The Annals of Applied Probability
2014, Vol. 24, No. 1, 337–377
DOI: 10.1214/13-AAP926

ERROR BOUNDS FOR METROPOLIS–HASTINGS ALGORITHMS APPLIED TO PERTURBATIONS OF GAUSSIAN MEASURES IN HIGH DIMENSIONS

BY ANDREAS EBERLE

Universität Bonn

The Metropolis-adjusted Langevin algorithm (MALA) is a Metropolis–Hastings method for approximate sampling from continuous distributions. We derive upper bounds for the contraction rate in Kantorovich–Rubinstein–Wasserstein distance of the MALA chain with semi-implicit Euler proposals applied to log-concave probability measures that have a density w.r.t. a Gaussian reference measure. For sufficiently "regular" densities, the estimates are dimension-independent, and they hold for sufficiently small step sizes h that do not depend on the dimension either. In the limit $h \downarrow 0$, the bounds approach the known optimal contraction rates for overdamped Langevin diffusions in a convex potential.

A similar approach also applies to Metropolis–Hastings chains with Ornstein–Uhlenbeck proposals. In this case, the resulting estimates are still independent of the dimension but less optimal, reflecting the fact that MALA is a higher order approximation of the diffusion limit than Metropolis–Hastings with Ornstein–Uhlenbeck proposals.

1. Introduction. The performance of Metropolis–Hastings (MH) methods [16, 23, 27] for sampling probability measures on high-dimensional continuous state spaces has attracted growing attention in recent years. The pioneering works by Roberts, Gelman and Gilks [28] and Roberts and Rosenthal [29] show in particular that for product measures π^d on $\mathbb{R}^d$, the average acceptance probabilities for the Random Walk Metropolis algorithm (RWM) and the Metropolis adjusted Langevin algorithm (MALA) converge to a strictly positive limit as $d \to \infty$ only if the step sizes h go to zero of order $O(d^{-1})$, $O(d^{-1/3})$, respectively. In this case, a diffusion limit as $d \to \infty$ has been derived, leading to an optimal scaling of the step sizes maximizing the speed of the limiting diffusion, and an asymptotically optimal acceptance probability.

Recently, the optimal scaling results for RWM and MALA have been extended significantly to targets that are not of product form but have a sufficiently regular density w.r.t. a Gaussian measure; cf. [22, 26]. On the other hand, it has been pointed out [3, 4, 8, 14] that for corresponding perturbations of Gaussian measures, the acceptance probability has a strictly positive limit as $d \to \infty$ for small

Received November 2012; revised February 2013.

MSC2010 subject classifications. Primary 60J22; secondary 60J05, 65C40, 65C05.

Key words and phrases. Metropolis algorithm, Markov chain Monte Carlo, Langevin diffusion, Euler scheme, coupling, contractivity of Markov kernels.

step sizes that do not depend on the dimension, provided the random walk or Euler proposals in RWM and MALA are replaced by Ornstein–Uhlenbeck or semi-implicit ("preconditioned") Euler proposals, respectively; cf. also below. Pillai, Stuart and Thiéry [25] show that in this case, the Metropolis–Hastings algorithm can be realized directly on an infinite-dimensional Hilbert space arising in the limit as $d \to \infty$, and the corresponding Markov chain converges weakly to an infinite-dimensional overdamped Langevin diffusion as $h \downarrow 0$.

Mixing properties and convergence to equilibrium of Langevin diffusions have been studied intensively [1, 2, 9, 15, 24, 31]. In particular, it is well-known that contractivity and exponential convergence to equilibrium in Wasserstein distance can be quantified if the stationary distribution is strictly log-concave [7, 35]; cf. also [11] for a recent extension to the nonlog-concave case. Because of the diffusion limit results, one might expect that the approximating Metropolis–Hastings chains have similar convergence properties. However, this heuristics may also be wrong, since the convergence of the Markov chains to the diffusion is known only in a weak and nonquantitative sense.

Although there is a huge number of results quantifying the speed of convergence to equilibrium for Markov chains on discrete state spaces (cf. [18, 32] for an overview), there are relatively few quantitative results on Metropolis–Hastings chains on $\mathbb{R}^d$ when d is large. The most remarkable exception are the well-known works [10, 17, 19–21] which prove an upper bound for the mixing time that is polynomial in the dimension for Metropolis chains with ball walk proposals for uniform measures on convex sets and more general log-concave measures.

Below, we develop an approach to quantify Wasserstein contractivity and convergence to equilibrium in a dimension-independent way for the Metropolis–Hastings chains with Ornstein–Uhlenbeck and semi-implicit Euler proposals. Our approach applies in the strictly log-concave case (or, more generally, if the measure is strictly log-concave on an appropriate ball) and yields bounds for small step sizes that are very precise. The results for semi-implicit Euler proposals require less restrictive assumptions than those for Ornstein–Uhlenbeck proposals, reflecting the fact that the corresponding Markov chain is a higher order approximation of the diffusion.

Our results are closely related and complementary to the recent work [13], and to the dimension-dependent geometric ergodicity results in [5]. In particular, in [13], Hairer, Stuart and Vollmer apply related methods to establish exponential convergence to equilibrium in Wasserstein distance for Metropolis–Hastings chains with Ornstein–Uhlenbeck proposals in a less quantitative way, but without assuming log-concavity. In the context of probability measures on function spaces, the techniques developed here are applied in the PhD Thesis of Gruhlke [12].

We now recall some basic facts on Metropolis–Hastings algorithms and describe our setup and the main results. Sections 2 and 3 contain basic results on Wasserstein contractivity of Metropolis–Hastings kernels, and contractivity of the proposal kernels. In Sections 4 and 5, we prove bounds quantifying rejection probabilities and the dependence of the rejection event on the current state for Ornstein–Uhlenbeck and semi-implicit Euler proposals. These bounds, combined with an upper bound for the exit probability of the corresponding Metropolis–Hastings chains from a given ball derived in Section 6 are crucial for the proof of the main results in Section 7.

1.1. *Metropolis–Hastings algorithms.* Let $U:\mathbb{R}^d \to \mathbb{R}$ be a lower bounded measurable function such that

$$\mathcal{Z} = \int_{\mathbb{R}^d} \exp(-U(x))\,dx < \infty,$$

and let μ denote the probability measure on $\mathbb{R}^d$ with density proportional to $\exp(-U)$. We use the same letter μ for the measure and its density, that is,

$$\mu(dx) = \mu(x)\,dx = \mathcal{Z}^{-1}\exp(-U(x))\,dx. \tag{1.1}$$

Below, we view the measure μ defined by (1.1) as a perturbation of the standard normal distribution γ^d in $\mathbb{R}^d$; that is, we decompose

$$U(x) = \tfrac{1}{2}|x|^2 + V(x), \qquad x \in \mathbb{R}^d, \tag{1.2}$$

with a measurable function $V:\mathbb{R}^d \to \mathbb{R}$, and obtain the representation

$$\mu(dx) = \widetilde{\mathcal{Z}}^{-1}\exp(-V(x))\gamma^d(dx) \tag{1.3}$$

with normalization constant $\widetilde{\mathcal{Z}} = \mathcal{Z}/(2\pi)^{d/2}$. Here $|\cdot|$ denotes the Euclidean norm.

Note that in $\mathbb{R}^d$, any probability measure with a strictly positive density can be represented as an absolutely continuous perturbation of γ^d as in (1.3). In an infinite-dimensional limit, however, the density may degenerate. Nevertheless, also on infinite-dimensional spaces, absolutely continuous perturbations of Gaussian measures form an important and widely used class of models.

EXAMPLE 1.1 (Transition path sampling). We briefly describe a typical application; cf. [14] and [12] for details. Suppose that we are interested in sampling a trajectory of a diffusion process in $\mathbb{R}^\ell$ conditioned to a given endpoint b at time $t = 1$. We assume that the unconditioned diffusion process $(Y_t, \mathbb{P})$ satisfies a stochastic differential equation of the form

$$dY_t = -\nabla H(Y_t)\,dt + dB_t, \tag{1.4}$$

where (B_t) is an ℓ-dimensional Brownian motion, and $H \in C^2(\mathbb{R}^\ell)$ is bounded from below. Then, by Girsanov's theorem and Itô's formula, a regular version of the law of the conditioned process satisfying $Y_0 = a$ and $Y_1 = b$ on the path space $E = \{y \in C([0,1],\mathbb{R}^\ell) : y_0 = a,\, y_1 = b\}$ is given by

$$\mu(dy) = C^{-1}\exp(-V(y))\gamma(dy), \tag{1.5}$$

where γ is the law of the Brownian bridge from a to b,

$$V(y) = \frac{1}{2}\int_0^1 \phi(y_s)\,ds \qquad \text{with } \phi(x) = \left|\nabla H(x)\right|^2 - \Delta H(x), \tag{1.6}$$

and $C = \exp(H(b) - H(a))$; cf. [31]. In order to obtain finite-dimensional approximations of the measure μ on E, we consider the *Wiener–Lévy expansion*

$$y_t = e_t + \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1}\sum_{i=1}^{\ell} x_{n,k,i}e_t^{n,k,i}, \qquad t \in [0, 1], \tag{1.7}$$

of a path $y \in E$ in terms of the basis functions $e_t = (1-t)a + tb$ and $e_t^{n,k,i} = 2^{-n/2}g(2^n t - k)e^i$ with $g(s) = \min(s, 1-s)^+$. Here the coefficients $x_{n,k,i}$, $n \geq 0$, $0 \leq k < 2^n$, $1 \leq i \leq \ell$, are real numbers. Recall that truncating the series at $n = m-1$ corresponds to taking the polygonal interpolation of the path y adapted to the dyadic partition $\mathcal{D}_m = \{k2^{-m} : k = 0, 1, \ldots, 2^m\}$ of the interval [0, 1]. Now fix $m \in \mathbb{N}$, let $d = (2^m - 1)\ell$ and let

$$x^d = \left(x_{n,k,i} : 0 \leq n < m, 0 \leq k < 2^n, 1 \leq i \leq l\right) \in \mathbb{R}^d$$

denote the vector consisting of the first d components in the basis expansion of a path $y \in E$. Then the image of the Brownian bridge measure γ under the projection $\pi_d : E \to \mathbb{R}^d$ that maps y to x^d is the d-dimensional standard normal distribution γ^d; for example, cf. [33]. Therefore, a natural finite-dimensional approximation to the infinite-dimensional sampling problem described above consists in sampling from the probability measure

$$\mu_d(dx) = \widetilde{\mathcal{Z}}_d^{-1}\exp\left(-V_d(x)\right)\gamma^d(dx) \tag{1.8}$$

on $\mathbb{R}^d$ where $\widetilde{\mathcal{Z}}_d$ is a normalization constant, and

$$V_d(x) = 2^{-m-1}\left(\frac{1}{2}\phi(y_0) + \sum_{k=1}^{2^m-1}\phi(y_{k2^{-m}}) + \frac{1}{2}\phi(y_1)\right); \tag{1.9}$$

with $y = e + \sum_{n<m}\sum_k\sum_i x_{n,k,i}e^{n,k,i}$ denoting the polygonal path corresponding to $x^d = (x_{n,k,i}) \in \mathbb{R}^d$.

Returning to our general setup, suppose that $p(x, dy) = p(x, y)\,dy$ is an absolutely continuous transition kernel on $\mathbb{R}^d$ with strictly positive densities $p(x, y)$. Let

$$\alpha(x, y) = \min\left(\frac{\mu(y)p(y, x)}{\mu(x)p(x, y)}, 1\right), \qquad x, y \in \mathbb{R}^d. \tag{1.10}$$

Note that $\alpha(x, y)$ does not depend on $\mathcal{Z}$. The Metropolis–Hastings algorithm with proposal kernel p is the following Markov chain Monte Carlo method for approximate sampling and Monte Carlo integration w.r.t. μ:

(1) Choose an initial state X_0.
(2) For $n := 0, 1, 2, \ldots$:

- Sample $Y_n \sim p(X_n, dy)$ and $U_n \sim \text{Unif}(0, 1)$ independently.
- If $U_n < \alpha(X_n, Y_n)$, then accept the proposal, and set $X_{n+1} := Y_n$, else reject the proposal and set $X_{n+1} := X_n$.

The algorithm generates a time-homogeneous Markov chain $(X_n)_{n=0,1,2,\ldots}$ with initial state X_0 and transition kernel

$$q(x, dy) = \alpha(x, y)p(x, y)\,dy + r(x) \cdot \delta_x(dy). \tag{1.11}$$

Here

$$r(x) = 1 - q\big(x, \mathbb{R}^d \setminus \{x\}\big) = 1 - \int_{\mathbb{R}^d} \alpha(x, y)p(x, y)\,dy \tag{1.12}$$

is the average rejection probability for the proposal when the Markov chain is at x. Note that $q(x, dy)$ restricted to $\mathbb{R}^d \setminus \{x\}$ is again absolutely continuous with density

$$q(x, y) = \alpha(x, y)p(x, y).$$

Since

$$\mu(x)q(x, y) = \alpha(x, y)\mu(x)p(x, y) = \min\big(\mu(y)p(y, x), \mu(x)p(x, y)\big)$$

is a symmetric function in x and y, the kernel $q(x, dy)$ satisfies the *detailed balance condition*

$$\mu(dx)q(x, dy) = \mu(dy)q(y, dx). \tag{1.13}$$

In particular, μ is a stationary distribution for the Metropolis–Hastings chain, and the chain with initial distribution μ is reversible. Therefore, under appropriate ergodicity assumptions, the distribution of X_n will converge to μ as $n \to \infty$.

To analyze Metropolis–Hastings algorithms it is convenient to introduce the function

$$G(x, y) = \log \frac{\mu(x)p(x, y)}{\mu(y)p(y, x)} = U(y) - U(x) + \log \frac{p(x, y)}{p(y, x)}. \tag{1.14}$$

For any $x, y \in \mathbb{R}^d$,

$$\alpha(x, y) = \exp\big(-G(x, y)^+\big). \tag{1.15}$$

In particular, for any $x, y, \widetilde{x}, \widetilde{y} \in \mathbb{R}^d$,

$$1 - \alpha(x, y) \le G(x, y)^+, \tag{1.16}$$

$$\big(\alpha(x, y) - \alpha(\widetilde{x}, \widetilde{y})\big)^+ \le \big(G(x, y) - G(\widetilde{x}, \widetilde{y})\big)^- \quad \text{and} \tag{1.17}$$

$$\big(\alpha(x, y) - \alpha(\widetilde{x}, \widetilde{y})\big)^- \le \big(G(x, y) - G(\widetilde{x}, \widetilde{y})\big)^+. \tag{1.18}$$

The function $G(x, y)$ defined by (1.14) can also be represented in terms of V: Indeed, since

$$\log \frac{\gamma^d(x)}{\gamma^d(y)} = \frac{1}{2}\big(|y|^2 - |x|^2\big),$$

we have

$$G(x, y) = V(y) - V(x) + \log \frac{\gamma^d(x)p(x, y)}{\gamma^d(y)p(y, x)}, \tag{1.19}$$

where $\gamma^d(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$ denotes the standard normal density in $\mathbb{R}^d$.

1.2. *Metropolis–Hastings algorithms with Gaussian proposals*. We aim at proving contractivity of Metropolis–Hastings kernels w.r.t. appropriate Kantorovich–Rubinstein–Wasserstein distances. For this purpose, we are looking for a proposal kernel that has adequate contractivity properties and sufficiently small rejection probabilities. The rejection probability is small if the proposal kernel approximately satisfies the detailed balance condition w.r.t. μ.

1.2.1. *Ornstein–Uhlenbeck proposals*. A straightforward approach would be to use a proposal density that satisfies the detailed balance condition

$$\gamma^d(x)p(x, y) = \gamma^d(y)p(y, x) \qquad \text{for any } x, y \in \mathbb{R}^d \tag{1.20}$$

w.r.t. the standard normal distribution. In this case,

$$G(x, y) = V(y) - V(x). \tag{1.21}$$

The simplest form of proposal distributions satisfying (1.20) are the transition kernels of AR(1) (discrete Ornstein–Uhlenbeck) processes given by

$$p_h^{\mathrm{OU}}(x, dy) = N\left(\left(1 - \frac{h}{2}\right)x, \left(h - \frac{h^2}{4}\right)I_d\right) \tag{1.22}$$

for some constant $h \in (0, 2)$. If Z is a standard normally distributed $\mathbb{R}^d$-valued random variable, then the random variables

$$Y_h^{\mathrm{OU}}(x) := \left(1 - \frac{h}{2}\right)x + \sqrt{h - \frac{h^2}{4}}\,Z, \qquad x \in \mathbb{R}^d, \tag{1.23}$$

have distributions $p_h^{\mathrm{OU}}(x, dy)$. Note that by (1.21), the acceptance probabilities

$$\alpha^{\mathrm{OU}}(x, y) = \exp\big(-G^{\mathrm{OU}}(x, y)^+\big) = \exp\big(-\big(V(y) - V(x)\big)^+\big) \tag{1.24}$$

for Ornstein–Uhlenbeck proposals do not depend on h.

