011) demonstrate its favourable properties in comparison with the RWM algorithm. In contrast to RWM, the acceptance probability is well defined on a Hilbert space, and this fact gives an intuitive explanation for the theoretical results derived in this paper in which we develop a theory explaining the superiority of pCN over RWM when applied on sequences of approximating spaces of increasing dimension. Our main positive results about pCN can be summarised in the following way (rigorous statements in Theorems 2.14, 2.15, 4.3 and 4.4):

CLAIM. *Suppose that both Φ and its local Lipschitz constant satisfy a growth assumption at infinity. Then for a fixed $0 < \delta \leq \frac{1}{2}$, the pCN algorithm applied to $\mu_m(\mu)$:*

I. *has a unique invariant measure μ_m (μ);*
II. *has a Wasserstein spectral gap uniformly in m;*

III. *has an* L^2*-spectral gap* $1-\beta$ *uniform in* m.

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

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

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

VI. *has an explicit bound on the mean square error* (*MSE*) *between itself and* $\mu(f)$ *for certain initial distributions* ν.

These positive results about pCN clearly apply to $\Phi = 0$ which corresponds to the target measures 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 Ornstein–Uhlenbeck (OU) process on $\mathcal{H}$ in this case. On the other hand the RWM algorithm applied to a specific Gaussian target measure γ_m has an L^2_μ-spectral gap which 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 establish the results I, II and IV for pCN and to establish the negative results for RWM, the statements III, V and VI follow by verification of the conditions of known results.

In addition to the significance of these results in their own right for the understanding of MCMC methods, we would also like to highlight the techniques that we use in the proofs. We apply recently developed tools for the study of Markov chains on infinite dimensional spaces; see Hairer, Mattingly and Scheutzow (2011). A weak version of Harris's theorem [proved in Hairer, Mattingly and Scheutzow (2011)] makes a Wasserstein spectral gap verifiable in practice, and for reversible Markov processes it even implies an L^2-spectral gap. Henceforth, we shall refer to this as the weak Harris theorem.

1.1. *Literature review.* The results in the literature can broadly be classified as follows [Rudolf (2012), Meyn and Tweedie (2009)]:

(1) For a metric on the space of measures such as the total variation or the Wasserstein metric, the rate of convergence to equilibrium can be characterised through the decay of $d(\nu\mathcal{P}^n, \mu)$ where ν is the initial distribution of the Markov chain.

(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 X to $\mathbb{R}$ modulo constants. The most prominent example here is the L^2-spectral gap.

(3) Direct methods like regeneration and the so-called split-chain which use the dynamics of the algorithm to introduce independence. The independence can be used to prove central limit theorem. Previous results have been formulated in terms of the following three main types of convergence:

Between these notions of convergence, there are many fruitful relations; for details consult Rudolf (2012). 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 Lawler and Sokal (1988) using the conductance concept due to Cheeger (1970). 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 Lovász and Simonovits (1993). The consequences of a spectral gap for the ergodic average in terms of a CLT and the MSE have been investigated in Cuny and Lin (2009), Kipnis and Varadhan (1986) and Rudolf (2012), respectively, and were first brought up in the MCMC literature in Chan and Geyer (1994), Geyer (1992).

For finite state Markov chains the spectral gap can be bounded in terms of quantities associated with its graph [Diaconis and Stroock (1991)]. This idea has also been applied to the Metropolis-algorithm in Sinclair and Jerrum (1989) and Frigessi et al. (1993).

A different approach using the splitting chain technique mentioned above was independently developed in Nummelin (1978) and Athreya and Ney (1978) 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 Meyn and Tweedie (2009) and again adapted and applied to the Metropolis–Hastings algorithm in Roberts and Tweedie (1996) resulting in a criterion for geometric ergodicity

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

Moreover, they established a criterion for a CLT. Extending this method, it was also possible to derive rigorous confidence intervals in Łatuszyński and Niemiro (2011).

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 pCN algorithm is not irreducible in the sense of Meyn and Tweedie (2009), hence there is no nontrivial 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. More precisely if $x - y$ is not in the Cameron–Martin space $Q(x, dz)$ and $Q(y, dz)$ are mutually singular, consequently the same is true for $P(x, dz)$ and $P(y, dz)$. 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 [Da Prato and Zabczyk (1992)].

For these reasons, existing theoretical results concerning RWM and pCN in high dimensions have been confined to scaling results and derivations of diffusion limits. In Beskos, Roberts and Stuart (2009) the RWM algorithm with a target that is absolutely continuous with respect to a product measure has been analysed for its dependence on the dimension. The proposal distribution is a centred 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 to zero. In a similar setup it was recently shown that there is a μ-reversible SPDE limit if the product law is a truncated Karhunen–Loève expansion [Mattingly, Pillai and Stuart (2012)]. This SPDE limit suggests that the number of steps necessary for a certain level of accuracy grows like $\mathcal{O}(m)$ because $\mathcal{O}(m)$ steps are necessary in order to approximate the SPDE limit on $[0, T]$. A similar result in Pillai, Stuart and Thiéry (2011) suggests that the pCN algorithm only needs $\mathcal{O}(1)$ steps.

Uniform contraction in a Wasserstein distance was first applied to MCMC in Joulin and Ollivier (2010) in order to get bounds on the variance and bias of the sample average of Lipschitz functionals. We use the weak Harris theorem to verify this contraction, and by using the results from Rudolf (2012), we obtain nonasymptotic bounds on the sample average of L^2_μ-functionals. In Eberle (2014) exponential convergence for a Wasserstein distance is proved for the Metropolis-adjusted-Langevin (MALA) and pCN algorithm for log-concave measures having a density with respect to a Gaussian measure. The rates obtained in this article are explicit in terms of additional bounds on the derivates of the density. In our proofs we do not assume log-concavity. However, the rate obtained here is less explicit.

Similarly, approaches based on the Bakry–Emery criterion [Bakry and Émery (1985)] seem to be only applicable if the measure is log-concave.

1.2. *Outline*. In this paper we substantiate these ideas by using spectral gaps derived by an application of the weak Harris theorem [Hairer, Mattingly and Scheutzow (2011)]. Section 2 contains the statements of our main results, namely Theorems 2.12, 2.14 and 2.15 concerning the desirable dimension-independence properties of the pCN method and Theorem 2.17 dealing with the undesirable dimension dependence of the RWM method. Section 2 starts by specifying the RWM and pCN algorithms as Markov chains, the 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. The proofs of the theorems in Section 2 are given in Section 3. We highlight that the key steps can be found in the Sections 3.1.2 and 3.2.2 where we dealt with the cases of global and local Lipschitz Φ, respectively. 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 emphasising the dimension independence of these results. We draw overall conclusions in Section 5.

2. Main results. In Section 2.1 we specify the RWM and pCN algorithms before summarising the weak Harris theorem in Section 2.2. Subsequently, we describe how a Wasserstein spectral gap implies an L^2_μ-spectral gap. Based on the

Algorithm 1 Preconditioned Crank–Nicolson

Initialise X_0.
For $n \geq 0$ do:

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

$$X_{n+1} = \begin{cases} p_{X_n}, & \text{with probability } \alpha(X_n, p_{X_n}), \\ X_n, & \text{otherwise.} \end{cases}$$

Here, $\alpha(x, y) = 1 \wedge \exp(\Phi(x) - \Phi(y))$.

weak Harris theorem, we give necessary conditions on the target measure for the pCN algorithm in order to have a dimension independent spectral gap in a Wasserstein distance in Section 2.3. In Section 2.4 we highlight one of the disadvantages of the RWM by giving an example satisfying our assumptions for the pCN algorithm for which the spectral gap of the RWM algorithm converges to zero as fast as any negative power of m as $m \to \infty$.

