The Annals of Applied Probability
2012, Vol. 22, No. 5, 1880–1927
DOI: 10.1214/11-AAP817

OPTIMAL SCALING OF RANDOM WALK METROPOLIS ALGORITHMS WITH DISCONTINUOUS TARGET DENSITIES

BY PETER NEAL, GARETH ROBERTS AND WAI KONG YUEN

University of Manchester, University of Warwick and Brock University

We consider the optimal scaling problem for high-dimensional random walk Metropolis (RWM) algorithms where the target distribution has a discontinuous probability density function. Almost all previous analysis has focused upon continuous target densities. The main result is a weak convergence result as the dimensionality d of the target densities converges to ∞. In particular, when the proposal variance is scaled by d^{-2}, the sequence of stochastic processes formed by the first component of each Markov chain converges to an appropriate Langevin diffusion process. Therefore optimizing the efficiency of the RWM algorithm is equivalent to maximizing the speed of the limiting diffusion. This leads to an asymptotic optimal acceptance rate of e^{-2} (=0.1353) under quite general conditions. The results have major practical implications for the implementation of RWM algorithms by highlighting the detrimental effect of choosing RWM algorithms over Metropolis-within-Gibbs algorithms.

1. Introduction. Random walk Metropolis (RWM) algorithms are widely used generic Markov chain Monte Carlo (MCMC) algorithms. The ease with which RWM algorithms can be constructed has no doubt played a pivotal role in their popularity. The efficiency of a RWM algorithm depends fundamentally upon the scaling of the proposal density. Choose the variance of the proposal to be too small and the RWM will converge slowly since all its increments are small. Conversely, choose the variance of the proposal to be too large and too high a proportion of proposed moves will be rejected. Of particular interest is how the scaling of the proposal variance depends upon the dimensionality of the target distribution. The target distribution is the distribution of interest and the MCMC algorithm is constructed such that the stationary distribution of the Markov chain is the target distribution.

The Introduction is structured as follows. We outline known results for continuous independent and identically distributed product densities from [14] and subsequent work. We highlight the scope and limitations of the results before introducing the discontinuous target densities to be studied in this paper. While the statements of the key results (Theorem 2.1) in this paper are similar to those given for continuous target densities, the proofs are markedly different. A discussion of

Received February 2011; revised September 2011.
MSC2010 subject classifications. Primary 60F05; secondary 65C05.
Key words and phrases. Random walk Metropolis, Markov chain Monte Carlo, optimal scaling.

why a new method of proof is required for discontinuous target densities is given. Finally, we give an outline of the remainder of the paper.

The results of this paper have quite general consequences for the implementation of Metropolis algorithms on discontinuous densities (as are commonly applied in many Bayesian Statistics problems), namely:

(1) Full- (high-) dimensional update rules can be an order of magnitude slower than strategies involving smaller dimensional updates. (See Theorem 3.3 below.)

(2) For target densities with bounded support, Metropolis algorithms can be an order of magnitude slower than algorithms which first transform the target support to $\mathbb{R}^d$ for some d.

In [14], a sequence of target densities of the form

$$\pi_d(\mathbf{x}^d) = \prod_{i=1}^{d} f(x_i^d) \tag{1.1}$$

were considered as $d \to \infty$, where $f(\cdot)$ is twice differentiable and satisfies certain mild moment conditions; see [14], (A1) and (A2). The following random walk Metropolis algorithm was used to obtain a sample $\mathbf{X}_0^d, \mathbf{X}_1^d, \dots$ from $\pi_d(\cdot)$. Draw $\mathbf{X}_0^d$ from $\pi_d(\cdot)$. For $t \geq 0$ and $i = 1, 2, \dots$, let $Z_{t,i}$ be independent and identically distributed (i.i.d.) according to $Z \sim N(0, 1)$ and $\mathbf{Z}_t^d = (Z_{t,1}, Z_{t,2}, \dots, Z_{t,d})$. At time t, propose

$$\mathbf{Y}^d = \mathbf{X}_t^d + \sigma_d \mathbf{Z}_t^d, \tag{1.2}$$

where σ_d is the proposal standard deviation to be discussed shortly. Set $\mathbf{X}_{t+1}^d = \mathbf{Y}^d$ with probability

$$\alpha(\mathbf{X}_t^d, \mathbf{Y}^d) \equiv 1 \wedge \frac{\pi_d(\mathbf{Y}^d)}{\pi_d(\mathbf{X}_t^d)}. \tag{1.3}$$