1.2.2. *Euler proposals*. In continuous time, under appropriate regularity and growth conditions on V, detailed balance w.r.t. μ is satsfied exactly by the transition functions of the diffusion process solving the over-damped Langevin stochastic differential equation

$$dX_t = -\tfrac{1}{2} X_t \, dt - \tfrac{1}{2} \nabla V(X_t) \, dt + dB_t, \tag{1.25}$$

because the generator

$$\mathscr{L} = \tfrac{1}{2}\Delta - \tfrac{1}{2} x \cdot \nabla - \tfrac{1}{2} \nabla V \cdot \nabla = \tfrac{1}{2}(\Delta - \nabla U \cdot \nabla)$$

is a self-adjoint operator on an appropriate dense subspace of $L^2(\mathbb{R}^d; \mu)$; cf. [31]. Although we cannot compute and sample from the transition functions exactly, we can use approximations as proposals in a Metropolis–Hastings algorithm. A corresponding MH algorithm where the proposals are obtained from a discretization scheme for the SDE (1.25) is called a *Metropolis-adjusted Langevin algorithm* (MALA); cf. [27, 30].

In this paper, we focus on the MALA scheme with proposal kernel

$$p_h(x, \cdot) = N\left(\left(1 - \frac{h}{2}\right) x - \frac{h}{2} \nabla V(x), \left(h - \frac{h^2}{4}\right) \cdot I_d \right) \tag{1.26}$$

for some constant $h \in (0, 2)$; that is, $p_h(x, \cdot)$ is the distribution of

$$\begin{aligned} Y_h(x) &= x - \frac{h}{2} \nabla U(x) + \sqrt{h - \frac{h^2}{4}} Z \\ &= \left(1 - \frac{h}{2}\right) x - \frac{h}{2} \nabla V(x) + \sqrt{h - \frac{h^2}{4}} Z, \end{aligned} \tag{1.27}$$

where $Z \sim \gamma^d$ is a standard normal random variable with values in $\mathbb{R}^d$.

Note that if $h - h^2/4$ is replaced by h, then (1.27) is a standard Euler discretization step for the SDE (1.25). Replacing h by $h - h^2/4$ ensures that detailed balance is satisfied exactly for $V \equiv 0$. Alternatively, (1.27) can be viewed as a semi-implicit Euler discretization step for (1.25):

REMARK 1.2 (Euler schemes). The *explicit Euler discretization* of the over-damped Langevin equation (1.25) with time step size $h > 0$ is given by

$$X_{n+1} = \left(1 - \frac{h}{2}\right) X_n - \frac{h}{2} \nabla V(X_n) + \sqrt{h} Z_{n+1}, \qquad n = 0, 1, 2, \ldots, \tag{1.28}$$

where $Z_n, n \in \mathbb{N}$, are i.i.d. random variables with distribution γ^d. The process (X_n) defined by (1.28) is a time-homogeneous Markov chain with transition kernel

$$p_h^{\mathrm{Euler}}(x, \cdot) = N\left(\left(1 - \frac{h}{2}\right) x - \frac{h}{2} \nabla V(x), h \cdot I_d \right). \tag{1.29}$$

Even for $V \equiv 0$, the measure μ is not a stationary distribution for the kernel p_h^{Euler}. A *semi-implicit Euler scheme* for (1.25) with time-step size $\varepsilon > 0$ is given by

$$X_{n+1} - X_n = -\frac{\varepsilon}{2} \cdot \frac{X_{n+1} + X_n}{2} - \frac{\varepsilon}{2} \nabla V(X_n) + \sqrt{\varepsilon} Z_{n+1} \tag{1.30}$$

with Z_n i.i.d. with distribution γ^d; cf. [14]. Note that the scheme is implicit only in the linear part of the drift but explicit in ∇V. Solving for X_{n+1} in (1.30) and substituting $h = \varepsilon/(1 + \frac{\varepsilon}{4})$ with $h \in (0, 2)$ yields the equivalent equation

$$X_{n+1} = \left(1 - \frac{h}{2}\right) X_n - \frac{h}{2} \nabla V(X_n) + \sqrt{h - \frac{h^2}{4}} Z_{n+1}. \tag{1.31}$$

We call the Metropolis–Hastings algorithm with proposal kernel $p_h(x, \cdot)$ a *semi-implicit MALA scheme with step size h*.

PROPOSITION 1.3 (Acceptance probabilities for semi-implicit MALA). *Let $V \in C^1(\mathbb{R}^d)$ and $h \in (0, 2)$. Then the acceptance probabilities for the Metropolis-adjusted Langevin algorithm with proposal kernels p_h are given by $\alpha_h(x, y) = \exp(-G_h(x, y)^+)$ with*

$$\begin{aligned} &G_h(x, y) \\ &\quad = V(y) - V(x) - \frac{y - x}{2} \cdot \bigl(\nabla V(y) + \nabla V(x)\bigr) \\ &\qquad + \frac{h}{8 - 2h} \bigl[(y + x) \cdot \bigl(\nabla V(y) - \nabla V(x)\bigr) + \bigl|\nabla V(y)\bigr|^2 - \bigl|\nabla V(x)\bigr|^2\bigr]. \end{aligned} \tag{1.32}$$

For explicit Euler proposals with step size $h > 0$, a corresponding representation holds with

$$\begin{aligned} G_h^{\text{Euler}}(x, y) &= V(y) - V(x) - \frac{y - x}{2} \cdot \bigl(\nabla V(y) + \nabla V(x)\bigr) \\ &\quad + \frac{h}{8} \bigl[\bigl|y + \nabla V(y)\bigr|^2 - \bigl|x + \nabla V(x)\bigr|^2\bigr]. \end{aligned} \tag{1.33}$$

The proof of the proposition is given in Section 4 below.

REMARK 1.4. For explicit Euler proposals, the $O(h)$ correction term in (1.33) does not vanish for $V \equiv 0$. More significantly, this term goes to infinity as $|y - x| \to \infty$, and the variance of $y - x$ w.r.t. the proposal distribution is of order $O(d)$.

1.3. *Bounds for rejection probabilities.* We fix a norm $\|\cdot\|_-$ on $\mathbb{R}^d$ such that

$$\|x\|_- \le |x| \qquad \text{for any } x \in \mathbb{R}^d. \tag{1.34}$$

We assume that V is sufficiently smooth w.r.t. $\|\cdot\|_-$ with derivatives growing at most polynomially:

ASSUMPTION 1.5. The function V is in $C^4(\mathbb{R}^d)$, and for any $n \in \{1,2,3,4\}$, there exist finite constants $C_n \in [0,\infty)$, $p_n \in \{0,1,2,\dots\}$ such that

$$\left|\left(\partial^n_{\xi_1,\dots,\xi_n} V\right)(x)\right| \le C_n \max\left(1, \|x\|_-\right)^{p_n} \|\xi_1\|_- \cdot \dots \cdot \|\xi_n\|_-$$

holds for any $x \in \mathbb{R}^d$ and $\xi_1,\dots,\xi_n \in \mathbb{R}^d$.

For discretizations of infinite-dimensional models, $\|\cdot\|_-$ will typically be a finite-dimensional approximation of a norm that is almost surely finite w.r.t. the limit measure in infinite-dimensions.

EXAMPLE 1.6 (Transition path sampling). Consider the situation of Example 1.1, and assume that H is in $C^6(\mathbb{R}^d)$. Then by (1.9) and (1.6), V_d is C^4. For $n \le 4$ and $x, \xi_1, \dots, \xi_n \in \mathbb{R}^d$, the directional derivatives of V_d are given by

$$\partial^n_{\xi_1\cdots\xi_n} V_d(x) = 2^{-m-1} \sum_{k=0}^{2^m} w_k D^n \phi(y_{k2^{-m}})[\eta_{1,k2^{-m}}, \dots, \eta_{n,k2^{-m}}], \tag{1.35}$$

where $y, \eta_1, \dots, \eta_n$ are the polygonal paths in E corresponding to $x, \xi_1, \dots, \xi_n$, respectively, $w_k = 1$ for $k = 1, \dots, 2^m - 1$ and $w_0 = w_1 = 1/2$. Assuming $\|D^4\phi(z)\| = O(|z|^r)$ for some integer $r \ge 0$ as $|z| \to \infty$, we can estimate

$$\left|\partial^n_{\xi_1\cdots\xi_n} V_d(x)\right| \le C_n \max\left(1, \|y\|_{L^q}\right)^{p_n} \|\eta_1\|_{L^q} \cdot \dots \cdot \|\eta_n\|_{L^q},$$

where $q = r+4$, $p_n = r + (4-n)$, $\|y\|_{L^q} = 2^{-m} \sum_{k=0}^{2^m} w_k |y_k|^q$ is a discrete L^q norm of the polygonal path y and $C_1, \dots, C_4$ are finite constants that *do not depend on the dimension* d. One could now choose for the minus norm the norm on $\mathbb{R}^d$ corresponding to the discrete L^q norm on polygonal paths. However, it is more convenient to choose a norm coming from an inner product. To this end, we consider the norms

$$\|y\|_\alpha = \left(\sum_{n,k,i} 2^{-2\alpha n} x_{n,k,i}^2\right)^{1/2}, \qquad y = e + \sum x_{n,k,i} e^{n,k,i},$$

on path space E, and the induced norms

$$\|x\|_\alpha = \left(\sum_{n<m} \sum_{k,i} 2^{-2\alpha n} x_{n,k,i}^2\right)^{1/2}, \qquad x \in \mathbb{R}^d,$$

on $\mathbb{R}^d$ where $d = (2^m - 1)\ell$. One can show that for $\alpha < 1/2 + 1/q$, the L^q norm can be bounded from above by $\|\cdot\|_\alpha$ independently of the dimension; cf. [12]. On the other hand, if $\alpha > 1/2$, then $\|y\|_\alpha < \infty$ for γ-almost every path y of the Brownian bridge. This property will be crucial when restricting to balls w.r.t. $\|\cdot\|_\alpha$. For $\|\cdot\|_- = \|\cdot\|_\alpha$ with $\alpha \in (1/2, 1/2 + 1/q)$, both requirements are satisfied, and Assumption 1.5 holds with constants that do not depend on the dimension.

The next proposition yields in particular an upper bound for the average rejection probability w.r.t. both Ornstein–Uhlenbeck and semi-implicit Euler proposals at a given position $x \in \mathbb{R}^d$; cf. [6] for an analogue result:

PROPOSITION 1.7 (Upper bounds for MH rejection probabilities). *Suppose that Assumption 1.5 is satisfied and let $k \in \mathbb{N}$. Then there exist polynomials $\mathcal{P}_k^{\mathrm{OU}} : \mathbb{R} \to \mathbb{R}_+$ and $\mathcal{P}_k : \mathbb{R}^2 \to \mathbb{R}_+$ of degrees $p_1 + 1$, $\max(p_3 + 3, 2p_2 + 2)$, respectively, such that for any $x \in \mathbb{R}^d$ and $h \in (0, 2)$,*

$$E\big[\big(1 - \alpha^{\mathrm{OU}}\big(x, Y_h^{\mathrm{OU}}(x)\big)\big)^k\big]^{1/k} \le \mathcal{P}_k^{\mathrm{OU}}\big(\|x\|_-\big) \cdot h^{1/2} \quad \textit{and}$$

$$E\big[\big(1 - \alpha_h\big(x, Y_h(x)\big)\big)^k\big]^{1/k} \le \mathcal{P}_k\big(\|x\|_-, \|\nabla U(x)\|_-\big) \cdot h^{3/2}.$$

The result is a consequence of Proposition 1.3. The proof is given in Section 4 below.

REMARK 1.8. (1) The polynomials $\mathcal{P}_k^{\mathrm{OU}}$ and $\mathcal{P}_k$ in Proposition 1.7 are explicit; cf. the proof below. They depend only on the values C_n, p_n in Assumption 1.5 for $n = 1$, $n = 2, 3$, respectively, and on the moments

$$m_n = E\big[\|Z\|_-^n\big], \qquad n \le k \cdot (p_1 + 1), \qquad n \le k \cdot \max(p_3 + 3, 2p_2 + 2),$$

respectively,

but they do not depend on the dimension d. For semi-implicit Euler proposals, the upper bound in Proposition 1.7 is stated in explicit form for the case $k = 1$ and $p_2 = p_3 = 0$ in (4.6) below.

(2) For *explicit Euler proposals*, corresponding estimates hold with m_n replaced by $\widetilde{m}_n = E[|Z|^n]$; cf. Remark 4.3 below. Note, however, that $\widetilde{m}_n \to \infty$ as $d \to \infty$.

Our next result is a bound of order $O(h^{1/2})$, $O(h^{3/2})$, respectively, for the average dependence of the acceptance event on the current state w.r.t. Ornstein–Uhlenbeck and semi-implicit Euler proposals. Let $\|\cdot\|_+$ denote the dual norm of $\|\cdot\|_-$ on $\mathbb{R}^d$, that is,

$$\|\xi\|_+ = \sup\{\xi \cdot \eta \,|\, \eta \in \mathbb{R}^d \text{ with } \|\eta\|_- \le 1\}.$$

Note that

$$\|\xi\|_- \le |\xi| \le \|\xi\|_+ \qquad \forall \xi \in \mathbb{R}^d.$$

For a function $F \in C^1(\mathbb{R}^d)$,

$$\begin{aligned}\left|F(y) - F(x)\right| &= \left|\int_0^1 (y-x)\cdot \nabla F\big((1-t)x + ty\big)\,dt\right| \\ &\le \|y-x\|_- \cdot \sup_{z\in[x,y]} \left\|\nabla F(z)\right\|_+,\end{aligned}$$

that is, the plus norm of ∇F determines the Lipschitz constant w.r.t. the minus norm.

PROPOSITION 1.9 (Dependence of rejection on the current state). *Suppose that Assumption 1.5 is satisfied, and let $k \in \mathbb{N}$. Then there exist polynomials $\mathcal{Q}_k^{\mathrm{OU}} : \mathbb{R} \to \mathbb{R}_+$ and $\mathcal{Q}_k : \mathbb{R}^2 \to \mathbb{R}_+$ of degrees* $p_2 + 1$, $\max(p_4 + 3, p_3 + p_2 + 2, 3p_2 + 1)$, *respectively, such that for any $x, \widetilde{x} \in \mathbb{R}^d$ and $h \in (0, 2)$,*

$$E\big[\big\|\nabla_x G^{\mathrm{OU}}\big(x, Y_h^{\mathrm{OU}}(x)\big)\big\|_+^k\big]^{1/k} \le \mathcal{Q}_k^{\mathrm{OU}}\big(\|x\|_-\big)\cdot h^{1/2}, \tag{1.36}$$

$$E\big[\big\|\nabla_x G_h\big(x, Y_h(x)\big)\big\|_+^k\big]^{1/k} \le \mathcal{Q}_k\big(\|x\|_-, \big\|\nabla U(x)\big\|_-\big)\cdot h^{3/2}, \tag{1.37}$$

$$\begin{aligned}&E\big[\big|\alpha^{\mathrm{OU}}\big(x, Y_h^{\mathrm{OU}}(x)\big) - \alpha^{\mathrm{OU}}\big(\widetilde{x}, Y_h^{\mathrm{OU}}(\widetilde{x})\big)\big|^k\big]^{1/k} \\ &\qquad \le \mathcal{Q}_k^{\mathrm{OU}}\big(\max\big(\|x\|_-, \|\widetilde{x}\|_-\big)\big)\cdot \|x - \widetilde{x}\|_- \cdot h^{1/2} \quad \textit{and}\end{aligned} \tag{1.38}$$

$$\begin{aligned}&E\big[\big|\alpha_h\big(x, Y_h(x)\big) - \alpha_h\big(\widetilde{x}, Y_h(\widetilde{x})\big)\big|^k\big]^{1/k} \\ &\qquad \le \mathcal{Q}_k\Big(\max\big(\|x\|_-, \|\widetilde{x}\|_-\big), \sup_{z\in[x,\widetilde{x}]}\big\|\nabla U(z)\big\|_-\Big)\cdot \|x - \widetilde{x}\|_- \cdot h^{3/2},\end{aligned} \tag{1.39}$$

where $[x, \widetilde{x}]$ denotes the line segment between x and $\widetilde{x}$.

The proof of the proposition is given in Section 5 below.

REMARK 1.10. Again, the polynomials $\mathcal{Q}_k^{\mathrm{OU}}$ and $\mathcal{Q}_k$ are explicit. They depend only on the values C_n, p_n in Assumption 1.5 for $n = 1, 2$, $n = 2, 3, 4$, respectively, and on the moments $m_n = E[\|Z\|_-^n]$ for $n \le k \cdot (p_2 + 1)$, $n \le k \cdot \max(p_4 + 3, p_3 + p_2 + 2, 2p_2 + 1)$, respectively, but they do not depend on the dimension d. For semi-implicit Euler proposals, the upper bound in Proposition 1.9 is made explicit for the case $k = 1$ and $p_2 = p_3 = p_4 = 0$ in (5.18) below.