2.1. *Algorithms*. We focus on convergence results for the pCN algorithm (Algorithm 1) which generates a Markov chain $\{X^n\}_{n\in\mathbb{N}}$ with $X^n \in H$ and $\{X_m^n\}_{n\in\mathbb{N}}$ when it is applied to the measures 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 the statements concerning the pCN and RWM algorithms occur in separate sections.

2.2. *Preliminaries*. In this section we review Lyapunov functions, Wasserstein distances, d-small sets and d-contracting Markov kernels in order to state the weak Harris theorem of Hairer, Mattingly and Scheutzow (2011). By weakening the notion of small sets, this theorem gives a sufficient condition for exponential con-

Algorithm 2 Random walk Metropolis

Initialise X_0.
For $n \geq 0$ do:

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

$$X_{n+1} = \begin{cases} p_{X_n}, & \text{with probability } \alpha(X_n, p_{X_n}), \\ X_n, & \text{otherwise.} \end{cases}$$

Here, $\alpha(x, y) = 1 \wedge \exp(\Phi(x) - \Phi(y) + \frac{1}{2}\langle x, \mathcal{C}^{-1}x\rangle - \frac{1}{2}\langle y, \mathcal{C}^{-1}y\rangle)$.

vergence in a Wasserstein distance. Moreover, we explain how this implies an L^2-spectral gap.

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 the 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 that

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

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

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

DEFINITION 2.3. Let $\mathcal{P}$ be a Markov operator on 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 $x, y \in S$

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

REMARK. The d-Wasserstein distance associated with

$$d(x, y) = \chi_{\{x \neq y\}}(x, y)$$

coincides with the total variation distance (up to a factor 2). If S is a small set Meyn and Tweedie (2009), then there exists a probability measure ν such that $\mathcal{P}$ can be decomposed into

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

This implies that $d_{\mathrm{TV}}(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) \leq s$ and 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

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

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

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

(Note that the bound for $n = 1$ implies all other bounds but with a different constant K.)

PROPOSITION 2.6 (Weak Harris theorem [Hairer, Mattingly and Scheutzow (2011)]). *Let $\mathcal{P}$ be a Markov kernel over a Polish space $\mathbf{E}$, and 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 $S = \{x \in \mathbf{E} : V(x) \le 4K\}$ is d-small.*

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

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

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. In particular there is at most one invariant measure. Moreover, if there exists a complete metric d_0 on $\mathbf{E}$ such that $d_0 \le \sqrt{d}$ and such that $\mathcal{P}$ is Feller on $\mathbf{E}$, then there exists a unique invariant measure μ for $\mathcal{P}$.

REMARK. Setting $\nu_2 = \mu$ we obtain the convergence rate to the invariant measure.

2.2.2. *The Wasserstein spectral gap implies an L^2-spectral gap.* In this section we give reasons why a Wasserstein spectral gap implies an L^2_μ-spectral gap under mild assumptions for a Markov kernel $\mathcal{P}$. The proof is based on a comparison of different powers of $\mathcal{P}$ using the spectral theorem.

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

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

The following proposition is a discrete-time version of Theorem 2.1(2) in Wang (2003). The proof given below is from private communication with Wang and is presented because of its beauty and the tremendous consequences in combination with the weak Harris theorem.

PROPOSITION 2.8 (Private communication [Röckner and Wang (2001)]). *Let $\mathcal{P}$ be a Markov transition operator which is reversible with respect to μ and suppose that* $\mathrm{Lip}(\tilde{d}) \cap L^\infty_\mu$ *is dense in L^2_μ. If for every such f there exists a constant $C(f)$ such that*

$$\tilde{d}\big((\mathcal{P}^n f)\mu, \mu\big) \le C(f)\exp(-\lambda n)\tilde{d}(f\mu, \mu),$$

then this implies the L^2_μ-spectral gap

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

PROOF. First assume that $0 \le f \in \mathrm{Lip}(\tilde{d}) \cap L^\infty(\mu)$ with $\mu(f) = 1$ and π being 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

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

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

$$\|P_t f - \mu(f)\|_2^2 \le 2\|P_t f^+ - \mu(f^+)\|_2^2 + 2\|P_t f^- - \mu(f^-)\|_2^2.$$

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

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

$$\|\mathcal{P}^n f - \mu(f)\|_2^2 \le C(f)\exp(-\lambda n) \qquad \text{for all } n,$$

then for the same f,

$$\|\mathcal{P}^n f - \mu(f)\|_2^2 \le \|f - \mu(f)\|_2^2 \exp(-\lambda n) \qquad \text{for all } n.$$

PROOF. Without loss of generality we assume that $\mu(\hat{f}^2) = 1$ 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. Thus for $k \in \mathbb{N}$,

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

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

2.3. *Dimension-independent spectral gaps for the pCN-algorithm.* 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 $\mathcal{P}$ and $\mathcal{P}^m$. 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 $\rho = 1-(1-2\delta)^{1/2}$, we can express the proposal of the pCN algorithm as

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

The following results do all hold for δ in $(0, \frac{1}{2}]$:

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). This relies on having a lower bound on the probability of X_{n+1} being in a ball around the mean. In fact, our assumptions are stronger because 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 $R > 0$, $\alpha_l > -\infty$ 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 all $x \in B_R(0)^c$,

$$\inf_{z\in B_{r(\|x\|)}((1-\rho)x)} \alpha(x,z) = \inf_{z\in B_{r(\|x\|)}((1-\rho)x)} \exp\bigl(-\Phi(z)+\Phi(x)\bigr) > \exp(\alpha_l). \tag{2.4}$$

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 Assumptions 2.10 and 2.11 be satisfied with either*:

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

(2) $r(\|x\|) = r \in R$ *for* $r \in \mathbb{R}^+$, *and then we take* $V = \|x\|^i$ *with* $i \in \mathbb{N}$.

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

$$\tilde{d}(x,y) = \sqrt{d(x,y)\bigl(1+V(x)+V(y)\bigr)} \quad \textit{with}$$

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

Then for ε *small enough there exists an* $\tilde{n}$ *such that for all probability measures* ν_1 *and* ν_2 *on* $\mathcal{H}$ *and* $P_m\mathcal{H}$, *respectively*,

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

$$\tilde{d}\bigl(\nu_1\mathcal{P}_m^{\tilde{n}}, \nu_2\mathcal{P}_m^{\tilde{n}}\bigr) \leq \tfrac{1}{2}\tilde{d}(\nu_1,\nu_2)$$

for all $m \in \mathbb{N}$.

PROOF. The conditions of the weak Harris theorem (Proposition 2.6) are satisfied by the Lemmata 3.2, 3.3 and 3.4. $\square$

A key step in the proof is to verify the d-contraction. In order to obtain 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 is accepted is the most difficult to tackle. By choosing $d = 1 \wedge \frac{\|x-y\|}{\varepsilon}$ with ε small enough, it turns out that the Lipschitz constant of $\alpha(x,y)$ can be brought under control.

