EXPLICIT CONVERGENCE BOUNDS FOR METROPOLIS MARKOV CHAINS: ISOPERIMETRY, SPECTRAL GAPS AND PROFILES

BY CHRISTOPHE ANDRIEU[1,a], ANTHONY LEE[1,b], SAM POWER[1,c] AND ANDI Q. WANG[2,d]

[1]*School of Mathematics, University of Bristol,* [a]*c.andrieu@bristol.ac.uk,* [b]*anthony.lee@bristol.ac.uk,* [c]*sam.power@bristol.ac.uk*

[2]*Department of Statistics, University of Warwick,* [d]*andi.wang@warwick.ac.uk*

We derive the first explicit bounds for the spectral gap of a random walk Metropolis algorithm on $\mathbb{R}^d$ for any value of the proposal variance, which when scaled appropriately recovers the correct d^{-1} dependence on dimension for suitably regular invariant distributions. We also obtain explicit bounds on the L^2-mixing time for a broad class of models. In obtaining these results, we refine the use of isoperimetric profile inequalities to obtain conductance profile bounds, which also enable the derivation of explicit bounds in a much broader class of models. We also obtain similar results for the preconditioned Crank–Nicolson Markov chain, obtaining dimension-independent bounds under suitable assumptions.

CONTENTS

1. Introduction.

1.1. *Results for Metropolis Markov chains.* Let π be a probability distribution on $(\mathsf{E}, \mathscr{E})$, where $\mathsf{E} = \mathbb{R}^d$ and $\mathscr{E}$ denotes its Borel σ-algebra, and suppose we seek to approximately sample from π. Markov chain Monte Carlo (MCMC) algorithms address this problem by simulating an ergodic, time-homogeneous Markov chain $(X_n)_{n \in \mathbb{N}}$ with invariant distribution π. One of the most simple and yet enduringly popular MCMC algorithms is the *Metropolis algorithm* of Metropolis et al. (1953). Assuming that π has density $\varpi = d\pi/d\nu$ with respect to some σ-finite measure ν, and Q is a ν-reversible Markov kernel, a *Metropolis Markov kernel* may be written

$$(1) \qquad P(x, A) = \int_A Q(x, dy)\alpha(x, y) + \mathbf{1}_A(x)\bar{\alpha}(x), \quad x \in \mathsf{E}, A \in \mathscr{E},$$

where for $x, y \in \mathsf{E}$,

$$(2) \quad \alpha(x, y) := \min\left\{1, \frac{\varpi(y)}{\varpi(x)}\right\}, \qquad \bar{\alpha}(x) := 1 - \alpha(x), \qquad \alpha(x) := \int_\mathsf{E} Q(x, dy)\alpha(x, y).$$

In many applications, $Q(x, \cdot)$ is a multivariate normal distribution with mean x and covariance matrix $\sigma^2 \cdot I_d$, where I_d is the identity matrix, in which case ν is the Lebesgue measure and P is the *Random–Walk Metropolis* (RWM) Markov kernel. Despite its simplicity, the RWM algorithm is known to perform very well for certain classes of target distributions, and furthermore to be a *robust* algorithm (see, e.g., Chen et al. (2020), Christensen, Roberts and Rosenthal (2005), Livingstone and Zanella (2022), Roberts, Gelman and Gilks (1997)). In this paper, quantitative analysis of the L^2-mixing time and spectral gap of the RWM Markov chain is the primary application, with a particular emphasis on the dependence of these quantities on dimension. This analysis relies on a more general theory applicable beyond the specific scenarios considered here; see Section 1.2.

For a given target distribution π, after fixing the coordinate system, the only tuning parameter of the RWM kernel is the proposal variance σ^2. It is well known that if σ^2 is too large, then the acceptance function α will deteriorate, and the Markov chain will tend to get "stuck" for long periods. On the other hand, if σ^2 is too small, then the Markov chain will tend to make very small steps. Both of these regimes correspond intuitively to slow convergence of the Markov chain. In the celebrated optimal scaling paper of Roberts, Gelman and Gilks (1997), it was shown, for a fairly restrictive class of target distributions, that the proposal variance σ^2 of RWM on $\mathbb{R}^d$ should scale like d^{-1} to obtain a stable acceptance ratio in the high-dimensional limit, and that the complexity of sampling depends linearly on dimension, via a particular but indicative weak convergence result to a Langevin diffusion. In this paper, we study the high-dimensional properties of the RWM algorithm from a different angle: we seek to explicitly bound the spectral gap of the RWM kernel in arbitrary dimension d and for *any* value of σ^2. For appropriately regular distributions, we find that scaling σ^2 as d^{-1} does indeed imply a spectral gap that is precisely of order d^{-1}, and that this choice of polynomial scaling is optimal. The following is a combination of Corollary 35 and Theorem 46:

THEOREM 1. *Let π have density $\pi(x) \propto \exp(-U(x))$ with respect to Lebesgue measure on $\mathbb{R}^d$, where the potential U is L-smooth, m-strongly convex and twice continuously differentiable. If P is the π-reversible RWM kernel with $\mathcal{N}(0, \sigma^2 \cdot I_d)$ proposal increments, then the spectral gap γ_P of P satisfies*

$$(3) \quad C \cdot L \cdot d \cdot \sigma^2 \cdot \exp(-2 \cdot L \cdot d \cdot \sigma^2) \cdot \frac{m}{L} \cdot \frac{1}{d} \leq \gamma_P \leq \min\left\{\frac{1}{2} \cdot L \cdot \sigma^2, (1 + m \cdot \sigma^2)^{-d/2}\right\},$$

where $C = 1.972 \times 10^{-4}$.

Twice continuous differentiability of U is only used to obtain the upper bound. For some intuition, densities with m-strongly convex and L-smooth potentials U can be sandwiched between $\mathcal{N}(x_*, L^{-1} \cdot I_d)$ and $\mathcal{N}(x_*, m^{-1} \cdot I_d)$, up to constant factors, where x_* is the maximizer of the density of π; see Lemma 42.

Both the lower and upper bounds in (3) demonstrate that taking σ^2 too small or too large causes γ_P to decrease. The lower bound in (3) is maximized by taking $\sigma^2 = 1/(2 \cdot L \cdot d)$, while the rate at which the upper bound decreases with d is also minimized, among polynomial scalings, by scaling σ^2 with d^{-1}. Taking $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-1/2}$ for any constant $\varsigma > 0$, we obtain

$$(4) \qquad\qquad\qquad C \cdot \varsigma^2 \cdot \exp(-2 \cdot \varsigma^2) \cdot \frac{m}{L} \cdot \frac{1}{d} \leq \gamma_P \leq \frac{\varsigma^2}{2} \cdot \frac{1}{d},$$

so the $\mathcal{O}(d^{-1})$ dimension dependence is tight. The lower bound is maximized by taking $\varsigma^2 = \frac{1}{2}$, although it is unlikely that this is optimal in practice due to the results of Roberts, Gelman and Gilks (1997). Similarly, it seems likely that the optimal value of C is possibly a few orders of magnitude larger.

We also study the L^2-convergence complexity of the RWM Markov chain, noting that convergence can initially be *faster* than that indicated by the spectral gap alone and this turns out to be crucial to establish our dimension dependence results for m-strongly convex and L-smooth potentials. Under the same conditions as Theorem 1 and taking $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-1/2}$ as above, we obtain that for at least two types of feasible initial distribution μ (see Theorem 49 and Remarks 50 and 51) one may take

$$n \in \mathcal{O}\left(\exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d \cdot \{\log d + \log \kappa + \log(\varepsilon_{\mathrm{Mix}}^{-1})\}\right),$$

and obtain $\chi^2(\mu P^n, \pi) \leq \varepsilon_{\mathrm{Mix}}$, where $\chi^2(\mu, \nu)$ denotes the χ^2 divergence between μ and ν and $\kappa := L/m$ is the condition number. In contrast, an analysis based only on the spectral gap bound $\gamma_P \in \Omega(1/(\kappa \cdot d))$ would suggest a mixing time in $\mathcal{O}(d^2 \kappa \log \kappa)$.

In practice, fluctuations of ergodic averages of $f \in L^2(\pi)$ are also of interest, and one may consider the *asymptotic variance*, given by

$$\mathrm{var}(P, f) := \lim_{n \to \infty} n \cdot \mathrm{var}\left(\frac{1}{n} \sum_{i=1}^{n} f(X_i)\right),$$

where $X_0 \sim \pi$. We show in Proposition 48 that with $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-1/2}$,

$$\mathrm{var}(P, f) \leq 10141 \cdot \varsigma^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \kappa \cdot d \cdot \|f\|_2^2, \quad \varsigma > 0.$$

We also show that linear functions satisfy $\mathrm{var}(P, f) \geq 2 \cdot \varsigma^{-2} \cdot d \cdot \|f\|_2^2$.

We also analyze the *preconditioned Crank–Nicolson* (pCN) Markov chain via essentially analogous theory to the RWM chain, since it is also a Metropolis Markov chain. For example, we show in Theorem 54 if $\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, \mathsf{C}) \exp(-\Psi(x))$ with Ψ convex, L-smooth and minimized at $x = 0$ then an appropriately tuned pCN Markov chain's spectral gap satisfies

$$\gamma_P \geq 3.62784 \times 10^{-5} \cdot (L \cdot \mathrm{Tr}(\mathsf{C}))^{-1},$$

giving dimension-independent bounds when $L \cdot \mathrm{Tr}(\mathsf{C})$ is bounded independent of dimension.

To prove these results we apply a general result, Theorem 18, which requires quantitative lower bounds on the *isoperimetric profile* of π for some metric d, complemented with a quantitative *close coupling* condition for P:

DEFINITION 2 (Close coupling). For a metric d on E and $\epsilon, \delta > 0$, a Markov kernel P evolving on E is (d, δ, ε)-*close coupling* if

$$d(x, y) \leq \delta \quad \Rightarrow \quad \|P(x, \cdot) - P(y, \cdot)\|_{\text{TV}} \leq 1 - \varepsilon, \quad x, y \in \mathsf{E}.$$

This is to be contrasted with what is known about the overdamped Langevin diffusion, which solves the stochastic differential equation

$$dX_t = \nabla \log \pi(X_t)\, dt + \sqrt{2} \cdot dW_t,$$

for which knowledge of the isoperimetric profile *alone* can provide information on its convergence. For example, the overdamped Langevin diffusion is associated with the classical Dirichlet form $f \mapsto \pi(|\nabla f|^2)$; (see, e.g., Pavliotis (2014), Section 4.5), and this allows one to deduce Poincaré and log-Sobolev inequalities in the presence of appropriate isoperimetric inequalities (see, e.g., Milman (2012), Section 2.2). The RWM chain may indeed be viewed as a discretization of this diffusion, but our results do not explicitly compare the diffusion with the Markov chain; indeed our quantitative bounds are valid in *any* dimension and for *any* value of σ^2. The additional close coupling condition required for RWM in fact introduces a penalty in the convergence bounds, by which convergence degrades as the product $\delta \cdot \varepsilon$ decreases. To demonstrate close coupling for Metropolis chains, we show that for $\alpha_0 := \inf_{z \in \mathsf{E}} \alpha(z)$, with α as in (2),

$$\|P(x, \cdot) - P(y, \cdot)\|_{\text{TV}} \leq \|Q(x, \cdot) - Q(y, \cdot)\|_{\text{TV}} + 1 - \alpha_0,$$

and we show that α_0 can be lower bounded for any σ^2 under the assumption of L-smoothness. One may then take δ such that $|x - y| \leq \delta \Rightarrow \|Q(x, \cdot) - Q(y, \cdot)\|_{\text{TV}} \leq \frac{1}{2} \cdot \alpha_0$ to obtain that P is close coupling with $\varepsilon \geq \frac{1}{2} \cdot \alpha_0$. In our analysis, we find that to maximize the spectral gap of P as a function of dimension, it is sufficient to scale σ^2 as d^{-1}. Ultimately, one may view the penalty for running an appropriately tuned RWM instead of Langevin as being of order d^{-1} in terms of the spectral gap.

1.2. *Roadmap.* In Section 2, we review the notions of *conductance profile* and *spectral profile* for Markov chains, and show how these can be used to establish bounds on the spectral gap and mixing time of the chain.

In Section 3, we introduce notions of *isoperimetric profiles* of probability measures with respect to a given metric. We show that when combined with the *close coupling* condition for an invariant Markov kernel, one can deduce bounds on the conductance profile of the chain, and hence on the spectral gap and mixing time. We then give a number of concrete examples in which the isoperimetric profile can be well-controlled, and discuss the implications on convergence. See Figure 1 for a diagram of these results. We also prove a general close-coupling result for Metropolis algorithms.

In Sections 4–5, we apply these tools to study the RWM algorithm. Although one can obtain nonasymptotic bounds for the spectral gap and mixing time under various isoperimetric and smoothness assumptions using our techniques, we focus on obtaining concrete results when U is m-strongly convex and L-smooth. In this case, we obtain nonasymptotic estimates on both the spectral gap and the mixing time. Furthermore, we show that for this class of densities the dependence of the spectral gap on dimension is sharp, up to sub-polynomial factors. We also demonstrate that, when appropriately tuned, the asymptotic variance of a
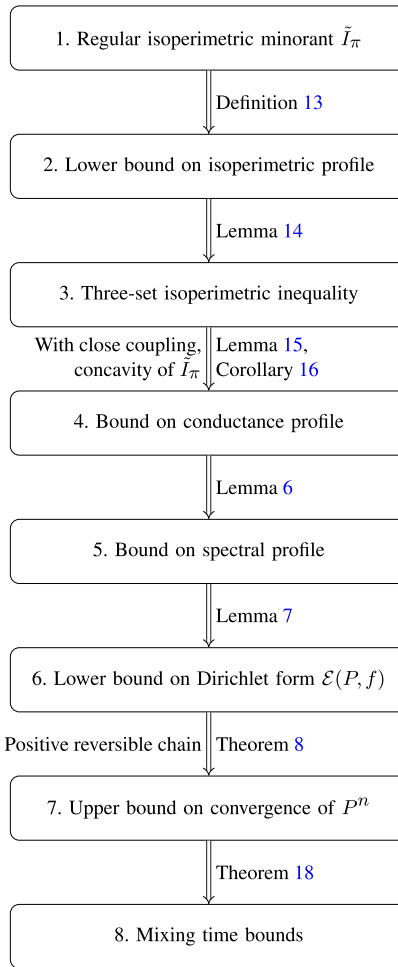
```
┌─────────────────────────────────────────┐
│  1. Regular isoperimetric minorant Ĩ_π   │
└─────────────────────────────────────────┘
                    │ Definition 13
                    ▼
┌─────────────────────────────────────────┐
│  2. Lower bound on isoperimetric profile │
└─────────────────────────────────────────┘
                    │ Lemma 14
                    ▼
┌─────────────────────────────────────────┐
│  3. Three-set isoperimetric inequality   │
└─────────────────────────────────────────┘
  With close coupling, │ Lemma 15,
     concavity of Ĩ_π  │ Corollary 16
                       ▼
┌─────────────────────────────────────────┐
│  4. Bound on conductance profile         │
└─────────────────────────────────────────┘
                    │ Lemma 6
                    ▼
┌─────────────────────────────────────────┐
│  5. Bound on spectral profile            │
└─────────────────────────────────────────┘
                    │ Lemma 7
                    ▼
┌─────────────────────────────────────────┐
│  6. Lower bound on Dirichlet form E(P, f)│
└─────────────────────────────────────────┘
  Positive reversible chain │ Theorem 8
                            ▼
┌─────────────────────────────────────────┐
│  7. Upper bound on convergence of P^n    │
└─────────────────────────────────────────┘
                    │ Theorem 18
                    ▼
┌─────────────────────────────────────────┐
│  8. Mixing time bounds                   │
└─────────────────────────────────────────┘
```

FIG. 1. *Outline of the results of Sections 2, 3.*

RWM chain is upper bounded by a linear factor d times the ideal variance, and that there is a matching linear lower bound for linear functions.

In Section 6, we apply the same tools to the study of the pCN algorithm for sampling perturbations of Gaussian measures. We again obtain nonasymptotic estimates on both the spectral gap and mixing time of the chain.

1.3. *Related work.* This paper develops in a systematic manner a comparatively crude analysis in a technical report by the authors (Andrieu et al. (2022), Sections 5.3–5.4).

One of the first attempts to establish directly the existence of a (right) L^2-spectral gap for RWM on $\mathbb{R}^d$ is Miclo and Roberto (2000), where a quantization approach is used to approximate the initial problem, and results concerning Markov chains on graphs are leveraged. The authors focus on the scenario $d = 1$, although it is argued that the results could be generalized to multiple dimensions. The assumptions made on the negative log-density U are less stringent than those considered here, however no quantitative bounds were provided. Another contribution in this direction is Jarner and Yuen (2004) where a lower bound on the conductance was obtained for monotone log-concave distributions on $\mathbb{R}$. The existence of an L^2 spectral gap can also be established using drift and minorization techniques: for example, Jarner and Hansen (2000) demonstrate that the RWM chain is L^1- geometrically ergodic under fairly mild conditions on π, and since RWM chains are reversible those conditions also

imply L^2-geometric ergodicity and the existence of a spectral gap (Roberts and Rosenthal (1997), Roberts and Tweedie (2001)). However, these techniques typically do not provide accurate quantitative bounds on the size of the gap; see for instance, Qin and Hobert (2021).

The results and the approach we take here are inspired by several papers. The recent papers Belloni and Chernozhukov (2009), Dwivedi et al. (2019) and Chen et al. (2020) are most closely related to our approach. All three of these papers also consider bounds on the conductance or conductance profile of the RWM kernel P, but *restricted* to some compact subset K of $\mathbb{R}^d$. As a consequence, they do not provide a positive lower bound on the spectral gap of the unrestricted P. More precisely, Belloni and Chernozhukov (2009) and Dwivedi et al. (2019) prove a restricted variant of Corollary 16 for the conductance, corresponding to connection 3–4 in Figure 1. Chen et al. (2020) prove restricted variants of connections 3–8. Dwivedi et al. (2019) and Chen et al. (2020) obtain complexity bounds for convergence of μP^n to π on $\mathbb{R}^d$ using the notion of s-conductance, which entails delicate balancing of the desired final error, the size of K, properties of μ and even σ^2. In particular, we emphasize that all three prior complexity analyses involve using specific, theoretically-motivated and typically unknown values of σ^2, so the results do not cover the arbitrary values of σ^2 used in practice. The restriction to K in these papers is necessary since the authors only verify the close coupling condition for P on K. In contrast, we are able to verify this condition globally, and hence there is no need to consider restrictions. As a result, we can obtain a positive lower bound on the spectral gap, and the convergence analysis does not require the same type of fine balancing. In particular, we also find an improved dependence of the mixing time on the condition number κ, in comparison to the dependence in Chen et al. (2020). We also mention Mathé and Novak (2007), who proved that the Metropolis chain with a ball-walk proposal for π log-concave with Lipschitz potential and restricted to a ball has a spectral gap in $\Omega(d^{-2})$; see also Rudolf (2009).

Belloni and Chernozhukov (2009) and Dwivedi et al. (2019) use a type of 3-set exponential isoperimetric profile inequality to infer a bound on the conductance of the chain restricted to K, in the presence of the close coupling condition. The isoperimetric inequality is verified for (perturbations of) m-strongly convex potentials. In Chen et al. (2020), a Gaussian 3-set isoperimetric profile inequality is used to infer a bound on the conductance profile of the restricted chain in the presence of the same coupling condition, and an isoperimetric profile inequality is verified for strongly convex potentials. Our main contribution in relation to this part of the theory is to show that any sufficiently regular isoperimetric profile implies a corresponding 3-set isoperimetric inequality. In fact, Chen et al. (2020)'s consideration of the Gaussian 3-set isoperimetric inequality and its implication for rapid convergence far from equilibrium was the main inspiration for our results relating classical isoperimetric profiles and conductance profiles more generally. Our subsequent mixing time results are mostly direct consequences of the relationships between the conductance profile, spectral profile and L^2-convergence, as developed by Goel, Montenegro and Tetali (2006).

Hairer, Stuart and Vollmer (2014) show the existence and stability of the spectral gap of pCN as $d \to \infty$ under quite general conditions, but the bounds so obtained are understood to be somewhat loose numerically, and their dependence on the various parameters of the target measure is implicit. Here, we make more restrictive assumptions on the target measure, which allows us to obtain bounds which are more interpretable and perhaps sharper.

1.4. *Notation.*   Notation is collected for convenience in Appendix A.

2. Conductance profile, spectral profile and mixing time bounds.   The spectral gap γ_P of a π-reversible Markov kernel P provides important information on the convergence of the chain. Indeed, for any $n \in \mathbb{N}_0$,

$$(5) \qquad \qquad \left\| P^n f \right\|_2 \leq \|f\|_2 \cdot (1 - \gamma_P)^n, \quad f \in L_0^2(\pi),$$

and the factor $(1 - \gamma_P)$ cannot be reduced in general, motivating quantitative lower bounds on γ_P. By taking $f = d\mu/d\pi - 1$ in (5) we may deduce bounds on $\chi^2(\mu P^n, \pi)$, the chi-squared divergence between μP^n and π, and thereby upper bound mixing times. However, using only this bound can give very conservative bounds when $\chi^2(\mu, \pi)$ is large, and so we will use more refined techniques to control the convergence behaviour of the chain when it is far from equilibrium. In particular, we make use of the *spectral profile* (Goel, Montenegro and Tetali (2006)) and *conductance profile* of the Markov chain (Lovász and Kannan (1999), Morris and Peres (2005)). These techniques are able to capture the following phenomenon: many Markov chains, when far from equilibrium, are able to mix at *faster* than exponential rates, or equally, that sets of small measure in the state space are comparatively easier to escape from. Moreover, these techniques are capable of providing greatly-improved bounds on mixing times, and in some cases, nearly-optimal bounds (see, e.g., Kozma (2007)).

DEFINITION 3 (Conductance and conductance profile).   The *conductance profile* of a π-invariant Markov kernel P is

$$\Phi_P(v) := \inf\left\{\frac{(\pi \otimes P)(A \times A^\complement)}{\pi(A)} : A \in \mathscr{E}, 0 < \pi(A) \le v\right\}, \quad v \in \left(0, \frac{1}{2}\right].$$

The *conductance* of P is $\Phi_P^* := \Phi_P(\frac{1}{2})$.

DEFINITION 4 (Spectral profile).   Let P be a π-invariant Markov kernel, then we define

$$C_0^+(A) := \{g : \mathsf{E} \to \mathbb{R} | \mathrm{supp}g \subseteq A, g \ge 0, g \ne \mathrm{const.}\ \pi\text{-a.s.}\}, \quad A \in \mathscr{E},$$

where $\mathrm{supp}g$ is the closure of $\{x \in \mathsf{E} : |g(x)| > 0\}$, and

$$\lambda_P(A) := \inf_{g \in C_0^+(A)} \frac{\mathcal{E}(P, g)}{\mathrm{Var}_\pi(g)}, \quad A \in \mathscr{E}, \pi(A) > 0.$$

The *spectral profile* of P is

$$\Lambda_P(v) := \inf\{\lambda_P(A) : A \in \mathscr{E}, 0 < \pi(A) \le v\}, \quad v > 0.$$

We note that for all $v > 0$ and π-reversible P, we have that $\Lambda_P(v) \ge \mathrm{Gap_R}(P) \ge \gamma_P$. To proceed from here, we first use a Cheeger-type argument to bound the spectral profile using the conductance profile. The statement and proof of Lemma 6 are very similar to Chen et al. (2020), Lemma 12, with one difference being that we do not restrict the state space. The proof can be found in Appendix B. We also recall Cheeger's inequalities.

LEMMA 5 (Lawler and Sokal (1988), Theorem 3.5; Cheeger's inequalities).   *If P is a π-reversible Markov kernel, then*

$$\frac{1}{2} \cdot \left[\Phi_P^*\right]^2 \le \mathrm{Gap_R}(P) \le 2\Phi_P^*.$$

LEMMA 6.   *If P is a π-reversible Markov kernel, then*

$$\Lambda_P(v) \ge \begin{cases} \dfrac{1}{2} \cdot \Phi_P(v)^2, & 0 < v \le \dfrac{1}{2}, \\ \dfrac{1}{2} \cdot \left[\Phi_P^*\right]^2, & v > \dfrac{1}{2}. \end{cases}$$

We will make use of the following lower bound on the Dirichlet form in terms of the spectral profile.

LEMMA 7 (Goel, Montenegro and Tetali (2006), Lemma 2.1). *For $g \in L^2(\pi)$ nonnegative and not constant π-a.s.,*

$$\mathcal{E}(P, g) \geq \mathrm{Var}_\pi(g) \cdot \frac{1}{2} \cdot \Lambda_P\left(4 \cdot \frac{[\pi(g)]^2}{\mathrm{Var}_\pi(g)}\right).$$

Our final result in this section shows how the conductance profile can be used to deduce bounds on convergence of P. We build on previous work, particularly Goel, Montenegro and Tetali (2006) in the discrete setting and Chen et al. (2020) on general state spaces with "restricted" conductance profiles.