For Ornstein–Uhlenbeck proposals, it will be useful to state the bounds in Propositions 1.7 and 1.9 more explicitly for the case $p_2 = 0$, that is, when the second derivatives of V are uniformly bounded w.r.t. the minus norm:

PROPOSITION 1.11. *Suppose that Assumption 1.5 is satisfied for $n=1,2$ with $p_2=0$. Then for any $x,\widetilde{x}\in\mathbb{R}^d$ and $h\in(0,2)$,*

$$\begin{aligned}&\mathbb{E}\big[1-\alpha^{\mathrm{OU}}\big(x,Y_h^{\mathrm{OU}}(x)\big)\big]\\&\quad\le m_1\big(C_1+C_2\|x\|_-\big)\cdot h^{1/2}\\&\qquad+\tfrac{1}{2}\big(2m_2C_2+C_1\|x\|_-+C_2\|x\|_-^2\big)\cdot h+\tfrac{1}{2}m_1C_2\|x\|_-\cdot h^{3/2},\end{aligned}$$

and

$$\begin{aligned}&E\big[\big|\alpha^{\mathrm{OU}}\big(x,Y_h^{\mathrm{OU}}(x)\big)-\alpha^{\mathrm{OU}}\big(\widetilde{x},Y_h^{\mathrm{OU}}(\widetilde{x})\big)\big|^2\big]^{1/2}\\&\quad\le\big(m_2^{1/2}C_2\cdot h^{1/2}+\tfrac{1}{2}\big(C_1+2C_2\max\big(\|x\|_-,\|\widetilde{x}\|_-\big)\big)\cdot h\big)\cdot\|x-\widetilde{x}\|_-.\end{aligned}$$

The proof is given in Sections 4 and 5 below. Again, corresponding bounds also hold for L^k norms for $k\neq1,2$.

1.4. *Wasserstein contractivity.* The bounds in Propositions 1.7, 1.9 and 1.11 can be applied to study contractivity properties of Metropolis–Hastings transition kernels. Recall that the *Kantorovich–Rubinstein* or *L^1-Wasserstein distance* of two probability measures μ and ν on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ w.r.t. a given metric d on $\mathbb{R}^d$ is defined by

$$\mathcal{W}(\mu,\nu)=\inf_{\eta\in\Pi(\mu,\nu)}\int d(x,\widetilde{x})\eta(dx\,d\widetilde{x}),$$

where $\Pi(\mu,\nu)$ consists of all couplings η of μ and ν, that is, all probability measures η on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals μ and ν; cf., for example, [34]. Recall that a coupling of μ and ν can be realized by random variables W and $\widetilde{W}$ defined on a joint probability space such that $W\sim\mu$ and $\widetilde{W}\sim\nu$.

In order to derive upper bounds for the distances $\mathcal{W}(\mu q_h,\nu q_h)$, and, more generally, $\mathcal{W}(\mu q_h^n,\nu q_h^n)$, $n\in\mathbb{N}$, we define a coupling of the MALA transition probabilities $q_h(x,\cdot)$, $x\in\mathbb{R}^d$, by setting

$$W_h(x):=\begin{cases}Y_h(x), & \text{if } \mathcal{U}\le\alpha_h\big(x,Y_h(x)\big),\\ x, & \text{if } \mathcal{U}>\alpha_h\big(x,Y_h(x)\big).\end{cases}$$

Here $Y_h(x)$, $x\in\mathbb{R}^d$, is the basic coupling of the proposal distributions $p_h(x,\cdot)$ defined by (1.27) with $Z\sim\gamma^d$, and the random variable $\mathcal{U}$ is uniformly distributed in $(0,1)$ and independent of Z.

Correspondingly, we define a coupling of the Metropolis–Hastings transition kernels q_h^{OU} based on Ornstein–Uhlenbeck proposals by setting

$$W_h^{\mathrm{OU}}(x):=\begin{cases}Y_h^{\mathrm{OU}}(x), & \text{if } \mathcal{U}\le\alpha^{\mathrm{OU}}\big(x,Y_h^{\mathrm{OU}}(x)\big),\\ x, & \text{if } \mathcal{U}>\alpha^{\mathrm{OU}}\big(x,Y_h^{\mathrm{OU}}(x)\big).\end{cases}$$

Let

$$B_R^- := \{x \in \mathbb{R}^d : \|x\|_- < R\}$$

denote the centered ball of radius R w.r.t. $\|\cdot\|_-$. As a consequence of Proposition 1.11 above, we obtain the following upper bound for the Kantorovich–Rubinstein–Wasserstein distance of $q_h^{\mathrm{OU}}(x, \cdot)$ and $q_h^{\mathrm{OU}}(\widetilde{x}, \cdot)$ w.r.t. the metric $d(x, \widetilde{x}) = \|x - \widetilde{x}\|_-$:

THEOREM 1.12 (Contractivity of MH transitions based on OU proposals). *Suppose that Assumption 1.5 is satisfied for $n = 1, 2$ with $p_2 = 0$. Then for any $h \in (0, 2)$, $R \in (0, \infty)$, and $x, \widetilde{x} \in B_R^-$,*

$$\mathbb{E}[\|W_h^{\mathrm{OU}}(x) - W_h^{\mathrm{OU}}(\widetilde{x})\|_-] \le c_h^{\mathrm{OU}}(R) \cdot \|x - \widetilde{x}\|_-,$$

where

$$c_h^{\mathrm{OU}}(R) = 1 - \tfrac{1}{2}h + m_2 C_2 h + A(1 + R)(1 + h^{1/2} R)h^{3/2}$$

with an explicit constant A that only depends on the values m_1, m_2, C_1 and C_2.

The proof is given in Section 7 below.

Theorem 1.12 shows that Wasserstein contractivity holds on the ball B_R^- provided $2m_2C_2 < 1$ and h is chosen sufficiently small depending on R [with $h^{1/2} = O(R^{-1})$]. In this case, the contraction constant $c_h^{\mathrm{OU}}(R)$ depends on the dimension only through the values of the constants C_1, C_2, m_1 and m_2. On the other hand, the following one-dimensional example shows that for $m_2C_2 > 1$, the acceptance-rejection step may destroy the contraction properties of the OU proposals:

EXAMPLE 1.13. Suppose that $d = 1$ and $\|\cdot\|_- = |\cdot|$. If $V(x) = bx^2/2$ with a constant $b \in (-1/2, 1/2)$, then by Theorem 1.12, Wasserstein contractivity holds for the Metropolis–Hastings chain with Ornstein–Uhlenbeck proposals on the interval $(-R, R)$ provided h is chosen sufficiently small. On the other hand, if $V(x) = bx^2/2$ for $|x| \le 1$ with a constant $b < -1$, then the logarithmic density

$$U(x) = V(x) + x^2/2 = (b + 1) \cdot x^2/2$$

is strictly concave for $|x| \le 1$, and it can be easily seen that Wasserstein contractivity on $(-1, 1)$ does not hold for the MH chain with OU proposals if h is sufficiently small.

A disadvantage of the result for Ornstein–Uhlenbeck proposals stated above is that not only a lower bound on the second derivative of V is required (this would be a fairly natural condition as the example indicates), but also an upper bound of the same size. For semi-implicit Euler proposals, we can derive a better result that requires only a strictly positive lower bound on the second derivative of

$U(x) = V(x) + |x|^2/2$ and Assumption 1.5 with arbitrary constants to be satisfied. For this purpose we assume that

$$\|\cdot\|_- = \langle\cdot,\cdot\rangle^{1/2}$$

for an inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^d$, and we make the following assumption on U:

ASSUMPTION 1.14. There exists a strictly positive constant $K \in (0, 1]$ such that

$$\langle \eta, \nabla^2 U(x) \cdot \eta \rangle \geq K\langle \eta, \eta\rangle \qquad \text{for any } x, \eta \in \mathbb{R}^d. \tag{1.40}$$

Of course, Assumption 1.14 is still restrictive, and it will often be satisfied only in a suitable ball around a local minimum of U. Most of the results below are stated on a given ball B_R^- w.r.t. the minus norm. In this case it is enough to assume that 1.14 holds on that ball. If $\|\cdot\|_-$ coincides with the Euclidean norm $|\cdot|$, then the assumption is equivalent to convexity of $U(x) - K|x|^2$. Moreover, since $\nabla^2 U(x) = I_d + \nabla^2 V(x)$, a sufficient condition for (1.40) to hold is

$$\|\nabla^2 V(x) \cdot \eta\|_- \leq (1 - K)\|\eta\|_- \qquad \text{for any } x, \eta \in \mathbb{R}^d. \tag{1.41}$$

As a consequence of Propositions 1.7 and 1.9 above, we obtain the following upper bound for the Kantorovich–Rubinstein–Wasserstein distance of $q_h(x,\cdot)$ and $q_h(\tilde{x},\cdot)$ w.r.t. the metric $d(x,\tilde{x}) = \|x - \tilde{x}\|_-$:

THEOREM 1.15 (Contractivity of semi-implicit MALA transitions). *Suppose that Assumptions 1.5 and 1.14 are satisfied. Then for any $h \in (0, 2)$, $R \in (0,\infty)$ and $x, \tilde{x} \in B_R^-$,*

$$\mathbb{E}\big[\|W_h(x) - W_h(\tilde{x})\|_-\big] \leq c_h(R) \cdot \|x - \tilde{x}\|_-,$$

where

$$c_h(R) = 1 - \tfrac{1}{2}Kh + \big(\tfrac{1}{8}M(R)^2 + \gamma(R)\big)h^2 + \big(K\beta(R) + \tfrac{1}{2}\delta(R)\big)h^{5/2}$$

with

$$\begin{aligned}
M(R) &= \sup\{\|\nabla^2 U(z) \cdot \eta\|_- : \eta \in B_1^-, z \in B_R^-\},\\
\beta(R) &= \sup\{\mathcal{P}_1(\|z\|_-, \|\nabla U(z)\|_-) : z \in B_R^-\},\\
\gamma(R) &= m_2^{1/2} \cdot \sup\{\mathcal{Q}_2(\|z\|_-, \|\nabla U(z)\|_-) : z \in B_R^-\},\\
\delta(R) &= \sup\{\mathcal{Q}_2(\|z\|_-, \|\nabla U(z)\|_-)\|\nabla U(z)\|_- : z \in B_R^-\}.
\end{aligned}$$

The proof is given in Section 7 below.

REMARK 1.16. Theorem 1.15 shows in particular that under Assumptions 1.5 and 1.14, there exist constants $C, q \in (0, \infty)$ such that the contraction

$$\mathbb{E}[\| W_h(x) - W_h(\widetilde{x})\|_-] \le \left(1 - \frac{K}{4}h\right)\|x - \widetilde{x}\|$$

holds for $x, \widetilde{x} \in B_R^-$ whenever $h^{-1} \ge C \cdot (1 + R^q)$.

EXAMPLE 1.17 (Transition path sampling). In the situation of Examples 1.1 and 1.6 above, condition (1.41) and (hence) assumption 1.14 are satisfied on a ball B_R^- with K independent of d provided $\|D^2\phi(x)\| \le 1 - K$ for any $x \in B_R^-$; cf. (1.35). More generally, by modifying the metric in a suitable way if necessary, one may expect Assumption 1.14 to hold uniformly in the dimension in neighborhoods of local minima of U.

1.5. *Conclusions*. For $R \in (0, \infty)$, we denote by $\mathcal{W}_R$ the Kantorovich–Rubinstein–Wasserstein distance based on the distance function

$$d_R(x, \widetilde{x}) := \min(\|x - \widetilde{x}\|_-, 2R). \tag{1.42}$$

Note that d_R is a bounded metric that coincides with the distance function induced by the minus norm on B_R^-. The bounds resulting from Theorems 1.15 and 1.12 can be iterated to obtain estimates for the KRW distance $\mathcal{W}_R$ between the distributions of the corresponding Metropolis–Hastings chains after n steps w.r.t. two different initial distributions.

COROLLARY 1.18. *Suppose that Assumptions 1.5 and 1.14 are satisfied, and let $h \in (0, 2)$ and $R \in (0, \infty)$. Then for any $n \in \mathbb{N}$, and for any probability measures μ, ν on $\mathcal{B}(\mathbb{R}^d)$,*

$$\begin{aligned}\mathcal{W}_R(\mu q_h^n, \nu q_h^n) &\le c_h(R)^n \mathcal{W}_R(\mu, \nu)\\ &\quad + 2R \cdot (\mathbb{P}_\mu[\exists k < n : X_k \notin B_R^-] + \mathbb{P}_\nu[\exists k < n : X_k \notin B_R^-]).\end{aligned}$$

Here $c_h(R)$ is the constant in Theorem 1.15, and $(X_n, \mathbb{P}_\mu)$ and $(X_n, \mathbb{P}_\nu)$ are Markov chains with transition kernel q_h and initial distributions μ, ν, respectively. A corresponding result with c_h replaced by c_h^{OU} holds for the Metropolis–Hastings chain with Ornstein–Uhlenbeck proposals.

Since the joint law of $W_h(x)$ and $W_h(\widetilde{x})$ is a coupling of $q_h(x, \cdot)$ and $q_h(\widetilde{x}, \cdot)$ for any $x, \widetilde{x} \in \mathbb{R}^d$, Corollary 1.18 is a direct consequence of Theorems 1.15, 1.12, respectively, and Theorem 2.3 below. The corollary can be used to quantify the Wasserstein distance between the distribution of the Metropolis–Hastings chain after n steps w.r.t. two different initial distributions. For this purpose, one can estimate the exit probabilities from the ball B_R^- via an argument based on a Lyapunov function. For semi-implicit Euler proposals we eventually obtain the following main result:

THEOREM 1.19 (Quantitative convergence bound for semi-implicit MALA). *Suppose that Assumptions 1.5 and 1.14 are satisfied. Then there exist constants* $C, D, q \in (0, \infty)$ *such that the estimate*

$$\mathcal{W}_{2R}(\nu q_h^n, \pi q_h^n) \le \left(1 - \frac{K}{4}h\right)^n \mathcal{W}_{2R}(\nu, \pi) + DR\exp\left(-\frac{KR^2}{8}\right)nh$$

holds for any $n \in \mathbb{N}$, $h, R \in (0, \infty)$ *such that* $h^{-1} \ge C \cdot (1+R)^q$, *and for any probability measures* ν, π *on* $\mathbb{R}^d$ *with support in* B_R^-. *The constants* C, D *and* q *can be made explicit. They depend only on the values of the constants in Assumptions 1.5 and 1.14 and on the moments* m_k, $k \in \mathbb{N}$, *w.r.t. the minus norm, but they do not depend explicitly on the dimension.*

The proof of Theorem 1.19 is given in Section 7 below.

Let $\mu_R(A) = \mu(A|B_R^-)$ denote the conditional probability measure given B_R^-. Recalling that μ is a stationary distribution for the kernel q_h, we can apply Theorem 1.19 to derive a bound for the Wasserstein distance of the discretization of the MALA chain and μ_R after n steps:

THEOREM 1.20. *Suppose that Assumptions 1.5 and 1.14 are satisfied. Then there exist constants* $C, \bar{D}, q \in (0, \infty)$ *that do not depend explicitly on the dimension such that the estimate*

$$\mathcal{W}_{2R}(\nu q_h^n, \mu_R) \le 58R\left(1 - \frac{K}{4}h\right)^n + \bar{D}R\exp\left(-\frac{KR^2}{33}\right)nh$$

holds for any $n \in \mathbb{N}$, $h, R \in (0, \infty)$ *such that* $h^{-1} \ge C \cdot (1+R)^q$, *and for any probability measure* ν *on* $\mathbb{R}^d$ *with support in* B_R^-.

The proof is given in Section 7.

Given an error bound $\varepsilon \in (0, \infty)$ for the Kantorovich–Rubinstein–Wasserstein distance, we can now determine how many steps of the MALA chain are required such that

$$\mathcal{W}_{2R}(\nu q_h^n, \mu_R) < \varepsilon \qquad \text{for any } \nu \text{ with support in } B_R^-. \tag{1.43}$$

Assuming

$$nh \ge \frac{4}{K}\log\left(\frac{116R}{\varepsilon}\right), \tag{1.44}$$

we have $58R(1 - Kh/4)^n \le \varepsilon/2$. Hence (1.43) holds provided the assumptions in Theorem 1.20 are satisfied, and

$$\bar{D}R\exp(-KR^2/33)nh < \varepsilon/2. \tag{1.45}$$

For a minimal choice of n, all conditions are satisfied if R is of order $(\log\varepsilon^{-1})^{1/2}$ up to a log log correction, and the inverse step size h^{-1} is of order $(\log\varepsilon^{-1})^{q/2}$

up to a $\log\log$ correction. Hence if Assumption 1.14 holds on $\mathbb{R}^d$, then a number n of steps that is polynomial in $\log \varepsilon^{-1}$ is sufficient to bound the error by ε independently of the dimension.