By changing the distance function d, we can also handle the case when Φ is locally Lipschitz 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 the Assumptions 2.10 and 2.13 be satisfied with $r(\|x\|) = r\|x\|^a$ where $r \in \mathbb{R}, a \in (\frac{1}{2}, 1)$ and either $V = \|x\|^i$ with $i \in \mathbb{N}$ or $V = \exp(v\|x\|)$.*

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

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

$$\tilde{d}(x,y) = \sqrt{d(x,y)\bigl(1+V(x)+V(y)\bigr)} \quad \textit{with}$$

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

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

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

$$\tilde{d}\bigl(\nu_1\mathcal{P}_m^{\tilde{n}}, \nu_2\mathcal{P}_m^{\tilde{n}}\bigr) \leq \tfrac{1}{2}\tilde{d}(\nu_1,\nu_2).$$

PROOF. This time Lemmata 3.2, 3.6 and 3.7 verify the conditions of the weak Harris theorem (Proposition 2.6). $\square$

REMARK. Our arguments work for $\delta \in (0, \frac{1}{2}]$; for $\delta = \frac{1}{2}$, the pCN algorithm becomes the independence sampler, and the Markov transition kernel becomes irreducible so that this case we can use the theory of Meyn and Tweedie (2009).

In order to get the same lower bound for the L^2_μ-spectral gap, we just have to verify that $\mathrm{Lip}(\tilde{\delta}) \cap L^\infty_\mu$ 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 $\mathrm{Lip}(\tilde{d}) \cap L^\infty(\mu)$ is dense in $L^2(H, \mathcal{B}, \mu)$. Since $\tilde{d}(x, y) \geq C(1 \wedge \|x - y\|)$, one has $\mathrm{Lip}(\|\cdot\|) \cap L^\infty(\mu) \subseteq \mathrm{Lip}(\tilde{d})$, so that it is enough to show that $\mathrm{Lip}(\|\cdot\|) \cap L^\infty(\mu)$ is dense in $L^2(H, \mathcal{B}, \mu)$. Suppose not; then there is $0 \neq g \in L^2(\mu)$ such that

$$\int fg \, d\mu = 0 \qquad \text{for all } f \in \mathrm{Lip} \cap L^\infty(\mu).$$

Since all Borel probability measures on a separable Banach space are characterised by their Fourier transform [see, e.g., Bogachev (2007)], they are characterised by integrals against bounded Lipschitz functions. Hence $g \, d\mu$ is the zero measure and hence $g \equiv 0$ in L^2_μ. □

2.4. *Dimension-dependent spectral gaps for RWM.* So far we have shown convergence results for the pCN. Therefore we present an example subsequently 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_m on

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

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

$$\mu_m = \gamma_m = \mathcal{L}\left(\sum_{i=1}^m \frac{1}{i} \xi_i e_i \right), \qquad \xi \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1). \tag{2.5}$$

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 $\delta = sm^{-a}$ for any $a \in [0, 1)$.

Using the notion of conductance,

$$\mathsf{C} = \inf_{\mu(A) \leq 1/2} \frac{\int_A \mathcal{P}(x, A^c) \, d\mu(x)}{\mu(A)}, \tag{2.6}$$

we obtain an upper bound on the spectral gap by Cheeger's inequality [Lawler and Sokal (1988), Sinclair and Jerrum (1989)],

$$1 - \beta \leq 2\mathsf{C}. \tag{2.7}$$

Our main observation is that there is a simple upper bound for the conductance of a Metropolis–Hastings algorithm because it can only move from a set A if:

- the proposed move lies in A^c, and
- the proposed move is accepted.

Just considering either event gives rise to simple upper bounds that can be used to make many results from the scaling analysis rigorous. We denote the expected acceptance probability for a proposal from x as

$$\alpha(x) = \int_{\mathcal{H}} \alpha(x, y)\, dQ(x, dy).$$

Considering only the acceptance of the proposal gives rise to

$$C \le \inf_{\mu(A)\le 1/2} \frac{\int_A \alpha(x)\mu(dx)}{\mu(A)}.$$

In particular, for any set B such that $\mu(B) \le \frac{1}{2}$, it follows that

$$C \le \sup_{x\in B} \alpha(x)$$

and also that

$$C \le 2\mathbb{E}_\mu \alpha(x).$$

The last result allows us to make scaling results like those in Beskos, Roberts and Stuart (2009) rigorous. Similarly, just supposing that the Metropolis–Hastings algorithm accepts all proposals gives rise to the following bound:

$$C \le \inf_{\mu(A)\le 1/2} \frac{\int_A Q(x, A^c)\, d\mu(x)}{\mu(A)}.$$

We summarise these observations in the subsequent proposition.

PROPOSITION 2.16. *Let $\mathcal{P}$ be a Metropolis–Hastings transition kernel for a target measure μ with proposal kernel $Q(x, dy)$ and acceptance probability $\alpha(x, y)$. The L^2_μ-spectral gap can be bounded by*

$$1 - \beta \le 1 - \Lambda \le 2C \le 2 \begin{cases} \sup_{x\in B} \alpha(x), & \text{for any } \mu(B) \le \dfrac{1}{2}, \\ \mathbb{E}_\mu \alpha(x), \end{cases} \tag{2.8}$$

and

$$1 - \beta \le 1 - \Lambda \le 2C \le 2 \inf_{\mu(A)\le 1/2} \frac{\int_A Q(x, A^c)\, d\mu(x)}{\mu(A)}. \tag{2.9}$$

In the following theorem we use the Proposition 2.16 for the RWM algorithm applied to μ_m as in equation (2.5) in order to quantify the behaviour of the spectral gap as m goes to ∞. We consider polynomial scaling of the step size parameter of the form $\delta_m \sim m^{-a}$ to zero. For $a < 1$ the bound in equation (2.8) is most useful as the acceptance behaviour is the determining quantity. For $a \ge 1$ the bound in equation (2.9) is most useful as the properties of the proposal kernel are determining in this regime.

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 equation* (2.5).

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

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

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

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

Using this inequality, we can find an upper bound on the acceptance probability $\alpha(x)$.

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

Completing the square and using the normalisation constant yields

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

$$\le (1+2\lambda\delta)^{-m/2} \exp\left(\sum_{i=1}^m \frac{\delta\lambda^2 i^2 x_i^2}{(2\delta\lambda+1)}\right).$$

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

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

We want to choose a and b in the above equation such that the right-hand side goes to zero as $m \to \infty$. In order to obtain decay from the first factor, we need that $a+b<1$ and to prevent growth from the second $a+2b > 2-2\sigma$ 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

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

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 using a symmetry argument satisfies $\gamma_m(A) = \frac{1}{2}$ to bound the conductance

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

Combining Fernique's theorem and Markov's inequality yields

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

so that the claim follows again by an application of 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 the Sections 3.1 and 3.2, respectively. For each condition we use the following lemma for the dependence of the constants l, K, c and s in the weak Harris theorem on m. This allows us to conclude that there exists $\tilde{n}(m) \le \tilde{n}$ such that

$$\tilde{d}(\nu_1 \mathcal{P}^{\tilde{n}}, \nu_2 \mathcal{P}^{\tilde{n}}) \le \tfrac{1}{2}\tilde{d}(\nu_1, \nu_2),$$
$$\tilde{d}(\nu_1 \mathcal{P}_m^{\tilde{n}(m)}, \nu_2 \mathcal{P}_m^{\tilde{n}(m)}) \le \tfrac{1}{2}\tilde{d}(\nu_1, \nu_2)$$

for all measures ν_1, ν_2 probability measures on $\mathcal{H}$ and $P_m\mathcal{H}$, respectively.