THEOREM 8. *Let P be a positive, π-reversible Markov kernel with $\Phi_P^* > 0$, $\mu \ll \pi$ a probability measure, and $\varepsilon_{\mathrm{Mix}} \in (0, 8)$. To ensure $\chi^2(\mu P^n, \pi) \leq \varepsilon_{\mathrm{Mix}}$, it suffices to take*

$$n \geq 2 + 4 \cdot \int_{\min\{4 \cdot u_0^{-1}, 1/2\}}^{1/2} \frac{1}{v \cdot \Phi_P(v)^2} dv + [\Phi_P^*]^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\mathrm{Mix}}}, 1\right\}\right),$$

where $u_0 = \chi^2(\mu, \pi)$.

PROOF. Writing $h = \frac{d\mu}{d\pi}$ and $u_n := \mathrm{Var}_\pi(P^n h) = \chi^2(\mu P^n, \pi)$, compute that

$$u_n - u_{n+1} = \mathcal{E}(P^2, P^n h) \geq \mathcal{E}(P, P^n h) \geq u_n \cdot \frac{1}{2} \cdot \Lambda_P(4 \cdot u_n^{-1}),$$

where we have used the positivity of P to bound $\mathcal{E}(P^2, f) \geq \mathcal{E}(P, f)$, Lemma 7, and recalled that $\pi(P^n h) = 1$ for all n. Defining $L_P(\eta) := \frac{1}{2} \cdot \eta \cdot \Lambda_P(4 \cdot \eta^{-1})$ for $\eta > 0$, it thus holds that $u_n - u_{n+1} \geq L_P(u_n)$.

We now distinguish between whether u_0 is greater or smaller than 8, noting that in the latter case, using the spectral gap directly allows for tighter control of the increment $u_n - u_{n+1}$. Supposing that $u_0 \geq 8$, we will first estimate how long it takes for u_n to drop below 8. Recalling that Λ_P is a decreasing function, it is straightforward to see that L_P is an increasing function, and hence measurable. Additionally, since Λ_P is bounded below by $\gamma_P > 0$, it follows that L_P is bounded away from 0 on intervals not containing 0. Assuming that both u_n, u_{n+1} are at least 8, we can then write

$$\int_{u_{n+1}}^{u_n} \frac{d\eta}{L_P(\eta)} \geq \frac{u_n - u_{n+1}}{L_P(u_n)} \geq 1.$$

Moreover, if $u_0 \geq u_1 \geq \cdots \geq u_n \geq 8$, then we may sum up these inequalities to see that

$$\int_8^{u_0} \frac{d\eta}{L_P(\eta)} \geq \int_{u_n}^{u_0} \frac{d\eta}{L_P(\eta)} \geq n.$$

In particular, for $n \geq 1 + \int_8^{u_0} \frac{d\eta}{L_P(\eta)}$, it must hold that $u_n < 8$. Now, recall that for $\eta \geq 8$, we can bound

$$L_P(\eta) = \frac{1}{2} \cdot \eta \cdot \Lambda_P(4 \cdot \eta^{-1}) \geq \frac{1}{2} \cdot \eta \cdot \left\{\frac{1}{2} \cdot \Phi_P(4 \cdot \eta^{-1})^2\right\} = \frac{\Phi_P(4 \cdot \eta^{-1})^2}{4 \cdot \eta^{-1}}.$$

We then compute that

$$\int_8^{u_0} \frac{d\eta}{L_P(\eta)} \leq \int_8^{u_0} \frac{4 \cdot \eta^{-1}}{\Phi_P(4 \cdot \eta^{-1})^2} d\eta = 4 \cdot \int_{4 \cdot u_0^{-1}}^{1/2} \frac{1}{v \cdot \Phi_P(v)^2} dv,$$

noting that Φ_P is monotone, hence measurable and bounded below by $\Phi_P(\frac{1}{2}) > 0$, hence the integral exists.

For $u_0 < 8$, we control the decay of u_n more tightly by using the spectral gap of P, γ_P. We obtain

$$u_n - u_{n+1} = \mathcal{E}(P^2, P^n h) \geq (1 - (1 - \gamma_P)^2) \cdot \mathrm{Var}_\pi(P^n h)$$

$$\implies \quad u_{n+1} \leq (1 - \gamma_P)^2 \cdot u_n$$

and thus that

$$u_n \leq (1 - \gamma_P)^{2 \cdot n} \cdot u_0 \leq \exp(-2 \cdot \gamma_P \cdot n) \cdot u_0.$$

One can then deduce that for $n \geq 1 + \frac{1}{2} \cdot \gamma_P^{-1} \cdot \log(\frac{u_0}{\varepsilon_{\mathrm{Mix}}})$, $u_n \leq \varepsilon_{\mathrm{Mix}}$. By Lemma 5, we recall that $\gamma_P \geq \frac{1}{2} \cdot [\Phi_P^*]^2$. The result follows by assembling the various cases. □

We note a similarity between the consequences of the spectral profile and the so-called "super-Poincaré" inequalities of Wang (2000). See Proposition 62 in Appendix B for some details on this connection, which may be known among experts but does not appear to have been explicitly documented. It is well known that stronger functional inequalities than the Poincaré inequality allow improved dependence on dimension/initialization in Markov chain mixing time results (see, e.g., Diaconis and Saloff-Coste (1998)) and one perspective on what we pursue in the sequel is that one may combine bounds on the isoperimetric profile with the close coupling condition to deduce bounds on the conductance profile and hence spectral profile, which contains comparable information to functional inequalities like the super-Poincaré inequality. Theorem 8 demonstrates how this functional inequality can provide sharper mixing time bounds than those based on the spectral gap alone.

3. From isoperimetric profiles and close coupling to mixing time bounds.

3.1. *General results.* From this point onwards, the following assumption is in force. All statements are made with respect to a given metric d on E, the dependence on which may be suppressed when no ambiguity can result.

ASSUMPTION 9. The probability distribution π on $\mathsf{E} = \mathbb{R}^d$ has a positive density w.r.t. Lebesgue, given by $\pi \propto \exp(-U)$, for some potential $U : \mathbb{R}^d \to \mathbb{R}$.

DEFINITION 10 (Three-set isoperimetric inequality). A probability measure π satisfies a *three-set isoperimetric inequality* with metric d and function $F : (0, \frac{1}{2}] \to [0, \infty)$ if for all measurable partitions of the state space $\mathsf{E} = S_1 \sqcup S_2 \sqcup S_3$ with $\pi(S_1), \pi(S_2) > 0$,

$$(6) \qquad\qquad \pi(S_3) \geq \mathsf{d}(S_1, S_2) \cdot F(\min\{\pi(S_1), \pi(S_2)\}).$$

DEFINITION 11 (Isoperimetric profile). For $A \in \mathscr{E}$ and $r \geq 0$, let $A_r := \{x \in \mathsf{E} : \mathsf{d}(x, A) < r\}$, and define the *Minkowski content* of A under π with respect to d by

$$\pi^+(A) = \lim_{r \to 0^+} \inf \frac{\pi(A_r) - \pi(A)}{r}.$$

The *isoperimetric profile of* π with respect to the metric d is

$$(7) \qquad\qquad I_\pi(p) := \inf\{\pi^+(A) : A \in \mathscr{E}, \pi(A) = p\}, \quad p \in (0, 1).$$

We note briefly that the isoperimetric profile can be controlled explicitly in many cases of interest; examples to this effect are provided in Section 3.2. The following is a special case of a nontrivial result for distributions defined on Riemannian manifolds which is the product of extensive research by several authors, and holds specifically for d being the natural metric induced by the given Riemannian structure; we recall it here to emphasize that the notion of regularity on the isoperimetric profile that we assume is reasonable.

LEMMA 12 (Milman (2009a), Theorem 1.8). *If U is convex and twice-continuously differentiable, then I_π is symmetric about $\frac{1}{2}$ and concave.*

DEFINITION 13. We say that $\tilde{I}_\pi : (0, 1) \to (0, \infty)$ is an *isoperimetric minorant of π* if $\tilde{I}_\pi \leq I_\pi$ pointwise. We furthermore say that $\tilde{I}_\pi$ is *regular* if it is symmetric about $\frac{1}{2}$, continuous, and monotone increasing on $(0, \frac{1}{2}]$.

To begin with, we show that the existence of a regular isoperimetric minorant is equivalent to the existence of a corresponding three-set isoperimetric inequality.

LEMMA 14. π *has a regular isoperimetric minorant $\tilde{I}_\pi$ w.r.t. the metric* d $\iff$ π *satisfies a three-set isoperimetric inequality with metric* d *and function $F = \tilde{I}_\pi$ on $(0, \frac{1}{2}]$.*

PROOF. ($\Leftarrow$) Following Bobkov and Houdré (1997a), Section 2, we may consider only closed sets A in (7). For arbitrary closed $A \in \mathscr{E}$ with $\pi(A) \in (0, 1)$, for $r > 0$ let A_r be as defined in Definition 11. We may take $S_1 = A$, $S_3 = A_r \setminus A = \{x \in \mathsf{E} \setminus A : \mathsf{d}(x, A) < r\}$ and $S_2 = \mathsf{E} \setminus A_r$. From Definition 10,

$$\pi(A_r) - \pi(A) \geq r F(\min\{\pi(A), \pi(\mathsf{E} \setminus A_r)\}),$$

from which we obtain

$$\pi^+(A) = \lim_{r \to 0^+} \inf \frac{\pi(A_r) - \pi(A)}{r} \geq F(\min\{\pi(A), \pi(A^\complement)\}),$$

since for closed A, $\lim_{r \to 0^+} \pi(\mathsf{E} \setminus A_r) = 1 - \lim_{r \to 0^+} \pi(A_r) = \pi(A^\complement)$. Hence, $\tilde{I}_\pi(t) = F(\min\{t, 1 - t\})$ is an isoperimetric minorant, symmetric on $(0, 1)$, continuous and monotone increasing on $(0, \frac{1}{2}]$ and hence regular.

($\Rightarrow$) Following Bobkov and Houdré (1997a), Theorem 4.1 and Remark 4.2, for any Lipschitz $f : \mathsf{E} \to [0, 1]$ one may write

$$\pi(|\nabla f|) \geq \int_0^1 \pi^+(f > t)\, dt \geq \int_0^1 I_\pi(\pi(f > t))\, dt \geq \int_0^1 \tilde{I}_\pi(\pi(f > t))\, dt,$$

where we have written $(f > t)$ for the set $\{x \in \mathsf{E} : f(x) > t\}$, and

$$|\nabla f(x)| := \limsup_{\mathsf{d}(x, y) \to 0^+} \frac{|f(x) - f(y)|}{\mathsf{d}(x, y)} \in [0, \infty],$$

is the modulus of the gradient of f. Now let $\mathsf{E} = S_1 \sqcup S_2 \sqcup S_3$ with $\pi(S_1), \pi(S_2) > 0$. If $\mathsf{d}(S_1, S_2) = 0$ then (6) holds trivially, so henceforth we assume $\mathsf{d}(S_1, S_2) > 0$. Following the construction in the proof of Ledoux (2001), Proposition 1.7, we define $f : \mathsf{E} \to [0, 1]$ by $f(x) := \min\{1, \frac{\mathsf{d}(S_1, x)}{\mathsf{d}(S_1, S_2)}\}$. This function is $\mathsf{d}(S_1, S_2)^{-1}$-Lipschitz on S_3 and flat elsewhere. It thus holds for this f that

$$\pi(|\nabla f|) \leq \pi(S_3) \cdot \mathsf{d}(S_1, S_2)^{-1} \quad \implies \quad \pi(S_3) \geq \mathsf{d}(S_1, S_2) \cdot \pi(|\nabla f|).$$

We now seek a lower bound on $\pi(|\nabla f|)$, for which we will make use of the isoperimetric profile. Observe that for $t \in [0, 1)$, it holds that

$$\{x \in \mathsf{E} : f(x) > t\} = \{x \in S_2 : f(x) > t\} \sqcup \{x \in S_3 : f(x) > t\}$$

$$= S_2 \sqcup \{x \in S_3 : f(x) > t\},$$

and hence that $\pi(f > t) \in [\pi(S_2), \pi(S_2 \sqcup S_3)] = [\pi(S_2), \pi(S_1^\complement)]$.

Suppose now that $\max\{\pi(S_1), \pi(S_2)\} \geq \frac{1}{2}$, and without loss of generality that $\pi(S_1) \geq \frac{1}{2} \geq \pi(S_2)$. It then follows for $t \in (0, 1)$ that $\pi(f > t) \in [\pi(S_2), \pi(S_1^{\complement})] \subseteq [0, \frac{1}{2}]$. By monotonicity of $\tilde{I}_\pi$ on $(0, \frac{1}{2}]$, it then holds that $\tilde{I}_\pi(\pi(f > t)) \geq \tilde{I}_\pi(\pi(S_2))$ and thus that $\pi(|\nabla f|) \geq \tilde{I}_\pi(\pi(S_2))$, from which it follows that

$$\pi(S_3) \geq \mathsf{d}(S_1, S_2) \cdot \tilde{I}_\pi(\pi(S_2)) = \mathsf{d}(S_1, S_2) \cdot \tilde{I}_\pi(\min\{\pi(S_1), \pi(S_2)\}).$$

On the contrary, suppose that $\max\{\pi(S_1), \pi(S_2)\} < \frac{1}{2}$. It then holds that any median v of f under π lies in $(0, 1)$, so one can write that

$$t \in [0, v] \quad \Longrightarrow \quad \pi(f > t) \geq \frac{1}{2}, \quad t \in (v, 1] \quad \Longrightarrow \quad \pi(f > t) \leq \frac{1}{2}.$$

Letting v be such a median, observe that $t \mapsto \tilde{I}_\pi(\pi(f > t))$ is increasing on $[0, v]$ and decreasing on $[v, 1]$. In particular,

$$\tilde{I}_\pi(\pi(f > t)) \geq \tilde{I}_\pi(\pi(f > 0)) = \tilde{I}_\pi(\pi(S_1^{\complement})), \quad t \in [0, v],$$

making use of the fact that $\pi \ll \text{Leb}$. Similarly,

$$\tilde{I}_\pi(\pi(f > t)) \geq \tilde{I}_\pi(\pi(f > u)), \quad v < t \leq u \leq 1,$$

and therefore

$$\tilde{I}_\pi(\pi(f > t)) \geq \lim_{u \to 1^-} \tilde{I}_\pi(\pi(f > u)) = \tilde{I}_\pi\left(\lim_{u \to 1^-} \pi(f > u)\right) \geq \tilde{I}_\pi(\pi(S_2)),$$

taking limits in u and applying continuity of $\tilde{I}_\pi$.

We thus decompose

$$\int_0^1 \tilde{I}_\pi(\pi(f > t)) \, \mathrm{d}t = \int_0^v \tilde{I}_\pi(\pi(f > t)) \, \mathrm{d}t + \int_v^1 \tilde{I}_\pi(\pi(f > t)) \, \mathrm{d}t$$

$$\geq \int_0^v \tilde{I}_\pi(\pi(S_1^{\complement})) \, \mathrm{d}t + \int_v^1 \tilde{I}_\pi((\pi(S_2))) \, \mathrm{d}t$$

$$= v \cdot \tilde{I}_\pi(\pi(S_1^{\complement})) + (1 - v) \cdot \tilde{I}_\pi(\pi(S_2))$$

$$\geq \min\{\tilde{I}_\pi(\pi(S_1^{\complement})), \tilde{I}_\pi(\pi(S_2))\}$$

$$= \min\{\tilde{I}_\pi(\pi(S_1)), \tilde{I}_\pi(\pi(S_2))\},$$

from which we may conclude. $\quad\square$

We now show that given a three-set isoperimetric inequality with a monotone increasing F, together with the close coupling assumption on the Markov kernel, one may deduce a lower bound on the conductance of any set for that Markov kernel. The proof, housed in Appendix B, follows closely that of Dwivedi et al. (2019), Lemma 6, which itself is based on several earlier works.

LEMMA 15. *Suppose that π satisfies a three-set isoperimetric inequality with metric d and function F monotone increasing on $(0, \frac{1}{2}]$. Let P be a $(\mathsf{d}, \delta, \varepsilon)$-close coupling, π-invariant Markov kernel. Then for any $A \in \mathcal{E}$ with $0 < \pi(A) \leq \frac{1}{2}$,*

$$\frac{(\pi \otimes P)(A \times A^{\complement})}{\pi(A)} \geq \sup_{\theta \in (0, 1)} \min\left\{\frac{1}{2} \cdot (1 - \theta) \cdot \varepsilon, \frac{1}{4} \cdot \varepsilon \cdot \delta \cdot \theta \cdot (F/\text{id})(\theta \cdot \pi(A))\right\},$$

where F/id is the function $x \mapsto F(x)/x$.

Since $\tilde{I}_\pi / \mathrm{id}$ is decreasing for a concave, regular isoperimetric minorant $\tilde{I}_\pi$, we obtain the following bounds on the conductance profile by combining Lemma 14, Lemma 15 and Definition 3. We emphasize that concavity of $\tilde{I}_\pi$ is crucial for obtaining nonzero lower bounds. Considering functions on $(0, \frac{1}{2}]$ of the form $p \mapsto c \cdot p^k$ the critical case is $k = 1$ and any $k > 1$ implies only a conductance profile lower bound of 0; our examples give isoperimetric minorants of the form $p \mapsto c \cdot p \cdot \log(\frac{1}{p})^r$ on $(0, \frac{1}{2}]$ for $r \in [0, 1]$. Additionally, it is well known that the uniform measure on the sphere $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ satisfies $I(p) \asymp p^{\frac{d-1}{d}}$; see for example, Section 9 of Bobkov and Houdré (1997a).

COROLLARY 16. *Suppose that $\tilde{I}_\pi$ is a regular, concave isoperimetric minorant of π w.r.t. the metric* d. *Let P be a* $(\mathsf{d}, \delta, \varepsilon)$*-close coupling, π-invariant Markov kernel. Then for any $v \in (0, \frac{1}{2}]$,*

$$\Phi_P(v) \geq \sup_{\theta \in [0,1]} \min\left\{ \frac{1}{2} \cdot (1 - \theta) \cdot \varepsilon, \frac{1}{4} \cdot \varepsilon \cdot \delta \cdot \theta \cdot (\tilde{I}_\pi / \mathrm{id})(\theta \cdot v) \right\}$$

$$\geq \frac{1}{4} \cdot \varepsilon \cdot \min\left\{ 1, \frac{1}{2} \cdot \delta \cdot \frac{\tilde{I}_\pi(\frac{1}{2} \cdot v)}{\frac{1}{2} \cdot v} \right\}.$$

REMARK 17. For obtaining tighter bounds on the conductance when $\delta \cdot \tilde{I}_\pi(\frac{1}{2})$ is sufficiently small, one can take $\theta = 1 - \delta \cdot \tilde{I}_\pi(\frac{1}{2})$ to see that

$$\Phi_P^* \geq \frac{1}{2} \cdot \varepsilon \cdot \delta \cdot \tilde{I}_\pi\left(\frac{1}{2} \cdot \left(1 - \delta \cdot \tilde{I}_\pi\left(\frac{1}{2} \right) \right) \right) \to \frac{1}{2} \cdot \varepsilon \cdot \delta \cdot \tilde{I}_\pi\left(\frac{1}{2} \right),$$

as $\delta \cdot \tilde{I}_\pi(\frac{1}{2}) \to 0^+$, which is a slight improvement on the nonasymptotic bound $\Phi_P^* = \Phi_P(\frac{1}{2}) \geq \frac{1}{2} \cdot \varepsilon \cdot \delta \cdot \tilde{I}_\pi(\frac{1}{4})$.

The following theorem is the culmination of this and the preceding section; see Figure 1. The proof is in Appendix B. For the mixing time bound, v_*^{-1} will typically increase quite rapidly as δ decreases. The bounds suggest three-stage behaviour when $v_* < \frac{1}{2}$, recalling that $u_0 = \chi^2(\mu, \pi)$. The first term is active when $u_0 > 4 \cdot v_*^{-1} > 8$, either because u_0 is extremely large or δ is large, while the second term is active when $u_0 > 8$ and the third term is active when $u_0 > \varepsilon_{\mathrm{Mix}}$. Of course, if $u_0 \leq \varepsilon_{\mathrm{Mix}}$, then one may take $n = 0$.

THEOREM 18. *Let π have a regular, concave isoperimetric minorant $\tilde{I}_\pi$ w.r.t. the metric* d, *and P be a* $(\mathsf{d}, \delta, \varepsilon)$*-close coupling, π-reversible, positive Markov kernel. Then*

1. *For $v \in (0, \frac{1}{2}]$, $\Phi_P(v) \geq \frac{1}{4} \cdot \varepsilon \cdot \min\{1, \frac{1}{2} \cdot \delta \cdot (\tilde{I}_\pi / \mathrm{id})(\frac{1}{2} \cdot v)\}$,*
2. *$\Phi_P^* \geq \frac{1}{4} \cdot \varepsilon \cdot \min\{1, 2 \cdot \delta \cdot \tilde{I}_\pi(\frac{1}{4})\}$,*
3. *$\gamma_P \geq \frac{1}{2} \cdot [\Phi_P^*]^2 \geq 2^{-5} \cdot \varepsilon^2 \cdot \min\{1, 4 \cdot \delta^2 \cdot \tilde{I}_\pi(\frac{1}{4})^2\}$.*

Furthermore, let $\varepsilon_{\mathrm{Mix}} \in (0, 8)$, $\mu \ll \pi$ be a probability measure and $u_0 := \chi^2(\mu, \pi)$. To ensure that $\chi^2(\mu P^n, \pi) \leq \varepsilon_{\mathrm{Mix}}$, it suffices to take

$$n \geq 2 + 2^6 \cdot \varepsilon^{-2} \cdot \max\left\{ \log\left(\frac{u_0}{4 \cdot v_*^{-1}} \right), 0 \right\}$$

$$+ 2^8 \cdot \varepsilon^{-2} \cdot \delta^{-2} \cdot \int_{\max\{\min\{2 \cdot u_0^{-1}, 1/4\}, v_*/2\}}^{1/4} \frac{\xi}{\tilde{I}_\pi(\xi)^2} \, \mathrm{d}\xi$$

$$+ 2^4 \cdot \max\left\{ 1, 2^{-2} \cdot \delta^{-2} \cdot \tilde{I}_\pi\left(\frac{1}{4} \right)^{-2} \right\} \cdot \varepsilon^{-2} \cdot \log\left(\max\left\{ \frac{\min\{u_0, 8\}}{\varepsilon_{\mathrm{Mix}}}, 1 \right\} \right),$$

where

$$v_* := \min\left\{\frac{1}{2}, \max\left\{0, \sup\left\{v > 0 : 1 \le \frac{1}{2}\cdot\delta\cdot\frac{\tilde{I}_\pi(\frac{1}{2}\cdot v)}{\frac{1}{2}\cdot v}\right\}\right\}\right\}. \tag{8}$$

We conclude our general results with a result that can be used to establish the close-coupling condition for Metropolis kernels, which we will use in the sequel to analyze both the RWM and pCN Markov chains.

LEMMA 19. *Let Q be a ν-reversible Markov kernel where $\nu \gg \pi$ is a σ-finite measure, P be the π-reversible Metropolis kernel with proposal Q and $\alpha_0 := \inf_{x \in \mathsf{E}} \alpha(x)$. Then*

$$\|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \le \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} + 1 - \alpha_0, \quad x, y \in \mathsf{E}.$$

PROOF. Let $x, y \in \mathsf{E}$ be arbitrary. We write $\varpi = \frac{\mathrm{d}\pi}{\mathrm{d}\nu}$. We construct a specific coupling of (X', Y') such that $X' \sim P(x, \cdot)$ and $Y' \sim P(y, \cdot)$. Without loss of generality, we may assume that $\varpi(x) \ge \varpi(y)$. Let (W_x, W_y) be drawn from a maximal coupling of $Q(x, \cdot)$ and $Q(y, \cdot)$, so that

$$\mathbb{P}(W_x = W_y) = 1 - \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}}.$$

With $\mathcal{U} \sim \mathrm{Unif}(0, 1)$, we complete the specification of the distribution of (X', Y') via

$$X' | \{W_x = w_x, W_y = w_y, \mathcal{U} = u\} = \begin{cases} w_x, & u \le \varpi(w_x)/\varpi(x), \\ x, & u > \varpi(w_x)/\varpi(x), \end{cases}$$

and

$$Y' | \{W_x = w_x, W_y = w_y, \mathcal{U} = u\} = \begin{cases} w_y, & u \le \varpi(w_y)/\varpi(y), \\ y, & u > \varpi(w_y)/\varpi(y). \end{cases}$$

On the event $\{W_x = W_y\} \cap \{X' = W_x\}$, we have $X' = Y' = W_x$ since $\varpi(x) \ge \varpi(y)$. Hence,

$$\begin{aligned} \mathbb{P}(X' = Y') &\ge \mathbb{P}(W_x = W_y, X' = W_x) \\ &\ge \mathbb{P}(W_x = W_y) + \mathbb{P}(X' = W_x) - 1 \\ &= (1 - \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}}) + \mathbb{P}(X' = W_x) - 1 \\ &\ge (1 - \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}}) + \alpha_0 - 1 \\ &= \alpha_0 - \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}}. \end{aligned}$$

We conclude by the coupling inequality: $\|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \le \mathbb{P}(X' \ne Y')$. $\square$

3.2. *Examples of isoperimetric profiles.* In this section, we provide some concrete examples of probability measures for which the isoperimetric profile admits regular, concave and tractable isoperimetric profiles or minorants. We first describe explicitly the profiles associated to specific measures, and then describe some general results which hold over well-behaved families of measures. Most results in this subsection are not originally ours, and are included in order to provide context on the breadth of applicability of our main results.