Otherwise set $\mathbf{X}_{t+1}^d = \mathbf{X}_t^d$. It is straightforward to check that $\{\mathbf{X}_t^d\}$ has stationary distribution $\pi_d(\cdot)$, and hence, for all $t \geq 0$, $\mathbf{X}_t^d \sim \pi_d(\cdot)$. The key question addressed in [14] was: starting from the stationary distribution, how should σ_d be chosen to optimize the rate at which the RWM algorithm explores the stationary distribution? Since the components of $\mathbf{X}_t^d$ are i.i.d., it suffices to study the marginal behavior of the first component, $X_{t,1}^d$. In [14], it was shown that if $\sigma_d = l/\sqrt{d}$ $(l > 0)$ and $U_t^d = X_{[td],1}^d$ $(t \geq 0)$, then

$$U^d \Rightarrow U \qquad \text{as } d \to \infty, \tag{1.4}$$

where $U.$ satisfies the Langevin SDE

$$dU_t = \sqrt{h(l)}\, dB_t + \phi(l) \frac{f'(U_t)}{2f(U_t)}\, dt \tag{1.5}$$

with $U_0 \sim f(\cdot)$ and $h(l) = 2l^2\Phi(-l\sqrt{I}/2)$ with Φ being the standard normal c.d.f. and $I \equiv \mathbb{E}_f[\{f'(X)/f(X)\}^2]$. Note that the "speed measure" of the diffusion $\phi(l)$ only depends upon f through I. The diffusion limit for U^d is unsurprising in that for a time interval of length $s > 0$, $O(d)$ moves are made each of size $O(1/\sqrt{d})$. Therefore the movements in the first component (appropriately normalized) converge to those of a Langevin diffusion with the "most efficient" asymptotic diffusion having the largest speed measure $h(l)$. Since the diffusion limit involves speeding up time by a factor of d, we say that the mixing of the algorithm is $O(d)$. The optimal value of l is $\hat{l} = 2.38/\sqrt{I}$, which leads to an average optimal acceptance rate (aoar) of 0.234. This has major practical implications for practitioners, in that, to monitor the (asymptotic) efficiency of the RWM algorithm it is sufficient to study the proportion of proposed moves accepted.

There are three key assumptions made in [14]. First, $\mathbf{X}_0^d \sim \pi_d(\cdot)$, that is, the algorithm starts in the stationary distribution and σ_d is chosen to optimize exploration of the stationary distribution. This assumption has been made in virtually all subsequent optimal scaling work; see, for example, [3, 7, 10, 11] and [15]. The one exception is [8], where $\mathbf{X}_0^d$ is started from the mode of $\pi_d(\cdot)$ with explicit calculations given for a standard multivariate normal distribution. In [8], it is shown that $\sigma_d = O(1/\sqrt{d})$ is optimal for maximizing the rate of convergence to the stationary distribution. Since convergence is shown to occur within $O(\log d)$ iterations, the time taken to explore the stationary distribution dominates the time taken to converge to the stationary distribution, and thus overall it is optimal to choose $\sigma_d = \hat{l}/\sqrt{d}$. It is difficult to prove generic results for $\mathbf{X}_0^d \not\sim \pi_d$. However, the findings of [8] suggest that even when $\mathbf{X}_0^d \not\sim \pi_d$, it is best to scale the proposal distribution based upon $\mathbf{X}_0^d \sim \pi_d$. It is worth noting that in [8] it was found that for the Metropolis adjusted Langevin algorithm (MALA), the optimal scaling of σ_d for $\mathbf{X}_0^d$ started at the mode of a multivariate normal is $O(d^{-1/4})$ compared to $O(d^{-1/6})$ for $\mathbf{X}_0^d \sim \pi_d$.