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

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

$$\int f(\|\xi\|)\, d\gamma_m(\xi) \le \int f(\|\xi\|)\, d\gamma(\xi),$$

and in particular

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

PROOF. The truncated Karhunen–Loève expansion relates γ_m to 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 of the function f

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

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

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 2.6 for the distance-like functions

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

3.1.1. *Lyapunov functions.* Under Assumption 2.10 we show the existence of a Lyapunov function V. This follows from two facts: First, the decay of V on $B_{r(\|x\|)}((1-\rho)x)$ and second the probability of the next step of the algorithm lying in that ball can be bounded below by Fernique's theorem; see Proposition A.1. Similarly, we will use the second part of Proposition A.1 to deal with proposals outside $B_{r(\|x\|)}((1-\rho)x)$.

LEMMA 3.2. *If Assumption 2.10 is satisfied with*:

(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* $V(x) = \exp(\ell\|x\|)$ *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. Then there exists a constant K_1 such that

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

On the other hand, there exists $0 < \tilde{l} < 1$ such that for all $x \in B_R(0)^c$,

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

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

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

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

$$\begin{aligned}\mathbb{E}\big(V(p_x) \vee V(x); A^c\big) &\le \int_{\sqrt{2\delta}\|\xi\| \ge c} \|x\|^i \vee \big((1-\rho)\|x\| + \sqrt{2\delta}\|\xi\|\big)^i \, d\gamma(\xi)\\ &\le \int_{\|\xi\| \ge c/\sqrt{2\delta}} \big(\|x\|^i + K\|\xi\|^p\big)\, d\gamma(\xi)\\ &\le \mathbb{P}\big(A^c\big)V(x) + K_2.\end{aligned}$$

Because a ball around the mean of a Gaussian measure on a separable space always has positive mass [Theorem 3.6.1 in Bogachev (1998)], we note that

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

for some constant $l < 1$.

For the second case we estimate

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

The right-hand side of the 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 exists an $l < 1$ such that

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

For the m-dimensional approximation $\mathcal{P}_m$ the probability of the event A is larger than $\mathcal{P}$ by Lemma 3.1. Since there is a common lower bound for $\mathbb{P}(\text{accept}|A)$ $l(m)$ is smaller than or equal to l. Similarly, $K_i(m)$ is smaller than K_i by Lemma 3.1. □

3.1.2. *The d-contraction.* In this section we show that $\mathcal{P}$ is d-contracting for $d(x, y) = 1 \wedge \frac{\|x-y\|}{\varepsilon}$ 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. The situation is illustrated in Figure 1. 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:

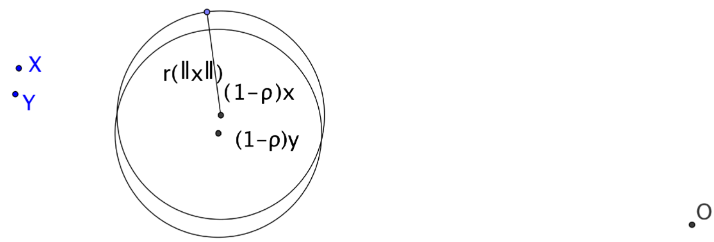

FIG. 1. *Contraction.*

(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 3.3. *If Φ in (1.1) satisfies Assumptions 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 that $\|x - y\| < \varepsilon$. Later we will choose $\varepsilon \ll 1$ so that if $\|x - y\| < \varepsilon$, 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 both cases separately. We assume without loss of generality that $\|y\| \geq \|x\|$.

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

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

where the last term bounds the case that only one of the proposals is accepted. Using the bound $\mathbb{P}(\text{both reject|A}) \leq 1 - \mathbb{P}(\text{both accept}|A)$ yields a nontrivial convex combination of d and $(1 - \rho)d$ because 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 nontrivial convex combination since $\mathbb{P}(A) > 0$ so that there is $\tilde{c} < 1$ with

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

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

$$\begin{aligned}
&\int_X \big|\alpha(x, p_x)(\xi) - \alpha(y, p_y)(\xi)\big| \, d\gamma(\xi) \\
&\quad \leq \int_{\mathcal{H}} \big|\Phi(p_x) - \Phi(p_y)\big| + \big|\Phi(x) - \Phi(y)\big| \, d\gamma(\xi) \qquad (3.5)\\
&\quad \leq 2L|x - y| \leq 2L\varepsilon d(x, y)
\end{aligned}$$

which yields an overall contraction for ε small enough.

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

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

The lower bound for $\mathbb{P}(\text{both accept}|B)$ follows this time from 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. $\square$

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, then d drops below $\frac{1}{2}$. Hence

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

LEMMA 3.4. *If S is bounded, then there exists 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),

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

PROOF. In order to obtain an upper bound for $d(\mathcal{P}^n(x,\cdot),\mathcal{P}^n(y,\cdot))$, 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, it follows by the definition of d [cf. (3.2)] that

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

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

$$d\big(\mathcal{P}^n(x,\cdot),\mathcal{P}^n(y,\cdot)\big) \le \mathbb{P}(B)\tfrac{1}{2} + \big(1-\mathbb{P}(B)\big)\cdot 1 < 1.$$

Writing ξ^i for the noise in the ith step, we bound

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

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

$$-\sup_{z\in B_{2R}(0)}\Phi(z) + \inf_{z\in B_{2R}(0)}\Phi(z) \le -\sup_{z\in B_{2R}(0)}\Phi(P_n z) + \inf_{z\in B_{2R}(0)}\Phi(P_n z).$$

Hence 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. We used that Φ is globally Lipschitz to prove that $\mathcal{P}$ and $\mathcal{P}_m$ is d-contracting; cf. equation (3.5). Now there is no one fixed ε that makes $\mathcal{P}$ d-contracting. Instead the idea is to change the metric in a way such that two points far out have to be closer in $\|\cdot\|_{\mathcal{H}}$ in order to be considered "close," that is, $d(x, y) < 1$. This is inspired by constructions in Hairer and Majda (2010), Hairer, Mattingly and Scheutzow (2011). 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 the two metrics d and $\bar{d}$ by

$$d(x, y) = 1 \wedge \bar{d}(x, y), \qquad \bar{d}(x, y) = \inf_{T, \psi \in \mathsf{A}(T, x, y)} \frac{1}{\varepsilon} \int_0^T \exp(\eta \|\psi\|)\, dt, \tag{3.6}$$

where ε and η will be chosen depending on Φ and γ in the subsequent proof. The situation is different from before because even in the case when "both accept," the distance can increase because of the weight. In order to control this, we notice the following:

LEMMA 3.5. *Let ψ be a path connecting x, y with $\|\dot{\psi}\| = 1$, then for $\bar{d}$ as in* (3.6):

(1) $\frac{1}{\varepsilon} \int_0^T \exp(\eta \|\psi\|)\, dt < 1$ *implies*

$$T \le J := \varepsilon \exp(-\eta(\|x\| \vee \|y\| - \varepsilon) \vee 0) \le \varepsilon;$$

(2) $\bar{d}(x, y) \le \frac{\|x-y\|}{\varepsilon} \exp(\eta(\|x\| \vee \|y\|))$ *and*

$$\frac{\|x - y\|}{\varepsilon} \exp(\eta(\|x\| \vee \|y\| - J) \vee 0) \le \bar{d}(x, y)$$

for all points such that $\bar{d}(x, y) < 1$;

(3) *for points such that* $\bar{d}(x, y) < 1$

$$\frac{\bar{d}(p_x, p_y)}{\bar{d}(x, y)} \le (1 - 2\delta)^{1/2} e^{-\eta\rho[\|x\| \vee \|y\| + \eta(\|\sqrt{2\delta}\xi\| + J)]}.$$

PROOF. In order to prove the first statement, we observe that

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

For the second part we denote by ψ the line segment connecting x and y in order to obtain an upper bound $d(x, y)$. For the lower bound we use $\|\psi\| \geq (\|x\| \vee \|y\| - J) \vee 0$ from the first part combined with the fact that $T \leq \varepsilon$. Using the second part we get

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

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 3.2 applies.