3.2.1. *Specific examples.*

EXAMPLE 20.   Let φ_γ, Φ_γ denote the PDF and CDF respectively of the one-dimensional standard Gaussian measure. For any $d \in \mathbb{N}$, the isoperimetric profile of $\pi = \mathcal{N}(0, I_d)$ w.r.t. $|\cdot|$ satisfies $I_\pi(p) = (\varphi_\gamma \circ \Phi_\gamma^{-1})(p)$ (Borell (1975), Sudakov and Tsirel'son (1978)), which satisfies

$$\lim_{p \to 0^+} \frac{I_\pi(p)}{p \cdot (\log \frac{1}{p})^{1/2}} = \sqrt{2},$$

see, for example, Barthe and Maurey (2000). Observing that I_π is concave and regular, one can deduce

$$I_\pi(p) \geq 2 \cdot I_\pi\left(\frac{1}{2}\right) \cdot \min\{p, 1 - p\} = \left(\frac{2}{\pi}\right)^{1/2} \cdot \min\{p, 1 - p\}.$$

EXAMPLE 21.   For the Laplace measure $\pi(\mathrm{d}x) \propto \exp(-|x|)\mathrm{d}x$ in one dimension, the isoperimetric profile w.r.t. $|\cdot|$ is given by $I_\pi(p) = \min\{p, 1 - p\}$ (see, e.g., Bobkov (1999)).

EXAMPLE 22.   For the Subbotin measure $\pi(\mathrm{d}x) \propto \exp(-|x|^\alpha)\mathrm{d}x$ in one dimension, with $\alpha \in (1, 2)$, it holds that the isoperimetric profile w.r.t. $|\cdot|$ can be bounded from below for $p \in (0, \frac{1}{2}]$ as $I_\pi(p) \geq K(\alpha) \cdot p \cdot (\log \frac{1}{p})^{1-1/\alpha}$ for some $K(\alpha) > 0$ (see, e.g., Barthe, Cattiaux and Roberto (2006), Latała and Oleszkiewicz (2000)).

3.2.2. *Functional inequalities.*

Many analyses of MCMC algorithms restrict to considering strongly log-concave targets π, that is, U is strongly convex, to obtain quantitative bounds. This means that for potentials with inhomogeneous local convexity properties, but good global isoperimetric properties (e.g., strongly convex in the tails, weakly convex in the center of the space), complexity bounds can be somewhat pessimistic. In this subsection, we give some examples of how to estimate isoperimetric profiles when π is only log-concave, given additional quantitative information about functional inequalities which they satisfy.

EXAMPLE 23.   For any log-concave π, there exists $\gamma_\pi > 0$ such that

$$\forall f \text{ locally Lipschitz}, \quad \pi(|\nabla f|^2) \geq \gamma_\pi \cdot \mathrm{Var}_\pi(f),$$

that is, π satisfies a Poincaré inequality (see, e.g., Bakry, Gentil and Ledoux (2014), Theorem 4.6.3). It is also known that Poincaré inequalities can be translated into L^1-Poincaré inequalities, with explicit control of the constants (Cattiaux and Guillin (2020)). Finally, L^1-Poincaré inequalities are equivalent to isoperimetric inequalities with respect to $|\cdot|$ of the form $I_\pi(p) \geq c \cdot \min\{p, 1 - p\}$, with the same constant (see, e.g., Kolesnikov (2007)). Combining these facts, if π is log-concave, one can deduce that

$$I_\pi(p) \geq \frac{1}{6} \cdot \gamma_\pi^{1/2} \cdot \min\{p, 1 - p\}.$$

EXAMPLE 24.   Consider log-concave π satisfying a logarithmic Sobolev inequality with constant λ_π, namely,

$$\forall f \text{ locally Lipschitz}, \quad \pi(|\nabla f|^2) \geq \lambda_\pi \cdot \mathrm{Ent}_\pi(f^2),$$

where for positive f, $\mathrm{Ent}_\pi(f) := \pi(f \cdot \log f) - \pi(f) \cdot \log \pi(f)$. A result of Ledoux (2011) then gives the following bound for the isoperimetric profile w.r.t. $|\cdot|$,

$$I_\pi(p) \geq \frac{1}{34} \cdot \lambda_\pi^{1/2} \cdot p \cdot \left(\log \frac{1}{p}\right)^{1/2}, \quad p \in \left(0, \frac{1}{2}\right].$$

EXAMPLE 25. For log-concave π satisfying a q-log-Sobolev inequality (see Bobkov and Zegarlinski (2005)), that is,

$$\forall f \text{ locally Lipschitz}, \quad D \cdot \text{Ent}_\pi\left(|f|^q\right) \leq \pi\left(|\nabla f|_q^q\right),$$

a result of Milman (2009b) establishes that for the isoperimetric profile w.r.t. $|\cdot|$,

$$I_\pi(p) \geq c_q \cdot D \cdot p \cdot \left(\log \frac{1}{p}\right)^{1/q} \quad \text{for } p \in (0, 1/2],$$

where $c_q > 0$ is universal. For $q = 2$, this entails the standard log-Sobolev inequality; for $q \in [1, 2)$, the assumption becomes stronger and corresponds to potentials which have tail behaviour like $U(x) \sim |x|^{q_*}$, where $q^{-1} + q_*^{-1} = 1$. For $q = 1$, the assumption is yet stronger and corresponds intuitively to potentials which have tail behaviour like $U(x) \sim \exp(c \cdot |x|)$ for some $c > 0$.

3.2.3. *Transfer principles.* Another practical aspect of working with isoperimetric profiles is that they are often preserved under suitably regular perturbations, some of which we detail here. These transfer principles can be used to accommodate potentials that are not convex. The first of these concerns the transfer of isoperimetric properties under Lipschitz transport; related statements are made in Barthe (2001).

LEMMA 26. *For $i = 1, 2$, let μ_i be a probability measure on the metric space $(\mathsf{E}_i, \mathsf{d}_i)$. Suppose that these measures are related through transport as*

$$\mu_2 = T_\# \mu_1,$$

where $T : \mathsf{E}_1 \to \mathsf{E}_2$ is a Lipschitz bijection. Then, for $\tilde{I}$ any isoperimetric minorant of μ_1 w.r.t. d_1, it holds that $|T|_{\text{Lip}}^{-1} \cdot \tilde{I}$ is an isoperimetric minorant of μ_2 w.r.t. the metric d_2. In particular, if T is also an isometry, then μ_1 and μ_2 have identical isoperimetric profiles w.r.t. their respective metrics.

PROOF. Let $A \subseteq \mathsf{E}_2$ be measurable and $A_r = \{x \in \mathsf{E}_2 : \mathsf{d}_2(x, A) \leq r\}$. Write $B := T^{-1}(A)$, and compute that

$$\begin{aligned} A_r &= \{x \in \mathsf{E}_2 : \mathsf{d}_2(x, A) \leq r\} \\ &= \{x \in \mathsf{E}_2 : \mathsf{d}_2(T(T^{-1}(x)), T(B)) \leq r\} \\ &\supseteq \{T(y) \in \mathsf{E}_2 : \mathsf{d}_1(y, B) \leq |T|_{\text{Lip}}^{-1} \cdot r\} \\ &= T(B_{|T|_{\text{Lip}}^{-1} \cdot r}), \end{aligned}$$

where $B_s := \{y \in \mathsf{E}_1 : \mathsf{d}_1(y, B) \leq s\}$. Then

$$\frac{\mu_2(A_r) - \mu_2(A)}{r} \geq \frac{\mu_2(T(B_{|T|_{\text{Lip}}^{-1} \cdot r})) - \mu_2(T(B))}{r} = \frac{\mu_1(B_{|T|_{\text{Lip}}^{-1} \cdot r}) - \mu_1(B)}{r},$$

whence it follows that

$$\begin{aligned} \mu_{2, \mathsf{d}_2}^+(A) &= \lim_{r \to 0^+} \inf \frac{\mu_2(A_r) - \mu_2(A)}{r} \\ &\geq \lim_{r \to 0^+} \inf \frac{\mu_1(B_{|T|_{\text{Lip}}^{-1} \cdot r}) - \mu_1(B)}{r} \\ &\geq \lim_{s \to 0^+} \inf \frac{\mu_1(B_s) - \mu_1(B)}{|T|_{\text{Lip}} \cdot s} \\ &= |T|_{\text{Lip}}^{-1} \cdot \mu_{1, \mathsf{d}_1}^+(B). \end{aligned}$$

By definition of isoperimetric minorants and μ_2, it holds that $\mu^+_{1,d_1}(B) \geq \tilde{I}(\mu_1(B)) = \tilde{I}(\mu_2(A))$, so the first result follows. For the second result, note that bijective isometries from E_1 to E_2 satisfy $|T|_{\text{Lip}} = |T^{-1}|_{\text{Lip}} = 1$ (noting that these Lipschitz norms are technically defined on different spaces), and that the isoperimetric profile is always an isoperimetric minorant. Applying the first result in both directions allows us to conclude. $\square$

The following result will be used frequently in the sequel.

LEMMA 27. *For an m-strongly convex potential U, with $\mathsf{d} = |\cdot|$, we have*

$$I_\pi(p) \geq m^{1/2} \cdot (\varphi_\gamma \circ \Phi_\gamma^{-1})(p) \geq C_\ell \cdot m^{1/2} \cdot \min\{p, 1-p\} \cdot \left(\log \frac{1}{\min\{p, 1-p\}}\right)^{1/2},$$

where $C_\ell := (\frac{2}{\pi \cdot \log(2)})^{1/2} \geq 0.958357$. Moreover, if we let $\tilde{I}_\pi(p) := m^{1/2} \cdot (\varphi_\gamma \circ \Phi_\gamma^{-1})(p)$, which is regular and concave, then

$$\tilde{I}_\pi\left(\frac{1}{4}\right) = m^{1/2} \cdot C_\gamma,$$

where $C_\gamma := (\varphi \circ \Phi^{-1})(\frac{1}{4}) \geq 0.3177765$.

PROOF. By the contraction principle of Caffarelli (2000), it is known that if one takes $\mu = \mathcal{N}(0, m^{-1} \cdot I_d)$, then there exists a 1-Lipschitz mapping T which pushes μ onto π, and hence a $m^{-1/2}$-Lipschitz mapping which pushes $\mathcal{N}(0, I_d)$ onto π. By Lemma 26, the first inequality follows. For the second inequality, observe that the function

$$g(p) = \frac{(\varphi_\gamma \circ \Phi_\gamma^{-1})(p)}{p \cdot (\log(\frac{1}{p}))^{1/2}},$$

is minimized on $(0, \frac{1}{2}]$ at $\frac{1}{2}$; see Figure 2, and defining $C_\ell = g(\frac{1}{2})$ gives the second inequality. The final claim is a direct computation. $\square$

Isoperimetric profiles also transfer under bounded changes of measure in the style of, for example, Holley and Strook (1987). We provide the following result that demonstrates this.

PROPOSITION 28. *Let μ be a probability measure on $\mathbb{R}^d$ with positive density with respect to Lebesgue. Suppose $\tilde{I}_\mu$ is a regular isoperimetric minorant of μ w.r.t. a metric d, and ν is a probability measure equivalent to μ with $\frac{d\nu}{d\mu} \in [c_1, c_2]$ for some $0 < c_1 \leq c_2 < \infty$. Then $p \mapsto c_1 \cdot I_\mu(c_2^{-1} \cdot \min\{p, 1-p\})$ for $p \in (0, 1)$ is a regular isoperimetric minorant of ν w.r.t. d.*

PROOF. We first deduce that μ satisfies a 3-set isoperimetric inequality with function $F_\mu(t) = \tilde{I}_\mu(t)$ for $t \in (0, \frac{1}{2}]$ by Lemma 14. Then, for any measurable partition $\mathsf{E} = S_1 \sqcup S_2 \sqcup S_3$ we have

$$\nu(S_3) \geq c_1 \mu(S_3)$$
$$\geq c_1 \mathsf{d}(S_1, S_2) F_\mu(\min\{\mu(S_1), \mu(S_2)\})$$
$$\geq c_1 \mathsf{d}(S_1, S_2) F_\mu(c_2^{-1} \min\{\nu(S_1), \nu(S_2)\}),$$

so ν satisfies a 3-set isoperimetric inequality with metric d and function $F_\nu(t) = c_1 F_\mu(c_2^{-1} t)$. We conclude by Lemma 14. $\square$
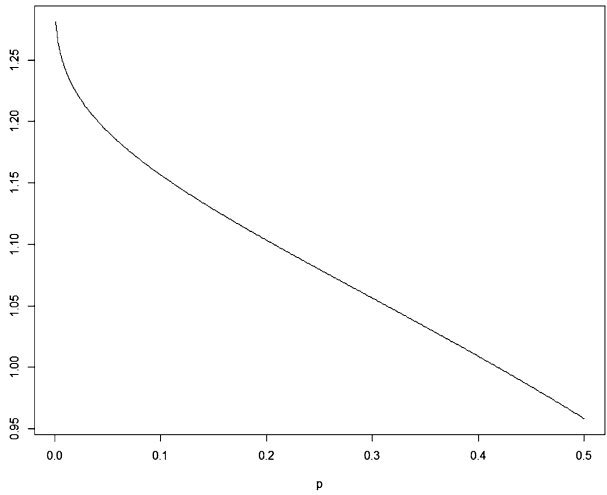
FIG. 2. *A plot of* $p \mapsto \frac{(\varphi_\gamma \circ \Phi_\gamma^{-1})(p)}{p \cdot [\log(\frac{1}{p})]^{1/2}}$.

REMARK 29. Proposition 28 provides an improvement of Belloni and Chernozhukov (2009), Lemma 2, who (essentially) work in the setting of probability measures whose density can be expressed as

$$\pi(x) = \exp\left(-\frac{1}{2}|x|^2 - V(x) - \xi(x)\right),$$

where V is convex, and ξ is uniformly bounded above and below. They deduce a "pseudo"-three-set isoperimetric inequality of the form

$$\pi(S_3) \geq c \cdot \exp(-\mathrm{Osc}(\xi)) \cdot t \cdot \exp\left(-\frac{1}{2} \cdot t^2\right) \cdot \min\{\pi(S_1), \pi(S_2)\},$$

where $\mathrm{Osc}(\xi) := \mathrm{ess\,sup}\,\xi - \mathrm{ess\,inf}\,\xi$, $t = \mathsf{d}(S_1, S_2)$, $\mathsf{d} = |\cdot|$, and $c > 0$ is an explicit constant.

In fact, using Caffarelli's result, one can see that $\exp(-\frac{1}{2}|x|^2 - V(x))\,dx$ will admit $\varphi_\gamma \circ \Phi_\gamma^{-1}$ as an isoperimetric minorant, and so combining this with Proposition 28, one obtains that π satisfies the inequality

$$\pi(S_3) \geq \exp(-\mathrm{Osc}(\xi)) \cdot \mathsf{d}(S_1, S_2) \cdot (\varphi_\gamma \circ \Phi_\gamma^{-1})(\min\{\pi(S_1), \pi(S_2)\}),$$

which relates to a result of Bobkov (2010) (which is stronger, but only valid in dimension 1). Combining this observation with the spectral profile approach and the other calculations of Belloni and Chernozhukov (2009), it seems likely that one could improve the dimension-dependence of their results. We do not pursue this claim further in this work.

EXAMPLE 30. It is known that under suitable convexity assumptions that isoperimetric profiles "almost" tensorize, that is, that the isoperimetric profile of $\pi^{\otimes n}$ satisfies $\inf_{n \geq 1} I_{\pi^{\otimes n}} \geq c \cdot I_\pi$ for some constant $c > 0$. In particular, the isoperimetric profile of product measures can be lower-bounded independently of dimension. We refer the reader to Bobkov and Houdré (1997b), Roberto (2010) for details.

3.3. *From spectral profile to mixing times*: *Examples*. In this section, we describe how to combine an isoperimetric profile for π with the close coupling condition assumption on P to estimate mixing times for the chain. Write $u_0 = \chi^2(\mu, \pi)$ for the initial χ^2-divergence.

In all of our examples, our isoperimetric minorants take the form $\tilde{I}_\pi(p) = c \cdot p \cdot \log(\frac{1}{p})^r$ for $p \in (0, \frac{1}{2}]$, with $r \in [0, 1]$. We briefly recap how the r parameter maps onto simple assumptions, before providing explicit calculations:

- $r = 0$ is "exponential-type" isoperimetry, which holds for any log-concave measure and corresponds roughly to potentials which have a tail growth of order or faster than $|x|$.
- $r = \frac{1}{2}$ is "Gaussian-type" isoperimetry, which holds for any log-concave measure with sub-Gaussian tails.
- $r \in (0, \frac{1}{2})$ corresponds to "intermediate" isoperimetry, and roughly corresponds to potentials which have tail growth $U(x) \sim |x|^{\frac{1}{1-r}} \in (|x|^1, |x|^2)$.
- $r \in (\frac{1}{2}, 1]$ corresponds to "light-tailed" isoperimetry, and roughly corresponds to potentials which have tail growth $U(x) \sim |x|^{\frac{1}{1-r}} \gg |x|^2$, with appropriate modifications for the case $r = 1$ as in Example 25.

Now, to compute: recall by Theorem 18 that in order to ensure that $\chi^2(\mu P^n, \pi) \le \varepsilon_{\text{Mix}} \le 8$, it suffices to take

$$n \ge 2 + 2^6 \cdot \varepsilon^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_*^{-1}}\right), 0\right\}$$

$$+ 2^8 \cdot \varepsilon^{-2} \cdot \delta^{-2} \cdot \int_{\max\{\min\{2 \cdot u_0^{-1}, 1/4\}, v_*/2\}}^{1/4} \frac{\xi}{\tilde{I}_\pi(\xi)^2} \, d\xi.$$

$$+ 2^4 \cdot \max\left\{1, 2^{-2} \cdot \delta^{-2} \cdot \tilde{I}_\pi\left(\frac{1}{4}\right)^{-2}\right\} \cdot \varepsilon^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right).$$

Assuming for now that $u_0 \ge 8$ is intermediate, and $v_* \le 4 \cdot u_0^{-1}$ is small, with v_* given in (8), we focus on the value of the middle integral,

$$\int_{2 \cdot u_0^{-1}}^{1/4} \frac{\xi}{\tilde{I}_\pi(\xi)^2} \, d\xi = c^{-2} \cdot \int_{2 \cdot u_0^{-1}}^{1/4} \frac{1}{\xi \cdot \log(\frac{1}{\xi})^{2r}} \, d\xi = c^{-2} \cdot \int_{\log(4)}^{\log(u_0/2)} \frac{du}{u^{2r}}.$$

There is now a trichotomy of behaviours based upon the relative positions of r and $\frac{1}{2}$ (recalling that in practice, the parameter ε is typically of constant order):

- If $r \in (\frac{1}{2}, 1]$, then the inner integral evaluates to $\frac{1}{2r-1} \cdot (\log(4)^{-(2r-1)} - \log(u_0/2)^{-(2r-1)}) \in \mathcal{O}(1)$. The total mixing time then scales roughly like

$$\max\left\{\log\left(\frac{u_0}{4 \cdot v_*^{-1}}\right), 0\right\} + \delta^{-2} \cdot \left\{\mathbf{1}_{u_0 > 8} + \log\frac{1}{\varepsilon_{\text{Mix}}}\right\}.$$

- If $r = \frac{1}{2}$, then the inner integral evaluates to $\log(\frac{\log(u_0/2)}{\log(4)}) \in \mathcal{O}(\log \log u_0)$. The total mixing time then scales roughly like

$$\max\left\{\log\left(\frac{u_0}{4 \cdot v_*^{-1}}\right), 0\right\} + \delta^{-2} \cdot \left\{\log \min\{\log u_0, v_*^{-1}\} \cdot \mathbf{1}_{u_0 > 8} + \log\frac{1}{\varepsilon_{\text{Mix}}}\right\}.$$

- If $r \in [0, \frac{1}{2})$, then the inner integral evaluates to $\frac{1}{1-2r} \cdot (\log(u_0/2)^{1-2r} - \log(4)^{1-2r}) \in \mathcal{O}(\log(u_0)^{1-2r})$. The total mixing time then scales roughly like

$$\max\left\{\log\left(\frac{u_0}{4 \cdot v_*^{-1}}\right), 0\right\} + \delta^{-2} \cdot \left\{(\min\{\log u_0, v_*^{-1}\})^{1-2r} \cdot \mathbf{1}_{u_0 > 8} + \log\frac{1}{\varepsilon_{\text{Mix}}}\right\}.$$

In particular, when $r = 0$, one obtains the same $\mathcal{O}(\log(u_0))$ dependence implied by a standard spectral gap approach, that is, the spectral profile provides no strict benefit. Indeed for

δ sufficiently small one checks that $v_* = 0$, and so the first term vanishes entirely. Moreover, the second and third terms are qualitatively identical, and thus there is no "three-phase" behaviour, but instead only *one* phase.

Recalling that v_*^{-1} grows rapidly as δ decreases we see that the effect of δ^{-2} is amplified for large u_0 or v_*^{-1} for $r \le 1/2$, roughly corresponding to distributions with tails heavier than Gaussians, and that this is not the case for lighter tailed distributions that is, $r > 1/2$.

4. Spectral gap of RWM in high dimensions. In Sections 4–5, we denote by P the RWM kernel defined by (1) with Q the Gaussian proposal kernel defined for a fixed but arbitrary $\sigma > 0$ by

$$Q(x, A) = \int \mathbf{1}_A(x + \sigma \cdot z) \mathcal{N}(dz; 0, I_d), \quad x \in \mathsf{E}, A \in \mathscr{E}.$$

In Appendix C we discuss how our analysis can be generalized to other proposal kernels with independent noise increments for each of the d components. Since Q is reversible w.r.t. the Lebesgue measure on $\mathbb{R}^d$, we have $\varpi = \pi \propto \exp(-U)$, following Assumption 9. It is standard to deduce by Baxendale (2005), Lemma 3.1, that P is a positive Markov kernel for this particular Q. We note the following useful expression

(9) $$\alpha(x) = \int \mathcal{N}(dz; 0, I_d) \cdot \min\{1, \exp(-(U(x + \sigma \cdot z) - U(x)))\},$$

and we also denote $\alpha_0 = \inf_{x \in \mathsf{E}} \alpha(x)$.

For the purposes of obtaining explicit bounds and matching negative results with dimension, we impose the following further assumption about π, noting that Assumption 9 is already in force. As will be discussed in Section 4.4, both m-strong convexity and L-smoothness can be weakened to obtain explicit bounds on the spectral gap.

ASSUMPTION 31.   For some $0 < m \le L$, U is m-strongly convex and L-smooth:

$$\frac{m}{2} \cdot |h|^2 \le U(x + h) - U(x) - \langle \nabla U(x), h \rangle \le \frac{L}{2} \cdot |h|^2, \quad x, h \in \mathsf{E}.$$

We write $\kappa := L/m$ for the condition number of the target measure.

EXAMPLE 32.   Let $\pi = \mathcal{N}(0, \sigma_0^2 \cdot I_d)$, so that $U(x) = \frac{1}{2 \cdot \sigma_0^2} \cdot |x|^2$. Then

$$U(x + h) - U(x) - \langle \nabla U(x), h \rangle = \frac{1}{2 \cdot \sigma_0^2} \cdot |h|^2, \quad x, h \in \mathsf{E},$$

so U is m-strongly convex and L-smooth with $L = m = 1/\sigma_0^2$ and $\kappa = 1$.

Another natural class of examples with strongly convex and smooth potentials comes from considering Bayesian posterior measures for which the prior is normal, and the log-likelihood is concave with bounded Hessian.

EXAMPLE 33.   Consider the task of Bayesian logistic regression, taking as prior $\pi_0 = \mathcal{N}(0, \sigma_0^2 \cdot I_d)$, and observing covariate-response pairs $\{(a_i, y_i)\}_{i=1}^N \subset \mathbb{R}^d \times \{0, 1\}$. The potential corresponding to the posterior measure is then given up to an additive constant by

$$U(x) = \frac{1}{2 \cdot \sigma_0^2} \cdot |x|^2 + \sum_{i=1}^N \{\log(1 + \exp(-\langle a_i, x \rangle)) - y_i \cdot \langle a_i, x \rangle\}.$$

Writing A for the $N \times d$ matrix with rows given by the $\{a_i\}$, one can check that U is m-strongly convex and L-smooth with $m \ge \frac{1}{\sigma_0^2}$ and $L \le \frac{1}{\sigma_0^2} + \frac{1}{4} \cdot \lambda_{\text{Max}}(AA^\top)$, λ_{Max} denoting the largest eigenvalue of a symmetric matrix. This gives $\kappa \le 1 + \frac{1}{4} \cdot \sigma_0^2 \cdot \lambda_{\text{Max}}(AA^\top)$.