On the other hand, if Assumption 1.14 is satisfied only on a ball B_R^- of given radius R, then a given error bound ε is definitely achieved only provided (1.45) holds with the minimal choice for nh satisfying (1.44), that is, if

$$8\bar{D}K^{-1}\log(116R\varepsilon^{-1})R\exp(-KR^2/33) < \varepsilon. \tag{1.46}$$

If ε is chosen smaller, then the chain may leave the ball B_R^- before sufficient mixing on B_R^- has taken place.

2. Wasserstein contractivity of Metropolis–Hastings kernels. In this section, we first consider an arbitrary stochastic kernel $q : S \times \mathcal{B}(S) \to [0, 1]$ on a metric space (S, d). Further below, we will choose $S = \mathbb{R}^d$ and $d(x, y) = \|x - y\|_- \wedge R$ for some constant $R \in (0, \infty]$, and we will assume that q is the transition kernel of a Metropolis–Hastings chain.

The *Kantorovich–Rubinstein* or L^1-*Wasserstein distance* of two probability measures μ and ν on the Borel-σ-algebra $\mathcal{B}(S)$ w.r.t. the metric d is defined by

$$\mathcal{W}_d(\mu, \nu) = \inf_\eta \int d(x, \tilde{x})\eta(dx\, d\tilde{x}),$$

where the infimum is over all couplings η of μ and ν, that is, over all probability measures η on $S \times S$ with marginals μ and ν; cf., for example, [34]. In order to derive upper bounds for the Kantorovich distances $\mathcal{W}_d(\mu q, \nu q)$, and more generally, $\mathcal{W}_d(\mu q^n, \nu q^n)$, $n \in \mathbb{N}$, we construct couplings between the measures $q(x, \cdot)$ for $x \in S$, and we derive bounds for the distances $\mathcal{W}_d(q(x, \cdot), q(\tilde{x}, \cdot))$, $x, \tilde{x} \in S$.

DEFINITION 2.1. A *Markovian coupling* of the probability measures $q(x, \cdot)$, $x \in S$, is a stochastic kernel c on the product space $(S \times S, \mathcal{B}(S \times S))$ such that for any $x, \tilde{x} \in S$, the distribution of the first and second component under $c((x, \tilde{x}), dy\, d\tilde{y})$ is $q(x, dy)$ and $q(\tilde{x}, d\tilde{y})$, respectively.

EXAMPLE 2.2. (1) Suppose that $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and let $(x, \tilde{x}, \omega) \mapsto Y(x, \tilde{x})(\omega)$, $(x, \tilde{x}, \omega) \mapsto \tilde{Y}(x, \tilde{x})(\omega)$ be product measurable functions from $S \times S \times \Omega$ to S such that $Y(x, \tilde{x}) \sim q(x, \cdot)$ and $\tilde{Y}(x, \tilde{x}) \sim q(\tilde{x}, \cdot)$ w.r.t. $\mathbb{P}$ for any $x, \tilde{x} \in S$. Then the joint distributions

$$c\big((x, \tilde{x}), \cdot\big) = \mathbb{P} \circ \big(Y(x, \tilde{x}), \tilde{Y}(x, \tilde{x})\big)^{-1}, \qquad x, \tilde{x} \in S,$$

define a Markovian coupling of the measures $q(x, \cdot)$, $x \in S$.

(2) In particular, if $(x, \omega) \mapsto Y(x)(\omega)$ is a product measurable function from $S \times \Omega$ to S such that $Y(x) \sim q(x, \cdot)$ for any $x \in S$, then

$$c\big((x, \tilde{x}), \cdot\big) = \mathbb{P} \circ \big(Y(x), Y(\tilde{x})\big)^{-1}$$

is a Markovian coupling of the measures $q(x, \cdot)$, $x \in S$.

Suppose that $(X_n, \widetilde{X}_n)$ on $(\Omega, \mathcal{A}, \mathbb{P})$ is a Markov chain with values in $S \times S$ and transition kernel c, where c is a Markovian coupling w.r.t. the kernel q. Then the components (X_n) and $(\widetilde{X}_n)$ are Markov chains with transition kernel q and initial distributions given by the marginals of the initial distribution of $(X_n, \widetilde{X}_n)$, that is, $(X_n, \widetilde{X}_n)$ is a coupling of these Markov chains. We will apply the following general theorem to quantify the deviation from equilibrium after n steps of the Markov chain with transition kernel q:

THEOREM 2.3. *Let $\gamma \in (0, 1)$, and let $c((x, \widetilde{x}), dy\, d\widetilde{y})$ be a Markovian coupling of the probability measures $q(x, \cdot)$, $x \in S$. Suppose that $\mathcal{O}$ is an open subset of S, and assume that the metric d is bounded. Let*

$$\Delta := \operatorname{diam} S = \sup\{d(x, \widetilde{x}) : x, \widetilde{x} \in S\}.$$

If the contractivity condition

$$\int d(y, \widetilde{y}) c\big((x, \widetilde{x}), dy\, d\widetilde{y}\big) \le \gamma \cdot d(x, \widetilde{x}) \tag{2.1}$$

holds for any $x, \widetilde{x} \in \mathcal{O}$, then

$$\begin{aligned} &\mathcal{W}_d\big(\mu q^n, \nu q^n\big) \\ &\quad \le \gamma^n \mathcal{W}_d(\mu, \nu) + \Delta \cdot \big(\mathbb{P}_\mu[\exists k < n : X_k \notin \mathcal{O}] + \mathbb{P}_\nu[\exists k < n : X_k \notin \mathcal{O}]\big) \end{aligned} \tag{2.2}$$

for any $n \in \mathbb{N}$ and for any probability measures μ, ν on $\mathcal{B}(S)$. Here $(X_n, \mathbb{P}_\mu)$ and $(X_n, \mathbb{P}_\nu)$ are Markov chains with transition kernel q and initial distributions μ, ν, respectively.

PROOF. Suppose that μ and ν are probability measures on $\mathcal{B}(S)$ and $\eta(dx\, d\widetilde{x})$ is a coupling of μ and ν. We consider the coupling chain $(X_n, \widetilde{X}_n)$ on $(\Omega, \mathcal{A}, \mathbb{P})$ with initial distribution η and transition kernel c. Since (X_n) and $(\widetilde{X}_n)$ are Markov chains with transition kernel q and initial distributions μ and ν, we have $\mathbb{P} \circ X_n^{-1} = \mu q^n$ and $\mathbb{P} \circ \widetilde{X}_n^{-1} = \nu q^n$ for any $n \in \mathbb{N}$. Moreover, by (2.1),

$$\begin{aligned} &\mathbb{E}\big[d(X_n, \widetilde{X}_n); (X_k, \widetilde{X}_k) \in \mathcal{O} \times \mathcal{O}\ \forall k < n\big] \\ &\quad = \mathbb{E}\left[\int d(x_n, \widetilde{x}_n) c\big((X_{n-1}, \widetilde{X}_{n-1}), dx_n\, d\widetilde{x}_n\big); (X_k, \widetilde{X}_k) \in \mathcal{O} \times \mathcal{O}\ \forall k < n\right] \\ &\quad \le \gamma \mathbb{E}\big[d(X_{n-1}, \widetilde{X}_{n-1}); (X_k, \widetilde{X}_k) \in \mathcal{O} \times \mathcal{O}\ \forall k < n-1\big]. \end{aligned}$$

Therefore, by induction,

$$\begin{aligned} \mathcal{W}_d\big(\mu q^n, \nu q^n\big) &\le \mathbb{E}\big[d(X_n, \widetilde{X}_n)\big] \\ &= \mathbb{E}\big[d(X_n, \widetilde{X}_n); (X_k, \widetilde{X}_k) \in \mathcal{O} \times \mathcal{O}\ \forall k < n\big] \\ &\quad + \mathbb{E}\big[d(X_n, \widetilde{X}_n); \exists k < n : (X_k, \widetilde{X}_k) \notin \mathcal{O} \times \mathcal{O}\big] \\ &\le \gamma^n d(x, \widetilde{x}) + \Delta \cdot \mathbb{P}\big[\exists k < n : (X_k, \widetilde{X}_k) \notin \mathcal{O} \times \mathcal{O}\big], \end{aligned}$$

which implies (2.2). □

REMARK 2.4. Theorem 2.3 may also be useful for studying local equilibration of a Markov chain within a metastable state. In fact, if $\mathcal{O}$ is a region of the state space where the process stays with high probability for a long time, and if a contractivity condition holds on $\mathcal{O}$, then the result can be used to bound the Kantorovich–Rubinstein–Wasserstein distance between the distribution after a finite number of steps and the stationary distribution conditioned to $\mathcal{O}$.

From now on, we assume that we are given a Markovian coupling of the proposal distributions $p(x,\cdot), x \in \mathbb{R}^d$, of a Metropolis–Hastings algorithm which is realized by product measurable functions $(x, \widetilde{x}, \omega) \mapsto Y(x,\widetilde{x})(\omega)$, $\widetilde{Y}(x,\widetilde{x})(\omega)$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that

$$Y(x,\widetilde{x}) \sim p(x,\cdot) \quad \text{and} \quad \widetilde{Y}(x,\widetilde{x}) \sim p(\widetilde{x},\cdot) \qquad \text{for any } x, \widetilde{x} \in \mathbb{R}^d.$$

Let $\alpha(x,y)$ and $q(x,dy)$ again denote the acceptance probabilities and the transition kernel of the Metropolis–Hastings chain with stationary distribution μ; cf. (1.10), (1.11) and (1.12). Moreover, suppose that $\mathcal{U}$ is a uniformly distributed random variable with values in $(0,1)$ that is independent of $\{Y(x,\widetilde{x}) : x, \widetilde{x} \in \mathbb{R}^d\}$. Then the functions $(x,\widetilde{x},\omega) \mapsto W(x,\widetilde{x})(\omega)$, $\widetilde{W}(x,\widetilde{x})(\omega)$ defined by

$$W(x,\widetilde{x}) := \begin{cases} Y(x,\widetilde{x}), & \text{if } \mathcal{U} \le \alpha(x, Y(x,\widetilde{x})), \\ x, & \text{if } \mathcal{U} > \alpha(x, Y(x,\widetilde{x})), \end{cases}$$

$$\widetilde{W}(x,\widetilde{x}) := \begin{cases} \widetilde{Y}(x,\widetilde{x}), & \text{if } \mathcal{U} \le \alpha(x, \widetilde{Y}(x,\widetilde{x})), \\ \widetilde{x}, & \text{if } \mathcal{U} > \alpha(x, \widetilde{Y}(x,\widetilde{x})), \end{cases}$$

realize a Markovian coupling between the Metropolis–Hastings transition functions $q(x,\cdot), x \in \mathbb{R}^d$, that is,

$$W(x,\widetilde{x}) \sim q(x,\cdot) \quad \text{and} \quad \widetilde{W}(x,\widetilde{x}) \sim q(\widetilde{x},\cdot)$$

for any $x, \widetilde{x} \in \mathbb{R}^d$. This coupling is optimal in the acceptance step in the sense that it minimizes the probability that a proposed move from x to $Y(x,\widetilde{x})$ is accepted and the corresponding proposed move from $\widetilde{x}$ to $\widetilde{Y}(x,\widetilde{x})$ is rejected or vice versa.

LEMMA 2.5 (Basic contractivity lemma for MH kernels). *For any* $x, \widetilde{x} \in \mathbb{R}^d$,

$$\begin{aligned} &\mathbb{E}\big[d\big(W(x,\widetilde{x}), \widetilde{W}(x,\widetilde{x})\big)\big] \\ &\quad \le \mathbb{E}\big[d\big(Y(x,\widetilde{x}), \widetilde{Y}(x,\widetilde{x})\big)\big] \\ &\qquad + \mathbb{E}\big[\big(d(x,\widetilde{x}) - d\big(Y(x,\widetilde{x}), \widetilde{Y}(x,\widetilde{x})\big)\big) \\ &\qquad\qquad \times \max\big(1 - \alpha\big(x, Y(x,\widetilde{x})\big), 1 - \alpha\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big)\big)\big] \\ &\qquad + \mathbb{E}\big[d(x, Y(x,\widetilde{x}) \cdot \big(\alpha\big(x, Y(x,\widetilde{x})\big) - \alpha\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big)\big)^+\big] \\ &\qquad + \mathbb{E}\big[d(\widetilde{x}, \widetilde{Y}(x,\widetilde{x}) \cdot \big(\alpha\big(x, Y(x,\widetilde{x})\big) - \alpha\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big)\big)^-\big]. \end{aligned}$$

PROOF. By the definition of W and by the triangle inequality, we obtain the estimate

$$\begin{aligned}
&\mathbb{E}\big[d\big(W(x,\widetilde{x}), \widetilde{W}(x,\widetilde{x})\big)\big] \\
&\quad \le \mathbb{E}\big[d\big(Y(x,\widetilde{x}), \widetilde{Y}(x,\widetilde{x})\big); \mathcal{U} < \min\big(\alpha\big(x, Y(x,\widetilde{x})\big), \alpha\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big)\big)\big] \\
&\qquad + d(x,\widetilde{x}) \cdot \mathbb{P}\big[\mathcal{U} \ge \min\big(\alpha\big(x, Y(x,\widetilde{x})\big), \alpha\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big)\big)\big] \\
&\qquad + \mathbb{E}\big[d\big(x, Y(x,\widetilde{x})\big); \alpha\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big) \le \mathcal{U} < \alpha\big(x, Y(x,\widetilde{x})\big)\big] \\
&\qquad + \mathbb{E}\big[d\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big); \alpha\big(x, Y(x,\widetilde{x})\big) \le \mathcal{U} < \alpha\big(\widetilde{x}, \widetilde{Y}(x,\widetilde{x})\big)\big].
\end{aligned}$$

The assertion now follows by conditioning on Y and $\widetilde{Y}$. $\square$

REMARK 2.6. (1) Note that the upper bound in Lemma 2.5 is close to an equality. Indeed, the only estimate in the proof is the triangle inequality that has been applied to bound $d(x,\widetilde{Y})$ by $d(x,\widetilde{x}) + d(\widetilde{x},\widetilde{Y})$ and $d(\widetilde{x},Y)$ by $d(x,\widetilde{x}) + d(x,Y)$.

(2) For the couplings and distances considered in this paper, $d(Y,\widetilde{Y})$ will always be deterministic. Therefore, the upper bound in the lemma simplifies to

$$\begin{aligned}
&\mathbb{E}\big[d(W,\widetilde{W})\big] \\
&\quad \le d(Y,\widetilde{Y}) + \big(d(x,\widetilde{x}) - d(Y,\widetilde{Y})\big) \cdot \mathbb{E}\big[\max\big(1-\alpha(x,Y), 1-\alpha(\widetilde{x},\widetilde{Y})\big)\big] \\
&\qquad + \mathbb{E}\big[d(x,Y)\big(\alpha(x,Y) - \alpha(\widetilde{x},\widetilde{Y})\big)^{+} + d(\widetilde{x},\widetilde{Y})\big(\alpha(x,Y) - \alpha(\widetilde{x},\widetilde{Y})\big)^{-}\big].
\end{aligned} \tag{2.3}$$

Here $\mathbb{E}[\max(1-\alpha(x,Y), 1-\alpha(\widetilde{x},\widetilde{Y}))]$ is the probability that at least one of the proposals is rejected.

(3) If the metric d is bounded with diameter Δ, then the last two expectations in the upper bound in Lemma 2.5 can be estimated by Δ times the probability $\mathbb{E}[|\alpha(x,Y) - \alpha(\widetilde{x},\widetilde{Y})|]$ that one of the proposals is rejected and the other one is accepted. Alternatively (and usually more efficiently), these terms can be estimated by Hölder's inequality.

3. Contractivity of the proposal step. In this section we assume $V \in C^2(\mathbb{R}^d)$. We study contractivity properties of the Metropolis–Hastings proposals defined in (1.23) and (1.27).

Note first that the *Ornstein–Uhlenbeck proposals* do not depend on V. For $h \in (0,2)$, the contractivity condition

$$\big\|Y_h^{\mathrm{OU}}(x) - Y_h^{\mathrm{OU}}(\widetilde{x})\big\| = \big\|(1-h/2)(x-\widetilde{x})\big\| = (1-h/2)\|x-\widetilde{x}\| \tag{3.1}$$

holds pointwise for any $x,\widetilde{x} \in \mathbb{R}^d$ w.r.t. an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$.

For the *semi-implicit Euler proposals*

$$Y_h(x) = x - \frac{h}{2}\nabla U(x) + \sqrt{h - \frac{h^2}{4}}Z, \qquad Z \sim \gamma^d.$$

Wasserstein contractivity does not necessarily hold. Close to optimal sufficient conditions for contractivity w.r.t. the minus norm can be obtained in a straightforward way by considering the derivative of Y_h w.r.t. x.