3.2.2. *The d-contraction.* The main difference between local and global Lipschitz Φ is proving that $\mathcal{P}$ and $\mathcal{P}_m$ is d-contracting.

LEMMA 3.6. *If Φ satisfies Assumptions 2.10 and 2.13, then $\mathcal{P}$ and $\mathcal{P}_m$ are d-contracting for d as in (3.6) 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 $A = \{\omega | \|\xi\| \leq \frac{2R}{\sqrt{2\delta}}\}$. First we choose R large, before dealing with the case when η is small and when ε is small. We have

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

where the first two lines deal with both accept and both reject in the case of A, the third and fourth line consider the same case in the event of A^c. The last line deals with the case when only one accepts. For the first two lines of equation (3.7) 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 3.5 that

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

where the last step follows for η small enough. Using the complementary probability, we obtain the following estimate:

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

Combining both estimates, it follows that $\mathbb{P}(A)(1 - \mathbb{P}(\text{both accept}|A)(1 - \tilde{\rho}))$ as coefficient in front of $d(x, y)$. In order to show that $\mathcal{P}$ is d-contracting, we have to prove that the expression in the third and fourth line of equation (3.7) is close to $\mathbb{P}(A^c) \cdot d(x, y)$. We notice that

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

where the last step followed by Lemma 3.5. 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 $\chi_{\sqrt{2\delta}\|\xi\|>2R}(1 \vee \exp(\eta(\sqrt{2\delta}\|\xi\| + J)))$ and applying Lemma 3.1, we conclude that this estimate holds uniformly in m. Combining 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 $d(x, y)$,

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

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

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

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

If "both accept," then the contraction factor associated to the event of A is smaller than $(1-\rho)$ because $r(\|x\|) \leq \frac{\rho}{2}\|x\|$ and by an application of Lemma 3.5. For the next term it follows that

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

Denoting the integral above by I, its integrand by $f(\xi)$ and $F > 0$, this yields

$$\begin{aligned}I \leq I_1 + I_2 = &\int_{\rho(\|y\|-J)+F \geq \|\sqrt{2\delta}\xi\| \geq r(\|x\| \wedge \|y\|)} f(\xi)\, d\gamma(\xi) \\ &+ \int_{\|\sqrt{2\delta}\xi\| \geq \rho(\|y\|-J)+F} f(\xi)\, d\gamma(\xi).\end{aligned}$$

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

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

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

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

For any $\tau > 0$ we choose F large enough and then η small enough so that $I \leq (1+\tau)\mathbb{P}(A^c)$. 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 $d(x, y)$. Calculating as above yields

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

where the last step follows using the upper bound for $\|x - y\|$ from Lemma 3.5. 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. □

3.2.3. *The d-smallness.* Analogously to the globally Lipschitz case, we have the following:

LEMMA 3.7. *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* (3.6),

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

PROOF. By Lemma 3.4, 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 which are all elements of $B_{2R}(0)$, then for n large enough,

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

Hence the result follows analogue to Lemma 3.4. □

4. Results concerning the sample-path average. In this section we focus on sample path properties of the pCN algorithm which can be derived from the Wasserstein and the L^2_μ-spectral gap. We prove a strong law of large numbers, a CLT and a bound on the MSE. This allows us to quantify the approximation of $\mu(f)$ by

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

4.1. *Consequences of the Wasserstein spectral gap.* The immediate consequences of a Wasserstein spectral gap are weaker than the results from the L^2-spectral gap because they apply to a smaller class of observables, but they hold for the algorithm started at any deterministic point.

4.1.1. *Change to a proper metric and implications for Lipschitz functionals.* For the Wasserstein CLT [Komorowski and Walczuk (2012)] we need a Wasserstein spectral gap with respect to a metric. The reason for this is that the Monge–Kantorovich duality is used for its proof Komorowski and Walczuk (2012). Recall that Theorem 2.14 yields a Wasserstein spectral gap for the "distance"

$$\tilde{d}(x,y) = \sqrt{\big(1 + \|x\|^i + \|y\|^i\big)(1 \wedge d)} \qquad \text{where}$$
$$d(x,y) = \inf_{T,\psi \in \mathsf{A}(T,x,y)} \frac{1}{\varepsilon}\int_0^T \exp\big(\eta\|\psi\|\big).$$

Because $\tilde{d}$ does not necessarily satisfy the triangle inequality, we introduce

$$
\begin{aligned}
d'(x,y) &= \sqrt{\inf_{\substack{x=z_1,\dots,z_n=y\\ n\geq 2}} \sum_{j=1}^{n-1} d_0(z_j,z_{j+1})},\\
d_0(x,y) &= d_1(x,y)\wedge d_2(x,y),\\
d_1(x,y) &= \begin{cases} 0, & x=y,\\ (1+\|x\|^i+\|y\|^i), & \text{otherwise},\end{cases}\\
d_2(x,y) &= \inf_{T,\psi\in\mathsf{A}(T,x,y)} F(\psi),\\
F(\psi) &= \frac{1}{\varepsilon}\int_0^T \exp(\eta\|\dot\psi\|)(1+\|\psi\|^i)\,dt.
\end{aligned}
\tag{4.1}
$$

It is straightforward to verify that d' is a metric by first showing that the expression inside the square root is a metric (triangle inequality is satisfied because of the infimum) and using that a square root of a metric is again a metric.

Moreover, $\mathcal{P}$ and $\mathcal{P}^m$ have a Wasserstein spectral gap with respect to d' because of the following lemma:

LEMMA 4.1. *Provided that ε is small enough, there exists a constant $C>0$ such that*

$$d'(x,y)\leq \tilde{d}(x,y)\leq Cd'(x,y)$$

for all pairs of points x, y in $\mathcal{H}$.

PROOF. Without loss of generality we assume that $\|y\|\geq\|x\|$. The inequality $d'\leq\tilde{d}$ follows from Lemma 4.2 since $d'\leq\sqrt{d_0}$ by definition.

In order show that $\tilde{d}\leq Cd'$, we will use Lemma 4.2 and reduce the number of summands appearing in equation (4.1) for d'. We can certainly assume that there is at most one index j in (4.1) such that $d_0(z_j,z_{j+1})=d_1(z_j,z_{j+1})$ because otherwise there are $1\leq j<k\leq n$ such that

$$d_0(z_j,z_{j+1})=d_1(z_j,z_{j+1}),\qquad d_0(z_k,z_{k+1})=d_1(z_k,z_{k+1})$$

which would lead to

$$d_0(z_j,z_{j+1})+\cdots+d_0(z_k,z_{k+1})\geq 2+\|z_j\|^i+\|z_{k+1}\|^i>d_1(z_j,z_{k+1}).$$

Hence the expression could be made smaller by removing all intermediate points between z_j and z_{k+1}, a contradiction.