Second, $\pi_d(\cdot)$ is an i.i.d. product density. This assumption has been relaxed by a number of authors with $\sigma_d = O(1/\sqrt{d})$ and an aoar of 0.234 still being the case, for example, independent, scaled product densities ([15] and [3]), Gibbs random fields [7], exchangeable normals [10] and elliptical densities [17]. Thus the simple rule of thumb of tuning σ_d such that one in four proposed moves are accepted holds quite generally. In [4] and [17], examples where the aoar is strictly less than 0.234 are given. These correspond to different orders of magnitude being appropriate for the scaling of the proposed moves in different components.

Third, the results are asymptotic as $d \to \infty$. However, simulations have shown that for i.i.d. product densities an acceptance rate of 0.234 is close to optimal for $d = 10$; see, for example, [10]. Departures from the i.i.d. product density require larger d for the asymptotic results to be optimal, but $d = 100$ is often seen in practical MCMC problems. In [12] and [16], optimal acceptance rates are obtained for finite d for some special cases.

With the exceptions of [11, 12] and [17], in the above works π_d is assumed to have a continuous (and suitably differentiable) probability density function (p.d.f.). The aim of the current work is to investigate the situation where the target distribution has a discontinuous p.d.f., and specifically, target distributions confined to the d-dimensional hypercube $[0, 1]^d$. That is, we consider target distributions of the form

$$\pi_d(\mathbf{x}^d) = \prod_{i=1}^{d} f(x_i^d), \tag{1.6}$$

where

$$f(x) \propto \exp(g(x)) 1_{\{0<x<1\}} \qquad (x \in \mathbb{R}) \tag{1.7}$$

and $g(\cdot)$ is twice differentiable upon $[0, 1]$ with

$$g^* = \sup_{0 \le y \le 1} |g'(y)| < \infty. \tag{1.8}$$

We then use the following random walk Metropolis algorithm to obtain a sample $\mathbf{X}_0^d, \mathbf{X}_1^d, \dots$ from $\pi_d(\cdot)$. Draw $\mathbf{X}_0^d$ from $\pi_d(\cdot)$. For $t \ge 0$ and $i = 1, 2, \dots$, let Z_{ti} be independent and identically distributed (i.i.d.) according to $Z \sim U[-1, 1]$ and $\mathbf{Z}_t^d = (Z_{t1}, Z_{t2}, \dots, Z_{td})$. At time t, propose

$$\mathbf{Y}^d = \mathbf{X}_t^d + \sigma_d \mathbf{Z}_t^d. \tag{1.9}$$

Set $\mathbf{X}_{t+1}^d = \mathbf{Y}^d$ with probability

$$\alpha(\mathbf{X}_t^d, \mathbf{Y}^d) \equiv 1 \wedge \frac{\pi_d(\mathbf{Y}^d)}{\pi_d(\mathbf{X}_t^d)}. \tag{1.10}$$

Otherwise set $\mathbf{X}_{t+1}^d = \mathbf{X}_t^d$.

In [11] and [17], spherical and elliptical densities are considered which have very different geometry to the hypercube restricted densities. Therefore different approaches are taken in these papers with results akin to those obtained for continuous target densities. Densities of the form (1.7) have previously been studied in [12], where the expected square jumping distance (ESJD) has been computed. The ESJD is

$$\mathbb{E}_{\pi_d}\left[\sum_{i=1}^{d} (X_{1,i}^d - X_{0,i}^d)^2\right] = d\mathbb{E}_{\pi_d}[(X_{1,1}^d - X_{0,1}^d)^2], \tag{1.11}$$

the mean squared distance between $\mathbf{X}_0^d$ and $\mathbf{X}_1^d$, where $\mathbf{X}_0^d \sim \pi_d$. In [12], Appendix B, it is shown that for $\sigma_d = l/d$ $(l > 0)$ and $f(x) = 1_{\{0<x<1\}}$,

$$d\mathbb{E}_{\pi_d}\left[\sum_{i=1}^{d} (X_{1,i}^d - X_{0,i}^d)^2\right] \to \frac{l^2}{3} \exp\left(-\frac{l}{2}\right) \qquad \text{as } d \to \infty. \tag{1.12}$$