LEMMA 3.1. *Let $h \in (0, 2)$, and let C be a convex subset of $\mathbb{R}^d$. If there exists a constant $\lambda \in (0, \infty)$ such that*

$$\left\| \left(I_d - \frac{h}{2} \nabla^2 U(x) \right) \cdot \eta \right\|_- \leq \lambda \|\eta\|_- \qquad \text{for any } \eta \in \mathbb{R}^d, x \in C, \tag{3.2}$$

then

$$\|Y_h(x) - Y_h(\widetilde{x})\|_- \leq \lambda \|x - \widetilde{x}\|_- \qquad \text{for any } x, \widetilde{x} \in C.$$

PROOF. If (3.2) holds, then

$$\|\partial_\eta Y_h(x)\|_- = \left\| \eta - \frac{h}{2} \nabla^2 U(x) \cdot \eta \right\|_- \leq \lambda \|\eta\|_-$$

for any $x \in C$ and $\eta \in \mathbb{R}^d$. Hence

$$\begin{aligned} \|Y_h(x) - Y_h(\widetilde{x})\|_- &= \left\| \int_0^1 \frac{d}{dt} Y_h\big(tx + (1-t)\widetilde{x}\big)\, dt \right\|_- \\ &\leq \int_0^1 \big\|\partial_{x-\widetilde{x}} Y_h\big(tx + (1-t)\widetilde{x}\big)\big\|_- dt \\ &\leq \lambda \|x - \widetilde{x}\|_- \qquad \text{for } x, \widetilde{x} \in C. \end{aligned}$$

□

REMARK 3.2. (1) Note that condition (3.2) requires a bound on $\nabla^2 U$ in both directions. This is in contrast to the continuous time case where a lower bound by a strictly positive constant is sufficient to guarantee contractivity of the derivative flow.

(2) Condition (3.2) is equivalent to

$$\xi \cdot \eta - \frac{h}{2} \partial^2_{\xi\eta} U(x) \leq \lambda \|\xi\|_+ \|\eta\|_- \qquad \text{for any } x \in C, \xi, \eta \in \mathbb{R}^d. \tag{3.3}$$

Recall that for $R \in (0, \infty]$,

$$M(R) = \sup\big\{ \|\nabla^2 U(x) \cdot \eta\|_- : \eta \in B_1^-, x \in B_R^- \big\}. \tag{3.4}$$

Hence $M(R)$ bounds the second derivative of U on B_R^- in both directions, whereas the constant K in Assumption 1.14 is a strictly positive lower bound for the second derivative. We also define

$$N(R) = \sup\big\{ \|\nabla^2 V(x) \cdot \eta\|_- : \eta \in B_1^-, x \in B_R^- \big\}. \tag{3.5}$$

Note that $M(R) \leq 1 + N(R)$. As a consequence of Lemma 3.1 we obtain:

PROPOSITION 3.3. *For any $h \in (0,2)$ and $x, \widetilde{x} \in B_R^-$,*

$$\|Y_h(x) - Y_h(\widetilde{x})\|_- \le \left(1 - \frac{1 - N(R)}{2} h\right) \cdot \|x - \widetilde{x}\|_-. \tag{3.6}$$

Moreover, if Assumption 1.14 holds, then

$$\|Y_h(x) - Y_h(\widetilde{x})\|_- \le \left(1 - \frac{K}{2} h + \frac{M(R)^2}{8} h^2\right) \cdot \|x - \widetilde{x}\|_-. \tag{3.7}$$

PROOF. Note that for $z \in [x, \widetilde{x}]$ and $\eta \in \mathbb{R}$,

$$\left(I - \frac{h}{2} \nabla^2 U(z)\right) \cdot \eta = \left(1 - \frac{h}{2}\right) \eta - \frac{h}{2} \nabla^2 V(z) \cdot \eta. \tag{3.8}$$

Therefore, by (3.5),

$$\left\| \left(I - \frac{h}{2} \nabla^2 U(z)\right) \cdot \eta \right\|_- \le \left(1 - \frac{h}{2}\right) \|\eta\|_- + \frac{h}{2} N(R) \cdot \|\eta\|_-.$$

Inequality (3.6) now follows by Lemma 3.1.

Moreover, if Assumption 1.14 holds, then for $z \in [x, \widetilde{x}]$ and $\eta \in \mathbb{R}^d$,

$$\begin{aligned} \left\| \left(I - \frac{h}{2} \nabla^2 U(z)\right) \cdot \eta \right\|_-^2 &= \|\eta\|_-^2 - h\langle \eta, \nabla^2 U(z) \cdot \eta\rangle + \frac{h^2}{4} \|\nabla^2 U(z) \cdot \eta\|_-^2 \\ &\le (1 - Kh + M(R)^2 h^2/4) \|\eta\|_-^2 \\ &= (1 - Kh/2 + M(R)^2 h^2/8)^2 \|\eta\|_-^2. \end{aligned}$$

Inequality (3.7) again follows by Lemma 3.1. $\square$

4. Upper bounds for rejection probabilities. In this section we derive the upper bounds for the MH rejection probabilities stated in Proposition 1.7. As a first step we prove the explicit formula for the MALA acceptance probabilities w.r.t. explicit and semi-implicit Euler proposals stated in Proposition 1.3:

PROOF OF PROPOSITION 1.3. For explicit Euler proposals with given step size $h > 0$,

$$\begin{aligned} &\log \gamma^d(x) p_h^{\mathrm{Euler}}(x, y) \\ &\quad = \frac{1}{2}|x|^2 + \frac{1}{2h}\left| y - \left(1 - \frac{h}{2}\right) x + \frac{h}{2} \nabla V(x)\right|^2 + C \\ &\quad = \frac{1}{2h}\biggl(h|x|^2 + |y - x|^2 - hx \cdot y + \frac{1}{4} h^2 |x|^2 + h(y - x) \cdot \nabla V(x) \\ &\qquad\qquad + \frac{1}{2} h^2 x \cdot \nabla V(x) + \frac{1}{4} h^2 |\nabla V(x)|^2\biggr) + C \\ &\quad = S(x, y) + \frac{1}{2}(y - x) \cdot \nabla V(x) + \frac{h}{8}|x + \nabla V(x)|^2 \end{aligned}$$

with a normalization constant C that does not depend on x and y, and a symmetric function $S:\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Therefore, by (1.19),

$$\begin{aligned}G_h^{\text{Euler}}(x,y) &= V(y) - V(x) + \log \gamma^d(x) p_h^{\text{Euler}}(x,y) - \log \gamma^d(y) p_h^{\text{Euler}}(y,x)\\ &= V(y) - V(x) - (y-x)\cdot\big(\nabla V(y) + \nabla V(x)\big)/2\\ &\quad + h\big(|y+\nabla V(y)|^2 - |x+\nabla V(x)|^2\big)/8.\end{aligned}$$

Similarly, for semi-implicit Euler proposals we obtain

$$\begin{aligned}-\log \gamma^d(x) p_h(x,y) &= \frac{1}{2}|x|^2 + \frac{1}{2}\left|y - \left(1-\frac{h}{2}\right)x + \frac{h}{2}\nabla V(x)\right|^2 \Big/ \left(h - \frac{h^2}{4}\right) + C\\ &= \frac{1}{2}\left(\left(h-\frac{h^2}{4}\right)|x|^2 + \left|y - \left(1-\frac{h}{2}\right)x + \frac{h}{2}\nabla V(x)\right|^2\right)\\ &\qquad \Big/\left(h - \frac{h^2}{4}\right) + C\\ &= -\frac{1}{2}\frac{h}{4-h}|x|^2 + \widetilde{S}(x,y) + \frac{1}{2}\cdot\frac{4}{4-h}(y-x)\nabla V(x)\\ &\qquad + \frac{1}{2}\cdot\frac{h}{4-h}|x+\nabla V(x)|^2\\ &= \widetilde{S}(x,y) + \frac{1}{2}(y-x)\cdot\nabla V(x)\\ &\qquad + \frac{1}{2}\frac{4}{4-h}\big[(y+x)\cdot\nabla V(x) + |\nabla V(x)|^2\big],\end{aligned}$$

and, therefore,

$$\begin{aligned}&G_h(x,y)\\ &\quad = V(y) - V(x) - (y-x)\cdot\big(\nabla V(y) + \nabla V(x)\big)/2\\ &\qquad + \frac{h}{8-2h}\big[(y+x)\cdot\big(\nabla V(y) - \nabla V(x)\big) + |\nabla V(y)|^2 - |\nabla V(x)|^2\big]. \qquad \square\end{aligned}$$

From now on we assume that Assumption 1.5 holds. We will derive upper bounds for the functions $G_h(x,y)$ computed in Proposition 1.3. By (1.16), these directly imply corresponding upper bounds for the MALA rejection probabilities.

Let $\partial^n_{\xi_1,\dots,\xi_n} V(z)$ denote the nth-order directional derivative of the function V at z in directions $\xi_1,\dots,\xi_n$. By $\partial^n V$ we denote the nth-order differential of V, that is, the n-form $(\xi_1,\dots,\xi_n) \mapsto \partial^n_{\xi_1,\dots,\xi_n} V$. For $x, \widetilde{x} \in \mathbb{R}^d$ and $n = 1,2,3,4$ let

$$L_n(x,\widetilde{x}) = \sup\big\{\big(\partial^n_{\xi_1,\dots,\xi_n} V\big)(z) : z \in [x,\widetilde{x}], \xi_1,\dots,\xi_n \in B_1^-\big\}. \tag{4.1}$$

In other words,

$$L_n(x,\widetilde{x}) = \sup_{z\in[x,\widetilde{x}]} \big\|\big(\partial^n V\big)(z)\big\|_-^*,$$

where $\|\cdot\|_-^*$ is the dual norm on n-forms defined by

$$\|l\|_-^* = \sup\{l(\xi_1, \dots, \xi_n) : \xi_1, \dots, \xi_n \in B_1^-\}.$$

In particular,

$$L_1(x, \widetilde{x}) = \sup_{z \in [x, \widetilde{x}]} \|\nabla V(z)\|_+.$$

By Assumption 1.5,

$$L_n(x, \widetilde{x}) \le C_n \cdot \max(1, \|x\|_-, \|\widetilde{x}\|_-)^{p_n} \qquad \forall x, \widetilde{x} \in \mathbb{R}^d, n = 1, 2, 3, 4. \tag{4.2}$$

We now derive upper bounds for the terms in the expression for $G_h(x, y)$ given in Proposition 1.3. We first express the leading order term in terms of 3rd derivatives of V:

LEMMA 4.1. *For $x, y \in \mathbb{R}^d$,*

$$\begin{aligned} V(y) - V(x) &- \frac{y-x}{2} \cdot (\nabla V(y) + \nabla V(x)) \\ &= -\frac{1}{2} \int_0^1 t(1-t) \partial_{y-x}^3 V((1-t)x + ty)\, dt. \end{aligned}$$

PROOF. A second-order expansion for $f(t) = V(x + t(y - x)), t \in [0, 1]$, yields

$$\begin{aligned} V(y) - V(x) &= \int_0^1 \partial_{y-x} V(x + t(y-x))\, dt \\ &= (y-x) \cdot \nabla V(x) + \int_0^1 \int_0^t \partial_{y-x}^2 V(x + s(y-x))\, ds\, dt \\ &= (y-x) \cdot \nabla V(x) + \int_0^1 (1-s) \partial_{y-x}^2 V(x + s(y-x))\, ds, \end{aligned}$$

and, similarly,

$$\begin{aligned} V(y) - V(x) &= (y-x) \cdot \nabla V(y) - \int_0^1 \int_t^1 \partial_{y-x}^2 V(x + s(y-x))\, ds\, dt \\ &= (y-x) \cdot \nabla V(y) - \int_0^1 s \partial_{y-x}^2 V(x + s(y-x))\, ds. \end{aligned}$$

By averaging both equations, we obtain

$$\begin{aligned} V(y) - V(x) &- \frac{y-x}{2} \cdot (\nabla V(y) + \nabla V(x)) \\ &= \frac{1}{2} \int_0^1 (1-2s) \partial_{y-x}^2 V(x + s(y-x))\, ds \\ &= \frac{1}{2} \int_0^1 t(1-t) \partial_{y-x}^3 V(x + t(y-x))\, dt. \end{aligned}$$

Here we have used that for any function $g \in C^1([0,1])$,

$$\int_0^1 (1-2s)g(s)\,ds = \int_0^1 (1-2s)\int_0^s g'(t)\,dt\,ds$$
$$= \int_0^1 \int_t^1 (1-2s)\,ds g'(t)\,dt = -\int_0^1 t(1-t)g'(t)\,dt. \qquad \square$$

LEMMA 4.2. *For $x, y \in \mathbb{R}^d$, the following estimates hold*:

(1) $|V(y) - V(x)| \le L_1(x,y) \cdot \|y-x\|_-$;
(2) $|V(y) - V(x) - \frac{y-x}{2} \cdot (\nabla V(y) + \nabla V(x))| \le \frac{1}{12} L_3(x,y) \cdot \|y-x\|_-^3$;
(3) $|(\nabla U(y) + \nabla U(x)) \cdot (\nabla V(y) - \nabla V(x))| \le L_2(x,y) \cdot \|\nabla U(y) + \nabla U(x)\|_- \cdot \|y-x\|_-$;
(4) $\|\nabla U(y) + \nabla U(x)\|_- \le 2\|\nabla U(x)\|_- + (1 + L_2(x,y)) \cdot \|y-x\|_-$.

REMARK 4.3. The estimates in Lemma 4.2 provide a bound for the terms in the expression (1.32) for $G_h(x,y)$ in the case of semi-implicit Euler proposals. For explicit Euler proposals, one also has to bound the term

$$\left|\nabla U(y)\right|^2 - \left|\nabla U(x)\right|^2 = \left|y + \nabla V(y)\right|^2 - \left|x + \nabla V(x)\right|^2.$$

Note that even when V vanishes, this term cannot be controlled in terms of $\|\cdot\|_-$ in general. A valid upper bound is

$$\left|\nabla U(y) + \nabla U(x)\right| \cdot |y-x| + L_2(x,y)\left\|\nabla U(y) + \nabla U(x)\right\|_- \cdot \|y-x\|_-.$$

PROOF OF LEMMA 4.2. By Lemma 4.1 and by definition of $L_n(x,y)$, we obtain

$$\left|V(y) - V(x)\right| \le \sup_{z \in [x,y]} \left|\partial_{y-x} V(z)\right| \le L_1(x,y) \cdot \|y-x\|_-,$$

$$\left|V(y) - V(x) - \frac{y-x}{2} \cdot \left(\nabla V(y) + \nabla V(x)\right)\right|$$
$$\le \frac{1}{2}\int_0^1 t(1-t)\,dt \sup_{z \in [x,y]} \left|\partial_{y-x}^3 V(z)\right| \le \frac{1}{12} L_3(x,y) \cdot \|y-x\|_-^3,$$

$$\left(\nabla U(y) + \nabla U(x)\right) \cdot \left(\nabla V(y) - \nabla V(x)\right)$$
$$= \partial_{\nabla U(y) + \nabla U(x)} V(y) - \partial_{\nabla U(y) + \nabla U(x)} V(x)$$
$$= \int_0^1 \partial^2_{\nabla U(y) + \nabla U(x), y-x} V\left((1-t)x + ty\right)dt$$
$$\le L_2(x,y) \cdot \left\|\nabla U(y) + \nabla U(x)\right\|_- \cdot \|y-x\|_-.$$

Moreover, the estimate

$$\begin{aligned}\|\nabla U(y)+\nabla U(x)\|_{-} &\le 2\|\nabla U(x)\|_{-}+\|\nabla U(y)-\nabla U(x)\|_{-}\\ &\le 2\|\nabla U(x)\|_{-}+\|y-x\|_{-}+\|\nabla V(y)-\nabla V(x)\|_{-}\\ &\le 2\|\nabla U(x)\|_{-}+\big(1+L_2(x,y)\big)\cdot\|y-x\|_{-}\end{aligned}$$

holds by definition of $L_2(x, y)$ and since

$$\begin{aligned}\|\nabla V(y)-\nabla V(x)\|_{-} \le |\nabla V(y)-\nabla V(x)| &= \sup_{|\xi|=1}\big(\partial_\xi V(y)-\partial_\xi V(x)\big)\\ &\le \sup_{\|\xi\|_{-}\le 1}\big(\partial_\xi V(y)-\partial_\xi V(x)\big).\end{aligned}$$

□

Recalling the definitions of $Y_h^{\mathrm{OU}}(x)$ and $Y_h(x)$ from (1.23) and (1.27), we obtain:

LEMMA 4.4. *For $x \in \mathbb{R}^d, h \in (0, 2)$ and $n \in \{1, 2, 3, 4\}$ with $p_n \ge 1$, we have*:

(1) $\|Y_h^{\mathrm{OU}}(x)-x\|_{-} \le \frac{h}{2}\|x\|_{-}+\sqrt{h}\|Z\|_{-}$;
(2) $\|Y_h(x)-x\|_{-} \le \frac{h}{2}\|\nabla U(x)\|_{-}+\sqrt{h}\|Z\|_{-}$;
(3) $\|Y_h^{\mathrm{OU}}(x)\|_{-} \le (1-\frac{h}{2})\|x\|_{-}+\sqrt{h}\|Z\|_{-}$;
(4) $\|Y_h(x)\|_{-} \le \|x\|_{-}+\frac{h}{2}\|\nabla U(x)\|_{-}+\sqrt{h}\|Z\|_{-}$;
(5) $L_n(x, Y_h^{\mathrm{OU}}(x)) \le C_n 2^{p_n-1}(\max(1,\|x\|_{-})^{p_n}+h^{p_n/2}\|Z\|_{-}^{p_n})$;
(6) $L_n(x, Y_h(x)) \le C_n 3^{p_n-1}(\max(1,\|x\|_{-})^{p_n} + (\frac{h}{2})^{p_n}\|\nabla U(x)\|_{-}^{p_n} + h^{p_n/2}\|Z\|_{-}^{p_n})$.