4.1. *Lower bound on the conductance and spectral gap.* The following is a general result, depending *only* on macroscopic properties of the target measure: if the measure has good isoperimetry, and the acceptance rate is lower bounded, then the chain will equilibrate at a rate commensurate with these properties (roughly in accordance with the convergence of the overdamped Langevin diffusion with the same target measure), attenuated only by the choice of σ.

THEOREM 34. *Let π admit a regular, concave isoperimetric minorant $\tilde{I}_\pi$ w.r.t. $|\cdot|$. Then*

$$\Phi_P^* \geq 2^{-3} \cdot \alpha_0 \cdot \min\left\{1, 2 \cdot \alpha_0 \cdot \sigma \cdot \tilde{I}_\pi\left(\frac{1}{4}\right)\right\},$$

and

$$\gamma_P \geq 2^{-7} \cdot \alpha_0^2 \cdot \min\left\{1, 4 \cdot \alpha_0^2 \cdot \sigma^2 \cdot \tilde{I}_\pi\left(\frac{1}{4}\right)^2\right\}.$$

PROOF. By Lemma 38, P is $(|\cdot|, \alpha_0 \cdot \sigma, \frac{1}{2} \cdot \alpha_0)$-close-coupling, and we conclude from Theorem 18. □

COROLLARY 35. *Under Assumption 31, let $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-1/2}$ for any $\varsigma > 0$, and C_γ as in Lemma 27. Then*

$$\Phi_P^* \geq 2^{-4} \cdot C_\gamma \cdot \varsigma \cdot \exp(-\varsigma^2) \cdot d^{-1/2} \cdot \left(\frac{m}{L}\right)^{1/2},$$

and

$$\gamma_P \geq 2^{-9} \cdot C_\gamma^2 \cdot \varsigma^2 \cdot \exp(-2 \cdot \varsigma^2) \cdot d^{-1} \cdot \frac{m}{L}.$$

PROOF. By Corollary 40, $\alpha_0 \geq \frac{1}{2} \cdot \exp(-\frac{1}{2} \cdot \varsigma^2)$ and we can take $\tilde{I}_\pi(\frac{1}{4}) \geq m^{1/2} \cdot C_\gamma$ by Lemma 27. Since

$$C_\gamma \cdot \exp\left(-\frac{1}{2} \cdot \varsigma^2\right) \cdot \varsigma \cdot L^{-1/2} \cdot d^{-1/2} \cdot m^{1/2} \leq C_\gamma \cdot e^{-1/2} \cdot 1 < 1,$$

for any $d \in \mathbb{N}$ and $\varsigma > 0$ (noting that $m \leq L$), we deduce that

$$\Phi_P^* \geq 2^{-3} \cdot \alpha_0 \cdot \min\left\{1, 2 \cdot \alpha_0 \cdot \sigma \cdot \tilde{I}_\pi\left(\frac{1}{4}\right)\right\}$$

$$\geq 2^{-4} \cdot \exp\left(-\frac{1}{2} \cdot \varsigma^2\right) \cdot \min\left\{1, \exp\left(-\frac{1}{2} \cdot \varsigma^2\right) \cdot \varsigma \cdot L^{-1/2} \cdot d^{-1/2} \cdot m^{1/2} \cdot C_\gamma\right\}$$

$$= 2^{-4} \cdot C_\gamma \cdot \varsigma \cdot \exp(-\varsigma^2) \cdot d^{-1/2} \cdot \left(\frac{m}{L}\right)^{1/2},$$

and the lower bound on γ_P follows similarly. □

REMARK 36. We find that $\Phi_P^* \in \Omega(d^{-1/2})$, and $\gamma_P \in \Omega(d^{-1})$. Fixing ς, we obtain (nonasymptotically) that

$$\Phi_P^* \cdot d^{1/2} \geq 2^{-4} \cdot C_\gamma \cdot \varsigma \cdot \exp(-\varsigma^2) \cdot \left(\frac{m}{L}\right)^{1/2}$$

$$\geq 0.019861 \cdot \varsigma \cdot \exp(-\varsigma^2) \cdot \left(\frac{m}{L}\right)^{1/2}.$$

This lower bound is maximized in ς by taking $\varsigma^2 = \frac{1}{2}$, which yields

$$\Phi_P^* \cdot d^{1/2} \geq 0.008518 \cdot \left(\frac{m}{L}\right)^{1/2}.$$

This particular bound-maximizing value of ς is likely an artifact of the proof technique; optimal scaling results of Roberts and Rosenthal (2001) suggest that $\varsigma \approx 2.38$ is optimal in high dimensions when $\pi = \mathcal{N}(0, I_d)$, although they do not provide a bound on the conductance or spectral gap of the associated Markov kernel. Similarly, taking $\varsigma^2 = \frac{1}{2}$ leads to the following bound for the spectral gap:

$$\gamma_P \cdot d \geq 3.62784 \times 10^{-5} \cdot \frac{m}{L}.$$

The following lemma gives a useful bound on the total variation distance between the proposals; the proof can be found in Appendix B.

LEMMA 37. *If $v > 0$ and $x, y \in \mathsf{E}$ satisfy $|x - y| \leq v \cdot \sigma$,*

$$\|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq \frac{1}{2} \cdot v.$$

The following is a key lemma establishing the close coupling condition for P.

LEMMA 38. *P is $(|\cdot|, \alpha_0 \cdot \sigma, \frac{1}{2} \cdot \alpha_0)$-close-coupling.*

PROOF. Assume $x, y \in \mathsf{E}$ are such that $|x - y| \leq \alpha_0 \cdot \sigma$. Then $\|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq \frac{1}{2} \cdot \alpha_0$ by Lemma 37. We conclude by applying Lemma 19. $\square$

To lower bound the acceptance rate for the RWM kernel, we first prove a result which holds under a general smoothness condition on the potential U, and then obtain the case of L-smoothness as a corollary.

LEMMA 39. *Suppose that for some nonnegative, nondecreasing ψ, the potential U satisfies the smoothness bound*

$$U(x + h) - U(x) - \langle \nabla U(x), h \rangle \leq \psi(|h|), \quad x, h \in \mathsf{E}.$$

1. *Then for any $\sigma \geq 0$,*

$$\alpha_0 \geq \frac{1}{2} \cdot \exp\left(-\int \mathcal{N}(dz; 0, I_d) \cdot \psi(\sigma \cdot |z|)\right).$$

2. *Let $\sigma = \varsigma \cdot d^{-1/2}$ where $\varsigma > 0$ is arbitrary, ψ be twice-differentiable, such that for some $c_0, c_1 > 0$,*

$$(\varsigma \cdot \psi'' - \psi')(t) \leq c_0 \cdot \exp(c_1 \cdot t).$$

Then

$$\alpha_0 \geq \frac{1}{2} \cdot \exp(-\psi(\varsigma) + \mathcal{O}(d^{-1})) \in \Omega(1).$$

PROOF. The proof proceeds by direct calculation. Let $x \in \mathsf{E}$ be arbitrary. Applying the growth bound assumption, it holds that

$$U(x + \sigma \cdot z) - U(x) \le \langle \nabla U(x), \sigma \cdot z \rangle + \psi(\sigma \cdot |z|),$$

and substituting this into (9) gives

$$\alpha(x) \ge \int \mathcal{N}(dz; 0, I_d) \cdot \min\{1, \exp(-\langle \nabla U(x), \sigma \cdot z \rangle - \psi(\sigma \cdot |z|))\}.$$

Applying the inequality $\min\{1, a \cdot b\} \ge \min\{1, a\} \cdot \min\{1, b\}$ establishes that

$$\alpha(x) \ge \int \mathcal{N}(dz; 0, I_d) \cdot \min\{1, \exp(-\langle \nabla U(x), \sigma \cdot z \rangle)\} \cdot \min\{1, \exp(-\psi(\sigma \cdot |z|))\}$$

$$\ge \int \mathcal{N}(dz; 0, I_d) \cdot \exp(-\psi(\sigma \cdot |z|)) \cdot \min\{1, \exp(-\langle \nabla U(x), \sigma \cdot z \rangle)\}$$

$$= \int \mathcal{N}(dz; 0, I_d) \cdot \exp(-\psi(\sigma \cdot |z|)) \cdot \min\{1, \exp(+\langle \nabla U(x), \sigma \cdot z \rangle)\},$$

where the last equality follows from the change of variables $z \mapsto -z$. Averaging these final two expressions and noting that $\min\{1, a\} + \min\{1, a^{-1}\} \ge 1$, it follows that

$$\alpha(x) \ge \frac{1}{2} \cdot \int \mathcal{N}(dz; 0, I_d) \cdot \exp(-\psi(\sigma \cdot |z|)).$$

By the convexity of $\psi \mapsto \exp(-\psi)$, we may apply Jensen's inequality to bound this integral from below, and so the first part follows.

Finally, take $\sigma = \varsigma \cdot d^{-1/2}$, and Taylor expand $s \mapsto \psi(s^{1/2})$ with a second-order remainder term around $s = \varsigma^2$. Applying the hypothesis on the derivatives of ψ then yields that for sufficiently large d, there exists $F(\varsigma) > 0$ such that

$$\int \mathcal{N}(dz; 0, I_d) \cdot \psi(\varsigma \cdot d^{-1/2} \cdot |z|) \le \psi(\varsigma) + F(\varsigma) \cdot d^{-1},$$

and we conclude. $\square$

COROLLARY 40. *Under Assumption* 31, *let* $\sigma = \varsigma \cdot d^{-1/2} \cdot L^{-1/2}$ *where* $\varsigma > 0$ *is arbitrary. Then*

$$\alpha_0 \ge \frac{1}{2} \cdot \exp\left(-\frac{1}{2} \cdot \varsigma^2\right).$$

PROOF. Assumption 31 implies that U satisfies the growth bound

$$U(x + h) - U(x) - \langle \nabla U(x), h \rangle \le \psi(|h|)$$

with $\psi : r \mapsto \frac{1}{2} \cdot L \cdot r^2$. Applying the first part of Lemma 39, we see that

$$\alpha_0 \ge \frac{1}{2} \cdot \exp\left(-\int \mathcal{N}(dz; 0, I_d) \cdot \frac{1}{2} \cdot L \cdot \sigma^2 \cdot |z|^2\right)$$

$$= \frac{1}{2} \cdot \exp\left(-\frac{1}{2} \cdot L \cdot \sigma^2 \cdot d\right)$$

$$= \frac{1}{2} \cdot \exp\left(-\frac{1}{2} \cdot \varsigma^2\right). \qquad \square$$

4.2. *Conductance and spectral gap upper bounds.*   To complement our lower bounds, we can show matching upper bounds with respect to dimension under Assumption 31. This shows that the conductance and spectral gap must decrease at least as fast as $\mathcal{O}(d^{-1/2})$ and $\mathcal{O}(d^{-1})$ respectively, and that these are the slowest polynomial decays possible. Hence, we may infer that in terms of optimizing conductance and spectral gap, d^{-1} is the correct polynomial scaling of σ^2.

We emphasize that the upper bounds are uniform over the class of m-strongly convex and L-smooth potentials, indicating that the dimension-dependence of this particular class of target distributions is well-characterized by the analysis. This is in contrast to bounds which rely only on specific examples exhibiting poor performance, as in minimax complexity analysis: for example, Wu, Schmidler and Chen (2022) show that the optimal scaling of stepsize with dimension in the Metropolis-adjusted Langevin algorithm is not uniform in this class.

4.2.1. *Conductance upper bounds.*

THEOREM 41.   *Under Assumption* 31 *and twice continuous-differentiability of U,*

$$\Phi_P^* \leq \min\{4 \cdot L^{1/2} \cdot \sigma, (1 + m \cdot \sigma^2)^{-d/2}\}.$$

Hence, among polynomial scalings of σ, the scaling $\sigma \sim d^{-1/2}$ is optimal with $\Phi_P^* \sim d^{-1/2}$.

PROOF.   The bounds follow from Propositions 44 and 45. Now let $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-\beta}$. If $\beta \geq 1/2$, then we obtain $\Phi_P^* \in \mathcal{O}(d^{-\beta})$ and this is maximized by taking $\beta = \frac{1}{2}$. Combined with Theorem 34 we may conclude that Φ_P^* decays as $d^{-1/2}$ as $d \to \infty$ when $\beta = \frac{1}{2}$. For $\beta < \frac{1}{2}$, we recall that by Proposition 45, the conductance decays faster than any polynomial in $1/d$, and in particular, faster than $d^{-1/2}$, from which the claim follows.   □

LEMMA 42.   *Under Assumption* 31, *let x_* be the* (*unique*) *minimizer of U. Then*

$$\left(\frac{m}{L}\right)^{d/2} \cdot \mathcal{N}(x; x_*, L^{-1} \cdot I_d) \leq \pi(x) \leq \left(\frac{L}{m}\right)^{d/2} \cdot \mathcal{N}(x; x_*, m^{-1} \cdot I_d).$$

PROOF.   Applying the definitions of m-strong convexity and L-smoothness, one sees that

$$\frac{m}{2} \cdot |x - x_*|^2 \leq U(x) - U(x_*) - \langle \nabla U(x_*), x - x_* \rangle \leq \frac{L}{2} \cdot |x - x_*|^2,$$

and since $\nabla U(x_*) = 0$, this can be simplified to

$$\frac{m}{2} \cdot |x - x_*|^2 \leq U(x) - U(x_*) \leq \frac{L}{2} \cdot |x - x_*|^2.$$

Recalling that $\pi(x) = \frac{1}{Z} \cdot \exp(-U(x))$ for some normalizing constant Z, this implies that

$$\left(\frac{2\pi}{L}\right)^{d/2} \cdot \mathcal{N}(x; x_*, L^{-1} \cdot I_d) \leq \pi(x) \cdot Z \cdot \exp(U(x_*)) \leq \left(\frac{2\pi}{m}\right)^{d/2} \cdot \mathcal{N}(x; x_*, m^{-1} \cdot I_d).$$

By integrating the above inequalities over space, one sees that

$$\left(\frac{2\pi}{L}\right)^{d/2} \leq Z \cdot \exp(U(x_*)) \leq \left(\frac{2\pi}{m}\right)^{d/2},$$

and substituting these bounds into the preceding display completes the proof.   □

LEMMA 43. *Under Assumption* 31 *and twice continuous-differentiability of* U, *Assumption* 31 *holds for any finite-dimensional marginal of* π.

PROOF. Preservation of m-strong log-concavity of the marginals of π is shown by Saumard and Wellner (2014), Theorem 3.8. For preservation of L-smoothness of the potential of any marginal of π, write the state as $x = (x_1, x_2)$, such that $\pi(x_1, x_2) = \exp(-U(x_1, x_2))$, and define the marginal $\pi(x_1) = \int \exp(-U(x_1, x_2))\, dx_2 = \exp(-V(x_1))$. Formal computations give that

$$\nabla V(x_1) = \mathbb{E}[\nabla_{x_1} U(x_1, X_2)|X_1 = x_1],$$

$$\nabla^2 V(x_1) = \mathbb{E}[\nabla^2_{x_1} U(x_1, X_2)|X_1 = x_1] - \mathbf{Cov}[\nabla_{x_1} U(x_1, X_2)|X_1 = x_1].$$

By the smoothness of U, for any fixed x_1 one can bound $|\nabla_{x_1} U(x_1, x_2)| \lesssim 1 + |x_2|$, $|\nabla^2_{x_1} U(x_1, x_2)| \lesssim 1$, and since log-concave measures admit moments of all orders, it is guaranteed that the integrals above do indeed exist and are finite. Combining this with the twice continuous-differentiability of U, we may validly interchange differentiation and integration, so that the formal identities described above hold.

Recalling that covariance matrices are positive-semidefinite and that U is L-smooth, compute that

$$\nabla^2 V(x_1) = \mathbb{E}[\nabla^2_{x_1} U(x_1, X_2)|X_1 = x_1] - \mathbf{Cov}[\nabla_{x_1} U(x_1, X_2)|X_1 = x_1]$$

$$\preceq \mathbb{E}[\nabla^2_{x_1} U(x_1, X_2)|X_1 = x_1]$$

$$\preceq L \cdot I_d,$$

from which the L-smoothness of V follows. $\square$

The proofs the following two propositions involve the identification of appropriate sets whose conductance can be bounded in terms of σ; the resulting calculations are somewhat involved so the proofs are housed in Appendix B.

PROPOSITION 44. *Under Assumption* 31 *and twice continuous-differentiability of* U,

$$\Phi_P^* \leq 2 \cdot L^{1/2} \cdot \sigma.$$

PROPOSITION 45. *Under Assumption* 31,

$$\Phi_P^* \leq (1 + m \cdot \sigma^2)^{-d/2}.$$

Furthermore, if $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-\beta}$ *with* $\beta < 1/2$, *then* $\Phi_P^* = \mathcal{O}(\exp(-c \cdot d^{1-2\cdot\beta}))$ *for any* $c \in (0, \frac{1}{2} \cdot \frac{m}{L} \cdot \varsigma^2)$, *and in particular, decays faster than any polynomial in* $1/d$.

4.2.2. *Spectral gap upper bounds.* A natural question is whether the lower bound for the spectral gap is of the correct order when $\sigma \sim d^{-1/2}$, that is, whether indeed $\gamma_P = \mathrm{Gap}_R(P)$ scales as d^{-1}. Under Assumption 31, we verify this directly and also show that this is the optimal polynomial scaling.

THEOREM 46. *Under Assumption* 31 *and twice continuous-differentiability of* U,

$$\gamma_P \leq \min\left\{\frac{1}{2} \cdot L \cdot \sigma^2, (1 + m \cdot \sigma^2)^{-d/2}\right\}.$$

Hence, among polynomial scalings of the σ, *the scaling* $\sigma \sim d^{-1/2}$ *is optimal with* $\gamma_P \sim d^{-1}$.

PROOF. The bounds follow from Lemma 47 and Proposition 45 combined with Lemma 5. Let $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-\beta}$. If $\beta \geq \frac{1}{2}$ then we obtain $\gamma_P \in \mathcal{O}(d^{-2 \cdot \beta})$ and this rate is maximized by taking $\beta = \frac{1}{2}$. Combined with Corollary 35, we may conclude that γ_P decays as d^{-1} as $d \to \infty$ when $\beta = \frac{1}{2}$. On the other hand, if $\beta < \frac{1}{2}$ then $\gamma_P \cdot d$ converges to 0 and hence $\beta < \frac{1}{2}$ leads to a faster decay of γ_P than $\beta = \frac{1}{2}$. $\square$

LEMMA 47. *Let U be L-smooth and twice continuously differentiable. Then*

$$\gamma_P \leq \frac{1}{2} \cdot L \cdot \sigma^2.$$

PROOF. By the Cramér–Rao inequality (see, e.g., Saumard and Wellner (2014), eq. 10.25), it holds for any $v \in \mathbb{R}^d$ that

$$\mathrm{Var}_\pi(\langle v, X \rangle) \geq v^\top \cdot \mathbb{E}_\pi[\nabla^2(-\log \pi(X))]^{-1} \cdot v \geq L^{-1} \cdot |v|^2,$$

by L-smoothness. For $v \neq 0$, define $g_v(x) := \langle v, x - \mathbb{E}_\pi[X] \rangle$, and compute

$$
\begin{aligned}
\mathcal{E}(P, g_v) &= \frac{1}{2} \int \pi(\mathrm{d}x) P(x, \mathrm{d}y) (g_v(y) - g_v(x))^2 \\
&= \frac{1}{2} \int \pi(\mathrm{d}x) P(x, \mathrm{d}y) \langle v, y - x \rangle^2 \\
&\leq \frac{1}{2} \int \pi(\mathrm{d}x) Q(x, \mathrm{d}y) \langle v, y - x \rangle^2 \\
&= \frac{1}{2} \int \pi(\mathrm{d}x) \mathcal{N}(\mathrm{d}z; 0, I_d) \langle v, \sigma \cdot z \rangle^2 \\
&= \frac{1}{2} \cdot \sigma^2 \cdot |v|^2.
\end{aligned}
$$

We obtain

$$\gamma_P = \mathrm{Gap}_R(P) = \inf_{f \in \mathrm{L}_0^2(\pi)} \frac{\mathcal{E}(P, f)}{\|f\|_2^2} \leq \frac{\mathcal{E}(P, g_v)}{\|g_v\|_2^2} \leq \frac{\frac{1}{2} \cdot \sigma^2 \cdot |v|^2}{L^{-1} \cdot |v|^2} = \frac{1}{2} \cdot L \cdot \sigma^2. \qquad \square$$

4.3. *Implications for the asymptotic variance.* In this sub-section, we address the asymptotic variance of MCMC estimators computed from RWM chains. We will show that when using appropriately-tuned RWM chains to compute expectations of functions under π, the asymptotic variance of these estimators is an inflation of the ideal variance by a factor which scales linearly with the dimension of the problem. Furthermore, we will exhibit that for a specific class of functions (in particular, affine functions) that this bound is tight in terms of its dimension-dependence.

PROPOSITION 48. *Let Assumption 31 hold, and $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-1/2}$ for any $\varsigma > 0$. Then, for any $f \in \mathrm{L}_0^2(\pi)$, the asymptotic variance of f can be bounded as*

$$\mathrm{var}(P, f) \leq 2^{10} \cdot C_\gamma^{-2} \cdot \varsigma^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \kappa \cdot d \cdot \|f\|_2^2.$$

Additionally, for any linear $f \in \mathrm{L}_0^2(\pi)$, $\mathrm{var}(P, f) \geq 2 \cdot \varsigma^{-2} \cdot d \cdot \|f\|_2^2$.

PROOF. Since P is reversible and $\mathrm{Gap}_R(P) > 0$, $\mathrm{Id} - P$ is invertible on $\mathrm{L}_0^2(\pi)$. We have

$$\mathrm{var}(P, f) = \langle f, (\mathrm{Id} + P) \cdot (\mathrm{Id} - P)^{-1} \cdot f \rangle.$$

Moreover, by considering the spectral resolution of f with respect to P, it is classical that

$$\langle f, (\mathrm{Id} + P) \cdot (\mathrm{Id} - P)^{-1} \cdot f \rangle \leq \frac{2}{\mathrm{Gap}_\mathrm{R}(P)} \cdot \|f\|_2^2,$$

where $\mathrm{Gap}_\mathrm{R}(P)$ is the right spectral gap of P. Recalling that $\mathrm{Gap}_\mathrm{R}(P) \geq \gamma_P$ (for any reversible P), we apply Corollary 35 to bound

$$\mathrm{Gap}_\mathrm{R}(P) \geq 2^{-9} \cdot C_\gamma^2 \cdot \varsigma^2 \cdot \exp(-2 \cdot \varsigma^2) \cdot d^{-1} \cdot \frac{m}{L},$$

and deduce the first bound. For the second bound, by positivity of P we may write

$$\mathrm{var}(P, f) = \langle f, (\mathrm{Id} + P) \cdot (\mathrm{Id} - P)^{-1} \cdot f \rangle \geq \langle f, (\mathrm{Id} - P)^{-1} \cdot f \rangle.$$

Recall the variational representation (see, e.g., Caracciolo, Pelissetto and Sokal (1990)),

$$\langle f, (\mathrm{Id} - P)^{-1} \cdot f \rangle = \sup_{g \in \mathrm{L}_0^2(\pi)} \{2 \cdot \langle f, g \rangle - \mathcal{E}(P, g)\}.$$

We will consider taking $f : x \mapsto \langle v, x - \mathbb{E}_\pi[X] \rangle$, and $g = c \cdot f$ for $c \in \mathbb{R}$. By the argument in the proof of Lemma 47, it holds that

$$\|f\|_2^2 \geq L^{-1} \cdot |v|^2, \qquad \mathcal{E}(P, f) \leq \frac{1}{2} \cdot \sigma^2 \cdot |v|^2.$$

We thus see that

$$\begin{aligned}
\mathrm{var}(P, f) &\geq \langle f, (\mathrm{Id} - P)^{-1} \cdot f \rangle \\
&= \sup_{g \in \mathrm{L}_0^2(\pi)} \{2 \cdot \langle f, g \rangle - \mathcal{E}(P, g)\} \\
&\geq \sup_{c \in \mathbb{R}} \{2 \cdot c \cdot \|f\|_2^2 - c^2 \cdot \mathcal{E}(P, f)\} \\
&= \frac{\|f\|_2^2}{\mathcal{E}(P, f)} \cdot \|f\|_2^2 \\
&\geq \frac{L^{-1} \cdot |v|^2}{\frac{1}{2} \cdot \sigma^2 \cdot |v|^2} \cdot \|f\|_2^2 \\
&= 2 \cdot \varsigma^{-2} \cdot d \cdot \|f\|_2^2. \qquad \square
\end{aligned}$$

4.4. *Discussion of the assumptions.* Under Assumption 31, Corollary 35 provides explicit bounds on the spectral gap of the RWM kernel for any $\sigma > 0$. Theorem 46 further shows that the optimal dependence on dimension is d^{-1} for this class of target measures. On the other hand, it is clear from the proof that the lower bounds on the spectral gap and conductance profile have m-strong convexity and L-smoothness of U as sufficient but not necessary conditions.