Thus asymptotically (as $d \to \infty$) the ESJD is maximized by taking $\hat{l} = 4$ which corresponds to an aoar of $\exp(-2)$ (=0.1353). In this paper, we show that $\sigma_d = l/d$ and an aoar of $\exp(-2)$ holds more generally for target distributions of the form given by (1.6) and (1.7). Moreover, we prove a much stronger result than that given in [12], in that, we prove that $V_s^d = X_{[sd^2],1}^d$ converges weakly to an appropriate Langevin diffusion V_s with speed measure $\phi(l) = (l^2/3)\exp(-l/(2f^*))$ as $d \to \infty$, where $f^* = \lim_{x \downarrow 0}\{(f(x) + f(1-x))/2\}$. This gives a clear indication of how the Markov chain explores the stationary distribution. By contrast the ESJD only gives a measure of average behavior and does not take account of the possibility of the Markov chain becoming "stuck." If $\mathbb{E}_{\mathbf{Z}^d}[\alpha(\mathbf{x}^d, \mathbf{x}^d + \mathbf{Z}^d)]$ is very low, the Markov chain started $\mathbf{X}_0^d = \mathbf{x}^d$ is likely to spend a large number of iterations at $\mathbf{x}^d$ before accepting a move away from $\mathbf{x}^d$. Note that since V_s^d involves speeding up time by a factor of d^2, we say that the mixing of the algorithm is $O(d^2)$. The ESJD is easy to compute and asymptotically, as $d \to \infty$, the ESJD (appropriately scaled) converges to $\phi(l)$. Thus in discussing possible extensions of the Langevin diffusion limit proved in Theorem 2.1 for i.i.d. product densities of the form given in (1.6) and (1.7), we make considerable use of the ESJD. However, we highlight the limitations of the ESJD in discussing extensions of Theorem 2.1.

In most previous work on optimal scaling, the components of $\mathbf{Z}^d$ are taken to be independent and identically distributed $Z \sim N(0, 1)$ random variables. The reason for choosing $Z \sim U[-1, 1]$ for discontinuous target densities is mathematical convenience. The results proved in this paper hold with Gaussian rather than uniform proposal distributions, but some elements of the proof are less straightforward. For discussion of the ESJD for densities (1.6) for general Z subject to $\mathbb{E}[Z^2] < \infty$, see [12], Appendix B.

While the key result, a Langevin diffusion limit for the movement in the first component, is the same as [14], the proof is markedly different. Note that, for finite d, U^d and V^d are not Markov chains since whether or not a proposed move is accepted depends upon all the components in $\pi_d(\cdot)$. In [14], it is shown that there exists $\{F_d\}$ such that $\mathbb{P}(\bigcup_{t=0}^{[Td]}\{\mathbf{X}_t^d \notin F_d\}) \to 0$ as $d \to \infty$ and

$$\sup_{\mathbf{x}^d \in F_d} \left| \mathbb{E}[\alpha(\mathbf{x}^d, \mathbf{x}^d + \sigma_d \mathbf{Z}^d)] - 2\Phi\left(-\frac{l\sqrt{I}}{2}\right) \right| \le \varepsilon_d, \tag{1.13}$$

where $\varepsilon_d \to 0$ as $d \to \infty$. While (1.13) is not explicitly stated in [14], it is the essence of the requirements of the sets $\{F_d\}$, stating that for large d, with high probability over the first Td iterations the acceptance probability of the Markov chain is approximately constant, being within ε_d of $2\Phi(-l\sqrt{I}/2)$. (Note n rather than d is used for dimensionality in [14].) Thus in the limit as $d \to \infty$ the effect of the other components on movements in the first component converges to a deterministic acceptance probability $2\Phi(-l\sqrt{I}/2)$. The situation is more complex for $\pi_d(\cdot)$ of the form given by (1.6) and (1.7) as the acceptance rate in the limit as