PROOF. Estimates (1)–(4) are direct consequences of the triangle inequality. Moreover, by (3) and (4),

$$\max\big(1,\|x\|_{-},\|Y_h^{\mathrm{OU}}(x)\|_{-}\big) \le \max\big(1,\|x\|_{-}\big)+\sqrt{h}\|Z\|_{-}$$

and

$$\max\big(1,\|x\|_{-},\|Y_h(x)\|_{-}\big) \le \max\big(1,\|x\|_{-}\big)+\frac{h}{2}\|\nabla U(x)\|_{-}+\sqrt{h}\|Z\|_{-}.$$

Estimates (5) and (6) now follow from (4.2) and Hölder's inequality. □

We now combine the estimates in Lemmas 4.2 and 4.4 in order to prove Proposition 1.7 and the first part of Proposition 1.11:

PROOF OF PROPOSITION 1.7. By (1.16) and Proposition 1.3, for $h \in (0, 2)$,

$$\mathbb{E}\big[\big(1-\alpha_h(x, Y_h(x))\big)^k\big]^{1/k} \le \|G_h(x, Y_h(x))^+\|_{L^k} \le I+\frac{h}{4}II, \tag{4.3}$$

where

$$I = \left\| V(Y_h(x)) - V(x) - \frac{Y_h(x) - x}{2} \cdot (\nabla V(Y_h(x)) - \nabla V(x)) \right\|_{L^k},$$
$$II = \|(\nabla U(Y_h(x)) + \nabla U(x)) \cdot (\nabla V(Y_h(x)) - \nabla V(x))\|_{L^k}.$$

By Lemma 4.2,

$$I \le \mathbb{E}[L_3(x, Y_h(x))^k \cdot \|Y_h(x) - x\|_-^{3k}]^{1/k}/12 \quad \text{and} \tag{4.4}$$

$$\begin{aligned} II &\le \mathbb{E}[L_2(x, Y_h(x))^k \cdot \|Y_h(x) - x\|_-^k \\ &\qquad \times (2\|\nabla U(x)\|_- + (1 + L_2(x, Y_h(x))) \cdot \|Y_h(x) - x\|_-)^k]^{1/k}. \end{aligned} \tag{4.5}$$

The assertion of Proposition 1.7 for semi-implicit Euler proposals is now a direct consequence of the estimates (2) and (6) in Lemma 4.4. The assertion for Ornstein–Uhlenbeck proposals follows similarly from (1.21), the estimates (1) in Lemma 4.2 and (1), (3) and (5) in Lemma 4.4. □

It is possible to write down the polynomial in Proposition 1.7 explicitly. For semi-implicit Euler proposals, we illustrate this in the case $k = 1$ and $p_2 = p_3 = 0$. Here, by (4.4) and (4.5) we obtain

$$I \le \frac{C_3}{12}\mathbb{E}[(h\|\nabla U(x)\|_-/2 + \sqrt{h}\|Z\|_-)^3] \le \frac{C_3}{4}(h^3\|\nabla U(x)\|_-^3/8 + h^{3/2}m_3),$$

$$\begin{aligned} II &\le C_2\mathbb{E}[(h\|\nabla U(x)\|_-/2 + \sqrt{h}\|Z\|_-) \\ &\qquad \times (2\|\nabla U(x)\|_- + (1 + C_2)(h\|\nabla U(x)\|_-/2 + \sqrt{h}\|Z\|_-))] \\ &\le C_2\Big(h\|\nabla U(x)\|_-^2 + 2\sqrt{h}\|\nabla U(x)\|_- m_1 \\ &\qquad + (1 + C_2)\Big(\frac{h^2}{2}\|\nabla U(x)\|_-^2 + 2hm_2\Big)\Big). \end{aligned}$$

Hence by (4.3),

$$\begin{aligned} \mathbb{E}[1 - \alpha_h(x, Y_h(x))] &\le h^{3/2} \cdot \Big(\frac{1}{4}C_3 m_3 + \frac{1}{2}C_2 m_1\|\nabla U(x)\|_-\Big) \\ &\quad + h^2 \cdot \Big(\frac{1}{4}C_2\|\nabla U(x)\|_-^2 + \frac{1}{2}C_2(1 + C_2)m_2\Big) \\ &\quad + h^3 \cdot \Big(\frac{1}{32}C_3\|\nabla U(x)\|_-^3 + \frac{1}{8}C_2(1 + C_2)\|\nabla U(x)\|_-^2\Big). \end{aligned} \tag{4.6}$$

For Ornstein–Uhlenbeck proposals, we derive the explicit bound for the rejection probabilities stated in Proposition 1.11 for the case $k = 1$ and $p_2 = 0$.

PROOF OF PROPOSITION 1.11, FIRST PART. If $p_2 = 0$, then for any $x \in \mathbb{R}^d$,

$$\|\nabla V(x)\|_+ \le \|\nabla V(0)\|_+ + \|\nabla V(x) - \nabla V(0)\|_+ \le C_1 + C_2 \cdot \|x\|_-. \tag{4.7}$$

Therefore, for any $x, y \in \mathbb{R}^d$,

$$|V(y) - V(x)| \le (C_1 + C_2 \cdot \max(\|x\|_-, \|y\|_-)) \cdot \|y - x\|_-.$$

Hence, by (1.21) and by (1) and (3) in Lemma 4.4,

$$\begin{aligned}
&\mathbb{E}[1 - \alpha^{\mathrm{OU}}(x, Y_h^{\mathrm{OU}}(x))] \\
&\quad \le \mathbb{E}[(V(Y_h^{\mathrm{OU}}(x)) - V(x))^+] \\
&\quad \le \mathbb{E}[(C_1 + C_2 \cdot \max(\|x\|_-, \|Y_h^{\mathrm{OU}}(x)\|_-)) \cdot \|Y_h^{\mathrm{OU}}(x) - x\|_-] \\
&\quad \le \mathbb{E}[(C_1 + C_2 \cdot (\|x\|_- + \sqrt{h}\|Z\|_-)) \cdot (h\|x\|_-/2 + \sqrt{h}\|Z\|_-)] \\
&\quad = m_1(C_1 + C_2\|x\|_-) \cdot h^{1/2} \\
&\qquad + \frac{1}{2}(2m_2C_2 + C_1\|x\|_- + C_2\|x\|_-^2) \cdot h + \frac{1}{2} m_1 C_2 \|x\|_- \cdot h^{3/2}. \qquad \square
\end{aligned}$$

5. Dependence of rejection on the current state. We now derive estimates for the derivatives of the functions

$$F_h(x, w) = G_h\left(x, x - \frac{h}{2}\nabla U(x) + w\right), \tag{5.1}$$

$$F_h^{\mathrm{OU}}(x, w) = G^{\mathrm{OU}}\left(x, x - \frac{h}{2}x + w\right), \qquad (x, w) \in \mathbb{R}^d \times \mathbb{R}^d, \tag{5.2}$$

w.r.t. x. Since

$$G_h(x, Y_h(x)) = F_h\left(x, \sqrt{h - \frac{h^2}{4}}Z\right) \qquad \text{with } Z \sim \gamma^d, \text{ and} \tag{5.3}$$

$$G^{\mathrm{OU}}(x, Y_h^{\mathrm{OU}}(x)) = F_h^{\mathrm{OU}}\left(x, \sqrt{h - \frac{h^2}{4}}Z\right) \qquad \text{with } Z \sim \gamma^d, \tag{5.4}$$

these estimates can then be applied to control the dependence of rejection on the current state x.

For *Ornstein–Uhlenbeck proposals*, by (1.21), we immediately obtain

$$\nabla_x F_h^{\mathrm{OU}}(x, w) = \left(1 - \frac{h}{2}\right)(\nabla V(y) - \nabla V(x)) - \frac{h}{2}\nabla V(x), \tag{5.5}$$

where $y := (1 - \frac{h}{2})x + w$.

For *semi-implicit Euler proposals*, the formula for the derivative is more involved. To simplify the notation we set for $x \in \mathbb{R}^d$ and fixed $h \in (0, 2)$,

$$x' := x - \frac{h}{2}\nabla U(x). \tag{5.6}$$

In the sequel, we use the conventions

$$v \cdot w = \sum_{i=1}^{d} v_i w_i, \qquad (v \cdot T)_j = \sum_{i=1}^{d} v_i T_{i,j}, \qquad (T \cdot v)_j = \sum_{j=1}^{d} T_{i,j} v_j,$$

$$(S \cdot T)_{ik} = \sum_{j=1}^{d} S_{i,j} T_{j,k}$$

for vectors $v, w \in \mathbb{R}^d$ and $(2, 0)$ tensors $S, T \in \mathbb{R}^d \otimes \mathbb{R}^d$. In particular,

$$v \cdot (S \cdot T) = (v \cdot S) \cdot T,$$

that is, the brackets may be omitted. We now give an explicit expression for the derivative of $F_h(x, w)$ w.r.t. the first variable:

PROPOSITION 5.1. *Suppose $V \in C^2(\mathbb{R}^d)$. Then for any $x, w \in \mathbb{R}^d$,*

$$\begin{aligned}\nabla_x F_h(x, w) = {} & \nabla V(y) - \nabla V(x) - \frac{y - x}{2} \cdot \big(\nabla^2 V(y) + \nabla^2 V(x)\big) \\ & + \frac{h}{4}(y - x) \cdot \nabla^2 V(y) \cdot \big(I_d + \nabla^2 V(x)\big) \\ & + \frac{h}{8 - 2h}\big(\nabla V(y) - \nabla V(x) + \nabla U(y) + \nabla U(x)\big) \\ & \times \big(\nabla^2 V(y) - \nabla^2 V(x)\big) \\ & - \frac{h^2}{16 - 4h}\big(\nabla V(y) - \nabla V(x) + \nabla U(y) + \nabla U(x)\big) \cdot \nabla^2 V(y) \\ & \times \big(I_d + \nabla^2 V(x)\big)\end{aligned}$$

with $y := x' + w = x - \frac{h}{2}\nabla U(x) + w$.

PROOF. Let

$$W(x) := \nabla V(x) = \nabla U(x) - x, \qquad x \in \mathbb{R}^d.$$

By Proposition 1.3,

$$F_h(x, w) = A_h(x, w) + \frac{h}{8 - 2h} B_h(x, w) \tag{5.7}$$

for any $x, w \in \mathbb{R}^d$, where

$$A_h(x, w) := V\big(x' + w\big) - V(x) - \frac{x' + w - x}{2} \cdot \big(\nabla V\big(x' + w\big) + \nabla V(x)\big) \quad \text{and}$$

$$B_h(x, w) := \big(\nabla U\big(x' + w\big) + \nabla U(x)\big) \cdot \big(\nabla V\big(x' + w\big) - \nabla V(x)\big).$$

Noting that by (5.6),

$$x - x' = \frac{h}{2}\nabla U(x) = \frac{h}{2}x + \frac{h}{2}\nabla V(x), \tag{5.8}$$

$$\nabla_x\big(x - x'\big) = \frac{h}{2}\nabla^2 U(x) = \frac{h}{2}I_d + \frac{h}{2}\nabla^2 V(x) \quad \text{and} \tag{5.9}$$

$$\nabla_x x' = I_d - \frac{h}{2}\nabla^2 U(x) = \left(1 - \frac{h}{2}\right)I_d - \frac{h}{2}\nabla^2 V(x), \tag{5.10}$$

we obtain with $y = x' + w$

$$\begin{aligned}
\nabla_x A_h(x, w) &= W\big(x' + w\big) \cdot \left(I_d - \frac{h}{2}\nabla^2 U(x)\right) - W(x) \\
&\quad - \frac{x' + w - x}{2} \cdot \left(\nabla W\big(x' + w\big) \cdot \left(I_d - \frac{h}{2}\nabla^2 U(x)\right) + \nabla W(x)\right) \\
&\quad + \frac{h}{4}\big(W\big(x' + w\big) + W(x)\big) \cdot \nabla^2 U(x) \\
&= W(y) - W(x) - \frac{y - x}{2} \cdot \big(\nabla W(y) + \nabla W(x)\big) \\
&\quad - \frac{h}{4}\big(W(y) - W(x)\big) \cdot \nabla^2 U(x) + \frac{h}{4}(y - x) \cdot \big(\nabla W(y) \cdot \nabla^2 U(x)\big), \\
\nabla_x B_h(x, w) &= \big(W\big(x' + w\big) - W(x)\big) \\
&\quad \times \left(\nabla^2 U\big(x' + w\big) \cdot \left(I_d - \frac{h}{2}\nabla^2 U(x)\right) + \nabla^2 U(x)\right) \\
&\quad + \big(\nabla U\big(x' + w\big) + \nabla U(x)\big) \\
&\quad \times \left(\nabla W\big(x' + w\big) \cdot \left(I_d - \frac{h}{2}\nabla^2 U(x)\right) - \nabla W(x)\right) \\
&= \big(W(y) - W(x)\big) \cdot \big(\nabla^2 U(y) \\
&\quad + \nabla^2 U(x)\big) + \big(\nabla U(y) + \nabla U(x)\big) \cdot \big(\nabla W(y) - \nabla W(x)\big) \\
&\quad - \frac{h}{2}\big(W(y) - W(x)\big) \cdot \big(\nabla^2 U(y) \cdot \nabla^2 U(x)\big) \\
&\quad - \frac{h}{2}\big(\nabla U(y) + \nabla U(x)\big) \cdot \big(\nabla W(y) \cdot \nabla^2 U(x)\big).
\end{aligned}$$

In total, we obtain

$$\begin{aligned}
\nabla_x F_h(x, w) &= \nabla_x A_h(x, w) + \frac{h}{8 - 2h}\nabla_x B_h(x, w) \\
&= W(y) - W(x) - \frac{y - x}{2} \cdot \big(\nabla W(y) + \nabla W(x)\big)
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{h}{8-2h}\big(W(y)-W(x)\big)\cdot\big(\nabla^2U(y)-\nabla^2U(x)\big)\\
&+ \frac{h}{4}(y-x)\cdot\nabla W(y)\cdot\nabla^2U(x)\\
&+ \frac{h}{8-2h}\big(\nabla U(y)+\nabla U(x)\big)\cdot\big(\nabla W(y)-\nabla W(x)\big)\\
&+ \left(\frac{2h}{8-2h}-\frac{h}{4}\right)\big(W(y)-W(x)\big)\cdot\nabla^2U(x)\\
&- \frac{h^2}{16-4h}\big(W(y)-W(x)\big)\cdot\nabla^2U(y)\cdot\nabla^2U(x)\\
&- \frac{h^2}{16-4h}\big(\nabla U(y)+\nabla U(x)\big)\cdot\nabla W(y)\cdot\nabla^2U(x)\\
={}& W(y)-W(x)-\frac{y-x}{2}\cdot\big(\nabla W(y)+\nabla W(x)\big)\\
&+ \frac{h}{4}(y-x)\cdot\nabla W(y)\cdot\nabla^2U(x)\\
&+ \frac{h}{8-2h}\big(W(y)-W(x)+\nabla U(y)+\nabla U(x)\big)\\
&\times\big(\nabla W(y)-\nabla W(x)\big)\\
&- \frac{h^2}{16-4h}\big(\big(W(y)-W(x)\big)\cdot\big(\nabla^2U(y)-I_d\big)\\
&+ \big(\nabla U(y)+\nabla U(x)\big)\cdot\nabla W(y)\big)\cdot\nabla^2U(x).
\end{aligned}
$$

Here we have used that

$$\nabla^2U = I_d+\nabla^2V = I_d+\nabla W.$$

The assertion follows by applying this identity to the remaining ∇^2U terms as well. □

Similar to Lemma 4.2 above, we now derive bounds for the individual summands in the expressions for $\nabla_x F_h^{\mathrm{OU}}$ and $\nabla_x F_h$ in (5.5) and Proposition 5.1.