Because d_2 is a Riemannian metric, it satisfies the triangle inequality in a sharp way in the sense that $d_2(x,y)=\inf_z(d_2(x,z)+d_2(z,y))$. As a consequence, the infimum is not changed by assuming that in equation (4.1) there is no index j such that

$$d_0(z_j,z_{j+1})=d_2(z_j,z_{j+1}),\qquad d_0(z_{j+1},z_{j+2})=d_2(z_{j+1},z_{j+2}).$$

Combining these two facts, equation (4.1) thus reduces to

$$\begin{aligned}(d'(x,y))^2 = \min\Big\{&d_0(x,y), \inf_{z_2,z_3} d_2(x,z_2)+d_1(z_2,z_3)+d_2(z_3,y),\\ &\inf_{z_2} d_2(x,z_2)+d_1(z_2,y), \inf_{z_2} d_1(x,z_2)+d_2(z_2,y)\Big\}.\end{aligned} \tag{4.2}$$

Recalling Lemma 4.2, it remains to show that $d' \geq C\sqrt{d_0}$ with d' given by (4.2). This is of course nontrivial only if (x,y) is such that $d'(x,y) < \sqrt{d_0(x,y)}$. Therefore we assume this fact from now on.

Suppose first that $\|y\| \leq Q$, for some constant $Q > 0$ to be determined later. Since $d'(x,y) \neq \sqrt{d_0(x,y)}$, there is at least one j such that $d_0(z_j,z_{j+1}) = d_1(z_j,z_{j+1})$ which leads to

$$1+2Q^i \geq d_0(x,y) \geq (d'(x,y))^2 \geq 1,$$

so that the bound $(1+2Q^i)d'(x,y) \geq \sqrt{d_0(x,y)}$ indeed follows in this case.

Suppose now that $\|y\| \geq Q$. Again, one summand $d_0(z_j,z_{j+1})$ in equation (4.2) satisfies

$$d_0(z_j,z_{j+1}) = d_1(z_j,z_{j+1}),$$

thus giving rise to a simple lower bound on d',

$$d'(x,y) \geq \sqrt{1+\|z_j\|^i}. \tag{4.3}$$

Because of (4.2), z_{j+1} is either equal to y or connected to y through a path $\psi_y \in \mathsf{A}(T, z_{j+1}, y)$ which is such that

$$F(\psi_y) \leq 1+2\|y\|^i, \tag{4.4}$$

where $F(\psi)$ is as in the definition of d_2. By the same reasoning as in the proof of Lemma 4.2, for Q large enough it is sufficient to consider paths starting in y and such that $\|\psi(t)\| \geq \|y\|/2$. The bound (4.4) thus yields an upper bound on $\|z_{j+1}-y\|$ by

$$1+2\|y\|^i \geq F(\psi_y) \geq \frac{1}{\varepsilon}\|z_{j+1}-y\| \exp(\eta\|y\|/2). \tag{4.5}$$

Combining this with (4.3), we have

$$\begin{aligned}d'(x,y)^2 &\geq 1+(\|y\|-\|z_{j+1}-y\|)^i \geq 1+\|y\|^i - i\|y\|^{i-1}\|z_{j+1}-y\|\\ &\geq 1+\frac{\|y\|^i}{2}+\left(\frac{\|y\|^i}{2}-\varepsilon(1+2\|y\|^i)\exp(-\eta\|y\|/2)\right),\end{aligned}$$

provided that $\varepsilon < 1/4$ and Q is large enough, the third summand is positive so that $d'(x,y)^2 \geq \frac{1}{4}d_1(x,y) \geq \frac{1}{4}d_0(x,y)$ completing the proof. □

LEMMA 4.2. *There is a $C > 0$ such that d_0 as defined in equation* (4.1) *satisfies*

$$d_0(x, y) \le \tilde{d}(x, y)^2 \le C d_0(x, y) \qquad \textit{for all } x, y.$$

PROOF. We assume again that $\|y\| \ge \|x\|$. In order to prove that $d_0(x, y) \le \tilde{d}(x, y)^2$, we only have to show that

$$\inf_{T, \psi \in \mathsf{A}(T,x,y)} F(\psi) \le \inf_{T, \psi \in \mathsf{A}(T,x,y)} \frac{1}{\varepsilon} \int_0^T \exp(\eta \|\psi\|)\, dt \big(1 + \|x\|^i + \|y\|^i\big).$$

Replacing $\psi(t)$ by

$$\big(1 \wedge \|y\| / \|\psi(t)\|\big) \psi(t)$$

in the expressions above does not cause an increase. Hence it is sufficient to consider paths ψ which satisfy

$$\|\psi(t)\| \le \|y\|, \qquad t \in [0, T]. \tag{4.6}$$

The bound $d_0 \le \tilde{d}^2$ then follows at once from

$$1 + \|\psi\|^i \le 1 + \|x\|^i + \|y\|^i.$$

We proceed now to show that $\tilde{d}(x, y)^2 \le C d_0(x, y)$ for which we only have to consider

$$\begin{aligned} d_2(x, y) &= \inf_{T, \psi \in \mathsf{A}(T,x,y)} \frac{1}{\varepsilon} \int_0^T \exp(\eta \|\psi\|) \big(1 + \|\psi\|^i\big)\, dt \\ &\le \big(1 + \|x\|^i + \|y\|^i\big) \end{aligned} \tag{4.7}$$

since the minimum expressions in $\tilde{d}^2$ and d_0 have $(1 + \|x\|^i + \|y\|^i)$ in common.

We will first use this to show that x and y have to be close if $\|y\|$ is large. We will show that any path ψ for which the expression in d_2 is close to the infimum has to satisfy $\|y\| \ge \psi \ge \frac{\|y\|}{2}$. Hence $1 + \|\psi\|^i$ and $(1 + \|x\|^i + \|y\|^i)$ are comparable. In order to gain a lower bound on $d_2(x, y)$, we distinguish between paths ψ which intersect or do not intersect $B_R(0)$. If the path lies completely outside the ball, we have

$$d_2(x, y) \ge \frac{1}{\varepsilon} \|x - y\| \exp(\eta R) \big(1 + R^i\big).$$

If ψ and $B_R(0)$ intersect, then $d_2(x, y)$ is larger than $d_2(B_R(0), y)$ which by symmetry corresponds to

$$\begin{aligned} d_2(x, y) &\ge \frac{1}{\varepsilon} \int_0^{\|y\| - R} \exp\big(\eta\big(\|y\| - t\big)\big) \big(1 + \big(\|y\| - t\big)^i\big)\, dt \\ &\ge \frac{1}{\varepsilon} \big(\|y\| - R\big) \exp\big(\eta\big(\|y\| - R\big) \big(1 + \big(\|y\| - R\big)^i\big)\big). \end{aligned}$$

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

$$d_2(x, y) \geq \frac{1}{4\varepsilon}\|x - y\| \exp(\eta\|y\|/2)\left(1 + \frac{\|y\|^i}{2}\right).$$

By (4.7) this implies

$$\|x - y\| \leq \frac{4\varepsilon \exp(-\eta\|y\|/2)}{1 + \|y\|^i/2}(1 + 2\|y\|^i). \tag{4.8}$$

For x and y in $B_Q(0)$ we have that $(\tilde{d})^2 \leq (2Q^i + 1)d_0$ because we can assume $\|\psi(t)\| \leq \|y\|$ as above. It is only left to consider $x, y \in B_{\tilde{Q}}(0)^c$ for $\tilde{Q} = Q - 4\varepsilon \exp(-\eta\frac{Q}{2})(1 + 2Q^i)$ because of equation (4.8). Subsequently, we will show that for Q and hence $\tilde{Q}$ large enough, it is sufficient for the infimum expression for d_2 to consider paths ψ that do not intersect $B_R(0)$ for $R = \frac{\|y\|}{2}$.