$d \to \infty$ is inherently stochastic. For example, suppose $\pi_d(\cdot)$ is the uniform distribution on the d-dimensional hypercube so that $\alpha(\mathbf{X}_t^d, \mathbf{Y}^d) = 1_{\{\mathbf{Y}^d \in [0,1]^d\}}$. Letting $R_d^L = (0, \sigma_d)$ and $R_d^U = (1 - \sigma_d, 1)$, this gives

$$\mathbb{E}[\alpha(\mathbf{x}^d, \mathbf{x}^d + \sigma_d \mathbf{Z}^d)] = \prod_{i \in R_d^L} \left(\frac{1}{2} + \frac{x_i}{2\sigma_d}\right) \times \prod_{i \in R_d^U} \left(\frac{1}{2} + \frac{1 - x_i}{2\sigma_d}\right). \tag{1.14}$$

Thus the acceptance probability is totally determined by the components at the boundary (within σ_d of 0 or 1). The total number of components in $R_d^L \cup R_d^U$ is $\mathrm{Bin}(d, 2l/d)$ which converges in distribution to $\mathrm{Po}(2l)$ as $d \to \infty$. Thus the number of components close to the boundary is inherently stochastic. Moreover, the location of the components within $R_d^L \cup R_d^U$ plays a crucial role in the acceptance probability; see (1.14). Therefore there is no hope of replicating directly the method of proof applied in [14] and subsequently, in [7] and [10].

We need a homogenization argument which involves looking at $\mathbf{X}_\cdot^d$ over $[d^\delta]$ steps; cf. [11]. In particular, we show that the acceptance probability converges very rapidly to its stationary measure, so that over $[d^\delta]$ iterations approximately $\exp(-lf^*/2)[d^\delta]$ proposed moves are accepted. By comparison, $|X_{[d^\delta],1}^d - X_{0,1}^d| \leq [d^\delta]\sigma_d$; thus the value of an individual component only makes small changes over $[d^\delta]$ iterations. That is, we show that there exists $\{\tilde{F}_d\}$ such that, for any $T > 0$, $\mathbb{P}(\bigcup_{t=0}^{[Td^2]} \{\mathbf{X}_t^d \notin \tilde{F}_d\}) \to 0$ as $d \to \infty$ and for $\delta > 0$,

$$\sup_{\mathbf{x}^d \in \tilde{F}_d} \left| \frac{1}{[d^\delta]} \sum_{t=0}^{[d^\delta]-1} \mathbb{E}[\alpha(\mathbf{X}_t^d, \mathbf{X}_t^d + \sigma_d \mathbf{Z}_t^d) | \mathbf{X}_0^d = \mathbf{x}^d] - \exp\left(-\frac{lf^*}{2}\right) \right| \leq \varepsilon_d \tag{1.15}$$

for some $\varepsilon_d \to 0$ as $d \to \infty$. For large d, with high probability over the first $[Td^2]$ iterations the Markov chain stays in $\tilde{F}_d$, where the average number of accepted proposed moves in the following $[d^\delta]$ iterations is $\exp(-lf^*/2)d^\delta + o(d^\delta)$. The arguments are considerably more involved than in [11], where spherically constrained target distributions were studied, due to the very different geometry of the hypercube and spherical constraints applied in this paper and [11], respectively. In particular, in [11], $\sigma_d = l/\sqrt{d}$ with an aoar of 0.234.