LEMMA 5.2. *For $V\in C^4(\mathbb{R}^d)$ and $x, y\in\mathbb{R}^d$ the following estimates hold*:

(1) $\|\nabla V(y)-\nabla V(x)\|_+ \le L_2(x,y)\cdot\|y-x\|_-$;
(2) $\|\nabla V(y)-\nabla V(x)-\frac{y-x}{2}\cdot(\nabla^2V(y)+\nabla^2V(x))\|_+ \le L_4(x,y)\cdot\|y-x\|_-^3/12$;
(3) $\|(y-x)\cdot\nabla^2V(y)\cdot(I_d+\nabla^2V(x))\|_+ \le L_2(y,y)\cdot(1+L_2(x,x))\cdot\|y-x\|_-$;

(4) $\|(\nabla V(y) - \nabla V(x) + \nabla U(y) + \nabla U(x)) \cdot (\nabla^2 V(y) - \nabla^2 V(x))\|_+ \le L_3(x,y) \cdot (L_2(x,y)\|y-x\|_- + \|\nabla U(y) + \nabla U(x)\|_-) \cdot \|y-x\|_-$;

(5) $\|(\nabla V(y) - \nabla V(x) + \nabla U(y) + \nabla U(x)) \cdot \nabla^2 V(y) \cdot (I_d + \nabla^2 V(x))\|_- \le L_2(y,y) \cdot (L_2(x,y)\|y-x\|_- + \|\nabla U(y) + \nabla U(x)\|_-) \cdot (1 + L_2(x,x))$.

PROOF. (1) For any $\xi \in \mathbb{R}^d$, we have

$$\left|\partial_\xi V(y) - \partial_\xi V(x)\right| \le \sup_{z\in[x,y]} \left|\partial^2_{y-x,\xi} V(z)\right| \le L_2(x,y)\|x-y\|_-\|\xi\|_-.$$

This proves (1) by definition of $\|\cdot\|_+$.

(2) By Lemma 4.1 applied to $\partial_\xi V$,

$$\left|\partial_\xi V(y) - \partial_\xi V(x) - \frac{y-x}{2} \cdot \big(\nabla\partial_\xi V(y) - \nabla\partial_\xi V(x)\big)\right|$$
$$\le \frac{1}{2}\int_0^1 t(1-t)\,dt \cdot \sup_{z\in[x,y]} \left|\partial^3_{y-x}\partial_\xi V(z)\right| \le \frac{1}{12} L_4(x,y)\|x-y\|_-^3\|\xi\|_-.$$

(3) For $v, w \in \mathbb{R}^d$, we have

$$\left|v \cdot \nabla^2 V(y) w\right| = \left|\partial^2_{vw} V(y)\right| \le L_2(x,y)\|v\|_-\|w\|_-. \tag{5.11}$$

Since $\|\cdot\|_-$ is weaker than the Euclidean norm, we obtain

$$\begin{aligned}
&\left|(y-x) \cdot \nabla^2 V(y) \cdot \big(I + \nabla^2 V(x)\big) \cdot \xi\right| \\
&\quad \le L_2(y,y)\|y-x\|_- \left\|\big(I + \nabla^2 V(x)\big) \cdot \xi\right\|_- \\
&\quad \le L_2(y,y)\|y-x\|_-\big(1 + L_2(x,x)\big)\|\xi\|_-.
\end{aligned}$$

(4), (5) For $v, w \in \mathbb{R}^d$,

$$\begin{aligned}
\left|v \cdot \big(\nabla^2 V(y) - \nabla^2 V(x)\big) \cdot w\right| &= \left|\int_0^1 \partial^3_{y-x,v,w} V\big((1-t)x + ty\big)\,dt\right| \\
&\le L_3(x,y)\|y-x\|_-\|v\|_-\|w\|_-.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\big(\nabla V(y) - \nabla V(x) + \nabla U(y) + \nabla U(x)\big) \cdot \big(\nabla^2 V(y) - \nabla^2 V(x)\big) \cdot \xi\right| \\
&\quad \le L_3(x,y)\|y-x\|_- \cdot \big(\left\|\nabla V(y) - \nabla V(x)\right\|_- + \left\|\nabla U(y) + \nabla U(x)\right\|_-\big) \cdot \|\xi\|_- \\
&\quad \le L_3(x,y)\|y-x\|_- \cdot \big(L_2(x,y)\|y-x\|_- + \left\|\nabla U(y) + \nabla U(x)\right\|_-\big) \cdot \|\xi\|_-,
\end{aligned}$$

and, correspondingly,

$$\begin{aligned}
&\left|\big(\nabla V(y) - \nabla V(x) + \nabla U(y) + \nabla U(x)\big) \cdot \nabla^2 V(y) \cdot \big(I + \nabla^2 V(x)\big) \cdot \xi\right| \\
&\quad \le L_2(y,y) \cdot \big(L_2(x,y)\|y-x\|_- + \left\|\nabla U(y) + \nabla U(x)\right\|_-\big) \\
&\qquad \times \big(1 + L_2(x,x)\big)\|\xi\|_-.
\end{aligned}$$

□

By combining Proposition 5.1 with the estimates in Lemma 5.2 and Lemma 4.4, we will now prove Proposition 1.9.

PROOF OF PROPOSITION 1.9. Fix $h \in (0,2)$. By (1.17) and (1.18), for any $x, \tilde{x} \in \mathbb{R}^d$,

$$\begin{aligned}
&\big\|\alpha_h\big(x, Y_h(x)\big) - \alpha_h\big(\tilde{x}, Y_h(\tilde{x})\big)\big\|_{L^k} \\
&\quad \le \big\|G_h\big(x, Y_h(x)\big) - G_h\big(\tilde{x}, Y_h(\tilde{x})\big)\big\|_{L^k} \\
&\quad \le \|x-\tilde{x}\|_- \cdot \sup_{z\in[x,\tilde{x}]} \big\|\big\|\nabla_x G_h\big(x, Y_h(x)\big)\big\|_+\big\|_{L^k}.
\end{aligned} \tag{5.12}$$

Moreover, by (5.3) and Proposition 5.1,

$$\begin{aligned}
&\big\|\big\|\nabla_x G_h\big(x, Y_h(x)\big)\big\|_+\big\|_{L^k} \\
&\quad = \big\|\big\|\nabla_x F_h\big(x, \sqrt{h-h^2/4}Z\big)\big\|_+\big\|_{L^k} \\
&\quad \le I + \frac{h}{4}\cdot II + \frac{h}{8-2h}\cdot III + \frac{h^2}{16-4h}\cdot IV,
\end{aligned} \tag{5.13}$$

where

$$\begin{aligned}
I &= \mathbb{E}\big[\big\|\nabla V\big(Y_h(x)\big) - \nabla V(x) - \tfrac{1}{2}\big(Y_h(x)-x\big) \\
&\qquad \times \big(\nabla^2 V\big(Y_h(x)\big) + \nabla^2 V(x)\big)\big\|_+^k\big]^{1/k},\\
II &= \mathbb{E}\big[\big\|\big(Y_h(x)-x\big)\cdot\nabla^2 V\big(Y_h(x)\big)\cdot\big(I+\nabla^2 V(x)\big)\big\|_+^k\big]^{1/k},\\
III &= \mathbb{E}\big[\big\|\big(\nabla V\big(Y_h(x)\big) - \nabla V(x) + \nabla U\big(Y_h(x)\big) + \nabla U(x)\big) \\
&\qquad \times \big(\nabla^2 V\big(Y_h(x)\big) - \nabla^2 V(x)\big)\big\|_+^k\big]^{1/k},\\
IV &= \mathbb{E}\big[\big\|\big(\nabla V\big(Y_h(x)\big) - \nabla V(x) + \nabla U\big(Y_h(x)\big) + \nabla U(x)\big) \\
&\qquad \times \nabla^2 V\big(Y_h(x)\big)\cdot\big(I+\nabla^2 V(x)\big)\big\|_+^k\big]^{1/k}.
\end{aligned}$$

By applying the estimates from Lemmas 5.2 and 4.2(4), we obtain

$$I \le \tfrac{1}{12}\mathbb{E}\big[L_4\big(x, Y_h(x)\big)^k \big\|Y_h(x)-x\big\|_-^{3k}\big]^{1/k}, \tag{5.14}$$

$$II \le \big(1+L_2(x,x)\big)\cdot\mathbb{E}\big[L_2\big(Y_h(x), Y_h(x)\big)^k \big\|Y_h(x)-x\big\|_-^k\big]^{1/k}, \tag{5.15}$$

$$\begin{aligned}
III &\le \mathbb{E}\big[L_3\big(x, Y_h(x)\big)^k \big\|Y_h(x)-x\big\|_-^k \\
&\qquad \times \big(\big(1+2L_2\big(x, Y_h(x)\big)\big)^k \big\|Y_h(x)-x\big\|_-^k + 2\big\|\nabla U(x)\big\|_-^k\big)\big]^{1/k},
\end{aligned} \tag{5.16}$$

$$\begin{aligned}
IV &\le \big(1+L_2(x,x)\big) \\
&\qquad \times \mathbb{E}\big[L_2\big(Y_h(x), Y_h(x)\big)^k \\
&\qquad\quad \times \big(\big(1+L_2\big(x, Y_h(x)\big)\big)^k \big\|Y_h(x)-x\big\|_-^k + 2\big\|\nabla U(x)\big\|_-^k\big)\big]^{1/k}.
\end{aligned} \tag{5.17}$$

The assertion for semi-implicit Euler proposals is now a direct consequence of the estimates in Lemma 4.4, (4.2) and (5.12).

The assertion for Ornstein–Uhlenbeck proposals follows in a similar way from (5.5), Lemma 5.2(1) and the estimates in Lemma 4.4. □

Again, it is possible to write down the polynomial in Proposition 1.9 explicitly. For semi-implicit Euler proposals, we illustrate this in the case $k = 1$ and $p_2 = p_3 = p_4 = 0$. Here, by (5.14), (5.15), (5.17) and (5.17) we obtain

$$\begin{aligned}
I &\le \frac{C_4}{12}\mathbb{E}\left[\left(\frac{h}{2}\|\nabla U(x)\|_- + \sqrt{h}\|Z\|_-\right)^3\right] \le \frac{C_4}{4}\left(\frac{h^3}{8}\|\nabla U(x)\|_-^3 + h^{3/2}m_3\right),\\
II &\le (C_2 + C_2^2)\mathbb{E}\left[\frac{h}{2}\|\nabla U(x)\|_- + \sqrt{h}\|Z\|_-\right]\\
&= (C_2 + C_2^2)\left(\frac{h}{2}\|\nabla U(x)\|_- + h^{1/2}m_1\right),\\
III &\le C_3\Bigg(2\|\nabla U(x)\|_- \cdot \mathbb{E}\left[\frac{h}{2}\|\nabla U(x)\|_- + \sqrt{h}\|Z\|_-\right]\\
&\qquad + (1 + 2C_2)\mathbb{E}\left[\left(\frac{h}{2}\|\nabla U(x)\|_- + \sqrt{h}\|Z\|_-\right)^2\right]\Bigg)\\
&\le 2C_3\|\nabla U(x)\|_-\left(\frac{h}{2}\|\nabla U(x)\|_- + \sqrt{h}m_1\right)\\
&\quad + C_3(1 + 2C_2)\left(\frac{h^2}{2}\|\nabla U(x)\|^2 + 2hm_2\right),\\
IV &\le (1 + C_2)C_2\left((1 + C_2)\mathbb{E}\left[\frac{h}{2}\|\nabla U(x)\|_- + \sqrt{h}\|Z\|_-\right] + 2\|\nabla U(x)\|_-\right)\\
&\le 2(1 + C_2)C_2\|\nabla U(x)\|_- + (1 + C_2)^2C_2\left(\frac{h}{2}\|\nabla U(x)\|_- + \sqrt{h}m_1\right).
\end{aligned}$$

Hence by (5.13), for $h \in (0, 2)$,

$$\begin{aligned}
&\mathbb{E}[\|\nabla_x G_h(x, Y_h(x))\|_+]\\
&\quad \le \tfrac{1}{4}h^{3/2}(C_4m_3 + (1 + C_2)C_2m_1 + 2C_3\|\nabla U(x)\|_-m_1)\\
&\qquad + \tfrac{1}{8}h^2(4C_2(1 + 2C_2)m_2 + 3C_2(1 + C_2)\|\nabla U(x)\|_- + 2C_3\|\nabla U(x)\|_-^2)\\
&\qquad + \tfrac{1}{16}h^{5/2}C_2(1 + C_2)^2(2m_1 + h^{1/2}\|\nabla U(x)\|_-)\\
&\qquad + \tfrac{1}{32}h^3(4C_3(1 + 2C_2)\|\nabla U(x)\|_-^2 + C_4\|\nabla U(x)\|_-^3).
\end{aligned} \tag{5.18}$$

For *Ornstein–Uhlenbeck proposals*, we prove the explicit bound for the dependence of the rejection probabilities on the current state for the case $k = 2$ and $p_2 = 0$ as stated in Proposition 1.11.

PROOF OF PROPOSITION 1.11, SECOND PART. If $p_2 = 0$, then by (5.4), (5.5) and (4.7),

$$\begin{aligned}\|\nabla_x G^{\mathrm{OU}}(x, Y_h^{\mathrm{OU}}(x))\|_+ &\le \|\nabla V(Y_h^{\mathrm{OU}}(x)) - \nabla V(x)\|_+ + \frac{h}{2}\|\nabla V(x)\|_+ \\ &\le C_2\|Y_h^{\mathrm{OU}}(x) - x\|_- + (C_1 + C_2\|x\|_-)h/2 \\ &\le C_2\|Z\|_- h^{1/2} + (C_1 + 2C_2\|x\|_-)h/2\end{aligned}$$

for any $x \in \mathbb{R}^d$. Therefore,

$$\mathbb{E}[\|\nabla_x G^{\mathrm{OU}}(x, Y_h^{\mathrm{OU}}(x))\|_+^2]^{1/2} \le C_2 m_2^{1/2} h^{1/2} + (C_1 + 2C_2\|x\|_-)h/2.$$

The assertion now follows similarly to (5.12). □

6. Upper bound for exit probabilities. In this section, we prove an upper bound for the exit probabilities of the MALA chain from the ball B_R^- that is required in the proof of Theorem 1.19; cf. [12] for a detailed proof of a more general result. Let

$$f(x) := \exp(K\|x\|_-^2/16). \tag{6.1}$$

The following lemma shows that $f(x)$ acts as a Lyapunov function for the MALA transition kernel on B_R^-:

LEMMA 6.1. *Suppose that Assumptions 1.5 and 1.14 hold. Then there exist constants $C_1, C_2, \rho_1 \in (0, \infty)$ such that*

$$q_h f \le f^{1-Kh/4} e^{C_2 h} \qquad \text{on } B_R^- \tag{6.2}$$

for any $R, h \in (0, \infty)$ such that $h^{-1} \ge C_1(1 + R)^{\rho_1}$.

PROOF. We first observe that a corresponding bound holds for the proposal kernel p_h. Indeed, by (1.27), and since $\|v\|_-^2 = v \cdot Gv$ with a nonnegative definite symmetric matrix $G \le I$, an explicit computation yields

$$\begin{aligned}(p_h f)(x) &= \mathbb{E}\left[\exp\left(K\left\|x - \frac{h}{2}\nabla U(x) + \sqrt{h - h^2/4}Z\right\|_-^2\Big/16\right)\right] \\ &\le \exp\left(K(1 + Kh/4)\left\|x - \frac{h}{2}\nabla U(x)\right\|_-^2\Big/16\right).\end{aligned}$$

Moreover, by Assumption 1.14,

$$\left\|x - \frac{h}{2}\nabla U(x)\right\|_-^2 \le \left(1 - \frac{Kh}{2}\right)\|x\|_-^2 + \frac{h}{2K}\|\nabla U(0)\|_-^2 + \frac{h^2}{4}\|\nabla U(x)\|_-^2.$$

Hence by Assumption 1.5, there exist constants $C_3, C_4, \rho_2 \in (0, \infty)$ such that

$$(p_h f)(x) \le f(x)^{1-Kh/4} e^{C_3 h}$$

for any $x \in \mathbb{R}^d$ and $h \in (0,\infty)$ such that $h^{-1} \geq C_4(1+\|x\|_-)^{\rho_2}$. By the upper bound for the rejection probabilities in Proposition 1.7, we conclude that there exists a polynomial s such that the corresponding upper bound

$$\begin{aligned} q_h f &\leq f^{1-Kh/4} e^{C_3 h} + s(R) h^{3/2} f \\ &= f^{1-Kh/4}\big(e^{C_3 h} + s(R) h^{3/2} f^{Kh/4}\big) \\ &\leq f^{1-Kh/4} e^{(C_3+1)h} \end{aligned}$$

holds on B_R^- whenever both $h^{-1} \geq C_4(1+R)^{\rho_2}$ and $s(R)h^{1/2} f^{Kh/4} \leq 1$. The assertion follows, since the second condition is satisfied if $K^2 h R^2/64 \leq 1$ and $s(R) e h^{1/2} \leq 1$. □

Now consider the first exit time

$$T_R := \inf\{n \geq 0 : X_n \notin B_R^-\},$$

where $(X_n, \mathbb{P}_x)$ is the Markov chain with transition kernel q_h and initial condition $X_0 = x$ P_x-a.s. We can estimate T_R by constructing a supermartingale based on the Lyapunov function f:

THEOREM 6.2. *If Assumptions 1.5 and 1.14 hold, then there exist constants* $C, \rho, D \in (0,\infty)$ *such that the upper bound*

$$\mathbb{P}_x[T_R \leq n] \leq Dnh \exp\big[K\big(\|x\|_-^2 - R^2\big)/24\big] \tag{6.3}$$

holds for any $n \geq 0$, $R, h \in (0,\infty)$ *such that* $h^{-1} \geq C(1+R)^\rho$, *and* $x \in B_R^-$.