For showing lower bounds, m-strong convexity of the potential is only used to verify an appropriate isoperimetric profile inequality. Hence, any other method for establishing such inequalities could be used instead. For instance, one may follow Example 23 to deduce an isoperimetric profile inequality for any log-concave probability measure; an explicit constant could be obtained by study of the corresponding overdamped Langevin diffusion. Example 28 could also be used to extend this reasoning to perturbations of log-concave probability measures. Similarly, for showing lower bounds, L-smoothness is only assumed to establish that the acceptance probability is uniformly lower bounded above 0, that is, $\alpha_0 = \inf_{x \in \mathsf{E}} \alpha(x) > 0$; our Lemma 39 shows that this is possible under far less restrictive smoothness conditions on

U, covering both light-tailed targets and potentials whose gradients are only, for example, Hölder-continuous. As such, L-smoothness should not be viewed as strictly necessary for the RWM to be performant; this stands in contrast to various gradient-based MCMC algorithms, whose performance can deteriorate in the absence of L-smoothness (see, e.g., Livingstone and Zanella (2022)).

On the other hand, if U has particularly poor regularity (e.g., non-Lipschitz-continuous gradients), then it can be necessary to scale σ differently in order to stabilize the acceptance probability away from 0, (see also Vogrinc and Kendall (2021), who study this phenomenon in the optimal scaling framework). For a concrete example, consider the target with potential given by $U(x) = |x|_p^p$ for $p \in [1, 2)$. Observing that $U(\sigma \cdot z) - U(0) = \sigma^p \cdot (\sum_{i=1}^d |z_i|^p)$, one sees that unless $\sigma^p \in \mathcal{O}(d^{-1})$, then $\alpha(0)$ will tend to 0 as $d \to \infty$. Similar cautionary examples which necessitate anomalous scalings of σ can be constructed by designing potentials with "sharp" growth around a local minimum, and following the strategy of Proposition 45.

5. Convergence and mixing time for RWM.
In this section, we analyze the mixing time for the RWM kernel P defined in Section 4 under Assumption 31.

5.1. *Three phase mixing.* The mixing time bound essentially consists of up to three phases. If the initial chi-squared divergence is very large, then there is an initial phase during which the convergence takes place at an exponential rate depending only on α_0 (and in particular, *not* directly on σ^2). Subsequently, there is a secondary phase during which convergence takes place at a faster-than-exponential rate, depending now on both σ^2 and α_0. Finally, once the chi-squared divergence drops below a universal constant (e.g., 8), the convergence is again exponential, again with a rate depending on both σ^2 and α_0. Qualitatively similar behaviour was also observed (for a different algorithm) in the work of Mou et al. (2022). In the proof of Theorem 49 in Appendix B we provide two additional different bounds, valid for any $\varsigma > 0$, which are stronger than the stated bound. In particular, the stated bound may be conservative when u_0 is sufficiently small, since the first phase can be nonexistent and the second phase is bounded fairly crudely.

THEOREM 49. *Under Assumption* 31, *let* $\mu \ll \pi$ *be a probability measure and* $u_0 = \chi^2(\mu, \pi)$. *Let* $\sigma = \varsigma \cdot L^{-1/2} \cdot d^{-1/2}$ *and* $\kappa = L/m$, *with* $\varsigma > 0$ *arbitrary and let the universal constants* C_ℓ, C_γ *be as defined in Lemma* 27. *Then, to guarantee that* $\chi^2(\mu P^n, \pi) \le \varepsilon_{\text{Mix}} \in (0, 8)$ *we may take*

$$n \ge 2 + 2^{10} \cdot \exp(\varsigma^2) \cdot \log(\max\{u_0, 1\})$$
$$+ 2^{14} \cdot C_\ell^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d \cdot \{\log(16 \cdot C_\ell^{-2} \varsigma^{-2}) + \log(\kappa \cdot d) + \varsigma^2\}$$
$$+ 2^6 \cdot C_\gamma^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d \cdot \log\left(\frac{8}{\varepsilon_{\text{Mix}}}\right),$$

that is, of order $\mathcal{O}(\log u_0 + \kappa \cdot d \cdot \log(\kappa \cdot d) + \kappa \cdot d \cdot \log(\varepsilon_{\text{Mix}}^{-1}))$.

5.2. *Two feasible "warm starts".* We provide two complexity bounds based on Theorem 49, corresponding to two different algorithmically feasible initial distributions μ. In Remark 50, we consider a Gaussian initial distribution centered at the mode of the density π with covariance $L^{-1} \cdot I_d$; when U is convex, identifying the mode numerically is feasible and hence the strategy relies on explicit knowledge of L. In Remark 51 we consider the approach suggested by Belloni and Chernozhukov (2009), for which the initial distribution is given by the distribution of the first "accepted" proposal from $Q(x_0, \cdot)$ where x_0 is arbitrary. This does not require explicit knowledge of L, only for example, a bound that allows one to tune

σ such that $\varsigma \in \Theta(1)$, and corresponds directly to how RWM chains are often initialized in practice. In particular, we observe that the RWM mixing time is fairly robust to the choice of initial point x_0 as long as it is not very far from the mode, and this is due to the fast phases of convergence identified in Theorem 49.

REMARK 50. If $\mu = \mathcal{N}(x_*, L^{-1} \cdot I_d)$, with x_* the mode of π, then by Lemma 42 we may obtain the bound $u_0 = \chi^2(\mu, \pi) \le \kappa^{d/2}$. It then holds that $\log u_0 = \mathcal{O}(d \cdot \log \kappa)$, and one obtains a mixing time bound which scales as $\mathcal{O}(\kappa \cdot d \cdot \log(\kappa \cdot d) + \kappa \cdot d \cdot \log(\varepsilon_{\text{Mix}}^{-1}))$. In relation to the comments preceding Theorem 49, we note that by taking ς sufficiently small, one may consider (13)–(14) in the proof of Theorem 49, ensure $u_0 \le 4 \cdot v_\circ^{-1}$ and thereby reduce the complexity to $\mathcal{O}(\kappa \cdot d \cdot \log(d \cdot \log \kappa) + \kappa \cdot d \cdot \log(\varepsilon_{\text{Mix}}^{-1}))$. For example, taking $\varsigma^2 = \frac{1}{2}$ as in Remark 36, one can deduce that

$$n \ge 2 + 96982 \cdot \kappa \cdot d \cdot \log\left(\frac{d}{2} \cdot \log \kappa\right) + 3446 \cdot \kappa \cdot d \cdot \log\left(\frac{8}{\varepsilon_{\text{Mix}}}\right),$$

is sufficient for $\chi^2(\mu P^n, \pi) \le \varepsilon_{\text{Mix}} \in (0, 8)$ when $\kappa^{d/2} > 8$.

In Example 33, we have $\kappa \le 1 + \frac{1}{4} \cdot \sigma_0^2 \cdot \lambda_{\text{Max}}(AA^\top)$ and if A is a random matrix with i.i.d. entries from a distribution with mean 0, variance 1 and finite 4th moment, then Yin, Bai and Krishnaiah (1988), Theorem 3.1, gives $\frac{1}{N} \lambda_{\text{Max}}(AA^\top) \to \frac{1}{\alpha}(1 + \sqrt{\alpha})^2$ a.s., where $d/N \to \alpha$. Hence, if σ_0^2 is $\mathcal{O}(d^{-1})$ then it is reasonable to expect κ independent of d in this regime and hence the $\mathcal{O}(d \log d)$ scaling of n given above. The precise bounds are likely quite loose, for example, $d = 100$, $\kappa = 10$, $\varepsilon_{\text{Mix}} = \frac{1}{2}$ give $n \sim 0.47 \times 10^9$, but yet are not astronomical. The analysis also does not take into account explicitly any concentration of the posterior distribution, and indeed the bounds hold irrespective of the values the data take.

REMARK 51. There is another approach to obtaining a warm start which neatly sidesteps the need for any preliminary optimization, suggested by Belloni and Chernozhukov (2009). The idea is to initialize the chain by a single accepted move of the Metropolis kernel from some arbitrary point x_0, that is,

$$\mu(\mathrm{d}x) = P^\alpha(x_0, \mathrm{d}x)$$

$$:= Q(x_0, \mathrm{d}x) \cdot \alpha(x_0, x) \cdot \alpha(x_0)^{-1}.$$

To this end, one can compute directly (in fact, for *any* Metropolis–Hastings chain) that

$$\pi(\mathrm{d}x) \cdot P^\alpha(x, \mathrm{d}y) = \pi(\mathrm{d}x) \cdot Q(x, \mathrm{d}y) \cdot \alpha(x, y) \cdot \alpha(x)^{-1}$$

$$= \pi(\mathrm{d}y) \cdot Q(y, \mathrm{d}x) \cdot \alpha(y, x) \cdot \alpha(x)^{-1}.$$

Disintegrating this joint measure appropriately, one sees for the RWM that

$$\frac{\mathrm{d}P^\alpha(x_0, \cdot)}{\mathrm{d}\pi}(x) = \alpha(x, x_0) \cdot \alpha(x_0)^{-1} \cdot \frac{\mathrm{d}Q(x, \cdot)}{\mathrm{d}\pi}(x_0)$$

$$\le 1 \cdot \alpha_0^{-1} \cdot \left(\sup_{y \in \mathsf{E}} \frac{\mathrm{d}Q(x, \cdot)}{\mathrm{d}\mathrm{Leb}}(y)\right) \cdot \left(\frac{\mathrm{d}\pi}{\mathrm{d}\mathrm{Leb}}(x_0)\right)^{-1}.$$

Computing that $\alpha_0^{-1} \le 2 \cdot \exp(\frac{1}{2} \cdot \varsigma^2)$, $\sup_{y \in \mathsf{E}} \frac{\mathrm{d}Q(x, \cdot)}{\mathrm{d}\mathrm{Leb}}(y) = (2 \cdot \pi \cdot \sigma^2)^{-d/2}$, $(\frac{\mathrm{d}\pi}{\mathrm{d}\mathrm{Leb}}(x_0))^{-1} \le \kappa^{d/2} \cdot (2 \cdot \pi \cdot L^{-1})^{d/2} \cdot \exp(\frac{1}{2} \cdot L \cdot |x_0 - x_*|^2)$, one obtains that

$$\sup_{x \in \mathsf{E}} \frac{\mathrm{d}P^\alpha(x_0, \cdot)}{\mathrm{d}\pi}(x) \le 2 \cdot \exp\left(\frac{1}{2} \cdot \varsigma^2\right) \cdot \left(\frac{\kappa \cdot d}{\varsigma^2}\right)^{d/2} \cdot \exp\left(\frac{1}{2} \cdot L \cdot |x_0 - x_*|^2\right)$$

$$\in \exp(\mathcal{O}(d \cdot \log(d \cdot \kappa) + L \cdot |x_0 - x_*|^2)),$$

when ς is of order 1. As such, provided that $L \cdot |x_0 - x_*|^2 \in \mathcal{O}(\kappa \cdot d \cdot \log(\kappa \cdot d))$, it holds that $\log u_0 \in \mathcal{O}(\kappa \cdot d \cdot \log(\kappa \cdot d))$, and one obtains a mixing time bound which again scales as $\mathcal{O}(\kappa \cdot d \cdot \log(\kappa \cdot d) + \kappa \cdot d \cdot \log(\varepsilon_{\mathrm{Mix}}^{-1}))$. From the perspective of complexity analysis, it is thus essentially sufficient to initialize the chain by identifying some point within a reasonable distance, that is, $\mathcal{O}((\frac{d}{m} \cdot \log(\kappa \cdot d))^{1/2})$ of the mode, and waiting until a single proposed move is accepted.

5.3. *Comparisons to existing results.* The prior work of Belloni and Chernozhukov (2009) is closely related; the authors study the application of MCMC techniques to frequentist estimation tasks for which direct optimization of the objective function is challenging. For their application, a different isoperimetric inequality is used, which allows them to handle potentials which are nonconvex, nonsmooth, or even both. However they assume that their target distributions are supported on a (large) closed ball. Their complexity bounds are essentially obtained by estimating the spectral gap, and constructing an appropriate "warm start", that is, an initial distribution μ which satisfies $\frac{d\mu}{d\pi} \leq W < \infty$, which implies a bound on the initial L^2 distance to equilibrium, recalled in Remark 51. Their result holds for a particular choice of σ^2, which depends on theoretical quantities that are typically unknown. In contrast, our bounds are valid for any σ^2. Their conductance bound is of the form $\Phi_P^* \in \Omega(d^{-1})$, as opposed to the $\Phi_P^* \in \Omega(d^{-1/2})$ that we find. This may appear suboptimal, but it may also be the case that their class of target distributions is sufficiently harder than those which we consider, to the point that the conductance is genuinely worse in this way.

Closest to our work are the twin papers Dwivedi et al. (2019) and Chen et al. (2020), which give complexity bounds for the RWM under Assumption 31. Under a "feasible" Gaussian initial law (roughly as in Remark 50), Dwivedi et al. (2019) obtain complexity bounds of $\mathcal{O}(d^2 \cdot \kappa^2 \cdot \log^{1.5}(\kappa \cdot \varepsilon_{\mathrm{Mix}}^{-1}))$ which Chen et al. (2020) refined to $\mathcal{O}(d \cdot \kappa^2 \cdot \log(d \cdot \varepsilon_{\mathrm{Mix}}^{-1}))$ by making use of the conductance profile framework; we observe that our complexity analysis implies a weaker dependence on κ. Their results also assume precise values of σ^2 that are typically unknown in practical applications, in contrast with ours.

6. Convergence and mixing time for pCN. In this section, we apply the same technique to analyze the convergence to equilibrium of the preconditioned Crank–Nicolson algorithm. We consider the following class of distributions, consistent with both Assumptions 9 and 31:

ASSUMPTION 52. The target measure π on $\mathsf{E} = \mathbb{R}^d$ can be written

$$\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, \mathsf{C}) \cdot \exp(-\Psi(x)),$$

where Ψ is convex, L-smooth, and minimized at $x = 0$, and C is a positive definite covariance matrix.

Letting $\nu = \mathcal{N}(0, \mathsf{C})$, throughout this section we denote by P the pCN kernel defined by (1), with Q the ν-reversible Gaussian kernel defined for a fixed but arbitrary $\rho \in (0, 1)$ by

$$Q(x, A) = \int \mathbf{1}_A(\rho \cdot x + \eta \cdot z) \mathcal{N}(\mathrm{d}z; 0, \mathsf{C}), \quad x \in \mathsf{E}, A \in \mathscr{E},$$

where $\rho^2 + \eta^2 = 1$. We will refer to η as the step-size of this kernel, and denote $\alpha_0 := \inf_{x \in \mathsf{E}} \alpha(x)$. In particular, we have that $\varpi \propto \exp(-\Psi)$ is the density of π w.r.t. the reference measure ν. By Doucet et al. (2015), Proposition 3(i), we may deduce that P is positive, taking in their notation $\chi = \nu$ and $r(u, v) = \mathcal{N}(u; \rho^{1/2} \cdot v, (1 - \rho) \cdot \mathsf{C})$.

The pCN algorithm is particularly popular in Bayesian inverse problems (see, e.g., Stuart (2010), Example 5.3), where C is typically a finite section of some infinite-dimensional trace-class covariance operator. As with RWM, one can handle relaxations of L-smoothness of Ψ by a suitable adaptation of Lemma 39.

REMARK 53. Assume π has an m-strongly log-concave density w.r.t. Lebesgue, with associated potential U minimized at 0. Then in Assumption 52 one may take $\mathsf{C} = m^{-1} \cdot I_d$ and $\Psi(x) = U(x) - \frac{1}{2} \cdot m \cdot |x|^2$. We observe that if U is L'-smooth, then Ψ is $(L' - m)$-smooth, and in particular we obtain the pCN condition number $\kappa = L/m = \kappa' - 1$, where $\kappa' = L'/m$ is the condition number one would associate with the RWM kernel. Our results for pCN with more general C can therefore be interpreted as applying to the class of densities with respect to Lebesgue which possess m-strongly convex and L-smooth potentials, where we see an improvement over RWM in terms of the condition number, at least if one is willing to use the minimizer of U to define an appropriate parameterization.

Throughout this section, we write $\mathrm{Tr}(\mathsf{C})$ for the trace of a matrix C, and define the norm

(10) $$|x|_{\mathsf{C}^{-1}} := |\mathsf{C}^{-1/2} \cdot x|, \quad x \in \mathsf{E}.$$

As with RWM, one can derive a spectral gap estimate for pCN by applying Theorem 18.

THEOREM 54. *Under Assumption 52, π admits a regular, concave isoperimetric minorant $\tilde{I}_\pi = \varphi_\gamma \circ \Phi_\gamma^{-1}$ w.r.t. $|\cdot|_{\mathsf{C}^{-1}}$ and*

$$\Phi_P^* \geq \frac{1}{4} \cdot C_\gamma \cdot \alpha_0^2 \cdot \eta, \qquad \gamma_P \geq 2^{-5} \cdot C_\gamma^2 \cdot \alpha_0^4 \cdot \eta^2,$$

where the constant C_γ is defined in Lemma 27. Writing $\eta = \varsigma \cdot (L \cdot \mathrm{Tr}(\mathsf{C}))^{-1/2} \in (0, 1)$, we have $\alpha_0 \geq \frac{1}{2} \cdot \exp(-\frac{1}{2} \cdot \varsigma^2)$, and hence

(11) $$\gamma_P \geq 2^{-9} \cdot C_\gamma^2 \cdot \exp(-2 \cdot \varsigma^2) \cdot \varsigma^2 \cdot (L \cdot \mathrm{Tr}(\mathsf{C}))^{-1}.$$

Optimizing over ς gives

$$\gamma_P \geq 2^{-10} \cdot C_\gamma^2 \cdot e^{-1} \cdot (L \cdot \mathrm{Tr}(\mathsf{C}))^{-1} \geq 3.62784 \times 10^{-5} \cdot (L \cdot \mathrm{Tr}(\mathsf{C}))^{-1}.$$

PROOF. This is proven in an analogous fashion to Theorem 34. The fact that $\tilde{I}_\pi = \varphi_\gamma \circ \Phi_\gamma^{-1}$ is an isoperimetric minorant for π is established in Lemma 55, and the close-coupling condition for pCN is established with $\varepsilon = \frac{1}{2} \cdot \alpha_0$ and $\delta = \alpha_0 \cdot \frac{\eta}{\rho}$ in Lemma 57. Thus we can apply Theorem 18 to deduce the conductance bound $\Phi_P^* \geq \frac{1}{8} \cdot \alpha_0 \cdot \min\{1, 2 \cdot C_\gamma \cdot \alpha_0 \cdot \frac{\eta}{\rho}\}$, which we then simplify by recalling that $\rho \leq 1$ and $2 \cdot C_\gamma \cdot \alpha_0 \cdot \eta \leq 2 \cdot C_\gamma \cdot 1 \cdot 1 < 1$. Finally, the lower bound on the acceptance rate is established in Lemma 58. □

6.1. *Lower bounds for pCN.* We now give appropriate results related to isoperimetry for the pCN algorithm. The key subtlety to establishing an appropriate isoperimetric inequality for the pCN kernel is the fact that we want to change metric from the flat Euclidean metric $|\cdot|$ to the metric $|\cdot|_{\mathsf{C}^{-1}}$.

LEMMA 55. *Under Assumption 52, π admits $\varphi_\gamma \circ \Phi_\gamma^{-1}$ as an isoperimetric minorant w.r.t. the metric $\mathsf{d}_\mathsf{C}(x, y) = |x - y|_{\mathsf{C}^{-1}}$.*

PROOF. First, define $\pi_W = (x \mapsto \mathsf{C}^{-1/2} \cdot x)_\# \pi$. One checks that the density of π_W is given by $\mathcal{N}(x; 0, I_d) \cdot \exp(-\Psi(\mathsf{C}^{1/2} \cdot x))$, for which the potential $\frac{1}{2} \cdot |x|^2 + \Psi(\mathsf{C}^{1/2} \cdot x)$ is 1-strongly convex. By Lemma 27, one sees that π_W admits $\varphi_\gamma \circ \Phi_\gamma^{-1}$ as an isoperimetric minorant w.r.t. $|\cdot|$. Defining $\mathsf{d}_\mathsf{C}(x, y) := |x - y|_{\mathsf{C}^{-1}}$, one can apply Lemma 26 with $(\mu_1, \mathsf{E}_1, \mathsf{d}_1) = (\pi_W, \mathsf{E}, |\cdot|)$, $(\mu_2, \mathsf{E}_2, \mathsf{d}_2) = (\pi, \mathsf{E}, \mathsf{d}_\mathsf{C})$, noting that $x \mapsto \mathsf{C}^{-1/2} \cdot x$ is an *isometry* between these two metric spaces, to conclude that π admits $\varphi_\gamma \circ \Phi_\gamma^{-1}$ as an isoperimetric minorant with respect to d_C. □

The following lemma gives a useful bound on the total variation distance between the proposals, analogous to Lemma 37; the proof can be found in Appendix B.

LEMMA 56.   *If $v > 0$ and $x, y \in \mathsf{E}$ satisfy $|x - y|_{\mathsf{C}^{-1}} \leq v \cdot \frac{\eta}{\rho}$,*

$$\|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq \frac{1}{2} \cdot v.$$

LEMMA 57.   *P is $(|\cdot|_{\mathsf{C}^{-1}}, \alpha_0 \cdot \frac{\eta}{\rho}, \frac{1}{2} \cdot \alpha_0)$-close-coupling.*

PROOF.   Assume $x, y \in \mathsf{E}$ are such that $|x - y|_{\mathsf{C}^{-1}} \leq \alpha_0 \cdot \frac{\eta}{\rho}$. Then $\|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq \frac{1}{2} \cdot \alpha_0$ by Lemma 56. Since Q is v-reversible, we may apply Lemma 19 to conclude.   □

LEMMA 58.   *Let $\tilde{\kappa} = L \cdot \mathrm{Tr}(\mathsf{C})$. The pCN chain satisfies*

$$\alpha(x) \geq \frac{1}{2} \cdot \exp\left(-\frac{1}{2} \cdot \eta^2 \cdot \tilde{\kappa}\right) > 0, \quad x \in \mathsf{E}.$$

In particular, take $\eta := \varsigma \cdot \tilde{\kappa}^{-1/2}$ where $\varsigma \in (0, \tilde{\kappa}^{1/2})$ is arbitrary. Then

$$\alpha_0 := \inf_{x \in \mathsf{E}} \alpha(x) \geq \frac{1}{2} \cdot \exp\left(-\frac{1}{2} \cdot \varsigma^2\right).$$

PROOF.   For $x \in \mathsf{E}$, let $\tilde{x} := \rho \cdot x$ and $w := \tilde{x} + \eta \cdot z$, with $z \sim \mathcal{N}(0, \mathsf{C})$. Recalling that $\rho \in (0, 1)$, we can apply convexity of Ψ to see that

$$\Psi(\tilde{x}) = \Psi\big((1 - \rho) \cdot 0 + \rho \cdot x\big)$$
$$\leq (1 - \rho) \cdot \Psi(0) + \rho \cdot \Psi(x)$$
$$\implies \quad \Psi(\tilde{x}) - \Psi(0) \leq \rho \cdot \big(\Psi(x) - \Psi(0)\big)$$
$$\leq \Psi(x) - \Psi(0)$$
$$\implies \quad \Psi(\tilde{x}) \leq \Psi(x).$$

Applying L-smoothness shows that

$$\Psi(w) \leq \Psi(\tilde{x}) + \langle \nabla \Psi(x), \eta \cdot z \rangle + \frac{1}{2} \cdot L \cdot \eta^2 \cdot |z|^2$$
$$\leq \Psi(x) + \langle \nabla \Psi(x), \eta \cdot z \rangle + \frac{1}{2} \cdot L \cdot \eta^2 \cdot |z|^2.$$

From here, we imitate the proof of Lemma 39, writing

$$\alpha(x) = \int \mathcal{N}(\mathrm{d}z; 0, \mathsf{C}) \cdot \min\{1, \exp(-[\Psi(w) - \Psi(x)])\}$$
$$\geq \int \mathcal{N}(\mathrm{d}z; 0, \mathsf{C}) \cdot \min\left\{1, \exp\left(-\left[\langle \nabla \Psi(x), \eta \cdot z \rangle + \frac{1}{2} \cdot L \cdot \eta^2 \cdot |z|^2\right]\right)\right\}$$
$$\geq \int \mathcal{N}(\mathrm{d}z; 0, \mathsf{C}) \cdot \min\{1, \exp(-[\langle \nabla \Psi(x), \eta \cdot z \rangle])\} \cdot \exp\left(-\frac{1}{2} \cdot L \cdot \eta^2 \cdot |z|^2\right)$$
$$\geq \frac{1}{2} \cdot \int \mathcal{N}(\mathrm{d}z; 0, \mathsf{C}) \cdot \exp\left(-\frac{1}{2} \cdot L \cdot \eta^2 \cdot |z|^2\right)$$
$$\geq \frac{1}{2} \cdot \exp\left(-\int \mathcal{N}(\mathrm{d}z; 0, \mathsf{C}) \cdot \frac{1}{2} \cdot L \cdot \eta^2 \cdot |z|^2\right)$$

$$= \frac{1}{2} \cdot \exp\left(-\frac{1}{2} \cdot L \cdot \eta^2 \cdot \mathrm{Tr}(\mathsf{C})\right).$$

The second inequality follows by algebraic substitution. □

6.2. *Mixing time for pCN.* We now give mixing time results for the pCN algorithm; the proof of Theorem 59 is in Appendix B.