Suppose that the path ψ would intersect $B_R(0)$. Then the functional is larger than the shortest path $\hat{\psi}$ to the boundary of the ball and hence

$$\begin{aligned} d_2 \geq F(\hat{\psi}) &\geq \frac{1}{\varepsilon}\int_0^{\|y\|-R} e^{\eta(\|y\|-t)}\left(1 + (\|y\| - t)^i\right) dt \\ &= \frac{1}{\varepsilon}\Bigg[\exp(\eta\|y\|)\left(\eta^{-1}(1 + \|y\|^i) + \sum_{j=1}^{i} \eta^{-1-j}\frac{i!}{(i-j)!}\|y\|^{i-j}\right) \\ &\qquad - \exp(\eta R)\left(\eta^{-1}(1 + R^i) + \sum_{j=1}^{i} \eta^{-1-j}\frac{i!}{(i-j)!}R^{i-j}\right)\Bigg] \end{aligned} \tag{4.9}$$

by $i + 1$ integrations by parts. Let l be the line connecting x and y. Then using (4.8) yields

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

For $R = \frac{\|y\|}{2}$ and $\tilde{Q}$ large enough we have $F(\psi) > F(l)$. Therefore for all $t \in [0, T]$ $\|y\| \geq \psi \geq \|y\|/2$ and thus

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

which yields that $\max(2Q^i + 1, 2^{i+1})d_0 \geq \tilde{d}^2$. $\square$

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_m (μ) are the unique invariant measures for $\mathcal{P}$ ($\mathcal{P}_m$) (resp.), μ_m (μ) is ergodic and Birkhoff's ergodic theorem applies. However, this theorem only applies to almost every initial condition, but we are able to extend it to every initial condition in this case which yields a strong law of large numbers.

THEOREM 4.3. *In the setting of Theorem 2.12 or 2.14, suppose that* $\operatorname{supp}\mu = \mathcal{H}$ *and* $h : \mathcal{H} \to \mathbb{R}$ *has Lipschitz constant* L *with respect to* $\tilde{d}$, *then for arbitrary* $X_0 \in \mathcal{H}$

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

PROOF. By Birkhoff's ergodic theorem, we know that this is true for measurable h and every initial condition in some set of full measure A. Because μ has full support, for any $t > 0$ we can choose $Y_0 \in A$ with $\tilde{d}(X_0, Y_0) \le t^2$. Hence

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

By the Wasserstein spectral gap, we can couple X^n and Y^n such that

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

for some $0 < r < 1$. An application of Markov's inequality yields

$$\mathbb{P}(\tilde{d}(X^n, Y^n) \ge c) \le C\frac{r^n \tilde{d}(X^0, Y^0)}{c}.$$

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

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

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

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

and because t was chosen arbitrarily, the result follows. □

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 to estimate the error for finite n. Because of Lemma 4.1 and arguments from Lemma 3.2, it is straightforward to verify that our assumptions imply those needed for the Wasserstein CLT in Komorowski and Walczuk (2012). This results in the following theorem:

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

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

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

4.2. *Consequences of L^2_μ-spectral gap.* Under the assumptions of Theorem 2.12 or 2.14, we have proved the existence of an L^2_μ-spectral gap in Section 2.2.2. Now we may use all existing consequences for the ergodic average with and without burn in ($n_0 = 0$)

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

The following result of Kipnis and Varadhan (1986) [see also Łatuszyński and Roberts (2013) whence the statement was adapted] then yields a CLT:

PROPOSITION 4.5. *Consider an ergodic Markov chain with transition operator P which is reversible with respect to a probability measure μ. Let $f \in L^2$ be such that*

$$\sigma^2_{f,P} = \left\langle \frac{1+P}{1-P} f, f \right\rangle < \infty,$$

and 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, provided that f is mean-zero, it follows from the L^2-spectral gap that

$$\sigma^2_{f,\mathcal{P}} \le \frac{2\mu(f^2)}{1-\beta}.$$

Due to Theorem 2.14, we have a lower bound on the spectral gap $1-\beta$ of $\mathcal{P}$ and $1-\beta_m$ of $\mathcal{P}_m$ which is uniform in m. Thus the ergodic average of the pCN algorithm applied to the target measures μ and μ_m has an m-independent upper bound on the asymptotic variance.

The result of Proposition 4.5 has been extended to μ for almost every initial condition in Cuny and Lin (2009) which also applies to our case.

A different approach due to Rudolf (2012) is to consider the MSE

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

Using Chebyscheff's inequality, this results in a confidence interval for $S(f)$. We can bound it by using the following proposition from Rudolf (2012):

PROPOSITION 4.6. *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*

$$\frac{1}{\log(\beta^{-1})}\begin{cases} \dfrac{p}{2(p-2)}\log\left(\dfrac{32p}{p-2}\right)\left\|\dfrac{d\nu}{d\mu}-1\right\|_{p/(p-2)}, & p \in (2,4), \\ \log(64)\left\|\dfrac{d\nu}{d\mu}-1\right\|_{p/(p-2)}, & p \in [4,\infty]. \end{cases} \tag{4.10}$$

Then

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

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

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

Using Fernique's theorem, this implies 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 a density with respect to a Gaussian. It is also possible to show that the function space version of the MALA Beskos, Kalogeropoulos and Pazos (2013) has a spectral gap if, in addition to the assumptions in this article, the gradient of Φ satisfies strong assumptions, and the gradient step is very small. There is also a variant of the hybrid Monte Carlo methods Beskos et al. (2011) adapted to the sampling of measures defined via a density with respect to a Gaussian and it would be interesting to employ the weak Harris theorem to study this algorithm.

Other classes of target measures, such as those arising from Besov prior measures [Dashti, Harris and Stuart (2012), Lassas, Saksman and Siltanen (2009)] or an infinite product of uniform measures in Schwab and Stuart (2012) would also provide interesting applications. The proposal of the pCN is reversible and has a spectral gap with respect to the Gaussian reference measure. For arbitrary reference and target measures, the third author has recently proved that for bounded Φ the Metropolis–Hastings algorithm has a spectral gap if the proposal is reversible and has a spectral gap with respect to the reference measure [Vollmer (2013)].

More generally, we expect that the weak Harris theorem 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 [Hairer, Mattingly and Scheutzow (2011)]. In contrast, the theory developed in Meyn and Tweedie (2009) 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 Hairer, Mattingly and Scheutzow (2011), 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 they do also hold on an arbitrary Banach space. This can be shown by using a Gaussian series [cf. Section 3.5 in Bogachev (1998)] instead of the Karhunen–Loève expansion and the m-independence is due to Theorem 3.3.6 in Bogachev (1998).

APPENDIX: GAUSSIAN MEASURES

As a consequence of Fernique's theorem, we have the following explicit bound on exponential moments of the norm of a Gaussian random variable, which is needed to show that $\mathcal{P}$ and $\mathcal{P}_m$ are d-contracting; see Section 3.2.2.