PROOF. Fix $n \in \mathbb{N}$, choose C_1, C_2, ρ_1 as in the lemma above, and let

$$M_j := f(X_j)^{(1-Kh/4)^{n-j}} \exp\left(-C_2 h \sum_{i=0}^{j} (1-Kh/4)^{n-i}\right)$$

for $j = 0, 1, \ldots, n$. If $h^{-1} \geq C_1(1+R)^{\rho_1}$, then by Jensen's inequality and (6.2),

$$\begin{aligned} \mathbb{E}_x[M_{j+1}|\mathcal{F}_j] &\leq (q_h f)(X_j)^{1-Kh/4} \exp\left(-C_2 h \sum_{i=0}^{j+1} (1-Kh/4)^{n-i}\right) \\ &\leq M_j \qquad \text{on } \{X_j \in B_R^-\} \text{ for any } j < n. \end{aligned}$$

Hence the stopped process $(M_j^{T_R})_{0 \leq j \leq n}$ is a supermartingale, and thus

$$\mathbb{E}_x[M_{T_R}; n-m \leq T_R \leq n] \leq \mathbb{E}_x[M_0] \qquad \text{for any } 0 \leq m \leq n.$$

Noting that $M_0 = f(x)^{(1-Kh/4)^n} = \exp((1-Kh/4)^n K\|x\|_-^2/16)$, and

$$\begin{aligned} M_{T_R} &\geq \big(f(X_{T_R}) \exp(-4C_2/K)\big)^{(1-Kh/4)^{n-T_R}} \\ &= \exp\left[\left(\frac{K}{16}R^2 - 4C_2/K\right) \cdot (1-Kh/4)^{n-T_R}\right], \end{aligned}$$

we obtain the bound

$$\begin{aligned}&\mathbb{P}_x[n-m \le T_R \le n]\\&\quad\le \exp\biggl[\biggl(1-\frac{Kh}{4}\biggr)^m\biggl(\frac{K}{16}\biggl(\biggl(1-\frac{Kh}{4}\biggr)^{n-m}\|x\|_-^2 - R^2\biggr)+\frac{4C_2}{K}\biggr)\biggr]\end{aligned}$$

for any $0 \le m \le n$ provided $R^2 \ge 64C_2/K^2$. In particular, if $mKh/2 \le \log(3/2)$, then $(1-Kh/4)^m \ge \exp(-mKh/2) \ge 2/3$, and hence

$$\mathbb{P}_x[n-m \le T_R \le n] \le \exp(4C_2K)\cdot\exp\bigl(K\bigl(\|x\|_-^2 - R^2\bigr)/24\bigr).$$

The assertion follows by partitioning $\{0,1,\dots,n\}$ into blocks of length $\le m$ where $m = \lfloor 2\log(3/2)K^{-1}h^{-1}\rfloor$. $\square$

7. Proof of the main results. In this section, we combine the results in order to derive the contraction properties for Metropolis–Hastings transition kernels stated in Theorems 1.15, 1.12 and 1.19, and we finally prove 1.20. Note that for $x, \tilde{x} \in \mathbb{R}^d$, the distances

$$\|Y_h^{\mathrm{OU}}(x) - Y_h^{\mathrm{OU}}(\tilde{x})\|_- = (1-h/2)\|x-\tilde{x}\|_- \quad\text{and} \tag{7.1}$$

$$\|Y_h(x) - Y_h(\tilde{x})\|_- = \|x-\tilde{x} - \bigl(\nabla U(x) - \nabla U(\tilde{x})\bigr)h/2\|_- \tag{7.2}$$

are deterministic. We now combine Lemma 2.5 with the estimates in Propositions 1.7 and 1.9:

PROOF OF THEOREM 1.15. We fix $h \in (0,2)$, $R \in (0,\infty)$ and $x, \tilde{x} \in B_R^-$. By the basic contractivity Lemma 2.5 and by (2.3), respectively,

$$\begin{aligned}&\mathbb{E}\bigl[\|W_h(x) - W_h(\tilde{x})\|_-\bigr]\\&\quad\le \|x-\tilde{x}\|_-\\&\qquad - \bigl(1-\mathbb{E}\bigl[\max\bigl(1-\alpha_h\bigl(x,Y_h(x)\bigr), 1-\alpha_h\bigl(\tilde{x},Y_h(\tilde{x})\bigr)\bigr)\bigr]\bigr)\\&\qquad \times\bigl(\|x-\tilde{x}\|_- - \|Y_h(x) - Y_h(\tilde{x})\|_-\bigr)\\&\qquad + \mathbb{E}\bigl[\max\bigl(\|x-Y_h(x)\|_-, \|\tilde{x}-Y_h(\tilde{x})\|_-\bigr)^2\bigr]^{1/2}\\&\qquad \times \mathbb{E}\bigl[\bigl(\alpha_h\bigl(x,Y_h(x)\bigr) - \alpha_h\bigl(\tilde{x},Y_h(\tilde{x})\bigr)\bigr)^2\bigr]^{1/2}.\end{aligned}$$

By Proposition 3.3,

$$\|Y_h(x) - Y_h(\tilde{x})\|_- \le \bigl(1-Kh/2 + M(R)^2h^2/8\bigr)\cdot\|x-\tilde{x}\|_-,$$

and by Lemma 4.4 (2),

$$\begin{aligned}&\mathbb{E}\bigl[\max\bigl(\|x-Y_h(x)\|_-, \|\tilde{x}-Y_h(\tilde{x})\|_-\bigr)^2\bigr]^{1/2}\\&\qquad\le m_2^{1/2}h^{1/2} + \max\bigl(\|\nabla U(x)\|_-, \|\nabla U(\tilde{x})\|_-\bigr)h/2.\end{aligned}$$

The assertion of Theorem 1.15 follows by combining these estimates with the bounds for the acceptance probabilities in Propositions 1.7 and 1.9. □

The corresponding bound for Ornstein–Uhlenbeck proposals follows similarly from Lemma 2.5 and Proposition 1.11:

PROOF OF THEOREM 1.12. We again fix $h \in (0,2)$, $R \in (0,\infty)$, and $x, \widetilde{x} \in B_R^-$. Since $Y_h^{\mathrm{OU}}(x) - Y_h^{\mathrm{OU}}(\widetilde{x}) = (1-h/2)(x-\widetilde{x})$ and $\|x - Y_h^{\mathrm{OU}}(x)\|_- \leq \|x\|_- h/2 + \|Z\|_- \sqrt{h}$, the basic contractivity Lemma 2.5 implies

$$\begin{aligned}
&\mathbb{E}[\|W_h^{\mathrm{OU}}(x) - W_h^{\mathrm{OU}}(\widetilde{x})\|_-] \\
&\quad \leq \left(1 - \frac{h}{2}\right)\|x - \widetilde{x}\|_- \\
&\qquad + \frac{h}{2}\|x - \widetilde{x}\|_- \mathbb{E}\big[\max\big(1 - \alpha^{\mathrm{OU}}\big(x, Y_h^{\mathrm{OU}}(x)\big), 1 - \alpha^{\mathrm{OU}}\big(\widetilde{x}, Y_h^{\mathrm{OU}}(\widetilde{x})\big)\big)\big] \\
&\qquad + \left(\frac{h}{2}\max\big(\|x\|_-, \|\widetilde{x}\|_-\big) + \sqrt{h} m_2^{1/2}\right) \\
&\qquad \times \mathbb{E}\big[\big(\alpha^{\mathrm{OU}}\big(x, Y_h^{\mathrm{OU}}(x)\big) - \alpha^{\mathrm{OU}}\big(\widetilde{x}, Y_h^{\mathrm{OU}}(\widetilde{x})\big)\big)^2\big]^{1/2}.
\end{aligned}$$

The assertion of Theorem 1.12 follows by combining this estimates with the bounds for the acceptance probabilities in Proposition 1.11. □

PROOF OF THEOREM 1.19. Noting that

$$\|x\|_- - (2R)^2 \leq -3R^2 \qquad \text{for any } x \in B_R^-,$$

the assertion is a direct consequence of Corollary 1.18 and Theorem 6.2 applied with R replaced by $2R$. □

Let $\mu_R = \mu(\cdot | B_R^-)$ denote the conditional measure on B_R^-. The fact that μ_R is a stationary distribution for the Metropolis–Hastings transition kernel q_h can be used to bound the Wasserstein distance between $\mu_R q_h^n$ and μ_R:

LEMMA 7.1. *For any $R \geq 0$ and $a \in (0,1)$,*

$$\begin{aligned}
\mathcal{W}_{2R}(\mu_R q_h^n, \mu_R) &\leq 8R(\mu_R q_h^n)(\mathbb{R}^d \setminus B_R^-) \\
&\leq 8(1-a)^{-1}\mathcal{W}_{2R}(\mu_R q_h^n, \delta_0 q_h^n) + 8R(\delta_0 q_h^n)(B_{aR}^-).
\end{aligned}$$

PROOF. The distance induced by the total variation norm $\|\cdot\|_{\mathrm{TV}}$ is the Wasserstein distance w.r.t. the metric $d(x,y) = I_{\{x \neq y\}}$. Since $d_{2R}(x,y) \leq 4Rd(x,y)$, we obtain

$$
\begin{aligned}
\mathcal{W}_{2R}(\mu_R q_h^n, \mu_R) &\le 4R\|\mu_R q_h^n - \mu_R\|_{\mathrm{TV}} \\
&= 8R\|(\mu_R q_h^n - \mu_R)^+\|_{\mathrm{TV}} \\
&\le 8R(\mu_R q_h^n)(\mathbb{R}^d \setminus B_R^-).
\end{aligned} \tag{7.3}
$$

Here we have used in the last step that $\mu q_h = \mu$, and hence

$$
\begin{aligned}
(\mu_R q_h^n)(A) &\le (\mu_R q_h^n)(A \cap B_R^-) + (\mu_R q_h^n)(\mathbb{R}^d \setminus B_R^-) \\
&\le \mu_R(A) + (\mu_R q_h^n)(\mathbb{R}^d \setminus B_R^-)
\end{aligned}
$$

for any Borel set $A \subseteq \mathbb{R}^d$. Moreover, for $a \in (0,1)$,

$$
\begin{aligned}
&\mathcal{W}_{2R}(\mu_R q_h^n, \delta_0 q_h^n) \\
&\quad \ge (R - aR) \cdot \big((\mu_R q_h^n)(\mathbb{R}^d \setminus B_R^-) - (\delta_0 q_h^n)(\mathbb{R}^d \setminus B_{aR}^-)\big).
\end{aligned} \tag{7.4}
$$

Indeed, for any coupling $\eta(dx\, d\widetilde{x})$ of the two measures,

$$
\eta(d_{2R}(x, \widetilde{x}) \ge R - aR) \ge (\mu_R q_h^n)(\mathbb{R}^d \setminus B_R^-) - (\delta_0 q_h^n)(\mathbb{R}^d \setminus B_{aR}^-).
$$

The assertion follows by combining the estimates in (7.3) and (7.4). $\square$

PROOF OF THEOREM 1.20. By combining the estimates in Theorem 1.19, Lemma 7.1 with $a = 6/7$, and Theorem 6.2, we obtain

$$
\begin{aligned}
\mathcal{W}_{2R}(\nu q_h^n, \mu_R) &\le \mathcal{W}_{2R}(\nu q_h^n, \mu_R q_h^n) + \mathcal{W}_{2R}(\mu_R q_h^n, \mu_R) \\
&\le \left(1 - \frac{K}{4}h\right)^n \mathcal{W}_{2R}(\nu, \mu_R) + DR\exp(-KR^2/8)nh \\
&\quad + 56 \cdot \left(1 - \frac{K}{4}h\right)^n \mathcal{W}_{2R}(\mu_R, \delta_0) + 56DR\exp(-KR^2/8)nh \\
&\quad + 8R\mathbb{P}_0[T_{6R/7} \le n] \\
&\le 58R \cdot \left(1 - \frac{K}{4}h\right)^n + 57DR\exp(-KR^2/8)nh \\
&\quad + 8DR\exp(-KR^2/33)nh.
\end{aligned}
$$

$\square$

Acknowledgments. I would like to thank in particular Daniel Gruhlke and Sebastian Vollmer for many fruitful discussions related to the contents of this paper.

REFERENCES

[1] BAKRY, D. (1994). L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on Probability Theory* (*Saint-Flour*, 1992). *Lecture Notes in Math.* **1581** 1–114. Springer, Berlin. MR1307413

[2] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. MR2381160

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

[4] BESKOS, A. and STUART, A. (2009). MCMC methods for sampling function space. In *ICIAM 07—6th International Congress on Industrial and Applied Mathematics* 337–364. Eur. Math. Soc., Zürich. MR2588600

[5] BOU-RABEE, N., HAIRER, M. and VANDEN-EIJNDEN, E. (2010). Non-asymptotic mixing of the MALA algorithm. Available at arXiv:1008.3514.

[6] BOU-RABEE, N. and VANDEN-EIJNDEN, E. (2010). Pathwise accuracy and ergodicity of metropolized integrators for SDEs. *Comm. Pure Appl. Math.* **63** 655–696. MR2583309

[7] CHEN, M. F. and LI, S. F. (1989). Coupling methods for multidimensional diffusion processes. *Ann. Probab.* **17** 151–177. MR0972776

[8] COTTER, S., ROBERTS, G., STUART, A. and WHITE, D. (2012). MCMC methods for functions: Modifying old algorithms to make them faster. Available at arXiv:1202.0709.

[9] DA PRATO, G. and ZABCZYK, J. (1996). *Ergodicity for Infinite-dimensional Systems. London Mathematical Society Lecture Note Series* **229**. Cambridge Univ. Press, Cambridge. MR1417491

[10] DYER, M., FRIEZE, A. and KANNAN, R. (1991). A random polynomial-time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.* **38** 1–17. MR1095916

[11] EBERLE, A. (2011). Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris* **349** 1101–1104. MR2843007

[12] GRUHLKE, D. (2013). Convergence of multilevel MCMC methods on path spaces. Ph.D. thesis, Univ. Bonn.

[13] HAIRER, M., STUART, A. and VOLLMER, S. (2011). Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. Available at arXiv:1112.1392.

[14] HAIRER, M., STUART, A. and VOSS, J. (2011). Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods. In *The Oxford Handbook of Nonlinear Filtering* 833–873. Oxford Univ. Press, Oxford. MR2884617

[15] HAS'MINSKIĬ, R. Z. (1980). *Stochastic Stability of Differential Equations. Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis* **7**. Sijthoff & Noordhoff, Alphen aan den Rijn. MR0600653

[16] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

[17] KANNAN, R., LOVÁSZ, L. and SIMONOVITS, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms* **11** 1–50. MR1608200

[18] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[19] LOVÁSZ, L. and VEMPALA, S. (2006). Hit-and-run from a corner. *SIAM J. Comput.* **35** 985–1005 (electronic). MR2203735

[20] LOVÁSZ, L. and VEMPALA, S. (2006). Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comput. System Sci.* **72** 392–417. MR2205290

[21] LOVÁSZ, L. and VEMPALA, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358. MR2309621

[22] 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

[23] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **6** 1087–1092.

[24] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London. MR1287609

[25] PILLAI, N. S., STUART, A. M. and THIERY, A. H. (2013). Gradient flow from a random walk in Hilbert space. Available at arXiv:1108.1494v3.

[26] PILLAI, N. S., STUART, A. M. and THIÉRY, A. H. (2012). Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.* **22** 2320–2356. MR3024970

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

[28] ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751

[29] ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 255–268. MR1625691

[30] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273

[31] ROYER, G. (2007). *An Initiation to Logarithmic Sobolev Inequalities*. *SMF/AMS Texts and Monographs* **14**. Amer. Math. Soc., Providence, RI. MR2352327

[32] SALOFF-COSTE, L. (1997). Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1996). *Lecture Notes in Math.* **1665** 301–413. Springer, Berlin. MR1490046

[33] STEELE, J. M. (2001). *Stochastic Calculus and Financial Applications*. *Applications of Mathematics* (*New York*) **45**. Springer, New York. MR1783083

[34] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin. MR2459454

[35] VON RENESSE, M.-K. and STURM, K.-T. (2005). Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.* **58** 923–940. MR2142879

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT BONN
ENDENICHER ALLEE 60
53115 BONN
GERMANY
E-MAIL: eberle@uni-bonn.de