THEOREM 59. *Under Assumption 52, let $\mu \ll \pi$ be a probability measure and $u_0 = \chi^2(\mu, \pi)$. Let $\tilde{\kappa} = L \cdot \mathrm{Tr}(\mathsf{C})$ and $\eta = \varsigma \cdot \tilde{\kappa}^{-1/2}$ with $\varsigma \in (0, \tilde{\kappa}^{1/2})$ arbitrary and let the universal constants C_ℓ, C_γ be as defined in Lemma 27. Then, to guarantee that $\chi^2(\mu P^n, \pi) \le \varepsilon_{\mathrm{Mix}} \in (0, 8)$ we may take*

$$n \ge 2 + 2^{10} \cdot \exp(\varsigma^2) \cdot \log(\max\{u_0, 1\})$$

$$+ 2^{14} \cdot C_\ell^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa} \cdot \{\log(16 \cdot C_\ell^{-2} \cdot \varsigma^{-2} \cdot \tilde{\kappa}) + \varsigma^2\}$$

$$+ 2^6 \cdot C_\gamma^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa} \cdot \log\left(\frac{8}{\varepsilon_{\mathrm{Mix}}}\right),$$

that is, of order $\mathcal{O}(\log u_0 + \tilde{\kappa} \log(\tilde{\kappa}) + \tilde{\kappa} \log(\varepsilon_{\mathrm{Mix}}^{-1}))$.

REMARK 60. We note that in contrast to the RWM, the assumptions made on the target for pCN allow for a dimension-independent control of the mixing behaviour. This phenomenon has been observed since at least Hairer, Stuart and Vollmer (2014), who establish the dimension-robustness of the spectral gap under similar assumptions, though with much less explicit quantitative results. On the other hand, one can still quantify the difficulty of navigating the target measure through the roughness of the potential Ψ, as summarized by L, and the effective dimension of the prior, as summarized by $\mathrm{Tr}(\mathsf{C})$; see Agapiou et al. (2017) for related notions.

REMARK 61. An analogous initial distribution to that in Remark 50 is $\mu = \mathcal{N}(0, \mathsf{C} \cdot (I_d + L \cdot \mathsf{C})^{-1})$, for which

$$\frac{\mathrm{d}\mu}{\mathrm{d}\pi}(x) \le \det(I_d + L \cdot \mathsf{C})^{1/2} \le \exp\left(\frac{1}{2} \cdot L \cdot \mathrm{Tr}(\mathsf{C})\right).$$

Hence, $\log u_0 \in \mathcal{O}(\tilde{\kappa})$, from which one concludes that the mixing time is bounded as $\mathcal{O}(\tilde{\kappa} \cdot \log \tilde{\kappa} + \tilde{\kappa} \cdot \log \varepsilon_{\mathrm{Mix}}^{-1})$.

6.3. *Comparison with independent Metropolis–Hastings.* A nonlocal pCN chain may be defined by taking $\rho = 0$, and hence $\eta = 1$. This corresponds to an independent Metropolis–Hastings (IMH) kernel with proposal distribution $q(A) = \mathcal{N}(A; 0, \mathsf{C})$. Theorem 54, strictly speaking, does not apply but does allow consideration of ρ arbitrarily close to 0 by taking ς^2 arbitrarily close to $L \cdot \mathrm{Tr}(\mathsf{C})$, resulting in a spectral gap bound of order $\alpha_0^4 \sim \exp(-2 \cdot L \cdot \mathrm{Tr}(C))$. This is somewhat crude, perhaps because the analysis here is more suitable for Markov chains with local behaviour.

On the other hand, we may deduce by Gåsemyr (2006), Theorem 2, that the spectral gap of the IMH is precisely

$$\gamma_P = \alpha_0 = \inf_{x \in \mathsf{E}} \frac{\mathrm{d}q}{\mathrm{d}\pi}(x) = \exp(\Psi(0)) \int \mathcal{N}(x; 0, \mathsf{C}) \exp(-\Psi(x)) \, \mathrm{d}x,$$

and we will see that when $\rho = 0$, the bound on α_0 deteriorates rapidly with d. By L-smoothness and $\nabla\Psi(0) = 0$, we have $\Psi(x) - \Psi(0) \le \frac{L}{2}|x|^2$, so that we have the estimate

$$\gamma_P \ge \int \mathcal{N}(x; 0, \mathsf{C}) \exp\left(-\frac{L}{2}|x|^2\right) \mathrm{d}x = \det(I_d + L\mathsf{C})^{-1/2}.$$

In the context of Remark 53, we obtain

$$\gamma_P \geq \left(1 + \frac{L}{m}\right)^{-d/2} = (1 + \kappa)^{-d/2},$$

which decreases much faster than d^{-1}; this also implies poor scaling of independent Metropolis–Hastings in comparison with RWM. In general, we only obtain the bound

$$\gamma_P \geq \exp\left(-\frac{1}{2} \cdot L \cdot \mathrm{Tr}(\mathsf{C})\right),$$

where the exponential dependence on $L \cdot \mathrm{Tr}(\mathsf{C})$ is much worse than the linear dependence in (11) when ς is chosen appropriately. Hence, significant improvements in the spectral gap bound are obtained by using appropriately tuned "local" Markov chains.

APPENDIX A: NOTATION

- The Euclidean norm on $\mathsf{E} = \mathbb{R}^d$ is denoted $|\cdot|$, which we will also use to denote the associated metric.
- We write Leb, and plainly dx, for the Lebesgue measure on $\mathbb{R}^d$.
- We write $\mathrm{L}^2(\pi)$ for the Hilbert space of (equivalence classes of) real-valued π-square-integrable measurable functions with inner product

$$\langle f, g \rangle = \int_{\mathsf{E}} f(x)g(x) \, d\pi(x),$$

and corresponding norm $\|\cdot\|_2$. For $g \in \mathrm{L}^2(\pi)$, $\mathrm{Var}_\pi(g) := \|g - \pi(g)\|_2^2$. We write $\mathrm{L}_0^2(\pi)$ for the set of functions $f \in \mathrm{L}^2(\pi)$ which also satisfy $\pi(f) = 0$.
- Given a set $A \in \mathcal{E}$ with $\pi(A) > 0$, we define the probability measure π_A on $(\mathsf{E}, \mathcal{E})$ via $\pi_A(\cdot) := \pi(\cdot \cap A)/\pi(A)$.
- Given a probability measure π and a function T on E, we define the pushforward measure of π under the action of T by $(T_{\#}\pi)(A) := \pi(T^{-1}(A))$.
- For a set $A \in \mathcal{E}$, its complement in E is denoted by $A^{\complement}$. We denote the corresponding indicator function by $\mathbf{1}_A : \mathsf{E} \to \{0, 1\}$.
- For two sets $A, B \in \mathcal{E}$ and a metric d on E, the distance between the two sets is given by

$$\mathsf{d}(A, B) := \inf\{\mathsf{d}(x, y) : x \in A, y \in B\}.$$

When one of the sets is a singleton, we will simply write $\mathsf{d}(x, B)$ for $\mathsf{d}(\{x\}, B)$, say.
- For two measures μ and ν, we write $\nu \ll \mu$ to mean that ν is absolutely continuous with respect to μ.
- For two probability measures μ and ν on $(\mathsf{E}, \mathcal{E})$, we let $\mu \otimes \nu(A \times B) := \mu(A) \cdot \nu(B)$ for $A, B \in \mathcal{E}$. For a Markov kernel $P(x, dy)$ on $\mathsf{E} \times \mathcal{E}$, we write for $\bar{A} \in \mathcal{E} \otimes \mathcal{E}$, the minimal product σ-algebra, $\mu \otimes P(\bar{A}) := \int_{\bar{A}} \mu(dx)P(x, dy)$.
- For a probability measure $\mu \ll \pi$, the chi-squared divergence between μ and π is given by $\chi^2(\mu, \pi) := \|\frac{d\mu}{d\pi} - 1\|_2^2$.
- A point mass distribution at x is denoted by δ_x.
- We associate, to a π-invariant Markov kernel P, the bounded linear operator also denoted $P : \mathrm{L}^2(\pi) \to \mathrm{L}^2(\pi)$, given by $Pf(x) = \int_{\mathsf{E}} P(x, dy)f(y)$. We may refer to P as a kernel or as an operator, the meaning being clear from the context.
- We write Id for the identity mapping on $\mathrm{L}^2(\pi)$, and id for the identity mapping on $\mathbb{R}$, and I_d for the $d \times d$ identity matrix.
- Given a bounded linear operator $P : \mathrm{L}^2(\pi) \to \mathrm{L}^2(\pi)$, we define the Dirichlet form $\mathcal{E}(P, f) := \langle (\mathrm{Id} - P)f, f \rangle$ for any $f \in \mathrm{L}^2(\pi)$.

- For a mapping $T : (\mathsf{E}, \mathsf{d}) \to (\mathsf{E}', \mathsf{d}')$ between metric spaces, the Lipschitz norm is defined as $|T|_{\mathrm{Lip}} := \sup_{x \neq y} \frac{\mathsf{d}'(T(x), T(y))}{\mathsf{d}(x, y)}$.
- The spectrum of a bounded linear operator P is the set

$$\mathcal{S}(P) := \{\lambda \in \mathbb{C} : P - \lambda \cdot \mathrm{Id} \text{ is not invertible}\}.$$

- For a π-invariant Markov kernel P, we denote by $\mathcal{S}_0(P)$ the spectrum of the restriction of P to $\mathrm{L}_0^2(\pi)$. We define the spectral gap of P to be $\gamma_P := 1 - \sup |\mathcal{S}_0(P)|$.
- If P is a π-reversible Markov kernel, then $\mathcal{S}_0(P) \subseteq [-1, 1]$ and we may define the right spectral gap as $\mathrm{Gap}_{\mathrm{R}}(P) := 1 - \sup \mathcal{S}_0(P)$, which satisfies (see, e.g., Douc et al. (2018), Theorem 22.A.19),

$$\mathrm{Gap}_{\mathrm{R}}(P) = \inf_{g \in \mathrm{L}_0^2(\pi), g \neq 0} \frac{\mathcal{E}(P, g)}{\|g\|_2^2}.$$

- We say that a π-reversible Markov kernel P is positive if $\langle f, Pf \rangle \geq 0$ for all $f \in \mathrm{L}^2(\pi)$. In this case, $\gamma_P = \mathrm{Gap}_{\mathrm{R}}(P)$.
- We write $\mathcal{N}(m, \Sigma)$ where $m \in \mathbb{R}^d$ and Σ is a $d \times d$ covariance matrix, for the corresponding Gaussian distribution on $\mathbb{R}^d$, $\mathcal{N}(x; m, \Sigma)$ for its density with respect to Lebesgue at $x \in \mathbb{R}^d$ and $\mathcal{N}(A; m, \Sigma)$ for the measure it assigns to $A \in \mathscr{E}$.
- We adopt the following $\mathcal{O}$ (resp. Ω) notation to indicate when functions grow no faster than (resp. no slower than) other functions. For $a \in \mathbb{R} \cup \{\infty\}$,
  - If $f(x) \in \mathcal{O}(g(x))$ as $x \to a$, this means that $\limsup_{x \to a} |\frac{f(x)}{g(x)}| < \infty$. When $a = +\infty$, then we may drop explicit mention of a.
  - If $f(x) \in \Omega(g(x))$ as $x \to a$, this means that $\liminf_{x \to a} |\frac{f(x)}{g(x)}| > 0$. In particular $f \in \mathcal{O}(g) \iff g \in \Omega(f)$.
  - We will write $f(x) \in \Theta(g(x))$ as $x \to a$ if both $f(x) \in \mathcal{O}(g(x))$ and $f(x) \in \Omega(g(x))$ as $x \to a$.

APPENDIX B: ADDITIONAL PROOFS

PROOF OF LEMMA 6. For the case $v > \frac{1}{2}$, we know that $\Lambda_P(v) \geq \mathrm{Gap}_{\mathrm{R}}(P)$, since we are taking an infimum over a smaller set of functions for any v. By Lemma 5, we have $\mathrm{Gap}_{\mathrm{R}}(P) \geq \frac{1}{2} \cdot [\Phi_P^*]^2$. Now consider the case $v \in (0, \frac{1}{2}]$, let A be a measurable set such that $0 < \pi(A) \leq v$, and let $h \in C_0^+(A)$. Consider the quantity

$$\mathcal{E}_1(P, h) := \frac{1}{2} \cdot \int \pi(\mathrm{d}x) \cdot P(x, \mathrm{d}y) \cdot |h(y) - h(x)|$$

and observe that, by symmetry of $\pi \otimes P$, one can write

$$\mathcal{E}_1(P, h) = \int \pi(\mathrm{d}x) \cdot P(x, \mathrm{d}y) \cdot |h(y) - h(x)| \cdot \mathbf{1}[h(x) < h(y)]$$

$$= \int \pi(\mathrm{d}x) \cdot P(x, \mathrm{d}y) \cdot \mathbf{1}[h(x) < h(y)] \cdot (h(y) - h(x))$$

$$= \int \pi(\mathrm{d}x) \cdot P(x, \mathrm{d}y) \cdot \int_{t \geq 0} \mathbf{1}[h(x) \leq t < h(y)] \, \mathrm{d}t$$

$$= \int_{t \geq 0} \left(\int \pi(\mathrm{d}x) \cdot P(x, \mathrm{d}y) \cdot \mathbf{1}[h(x) \leq t < h(y)] \right) \mathrm{d}t.$$

Now, observe that if one defines $H_t := \{x \in A : h(x) > t\}$, then

$$\int \pi(\mathrm{d}x) \cdot P(x, \mathrm{d}y) \cdot \mathbf{1}[h(x) \le t < h(y)] = (\pi \otimes P)(H_t^{\complement} \times H_t)$$

$$= (\pi \otimes P)(H_t \times H_t^{\complement}).$$

Recalling that $H_t \subseteq A$ and hence that $\pi(H_t) \le \pi(A)$, one sees that if $\pi(H_t) > 0$,

$$(\pi \otimes P)(H_t \times H_t^{\complement}) = \pi(H_t) \cdot \frac{(\pi \otimes P)(H_t \times H_t^{\complement})}{\pi(H_t)}$$

$$\ge \pi(H_t) \cdot \inf\left\{\frac{(\pi \otimes P)(S \times S^{\complement})}{\pi(S)} : 0 < \pi(S) \le \pi(A)\right\}$$

$$= \pi(H_t) \cdot \Phi_P(\pi(A)),$$

while the inequality holds trivially if $\pi(H_t) = 0$. It thus holds that

$$\mathcal{E}_1(P, h) = \int_{t \ge 0}\left(\int \pi(\mathrm{d}x) \cdot P(x, \mathrm{d}y) \cdot \mathbf{1}[h(x) \le t < h(y)]\right)\mathrm{d}t$$

$$= \int_{t \ge 0}(\pi \otimes P)(H_t \times H_t^{\complement})\,\mathrm{d}t$$

$$\ge \int_{t \ge 0} \pi(H_t) \cdot \Phi_P(\pi(A))\,\mathrm{d}t$$

$$= \left(\int_{t \ge 0}\int \pi(\mathrm{d}x) \cdot \mathbf{1}[h(x) > t]\,\mathrm{d}t\right) \cdot \Phi_P(\pi(A))$$

$$= \left(\int \pi(\mathrm{d}x) \cdot \int_{t \ge 0}\mathbf{1}[h(x) > t]\,\mathrm{d}t\right) \cdot \Phi_P(\pi(A))$$

$$= \left(\int \pi(\mathrm{d}x) \cdot h(x)\mathrm{d}t\right) \cdot \Phi_P(\pi(A))$$

$$= \pi(h) \cdot \Phi_P(\pi(A)).$$

Now, let $g \in C_0^+(A)$, and take $h = g^2$ in the above to see that

$$\pi(g^2) \cdot \Phi_P(\pi(A)) \le \frac{1}{2} \cdot \int \pi(\mathrm{d}x)P(x, \mathrm{d}y)\big|g(x)^2 - g(y)^2\big|$$

$$\le \frac{1}{2} \cdot \left(\int \pi(\mathrm{d}x)P(x, \mathrm{d}y)\big|g(x) - g(y)\big|^2\right)^{1/2}$$

$$\times \left(\int \pi(\mathrm{d}x)P(x, \mathrm{d}y)\big|g(x) + g(y)\big|^2\right)^{1/2}$$

$$\le \frac{1}{2} \cdot \left(2 \cdot \mathcal{E}(P, g)\right)^{1/2} \cdot \left(4 \cdot \pi(g^2)\right)^{1/2}$$

$$= 2^{1/2} \cdot \mathcal{E}(P, g)^{1/2} \cdot \pi(g^2)^{1/2},$$

from which we may deduce that

$$\frac{1}{2} \cdot \Phi_P(\pi(A))^2 \le \frac{\mathcal{E}(P, g)}{\pi(g^2)} \le \frac{\mathcal{E}(P, g)}{\mathrm{Var}_\pi(g)}.$$

Taking an infimum over g shows that $\lambda_P(A) \ge \frac{1}{2} \cdot \Phi_P(\pi(A))^2$, and taking an infimum over A shows that $\Lambda_P(v) \ge \frac{1}{2} \cdot \Phi_P(v)^2$. $\quad\square$

PROOF OF LEMMA 15. For $A \in \mathcal{E}$, define the sets

$$S_1 := \left\{ z \in A : P(z, A^{\complement}) < \frac{1}{2} \cdot \varepsilon \right\},$$

$$S_2 := \left\{ z \in A^{\complement} : P(z, A) < \frac{1}{2} \cdot \varepsilon \right\}$$

and $S_3 := (S_1 \cup S_2)^{\complement}$, and let $\theta \in (0, 1)$. We consider two cases. First, we establish that when either $\pi(S_1) \leq \theta \cdot \pi(A)$ or $\pi(S_2) \leq \theta \cdot \pi(A^{\complement})$, then

(12) $$(\pi \otimes P)(A \times A^{\complement}) \geq \frac{1}{2} \cdot (1 - \theta) \cdot \varepsilon \cdot \min\{\pi(A), \pi(A^{\complement})\}.$$

If $\pi(S_1) \leq \theta \cdot \pi(A)$ then

$$\begin{aligned}
\pi(A) &= \pi(S_1) + \pi(A \setminus S_1) \\
&\leq \theta \cdot \pi(A) + \pi(A \setminus S_1) \\
\implies \quad &\pi(A \setminus S_1) \geq (1 - \theta) \cdot \pi(A).
\end{aligned}$$

Now,

$$\begin{aligned}
(\pi \otimes P)(A \times A^{\complement}) &\geq (\pi \otimes P)((A \setminus S_1) \times A^{\complement}) \\
&\geq \frac{1}{2} \cdot \varepsilon \cdot \pi(A \setminus S_1) \\
&\geq \frac{1}{2} \cdot (1 - \theta) \cdot \varepsilon \cdot \pi(A).
\end{aligned}$$

Similarly, if $\pi(S_2) \leq \theta \cdot \pi(A^{\complement})$ then

$$\begin{aligned}
\pi(A^{\complement}) &= \pi(S_2) + \pi(A^{\complement} \setminus S_2). \\
&\leq \theta \cdot \pi(A^{\complement}) + \pi(A^{\complement} \setminus S_2) \\
\implies \quad &\pi(A^{\complement} \setminus S_2) \geq (1 - \theta) \cdot \pi(A^{\complement}),
\end{aligned}$$

and arguing as before:

$$\begin{aligned}
(\pi \otimes P)(A^{\complement} \times A) &\geq (\pi \otimes P)((A^{\complement} \setminus S_2) \times A) \\
&\geq \frac{1}{2} \cdot \varepsilon \cdot \pi(A^{\complement} \setminus S_2) \\
&\geq \frac{1}{2} \cdot (1 - \theta) \cdot \varepsilon \cdot \pi(A^{\complement}).
\end{aligned}$$

The first claim thus follows. In the second case, $\pi(S_1) > \theta \cdot \pi(A)$ and $\pi(S_2) > \theta \cdot \pi(A^{\complement})$. As noticed by Dwivedi et al. (2019), reversibility is not required to establish the following

$$\begin{aligned}
(\pi \otimes P)(A \times A^{\complement}) &= (\pi \otimes P)(\mathsf{E} \times A^{\complement}) - [(\pi \otimes P)(A^{\complement} \times \mathsf{E}) - (\pi \otimes P)(A^{\complement} \times A)] \\
&= \pi(A^{\complement}) - \pi(A^{\complement}) + (\pi \otimes P)(A^{\complement} \times A) \\
&= (\pi \otimes P)(A^{\complement} \times A).
\end{aligned}$$

We then compute

$$
\begin{aligned}
(\pi \otimes P)(A \times A^\complement) &= \frac{1}{2} \cdot (\pi \otimes P)(A \times A^\complement) + \frac{1}{2} \cdot (\pi \otimes P)(A^\complement \times A) \\
&\geq \frac{1}{2} \cdot (\pi \otimes P)((A \setminus S_1) \times A^\complement) + \frac{1}{2} \cdot (\pi \otimes P)((A^\complement \setminus S_2) \times A) \\
&\geq \frac{1}{4} \cdot \varepsilon \cdot \pi(A \setminus S_1) + \frac{1}{4} \cdot \varepsilon \cdot \pi(A^\complement \setminus S_2) \\
&= \frac{1}{4} \cdot \varepsilon \cdot \pi(S_3).
\end{aligned}
$$

Now for $(z, z') \in S_1 \times S_2$ we have

$$
\begin{aligned}
\|P(z, \cdot) - P(z', \cdot)\|_{\mathrm{TV}} &\geq P(z, A) - P(z', A) \\
&= 1 - P(z, A^\complement) - P(z', A) \\
&> 1 - \varepsilon.
\end{aligned}
$$

This implies that $\mathsf{d}(S_1, S_2) = \inf\{|z - z'| : (z, z') \in S_1 \times S_2\} \geq \delta$, since P is $(\mathsf{d}, \delta, \varepsilon)$-close coupling. Hence, using Definition 10 and monotonicity of F,

$$
\begin{aligned}
(\pi \otimes P)(A \times A^\complement) &\geq \frac{1}{4} \cdot \varepsilon \cdot \pi(S_3) \\
&\geq \frac{1}{4} \cdot \varepsilon \cdot \mathsf{d}(S_1, S_2) \cdot F(\min\{\pi(S_1), \pi(S_2)\}) \\
&\geq \frac{1}{4} \cdot \varepsilon \cdot \delta \cdot F(\min\{\theta \cdot \pi(A), \theta \cdot \pi(A^\complement)\}).
\end{aligned}
$$

We conclude by combining this inequality with (12) and considering A with $\pi(A) \leq \frac{1}{2}$. □

PROOF OF THEOREM 18.   By Corollary 16, we have

$$
\Phi_P(v) \geq \frac{1}{4} \cdot \varepsilon \cdot \min\left\{1, \frac{1}{2} \cdot \delta \cdot \frac{\tilde{I}_\pi(\frac{1}{2} \cdot v)}{\frac{1}{2} \cdot v}\right\}, \quad v \in (0, 1/2],
$$

and the bounds on Φ_P^* and γ_P follow then from Definition 3, Lemma 5 and positivity of P. Writing $h = \frac{d\mu}{d\pi}$ and $u_n := \mathrm{Var}_\pi(P^n h) = \chi^2(\mu P^n, \pi)$, we recall by Theorem 8 that in order to ensure that $u_n \leq \varepsilon_{\mathrm{Mix}}$, it suffices to take

$$
n \geq 2 + 4 \cdot \int_{\min\{4 \cdot u_0^{-1}, 1/2\}}^{1/2} \frac{dv}{v \cdot \Phi_P(v)^2} + [\Phi_P^*]^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\mathrm{Mix}}}, 1\right\}\right).
$$

Then for $v \in (0, v_*)$, as defined in (8), it holds that $\Phi_P(v) \geq \frac{1}{4} \cdot \varepsilon$, and for $v \in (v_*, \frac{1}{2})$, it holds that $\Phi_P(v) \geq \frac{1}{8} \cdot \varepsilon \cdot \delta \cdot \frac{\tilde{I}_\pi(\frac{1}{2} \cdot v)}{\frac{1}{2} \cdot v}$. One thus writes

$$
\begin{aligned}
\int_{\min\{4 \cdot u_0^{-1}, 1/2\}}^{1/2} \frac{1}{v \cdot \Phi_P(v)^2} \, dv &= \int_{\min\{4 \cdot u_0^{-1}, 1/2\}}^{\max\{\min\{4 \cdot u_0^{-1}, 1/2\}, v_*\}} \frac{1}{v \cdot \Phi_P(v)^2} \, dv \\
&\quad + \int_{\max\{\min\{4 \cdot u_0^{-1}, 1/2\}, v_*\}}^{1/2} \frac{1}{v \cdot \Phi_P(v)^2} \, dv.
\end{aligned}
$$