PROPOSITION A.1. *For β small enough, there exists a constant F_β such that*

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

Furthermore, for any $\alpha \in \mathbb{R}^+$ there is a constant $C_{\alpha,\beta}$ such that for $K > \frac{\alpha}{2\beta}$

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

PROOF. The first claim is just Fernique's theorem; see, for example, Bogachev (1998), Da Prato and Zabczyk (1992), Hairer (2010). Using integration by parts and Fubini, we get

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

Setting $f(x) = \exp(\alpha x)$ and applying Fernique's theorem yields

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

Since, for K as in the statement, one verifies that

$$\beta t^2 - \alpha t \geq \beta K^2 - \alpha K + \beta(t-K)^2,$$

and the required bound follows at once. $\square$

Acknowledgments. We thank Feng-Yu Wang for pointing out the connection between Wasserstein and L^2-spectral gaps and Professor Andreas Eberle for many fruitful discussions about this topic.

REFERENCES

ADLER, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **12**. IMS, Hayward, CA. MR1088478

ATHREYA, K. B. and NEY, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501. MR0511425

BAKRY, D. and ÉMERY, M. (1985). Diffusions hypercontractives. In *Séminaire de Probabilités, XIX*, 1983/84. *Lecture Notes in Math.* **1123** 177–206. Springer, Berlin. MR0889476

BESKOS, A., KALOGEROPOULOS, K. and PAZOS, E. (2013). Advanced MCMC methods for sampling on diffusion pathspace. *Stochastic Process. Appl.* **123** 1415–1453. MR3016228

BESKOS, A., ROBERTS, G. and STUART, A. (2009). Optimal scalings for local Metropolis–Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.* **19** 863–898. MR2537193

BESKOS, A., ROBERTS, G., STUART, A. and VOSS, J. (2008). MCMC methods for diffusion bridges. *Stoch. Dyn.* **8** 319–350. MR2444507

BESKOS, A., PINSKI, F., SANZ-SERNA, J. M. and STUART, A. M. (2011). Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Process. Appl.* **121** 2201–2230. MR2822774

BOGACHEV, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. MR1642391

BOGACHEV, V. I. (2007). *Measure Theory. Vol. I, II.* Springer, Berlin. MR2267655

CHAN, K. S. and GEYER, C. J. (1994). Discussion: Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1747–1758.

CHEEGER, J. (1970). A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in Analysis* (*Papers Dedicated to Salomon Bochner*, 1969) 195–199. Princeton Univ. Press, Princeton, NJ. MR0402831

COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446.

CUNY, C. and LIN, M. (2009). Pointwise ergodic theorems with rate and application to the CLT for Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 710–733. MR2548500

DASHTI, M., HARRIS, S. and STUART, A. (2012). Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging* **6** 183–200. MR2942737

DASHTI, M. and STUART, A. M. (2011). Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal.* **49** 2524–2542. MR2873245

DA PRATO, G. and ZABCZYK, J. (1992). *Stochastic Equations in Infinite Dimensions. Encyclopedia of Mathematics and Its Applications* **44**. Cambridge Univ. Press, Cambridge. MR1207136

DIACONIS, P. and STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61. MR1097463

EBERLE, A. (2014). Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Ann. Appl. Probab.* **24** 337–377.

FRIGESSI, A., DI STEFANO, P., HWANG, C.-R. and SHEU, S. J. (1993). Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B* **55** 205–219. MR1210432

GEYER, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7** 473–483.

GEYER, C. J. and THOMPSON, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Amer. Statist. Assoc.* **90** 909–920.

HAIRER, M. (2010). An introduction to stochastic PDEs. Lecture notes, University of Warwick.

HAIRER, M. and MAJDA, A. J. (2010). A simple framework to justify linear response theory. *Nonlinearity* **23** 909–922. MR2602020

HAIRER, M., MATTINGLY, J. C. and SCHEUTZOW, M. (2011). Asymptotic coupling and a general form of Harris' Theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149** 223–259. MR2773030

HAIRER, M., STUART, A. M. and VOSS, J. (2007). Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.* **17** 1657–1706. MR2358638

HASTINGS, W. K. (1970). Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97.

HJORT, N. L., HOLMES, C., MÜLLER, P. and WALKER, S. G., eds. (2010). *Bayesian Nonparametrics. Cambridge Series in Statistical and Probabilistic Mathematics* **28**. Cambridge Univ. Press, Cambridge. MR2722987

JOULIN, A. and OLLIVIER, Y. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** 2418–2442. MR2683634

KIPNIS, C. and VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19. MR0834478

KOMOROWSKI, T. and WALCZUK, A. (2012). Central limit theorem for Markov processes with spectral gap in the Wasserstein metric. *Stochastic Process. Appl.* **122** 2155–2184. MR2921976

LASSAS, M., SAKSMAN, E. and SILTANEN, S. (2009). Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging* **3** 87–122. MR2558305

ŁATUSZYŃSKI, K. and NIEMIRO, W. (2011). Rigorous confidence bounds for MCMC under a geometric drift condition. *J. Complexity* **27** 23–38. MR2745298

ŁATUSZYŃSKI, K. and ROBERTS, G. O. (2013). CLTs and asymptotic variance of time-sampled Markov chains. *Methodol. Comput. Appl. Probab.* **15** 237–247. MR3030220

LAWLER, G. F. and SOKAL, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309** 557–580. MR0930082

LEE, P. M. (2004). *Bayesian Statistics*: *An Introduction*, 3rd ed. Arnold, London. MR2069264

LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer, New York. MR2401592

LOVÁSZ, L. and SIMONOVITS, M. (1993). Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms* **4** 359–412. MR1238906

MATTINGLY, J. C., PILLAI, N. S. and STUART, A. M. (2012). Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Probab.* **22** 881–930. MR2977981

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., TELLER, E. et al. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087.

MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253

NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Probab. Theory Related Fields* **43** 309–318.

PILLAI, N. S., STUART, A. M. and THIÉRY, A. H. (2011). Optimal proposal design for random walk type Metropolis algorithms with Gaussian random field priors. ArXiv E-prints.

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer, New York. MR2080278

ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158

RÖCKNER, M. and WANG, F.-Y. (2001). Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups. *J. Funct. Anal.* **185** 564–603. MR1856277

RUDOLF, D. (2012). Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math.* (*Rozprawy Mat.*) **485** 1–93. MR2977521

SCHWAB, C. and STUART, A. M. (2012). Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems* **28** 045003, 32. MR2903278

SINCLAIR, A. and JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82** 93–133. MR1003059

STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785

TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401

VOLLMER, S. J. (2013). Dimension-independent MCMC sampling for inverse problems with non-Gaussian priors. Available at arXiv:1302.2213.

WANG, F.-Y. (2003). Functional inequalities for the decay of sub-Markov semigroups. *Potential Anal.* **18** 1–23. MR1953493

M. HAIRER
A. M. STUART
MATHEMATICAL INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
UNITED KINGDOM
E-MAIL: m.hairer@warwick.ac.uk
a.m.stuart@warwick.ac.uk

S. J. VOLLMER
MATHEMATICAL INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
UNITED KINGDOM
PRESENT ADDRESS:
DEPARTMENT OF STATISTICS
1 SOUTH PARKS ROAD
OXFORD, OX1 3TG
UNITED KINGDOM
E-MAIL: s.vollmer@warwick.ac.uk