We treat the two integrals separately. For the first, write

$$\int_{\min\{4\cdot u_0^{-1}, 1/2\}}^{\max\{\min\{4\cdot u_0^{-1}, 1/2\}, v_*\}} \frac{dv}{v \cdot \Phi_P(v)^2} \leq \int_{\min\{4\cdot u_0^{-1}, 1/2\}}^{\max\{\min\{4\cdot u_0^{-1}, 1/2\}, v_*\}} \frac{1}{v \cdot (\frac{1}{4} \cdot \varepsilon)^2} dv$$

$$= 2^4 \cdot \varepsilon^{-2} \cdot \log\left(\frac{\max\{\min\{4 \cdot u_0^{-1}, 1/2\}, v_*\}}{\min\{4 \cdot u_0^{-1}, \frac{1}{2}\}}\right)$$

$$= 2^4 \cdot \varepsilon^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_*^{-1}}\right), 0\right\},$$

where the final equality follows from a case-by-case analysis. For the second, write

$$\int_{\max\{\min\{4\cdot u_0^{-1}, 1/2\}, v_*\}}^{1/2} \frac{dv}{v \cdot \Phi_P(v)^2}$$

$$\leq \int_{\max\{\min\{4\cdot u_0^{-1}, 1/2\}, v_*\}}^{1/2} \frac{1}{v \cdot (\frac{1}{8} \cdot \varepsilon \cdot \delta \cdot \frac{\tilde{I}_\pi(\frac{1}{2} \cdot v)}{\frac{1}{2} \cdot v})^2} dv$$

$$= 2^6 \cdot \varepsilon^{-2} \cdot \delta^{-2} \cdot \int_{\max\{\min\{4\cdot u_0^{-1}, 1/2\}, v_*\}}^{1/2} \frac{(\frac{1}{2} \cdot v)^2}{v \cdot \tilde{I}_\pi(\frac{1}{2} \cdot v)^2} dv$$

$$= 2^6 \cdot \varepsilon^{-2} \cdot \delta^{-2} \cdot \int_{\max\{\min\{4\cdot u_0^{-1}, 1/2\}, v_*\}}^{1/2} \frac{(\frac{1}{2} \cdot v)}{\tilde{I}_\pi(\frac{1}{2} \cdot v)^2} d\left(\frac{1}{2} \cdot v\right)$$

$$= 2^6 \cdot \varepsilon^{-2} \cdot \delta^{-2} \cdot \int_{\max\{\min\{2\cdot u_0^{-1}, 1/4\}, v_*/2\}}^{1/4} \frac{\xi}{\tilde{I}_\pi(\xi)^2} d\xi.$$

For the final term, we have $\Phi_P^* = \Phi_P(\frac{1}{2}) \geq 2^{-2} \cdot \varepsilon \cdot \min\{1, 2 \cdot \delta \cdot \tilde{I}_\pi(\frac{1}{4})\}$. We conclude by combining these bounds. $\square$

PROOF OF LEMMA 37. This is obtained via Pinsker's inequality. First compute directly that

$$KL(Q(x, \cdot), Q(y, \cdot)) = \frac{1}{2 \cdot \sigma^2} \cdot |x - y|^2.$$

Recalling Pinsker's inequality, we deduce that

$$\|Q(x, \cdot) - Q(y, \cdot)\|_{TV} \leq \left(\frac{1}{2} \cdot KL(Q(x, \cdot), Q(y, \cdot))\right)^{1/2} = \frac{1}{2 \cdot \sigma} \cdot |x - y|. \qquad \square$$

PROOF OF PROPOSITION 44. First, let v_1 be a median of x_1, the first coordinate of x, under π, and let $x_{*,1}$ be the mode of the marginal law of x_1 under π, which exists and is unique as a consequence of Lemma 43. Now, define $A = \{x \in E : x_1 \geq v_1\}$, so that $\pi(A) = \frac{1}{2}$. We let $Z \sim \mathcal{N}(0, I_d)$, and by neglecting the acceptance probability, we obtain the bounds

$$(\pi \otimes P)(A \times A^\complement) = \int_A \pi(dx) P(x, A^\complement)$$

$$= \int_A \pi(dx) \int_E \mathcal{N}(dz; 0, I_d) \cdot \min\left\{1, \frac{\pi(x + \sigma \cdot z)}{\pi(x)}\right\} \cdot \mathbf{1}_{A^\complement}(x + \sigma \cdot z)$$

$$\leq \int_A \pi(dx) \mathbb{P}(x + \sigma \cdot Z \in A^\complement)$$

$$= \int_A \pi(dx)\mathbb{P}(x_1 + \sigma \cdot Z_1 < v_1)$$

$$= \int_{x_1 \geq v_1} \pi_1(dx_1)\mathbb{P}(x_1 + \sigma \cdot Z_1 < v_1),$$

where π_1 is the marginal law of x_1 under π. Recall that by Lemma 43, π_1 will also be m-strongly log-concave and admit an L-smooth potential, and so we may apply Lemma 42 to control the density of π_1 as $\pi_1(dx_1) \leq (\frac{L}{m})^{1/2} \cdot \mathcal{N}(dx_1; x_{*,1}, m^{-1})$. Substituting this, we may bound

$$(\pi \otimes P)(A \times A^{\complement}) \leq \left(\frac{L}{m}\right)^{1/2} \cdot \int_{x_1 \geq v_1} \mathcal{N}(dx_1; x_{*,1}, m^{-1})\mathbb{P}(x_1 + \sigma \cdot Z_1 < v_1)$$

$$= \left(\frac{L}{m}\right)^{1/2} \cdot \int_{x_1 \geq v_1} \mathcal{N}(dx_1; x_{*,1}, m^{-1})\mathbb{P}\left(Z_1 < \frac{v_1 - x_1}{\sigma}\right)$$

$$\leq \left(\frac{L}{m}\right)^{1/2} \cdot \int_{x_1 \geq v_1} \mathcal{N}(dx_1; x_{*,1}, m^{-1}) \exp\left(-\frac{1}{2 \cdot \sigma^2} \cdot (v_1 - x_1)^2\right)$$

$$= \left(\frac{L}{2\pi}\right)^{1/2} \cdot \int_{x_1 \geq v_1} \exp\left(-\frac{m}{2} \cdot (x_1 - x_{*,1})^2 - \frac{1}{2 \cdot \sigma^2} \cdot (v_1 - x_1)^2\right) dx_1,$$

where we have used the Chernoff bound $\mathbb{P}(Z_1 \leq -z) \leq \exp(-\frac{1}{2}z^2)$ for $z > 0$, to move from the second to the third line.

Computing directly that

$$\frac{m}{2}(x_1 - x_{*,1})^2 + \frac{1}{2 \cdot \sigma^2} \cdot (v_1 - x_1)^2 = \frac{1}{2} \cdot \frac{1 + m \cdot \sigma^2}{\sigma^2} \cdot \left(x_1 - \frac{v_1 + m \cdot \sigma^2 \cdot x_{*,1}}{1 + m \cdot \sigma^2}\right)^2$$

$$+ \frac{m}{2 \cdot (1 + m \cdot \sigma^2)} \cdot (v_1 - x_{*,1})^2,$$

one sees that

$$\pi \otimes P(A \times A^{\complement}) \leq \left(\frac{L}{2\pi}\right)^{1/2} \cdot \int_{x_1 \geq v_1} \exp\left(-\frac{1}{2} \cdot \frac{1 + m \cdot \sigma^2}{\sigma^2} \cdot \left(x_1 - \frac{v_1 + m \cdot \sigma^2 \cdot x_{*,1}}{1 + m \cdot \sigma^2}\right)^2\right.$$

$$\times \exp\left(-\frac{m}{2 \cdot (1 + m \cdot \sigma^2)} \cdot (v_1 - x_{*,1})^2\right) dx_1$$

$$= \left(\frac{L \cdot \sigma^2}{1 + m \cdot \sigma^2}\right)^{1/2} \cdot \exp\left(-\frac{m}{2 \cdot (1 + m \cdot \sigma^2)} \cdot (v_1 - x_{*,1})^2\right)$$

$$\times \int_{x_1 \geq v_1} \mathcal{N}\left(dx_1; \frac{v_1 + m \cdot \sigma^2 \cdot x_{*,1}}{1 + m \cdot \sigma^2}, \frac{\sigma^2}{1 + m \cdot \sigma^2}\right)$$

$$\leq \left(\frac{L \cdot \sigma^2}{1 + m \cdot \sigma^2}\right)^{1/2} \cdot 1 \cdot 1$$

$$\leq L^{1/2} \cdot \sigma.$$

Recalling that $\pi(A) = \pi(A^{\complement}) = \frac{1}{2}$, the result follows. $\square$

PROOF OF PROPOSITION 45. By m-strong convexity of the potential, it holds that

$$U(x + \sigma \cdot z) - U(x) \geq \langle \nabla U(x), \sigma \cdot z \rangle + \frac{1}{2} \cdot m \cdot \sigma^2 \cdot |z|^2,$$

and substituting this into (9) we obtain

$$\alpha(x) \le \int \mathcal{N}(dz; 0, I_d) \cdot \min\left\{1, \exp\left(-\langle \nabla U(x), \sigma \cdot z \rangle - \frac{1}{2} \cdot m \cdot \sigma^2 \cdot |z|^2\right)\right\}.$$

Applying the inequality $\min\{1, c\} \le c$ establishes that

$$\alpha(x) \le \int \mathcal{N}(dz; 0, I_d) \cdot \exp\left(-\langle \nabla U(x), \sigma \cdot z \rangle - \frac{1}{2} \cdot m \cdot \sigma^2 \cdot |z|^2\right).$$

Straightforward computations show that

$$\mathcal{N}(dz; 0, I_d) \cdot \exp\left(-\langle \nabla U(x), \sigma \cdot z \rangle - \frac{1}{2} \cdot m \cdot \sigma^2 \cdot |z|^2\right)$$

$$= (1 + m \cdot \sigma^2)^{-d/2} \exp\left(\frac{1}{2} \cdot \frac{\sigma^2 \cdot |\nabla U(x)|^2}{1 + m \cdot \sigma^2}\right)$$

$$\times \mathcal{N}\left(dz; -\frac{\sigma}{1 + m \cdot \sigma^2} \cdot \nabla U(x), \frac{1}{1 + m \cdot \sigma^2} \cdot I_d\right),$$

which allows us to write

$$\alpha(x) \le (1 + m \cdot \sigma^2)^{-d/2} \cdot \exp\left(\frac{1}{2} \cdot \frac{\sigma^2}{1 + m \cdot \sigma^2} \cdot |\nabla U(x)|^2\right).$$

Now, for $\rho > 0$, define the set

$$B_\rho = \left\{x : \frac{\sigma^2}{1 + m \cdot \sigma^2} \cdot |\nabla U(x)|^2 \le \rho^2\right\},$$

which for ρ small enough will have $0 < \pi(B_\rho) < \pi(B_\rho^C)$. It then follows that

$$\Phi_P^* \le \frac{\pi \otimes P(B_\rho \times B_\rho^C)}{\pi(B_\rho)}$$

$$= \int \pi_{B_\rho}(dx) \cdot P(x, B_\rho^C)$$

$$\le \int \pi_{B_\rho}(dx) \cdot P(x, \{x\}^C)$$

$$= \int \pi_{B_\rho}(dx) \cdot \alpha(x)$$

$$\le (1 + m \cdot \sigma^2)^{-d/2} \cdot \exp\left(\frac{1}{2} \cdot \rho^2\right),$$

and taking an infimum as $\rho \to 0^+$ gives that $\Phi_P \le (1 + m \cdot \sigma^2)^{-d/2}$, as claimed.

For the subsequent remark, compute that

$$\log \Phi_P^* \le -\frac{1}{2} \cdot d \cdot \log(1 + m \cdot \sigma^2)$$

$$= -\frac{1}{2} \cdot d \cdot \log\left(1 + \frac{m}{L} \cdot \frac{\varsigma^2}{d^{2 \cdot \beta}}\right)$$

$$= -\frac{1}{2} \cdot d^{1 - 2 \cdot \beta} \cdot \left(d^{2 \cdot \beta} \cdot \log\left(1 + \frac{m}{L} \cdot \frac{\varsigma^2}{d^{2 \cdot \beta}}\right)\right)$$

$$= -\frac{1}{2} \cdot d^{1 - 2 \cdot \beta} \cdot \left(\frac{m}{L} \cdot \varsigma^2 + \mathcal{O}(d^{-2 \cdot \beta})\right)$$

$$= -\frac{1}{2} \cdot d^{1-2\cdot\beta} \cdot \frac{m}{L} \cdot \varsigma^2 \cdot \left(1 + \mathcal{O}(d^{-2\cdot\beta})\right)$$

$$\leq -\frac{1}{2} \cdot \frac{m}{L} \cdot \varsigma^2 \cdot d^{1-2\cdot\beta} \cdot \left(1 + o(1)\right).$$

Fixing $c \in (0, \frac{1}{2} \cdot \frac{m}{L} \cdot \varsigma^2)$, it holds for sufficiently large d that $\log \Phi_P^* \leq -c \cdot d^{1-2\cdot\beta}$. $\quad\square$

PROOF OF THEOREM 49. From Lemma 27, we can write $I_\pi(p) \geq \tilde{I}_\pi(p) := m^{1/2} \cdot (\varphi_\gamma \circ \Phi_\gamma^{-1})(p)$, which admits the bounds

$$\tilde{I}_\pi(p) \geq C_\ell \cdot m^{1/2} \cdot p \cdot \left(\log \frac{1}{p}\right)^{1/2}, \quad p \in (0, 1/2],$$

and $\tilde{I}_\pi(\frac{1}{4}) = m^{1/2} \cdot C_\gamma$. From Theorem 18,

$$n \geq 2 + 2^6 \cdot \varepsilon^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_*^{-1}}\right), 0\right\}$$

$$+ 2^8 \cdot \varepsilon^{-2} \cdot \delta^{-2} \cdot \int_{\max\{\min\{2\cdot u_0^{-1}, 1/4\}, v_*/2\}}^{1/4} \frac{\xi}{\tilde{I}_\pi(\xi)^2} \, d\xi$$

$$+ 2^4 \cdot \max\left\{1, 2^{-2} \cdot \delta^{-2} \cdot \tilde{I}_\pi\left(\frac{1}{4}\right)^{-2}\right\} \cdot \varepsilon^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right)$$

is sufficient, where v_* is defined in (8). Additionally, from Lemma 38, P is $(|\cdot|, \alpha_0 \cdot \sigma, \frac{1}{2} \cdot \alpha_0)$-close coupling. Substituting these values and using the lower bound on $\tilde{I}_\pi(p)$ to upper bound the integrand and lower bound $v_* \geq v_\circ := \min\{\frac{1}{2}, 2 \cdot \exp(-4 \cdot C_\ell^{-2} \cdot \sigma^{-2} \cdot \alpha_0^{-2} \cdot m^{-1})\}$ it suffices to take

$$n \geq 2 + 2^8 \cdot \alpha_0^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_\circ^{-1}}\right), 0\right\}$$

$$+ 2^{10} \cdot C_\ell^{-2} \cdot \alpha_0^{-4} \cdot \sigma^{-2} \cdot m^{-1} \cdot \int_{\max\{\min\{2\cdot u_0^{-1}, 1/4\}, v_\circ/2\}}^{1/4} \frac{1}{\xi \cdot \log(\frac{1}{\xi})} \, d\xi$$

$$+ 2^4 \cdot \max\{1, 2^{-2} \cdot C_\gamma^{-2} \cdot \alpha_0^{-2} \cdot \sigma^{-2} \cdot m^{-1}\} \cdot \alpha_0^{-2} \cdot \sigma^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right),$$

and computing that

$$0 < a < b < 1 \quad \Longrightarrow \quad \int_a^b \frac{1}{\xi \cdot \log(\frac{1}{\xi})} \, d\xi = \log\left(\frac{\log(1/a)}{\log(1/b)}\right)$$

provides the bound

$$n \geq 2 + 2^8 \cdot \alpha_0^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_\circ^{-1}}\right), 0\right\}$$

$$+ 2^{10} \cdot C_\ell^{-2} \cdot \alpha_0^{-4} \cdot \sigma^{-2} \cdot m^{-1} \cdot \log\left(\frac{\log(\min\{\max\{\frac{1}{2} \cdot u_0, 4\}, 2 \cdot v_\circ^{-1}\})}{\log 4}\right)$$

$$+ 2^4 \cdot \max\{1, 2^{-2} \cdot C_\gamma^{-2} \cdot \alpha_0^{-2} \cdot \sigma^{-2} \cdot m^{-1}\} \cdot \alpha_0^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right).$$

By Corollary 40 and $\exp(\varsigma^2) \cdot \varsigma^{-2} \geq \exp(1)$, we obtain

$$\max\{1, 2^{-2} \cdot C_\gamma^{-2} \cdot \alpha_0^{-2} \cdot \sigma^{-2} \cdot m^{-1}\} \leq \max\{1, C_\gamma^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d\}$$

$$= C_\gamma^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d,$$

and simplifying the other terms provides the bound

$$n \geq 2 + 2^{10} \cdot \exp(\varsigma^2) \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_\circ^{-1}}\right), 0\right\}$$

$$(13) \qquad + 2^{14} \cdot C_\ell^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d \cdot \log\left(\frac{\log(\min\{\max\{\frac{1}{2} \cdot u_0, 4\}, 2 \cdot v_\circ^{-1}\})}{\log 4}\right)$$

$$+ 2^6 \cdot C_\gamma^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right),$$

where we may also bound

$$v_\circ^{-1} = \max\left\{2, \frac{1}{2}\exp(4 \cdot C_\ell^{-2} \cdot \sigma^{-2} \cdot \alpha_0^{-2} \cdot m^{-1})\right\}$$

$$(14) \qquad \leq \max\left\{2, \frac{1}{2} \cdot \exp(16 \cdot C_\ell^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d)\right\}$$

$$\leq \frac{1}{2} \cdot \exp(16 \cdot C_\ell^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \kappa \cdot d).$$

For the stated bound, since $v_\circ^{-1} \geq 2$, we have $\frac{u_0}{4 \cdot v_\circ^{-1}} \leq u_0$, leading to the bound on the first term. The second term follows from

$$\log\left(\frac{\log(\min\{\max\{\frac{1}{2} \cdot u_0, 4\}, 2 \cdot v_\circ^{-1}\})}{\log 4}\right) \leq \log\left(\frac{\log(2 \cdot v_\circ^{-1})}{\log 4}\right)$$

$$\leq \log(\log(2 \cdot v_\circ^{-1}))$$

$$\leq \log(16 \cdot C_\ell^{-2} \cdot \varsigma^{-2} \cdot \kappa \cdot d) + \varsigma^2,$$

where we have used the bound on $v_\circ^{-1}$ to arrive at the final inequality. □

PROOF OF LEMMA 56. Compute directly that

$$\text{KL}(Q(x, \cdot), Q(y, \cdot)) = \frac{1}{2} \cdot \frac{\rho^2}{\eta^2} \cdot |x - y|_{\mathsf{C}^{-1}}^2.$$

By Pinsker's inequality, it thus holds that

$$\|Q(x, \cdot) - Q(y, \cdot)\|_{\text{TV}} \leq \frac{1}{2} \cdot \frac{\rho}{\eta} \cdot |x - y|_{\mathsf{C}^{-1}}. \qquad \Box$$

PROOF OF THEOREM 59. The proof structure broadly follows that of Theorem 49; certain details which are omitted here are spelt out more clearly in that proof.

By Lemma 55, π admits the regular, concave $|\cdot|_{\mathsf{C}^{-1}}$-isoperimetric minorant $\varphi_\gamma \circ \Phi_\gamma^{-1}$, and by Lemma 57, P satisfies a close coupling inequality with $\delta = \alpha_0 \cdot \frac{\eta}{\rho}$, $\varepsilon = \frac{1}{2} \cdot \alpha_0$. By Theorem 18, it suffices to take

$$n \geq 2 + 2^6 \cdot \varepsilon^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_\circ^{-1}}\right), 0\right\}$$

$$+ 2^8 \cdot \varepsilon^{-2} \cdot \delta^{-2} \cdot \int_{\max\{\min\{2 \cdot u_0^{-1}, 1/4\}, v_\circ/2\}}^{1/4} \frac{\xi}{\tilde{I}_\pi(\xi)^2} \, d\xi$$

$$+ 2^4 \cdot \max\left\{1, 2^{-2} \cdot \delta^{-2} \cdot \tilde{I}_\pi\left(\frac{1}{4}\right)^{-2}\right\} \cdot \varepsilon^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right).$$

Recalling that $\varphi_\gamma \circ \Phi_\gamma^{-1} \geq C_\ell \cdot p \cdot (\log \frac{1}{p})^{1/2}$ for $p \in [0, \frac{1}{2}]$, we thus simplify to

$$n \geq 2 + 2^8 \cdot \alpha_0^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_\circ^{-1}}\right), 0\right\}$$

$$+ 2^{10} \cdot C_\ell^{-2} \cdot \alpha_0^{-4} \cdot \frac{\rho^2}{\eta^2} \cdot \int_{\max\{\min\{2 \cdot u_0^{-1}, 1/4\}, v_\circ/2\}}^{1/4} \frac{1}{\xi \cdot \log(\frac{1}{\xi})} \, d\xi$$

$$+ 2^6 \cdot \max\left\{1, 2^{-2} \cdot C_\gamma^{-2} \cdot \alpha_0^{-2} \cdot \frac{\rho^2}{\eta^2}\right\} \cdot \alpha_0^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right)$$

and repeating earlier calculations with the inner integral gives the bound

$$n \geq 2 + 2^8 \cdot \alpha_0^{-2} \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_\circ^{-1}}\right), 0\right\}$$

$$+ 2^{10} \cdot C_\ell^{-2} \cdot \alpha_0^{-4} \cdot \frac{\rho^2}{\eta^2} \cdot \log\left(\frac{\log(\min\{\max\{\frac{1}{2} \cdot u_0, 4\}, 2 \cdot v_\circ^{-1}\})}{\log 4}\right)$$

$$+ 2^6 \cdot \max\left\{1, 2^{-2} \cdot C_\gamma^{-2} \cdot \alpha_0^{-2} \cdot \frac{\rho^2}{\eta^2}\right\} \cdot \alpha_0^{-2} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right),$$

where $v_\circ := \min\{\frac{1}{2}, 2 \cdot \exp(-4 \cdot C_\ell^{-2} \cdot \alpha_0^{-2} \cdot \frac{\rho^2}{\eta^2})\}$.

We observe that

$$\inf_{\varsigma \in (0, L^{1/2} \cdot \text{Tr}(C)^{1/2})} \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa} = \inf_{\eta \in (0,1)} \exp(\eta^2 \cdot \tilde{\kappa}) \cdot \eta^{-2}$$

$$= \begin{cases} \exp(\tilde{\kappa}), & \tilde{\kappa} \leq 1, \\ \tilde{\kappa} \cdot \exp(1), & \tilde{\kappa} > 1, \end{cases}$$

$$\geq 1.$$

Using Lemma 58, the bound above and $\rho^2 \leq 1$, we obtain

$$\max\left\{1, 2^{-2} \cdot C_\gamma^{-2} \cdot \alpha_0^{-2} \cdot \frac{\rho^2}{\eta^2}\right\} \leq \max\{1, C_\gamma^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa}\}$$

$$= C_\gamma^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa},$$

providing the bound

$$n \geq 2 + 2^{10} \cdot \exp(\varsigma^2) \cdot \max\left\{\log\left(\frac{u_0}{4 \cdot v_\circ^{-1}}\right), 0\right\}$$

$$+ 2^{14} \cdot C_\ell^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa} \cdot \log\left(\frac{\log(\min\{\max\{\frac{1}{2} \cdot u_0, 4\}, 2 \cdot v_\circ^{-1}\})}{\log 4}\right)$$

$$+ 2^6 \cdot C_\gamma^{-2} \cdot \exp(2 \cdot \varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa} \cdot \log\left(\max\left\{\frac{\min\{u_0, 8\}}{\varepsilon_{\text{Mix}}}, 1\right\}\right),$$

where

$$v_\circ^{-1} = \max\left\{2, \frac{1}{2} \exp\left(4 \cdot C_\ell^{-2} \cdot \sigma^{-2} \cdot \alpha_0^{-2} \cdot \frac{\rho^2}{\eta^2}\right)\right\}$$

$$\leq \max\left\{2, \frac{1}{2} \cdot \exp(16 \cdot C_\ell^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa})\right\}$$

$$\leq \frac{1}{2} \cdot \exp(16 \cdot C_\ell^{-2} \cdot \exp(\varsigma^2) \cdot \varsigma^{-2} \cdot \tilde{\kappa}).$$

The stated bound is obtained by the following observations. Since $v_\circ^{-1} \geq 2$, $\frac{u_0}{4 \cdot v_\circ^{-1}} \leq u_0$, leading to the bound on the first term. The bound on the second term follows from

$$\log\left(\frac{\log(\min\{\max\{\frac{1}{2} \cdot u_0, 4\}, 2 \cdot v_\circ^{-1}\})}{\log 4}\right) \leq \log\left(\frac{\log(2 \cdot v_\circ^{-1})}{\log 4}\right)$$
$$\leq \log(\log(2 \cdot v_\circ^{-1}))$$
$$\leq \log(16 \cdot C_\ell^{-2} \cdot \varsigma^{-2} \cdot \tilde{\kappa}) + \varsigma^2,$$

where we have used the bound on $v_\circ^{-1}$ to arrive at the final inequality. $\quad\square$

PROPOSITION 62. *Suppose that the π-invariant Markov kernel P satisfies the optimized spectral profile inequality*

$$\mathcal{E}(P, g) \geq \mathrm{Var}_\pi(g) \cdot F\left(\frac{\pi(g)^2}{\mathrm{Var}_\pi(g)}\right)$$

for all nonnegative, nonconstant π-a.s. $g \in L^2(\pi)$, where F is positive, decreasing and $\limsup_{v \to 0^+} F(v) > 0$ (and may be infinite). Then P also satisfies a super-Poincaré inequality of the form

$$\mathrm{Var}_\pi(f) \leq s \cdot \mathcal{E}(P, f) + \beta(s) \cdot \pi(|f|)^2,$$

where β is positive, decreasing, and can be written explicitly.

PROOF. For $w \geq 0$, define

$$F^*(w) = \inf_{v \geq 0}\{F(v) + w \cdot v\}.$$

Note that F^* is positive, increasing, and id/F^* is increasing. By our assumption that $\limsup_{v \to 0^+} F(v) > 0$, it also holds for $w > 0$ that $F^*(w) > 0$. One can then write

$$F(v) \geq \sup_{w \geq 0}\{F^*(w) - w \cdot v\},$$

and hence for any $w > 0$ and nonnegative, nonconstant $g \in L^2(\pi)$ that

$$\mathcal{E}(P, g) \geq \mathrm{Var}_\pi(g) \cdot \left\{F^*(w) - w \cdot \frac{\pi(g)^2}{\mathrm{Var}_\pi(g)}\right\}$$
$$\implies \quad \mathrm{Var}_\pi(g) \leq \frac{1}{F^*(w)} \cdot \mathcal{E}(P, g) + \frac{w}{F^*(w)} \cdot \pi(g)^2.$$

Writing $s = \frac{1}{F^*(w)}$ and $\beta_0 = (\mathrm{Id}/F^*) \circ (1/F^*)$ (which is decreasing), we see for nonnegative $g \in L^2(\pi)$ that

$$\mathrm{Var}_\pi(g) \leq s \cdot \mathcal{E}(P, g) + \beta_0(s) \cdot \pi(g)^2.$$

Now, let $f \in L^2(\pi)$ and write $f_\pm = \max(\pm f, 0) \geq 0$, so that

$$\mathrm{Var}_\pi(f_\pm) \leq s \cdot \mathcal{E}(P, f_\pm) + \beta_0(s) \cdot \pi(f_\pm)^2.$$

Standard calculations give that $\mathrm{Var}_\pi(f) \leq 2 \cdot (\mathrm{Var}_\pi(f_+) + \mathrm{Var}_\pi(f_-))$, $\mathcal{E}(P, f_+) + \mathcal{E}(P, f_-) \leq \mathcal{E}(P, f)$ (Goel, Montenegro and Tetali (2006), Lemma 2.3), and $\pi(f_+)^2 + \pi(f_-)^2 \leq \pi(|f|)^2$. Assembling these inequalities yields that

$$\mathrm{Var}_\pi(f) \leq 2 \cdot \left(\mathrm{Var}_\pi(f_+) + \mathrm{Var}_\pi(f_-)\right)$$
$$\leq 2 \cdot s \cdot \left(\mathcal{E}(P, f_+) + \mathcal{E}(P, f_-)\right) + 2 \cdot \beta_0(s) \cdot \left(\pi(f_+)^2 + \pi(f_-)^2\right)$$
$$\leq 2 \cdot s \cdot \mathcal{E}(P, f) + 2 \cdot \beta_0(s) \cdot \pi(|f|)^2,$$

that is, that

$$\mathrm{Var}_\pi(f) \leq s \cdot \mathcal{E}(P, f) + \beta(s) \cdot \pi(|f|)^2,$$

where $\beta : s \mapsto 2 \cdot \beta_0(\frac{1}{2} \cdot s)$. The result follows. $\quad\square$

APPENDIX C: DIFFERENT PROPOSAL DISTRIBUTIONS FOR RWM

For simplicity, the analysis of the RWM in Sections 4–5 involved specifically Gaussian proposal distributions. We extend the results here to a wide class of proposal distributions with independent noise increments for each of the d components. That is, we define

$$Q(x, A) = \int \mathbf{1}_A(x + \sigma \cdot z) q^{\otimes d}(\mathrm{d}z), \quad x \in \mathsf{E}, A \in \mathscr{E},$$

where q is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that is symmetric, that is, $q(A) = q(-A)$ for all $A \in \mathcal{B}(\mathbb{R})$ where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra of $\mathbb{R}$. We define for $t \in \mathbb{R}$, q_t to be the distribution of $Z + t$ where $Z \sim q$. We define the squared Hellinger distance between P and Q as

$$d_\mathrm{H}(P, Q)^2 = \int \left(\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}(x)\right)^{1/2} - \left(\frac{\mathrm{d}Q}{\mathrm{d}\lambda}(x)\right)^{1/2} \right)^2 \lambda(\mathrm{d}x),$$

where λ is any common dominating reference measure. This is a suitable metric to use because of its tensorization properties and because by Le Cam's inequalities (see, e.g., Tsybakov (2009), Section 2.4), it holds that

$$\|P - Q\|_\mathrm{TV} \in \left[\frac{1}{2} \cdot d_\mathrm{H}(P, Q)^2, d_\mathrm{H}(P, Q) \right].$$

Hence, in order to control $\|P - Q\|_\mathrm{TV} \in \Theta(1)$, it is necessary and sufficient to control $d_\mathrm{H}(P, Q) \in \Theta(1)$.

PROPOSITION 63. *Let U be a potential such that*

$$(15) \qquad U(x + h) - U(x) - \langle \nabla U(x), h \rangle \leq \sum_{i=1}^d \psi(|h_i|),$$

for some $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ which is nondecreasing and satisfies $\psi(0) = 0$. Assume there exists $L_H > 0$ such that

$$d_\mathrm{H}(q, q_t)^2 \leq \frac{1}{2} \cdot L_H \cdot t^2 \quad \text{for } t \in \mathbb{R}.$$

Assume also for $\sigma > 0$ that

$$\int_\mathbb{R} q(\mathrm{d}z_1) \cdot \exp(-\psi(\sigma \cdot |z_1|)) \geq 1 - \xi(\sigma)$$

for some $\xi : \mathbb{R}_+ \to \mathbb{R}_+$ which is continuous, nondecreasing and satisfies $\xi(0) = 0$. Let $\eta > 0$ and define P_η to be the RWM kernel with proposal scaling σ satisfying $\xi(\sigma) = d^{-1} \cdot \eta$. It then holds that

1. *P_η has minimal acceptance rate $\alpha_0 \geq \frac{1}{2} \cdot (1 - \frac{\eta}{d})^d \geq \frac{1}{2} \cdot \exp(-\eta \cdot \frac{d}{d-\eta})$.*
2. *The kernel P_η is $(|\cdot|, \sigma \cdot (2 \cdot L_H)^{-1/2} \cdot \alpha_0, \frac{1}{2} \cdot \alpha_0)$-close coupling.*

PROOF. For the first part, following the same steps in the proof of Lemma 39, we have

$$\alpha_0 \geq \frac{1}{2} \int q^{\otimes d}(\mathrm{d}z) \cdot \exp\left(-\sum_{i=1}^{d} \psi(\sigma \cdot |z_i|)\right)$$

$$= \frac{1}{2}\left\{\int q(\mathrm{d}z_1) \cdot \exp(-\psi(\sigma \cdot |z_1|))\right\}^d,$$

from which the result follows. For the second part, we may write

$$\|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} = \left\|\bigotimes_{i=1}^{d} q_{\frac{x_i - y_i}{\sigma}} - \bigotimes_{i=1}^{d} q\right\|_{\mathrm{TV}}.$$

Moreover, we have in general that

$$1 - \frac{1}{2} \cdot d_{\mathrm{H}}\left(\bigotimes_{i=1}^{d} P_i, \bigotimes_{i=1}^{d} Q_i\right)^2 = \prod_{i=1}^{d}\left\{1 - \frac{1}{2} \cdot d_{\mathrm{H}}(P_i, Q_i)^2\right\}$$

$$\geq 1 - \sum_{i=1}^{d} \frac{1}{2} \cdot d_{\mathrm{H}}(P_i, Q_i)^2.$$

We then see that

$$\frac{1}{2} \cdot \left\|\bigotimes_{i=1}^{d} q_{\frac{x_i - y_i}{\sigma}} - \bigotimes_{i=1}^{d} q\right\|_{\mathrm{TV}}^2 \leq \sum_{i=1}^{d} \frac{1}{2} \cdot d_{\mathrm{H}}(q_{\frac{x_i - y_i}{\sigma}}, q)^2$$

$$\leq \sum_{i=1}^{d} \frac{L_H \cdot |x_i - y_i|^2}{4 \cdot \sigma^2}$$

$$= \frac{L_H \cdot |x - y|^2}{4 \cdot \sigma^2},$$

so that $\|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq (\frac{L_H}{2})^{1/2} \cdot \frac{|x-y|}{\sigma}$ in general. In particular, taking $\mathsf{d}(x, y) = |x - y| \leq \sigma \cdot (2 \cdot L_H)^{-1/2} \cdot \alpha_0$, we obtain that $\|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq \frac{1}{2}\alpha_0$. We may then conclude by Lemma 19. □

The conditions may be verified in various settings with specific choices of q. If $\psi(x) = x^\alpha$ for some $\alpha \in [0, 2]$ and q has finite αth moment then a bound on α_0 is straightforward: if $\sigma = \varsigma \cdot d^{-1/\alpha}$ then by Jensen's inequality,

$$\alpha_0 \geq \frac{1}{2} \cdot \exp\left(-d \int q(\mathrm{d}z) \cdot \sigma^\alpha |z|^\alpha\right) = \frac{1}{2} \cdot \exp\left(-\varsigma^\alpha \int q(\mathrm{d}z)|z|^\alpha\right).$$

If q has only a smaller moment, then the next Lemma shows that it is still possible to obtain a bound. For example, if $\alpha = 2$ and q has a finite second moment, we obtain the same scaling $\xi(\sigma) \sim \sigma^2$ as for Gaussian proposals. If q has fewer moments, or U is rougher, the scaling of σ with dimension is more severe. For example, if q is Cauchy and $\alpha = 2$ then one can only obtain s arbitrarily close to 1 in Lemma 64, and so a scaling of $\xi(\sigma) \sim \sigma$ which leads to scaling $\sigma \sim 1/d$. In fact, the calculations for the Cauchy can be done exactly, such that we have

$$\int q(\mathrm{d}z) \cdot \exp(-\sigma^2 \cdot z^2) = \exp(\sigma^2) \cdot \left\{1 - \frac{2}{\pi} \int_0^\sigma \exp(-t^2) \, \mathrm{d}t\right\},$$

which gives the same scaling.

LEMMA 64. *Suppose $\psi(x) = x^\alpha$ for some $\alpha \in [0, 2]$, and assume q has finite rth moment for some $r > 0$. Then*

$$\xi(\sigma) \leq \sigma^s \cdot \mathbb{E}_q[|Z|^s],$$

where $s = \min\{\alpha, r\}$.

PROOF. Observe that $\exp(-t) \geq 1 - t \geq 1 - t^\beta$ for any $\beta \in [0, 1]$. Then

$$\int q(\mathrm{d}z) \cdot \exp(-\psi(\sigma \cdot |z|)) \geq 1 - \int q(\mathrm{d}z) \cdot \psi(\sigma \cdot |z|)^\beta$$
$$= 1 - \sigma^{\alpha\beta} \mathbb{E}_q[|Z|^{\alpha\beta}],$$

for $\beta \in [0, 1]$. We conclude by taking $\beta = \min\{r/\alpha, 1\}$. □

One approach to verifying the Hellinger condition is to follow asymptotic statistical theory using a nonasymptotic variant of differentiability in quadratic mean (van der Vaart (1998), Section 7.2). In particular, we will assume that the proposal q satisfies

$$(16) \qquad \int \left(q(x + t)^{1/2} - q(x)^{1/2} \cdot \left(1 + \frac{1}{2} \cdot s(x) \cdot t \right) \right)^2 \leq \varphi(|t|) \quad \text{as } t \to 0,$$

for some s which is square-integrable under q and some φ which vanishes at least quadratically around 0. The triangle inequality then yields

$$d_{\mathrm{H}}(q, q_t) \leq \frac{1}{2} \cdot |s|_{\mathrm{L}^2(q)} \cdot |t| + \varphi(|t|)^{1/2} \in \Theta(t),$$

as $t \to 0^+$; boundedness of the Hellinger distance allows one to conclude the existence of a suitable L_H. Such an estimate should thus hold for all proposals corresponding to "regular" statistical (location) models; see van der Vaart (1998), Example 7.8, for further discussion. Alternatively, if $\log q$ is differentiable then it may be convenient to bound $d_{\mathrm{H}}(q, q_t)$ using the KL divergence. For example, if $\log q$ is L_q-smooth and one defines $K(t) = \mathrm{KL}(q, q_t)$, we have $K(0) = K'(0) = 0$ and

$$|K'(s)| = \left| \int q(x)\{(\log q)'(x - s) - (\log q)'(x)\} \, \mathrm{d}x \right| \leq L_q |s|,$$

from which we may deduce that $d_{\mathrm{H}}(q, q_t)^2 \leq K(t) = \int_0^t K'(s) \, \mathrm{d}Fs \leq \frac{1}{2} L_q t^2$.

Finally, the there remains an additional subtlety concerning the positivity of the RWM Markov operator P, which we use to ensure that the spectral gap and right spectral gap coincide. The approach of Baxendale (2005) that we have used for normal increments applies quite generally to increment distributions which are decomposable in a particular sense, but does not hold in complete generality. Hence one may require alternative arguments to bound the left spectral gap or consider instead the Markov chain associated with the operator $P_{\text{lazy}} = \frac{1}{2}(P + \mathrm{Id})$, which is necessarily positive. In particular, translating our main arguments appropriately would then establish that the same quantitative results (in terms of how the mixing time scales with d, κ) hold for a lazy chain, up to some absolute constant factors.

REFERENCES

AGAPIOU, S., PAPASPILIOPOULOS, O., SANZ-ALONSO, D. and STUART, A. M. (2017). Importance sampling: Intrinsic dimension and computational cost. *Statist. Sci.* **32** 405–431. MR3696003 https://doi.org/10.1214/17-STS611

ANDRIEU, C., LEE, A., POWER, S. and WANG, A. Q. (2022). Poincaré inequalities for Markov chains: A meeting with Cheeger, Lyapunov and Metropolis. Technical report. Available at https://arxiv.org/abs/2208.05239.

BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **348**. Springer, Cham. MR3155209 https://doi.org/10.1007/978-3-319-00227-9

BARTHE, F. (2001). Levels of concentration between exponential and Gaussian. *Ann. Fac. Sci. Toulouse Math.* (6) **10** 393–404. MR1923685

BARTHE, F., CATTIAUX, P. and ROBERTO, C. (2006). Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry. *Rev. Mat. Iberoam.* **22** 993–1067. MR2320410 https://doi.org/10.4171/RMI/482

BARTHE, F. and MAUREY, B. (2000). Some remarks on isoperimetry of Gaussian type. *Ann. Inst. Henri Poincaré Probab. Stat.* **36** 419–434. MR1785389 https://doi.org/10.1016/S0246-0203(00)00131-X

BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987 https://doi.org/10.1214/105051604000000710

BELLONI, A. and CHERNOZHUKOV, V. (2009). On the computational complexity of MCMC-based estimators in large samples. *Ann. Statist.* **37** 2011–2055. MR2533478 https://doi.org/10.1214/08-AOS634

BOBKOV, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27** 1903–1921. MR1742893 https://doi.org/10.1214/aop/1022874820

BOBKOV, S. (2010). Perturbations in the Gaussian isoperimetric inequality. *J. Math. Sci.* **166**.

BOBKOV, S. G. and HOUDRÉ, C. (1997a). *Some Connections Between Isoperimetric and Sobolev-Type Inequalities* **616**. American Mathematical Soc.

BOBKOV, S. G. and HOUDRÉ, C. (1997b). Isoperimetric constants for product probability measures. *Ann. Probab.* **25** 184–205. MR1428505 https://doi.org/10.1214/aop/1024404284

BOBKOV, S. G. and ZEGARLINSKI, B. (2005). *Entropy Bounds and Isoperimetry*. American Mathematical Soc.

BOREL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216. MR0399402 https://doi.org/10.1007/BF01425510

CAFFARELLI, L. A. (2000). Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** 547–563. MR1800860 https://doi.org/10.1007/s002200000257

CARACCIOLO, S., PELISSETTO, A. and SOKAL, A. D. (1990). Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *J. Stat. Phys.* **60** 1–53. MR1063212 https://doi.org/10.1007/BF01013668

CATTIAUX, P. and GUILLIN, A. (2020). On the Poincaré constant of log-concave measures. In *Geometric Aspects of Functional Analysis. Vol. I. Lecture Notes in Math.* **2256** 171–217. Springer, Cham. MR4175748 https://doi.org/10.1007/978-3-030-36020-7_9

CHEN, Y., DWIVEDI, R., WAINWRIGHT, M. J. and YU, B. (2020). Fast mixing of metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients. *J. Mach. Learn. Res.* **21** Paper No. 92. MR4119160

CHRISTENSEN, O. F., ROBERTS, G. O. and ROSENTHAL, J. S. (2005). Scaling limits for the transient phase of local Metropolis–Hastings algorithms. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 253–268. MR2137324 https://doi.org/10.1111/j.1467-9868.2005.00500.x

DIACONIS, P. and SALOFF-COSTE, L. (1998). What do we know about the Metropolis algorithm? *J. Comput. System Sci.* **57** 20–36.

DOUC, R., MOULINES, E., PRIOURET, P. and SOULIER, P. (2018). *Markov Chains. Springer Series in Operations Research and Financial Engineering*. Springer, Cham. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005 https://doi.org/10.1093/biomet/asu075

DWIVEDI, R., CHEN, Y., WAINWRIGHT, M. J. and YU, B. (2019). Log-concave sampling: Metropolis–Hastings algorithms are fast. *J. Mach. Learn. Res.* **20** Paper No. 183. MR4048994

GÅSEMYR, J. (2006). The spectrum of the independent Metropolis–Hastings algorithm. *J. Theoret. Probab.* **19** 152–165. MR2256484 https://doi.org/10.1007/s10959-006-0009-2

GOEL, S., MONTENEGRO, R. and TETALI, P. (2006). Mixing time bounds via the spectral profile. *Electron. J. Probab.* **11** 1–26. MR2199053 https://doi.org/10.1214/EJP.v11-300

HAIRER, M., STUART, A. M. and VOLLMER, S. J. (2014). Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.* **24** 2455–2490. MR3262508 https://doi.org/10.1214/13-AAP982

HOLLEY, R. and STROOCK, D. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.* **46** 1159–1194. MR0893137 https://doi.org/10.1007/BF01011161

JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4

JARNER, S. F. and YUEN, W. K. (2004). Conductance bounds on the L^2 convergence rate of Metropolis algorithms on unbounded state spaces. *Adv. in Appl. Probab.* **36** 243–266. MR2035782 https://doi.org/10.1239/aap/1077134472

KOLESNIKOV, A. V. (2007). Modified log-Sobolev inequalities and isoperimetry. *Atti Accad. Naz. Lincei, Rend. Lincei, Mat. Appl.* **18** 179–208. MR2314171 https://doi.org/10.4171/RLM/489

KOZMA, G. (2007). On the precision of the spectral profile. *ALEA Lat. Am. J. Probab. Math. Stat.* **3** 321–329. MR2372888

LATAŁA, R. and OLESZKIEWICZ, K. (2000). Between Sobolev and Poincaré. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **1745** 147–168. Springer, Berlin. MR1796718 https://doi.org/10.1007/BFb0107213

LAWLER, G. F. and SOKAL, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309** 557–580. MR0930082 https://doi.org/10.2307/2000925

LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347 https://doi.org/10.1090/surv/089

LEDOUX, M. (2011). From concentration to isoperimetry: Semigroup proofs. In *Concentration, Functional Inequalities and Isoperimetry*. *Contemp. Math.* **545** 155–166. Amer. Math. Soc., Providence, RI. MR2858471 https://doi.org/10.1090/conm/545/10770

LIVINGSTONE, S. and ZANELLA, G. (2022). The Barker proposal: Combining robustness and efficiency in gradient-based MCMC. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 496–523. MR4412995 https://doi.org/10.1111/rssb.12482

LOVÁSZ, L. and KANNAN, R. (1999). Faster mixing via average conductance. In *Annual ACM Symposium on Theory of Computing* (*Atlanta, GA*, 1999) 282–287. ACM, New York. MR1798047 https://doi.org/10.1145/301250.301317

MATHÉ, P. and NOVAK, E. (2007). Simple Monte Carlo and the Metropolis algorithm. *J. Complexity* **23** 673–696. MR2372022 https://doi.org/10.1016/j.jco.2007.05.002

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.* **21** 1087–1092.

MICLO, L. and ROBERTO, C. (2000). Trous spectraux pour certains algorithmes de Métropolis sur **R**. In *Séminaire de Probabilités, XXXIV*. *Lecture Notes in Math.* **1729** 336–352. Springer, Berlin. MR1768073 https://doi.org/10.1007/BFb0103812

MILMAN, E. (2009a). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** 1–43. MR2507637 https://doi.org/10.1007/s00222-009-0175-9

MILMAN, E. (2009b). On the role of convexity in functional and isoperimetric inequalities. *Proc. Lond. Math. Soc.* (3) **99** 32–66. MR2520350 https://doi.org/10.1112/plms/pdn045

MILMAN, E. (2012). Properties of isoperimetric, functional and transport-entropy inequalities via concentration. *Probab. Theory Related Fields* **152** 475–507. MR2892954 https://doi.org/10.1007/s00440-010-0328-1

MORRIS, B. and PERES, Y. (2005). Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields* **133** 245–266. MR2198701 https://doi.org/10.1007/s00440-005-0434-7

MOU, W., FLAMMARION, N., WAINWRIGHT, M. J. and BARTLETT, P. L. (2022). An efficient sampling algorithm for non-smooth composite potentials. *J. Mach. Learn. Res.* **23** Paper No. 233. MR4577672

PAVLIOTIS, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations. Texts in Applied Mathematics* **60**. Springer, New York. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7

QIN, Q. and HOBERT, J. P. (2021). On the limitations of single-step drift and minorization in Markov chain convergence analysis. *Ann. Appl. Probab.* **31** 1633–1659. MR4312841 https://doi.org/10.1214/20-aap1628

ROBERTO, C. (2010). Isoperimetry for product of probability measures: Recent results. *Markov Process. Related Fields* **16** 617–634. MR2895085

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 https://doi.org/10.1214/aoap/1034625254

ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. MR1448322 https://doi.org/10.1214/ECP.v2-981

ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450 https://doi.org/10.1214/ss/1015346320

ROBERTS, G. O. and TWEEDIE, R. L. (2001). Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains. *J. Appl. Probab.* **38** 37–41.

RUDOLF, D. (2009). Explicit error bounds for lazy reversible Markov chain Monte Carlo. *J. Complexity* **25** 11–24. MR2475305 https://doi.org/10.1016/j.jco.2008.05.005

SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785 https://doi.org/10.1017/S0962492910000061

SUDAKOV, V. N. and TSIREL'SON, B. S. (1978). Extremal properties of half-spaces for spherically invariant measures. *J. Sov. Math.* **9** 9–18.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

VOGRINC, J. and KENDALL, W. S. (2021). Counterexamples for optimal scaling of Metropolis–Hastings chains with rough target densities. *Ann. Appl. Probab.* **31** 972–1019. MR4254501 https://doi.org/10.1214/20-aap1612

WANG, F.-Y. (2000). Functional inequalities for empty essential spectrum. *J. Funct. Anal.* **170** 219–245. MR1736202 https://doi.org/10.1006/jfan.1999.3516

WU, K., SCHMIDLER, S. and CHEN, Y. (2022). Minimax mixing time of the Metropolis-adjusted Langevin algorithm for log-concave sampling. *J. Mach. Learn. Res.* **23** Paper No. 270. MR4577709

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344 https://doi.org/10.1007/BF00353874