By exploiting the homogenization argument it is possible to prove that V^d converges weakly to an appropriate Langevin diffusion V, given in Theorem 2.1. In Section 2, Theorem 2.1 is presented along with an outline of the proof. Also in Section 2, a description of the pseudo-RWM algorithm is given. The pseudo-RWM algorithm plays a key role in the proof of Theorem 2.1. The pseudo-RWM process moves at each iteration and the moves in the pseudo-RWM process are identical to those of the RWM process, conditioned upon a proposed move in the RWM process being accepted. The proof of Theorem 2.1 is long and technical with the details split into three key sections which are given in the Appendix; see Section 2 for more details. In Section 3, two interesting extensions of Theorem 2.1 are given.

In particular, Theorem 3.3 has major practical implications for the implementation of RWM algorithms by highlighting the detrimental effect of choosing RWM algorithms over Metropolis-within-Gibbs algorithms. The target densities for which theoretical results can be proved are limited, so discussion of possible extensions of Theorem 2.1 are given. In particular, we discuss general π_d restricted to the hypercube, general discontinuities in f and $\mathbf{X}_0^d \not\sim \pi_d$.

2. Pseudo-RWM algorithm and Theorem 2.1. We begin by defining the pseudo-random walk Metropolis (pseudo-RWM) process. We will then be in position to formally state the main theorem, Theorem 2.1. An outline of the proof of Theorem 2.1 is given, with the details, which are long and technical, placed in the Appendix.

For $d \geq 1$, let

$$h_d(\mathbf{z}^d) = \begin{cases} 2^{-d}, & \text{if } \mathbf{z}^d \in (-1,1)^d, \\ 0, & \text{otherwise.} \end{cases}$$

Let $J_d(\mathbf{x}^d)$ denote the probability of accepting a move in the RWM process given the current state of the process is $\mathbf{x}^d$. Then

$$J_d(\mathbf{x}^d) = \int h_d(\mathbf{z}^d)\left\{1 \wedge \frac{\pi_d(\mathbf{x}^d + \sigma_d \mathbf{z}^d)}{\pi_d(\mathbf{x}^d)}\right\} d\mathbf{z}^d. \tag{2.1}$$

Let $b_d^r(\mathbf{x}^d) = \sum_{j=1}^d 1_{\{x_j \in R_d^r\}}$, the total number of components of $\mathbf{x}^d$ in $R_d^r = (0, r/d) \cup (1 - r/d, 1)$. By Taylor's theorem for all $0 \leq x_i, x_i + \sigma_d z_i \leq 1$ and $-1 \leq z_i \leq 1$,

$$g(x_i + \sigma_d z_i) - g(x_i) \geq -g^* \sigma_d \tag{2.2}$$

with g^* defined in (1.8). Hence, for all $\mathbf{x}^d \in [0,1]^d$,

$$\begin{aligned} J_d(\mathbf{x}^d) &= \int h_d(\mathbf{z}^d)\left\{1 \wedge \prod_{i=1}^d \frac{\exp(g(x_i + \sigma_d z_i))}{\exp(g(x_i))}\right\} 1_{\{\mathbf{x}^d + \sigma_d \mathbf{z}^d \in [0,1]^d\}} d\mathbf{z}^d \\ &\geq \int h_d(\mathbf{z}^d)\{1 \wedge \exp(-d g^* \sigma_d)\} 1_{\{\mathbf{x}^d + \sigma_d \mathbf{z}^d \in [0,1]^d\}} d\mathbf{z}^d \\ &= \exp(-l g^*) \int h_d(\mathbf{z}^d) 1_{\{\mathbf{x}^d + \sigma_d \mathbf{z}^d \in [0,1]^d\}} d\mathbf{z}^d \\ &\geq \exp(-l g^*) \left(\frac{1}{2}\right)^{b_d^l(\mathbf{x}^d)}. \end{aligned} \tag{2.3}$$

This lower bound for $J_d(\mathbf{x}^d)$ will be used repeatedly.

The pseudo-RWM process moves at each iteration, which is the key difference to the RWM process. Furthermore, the moves in the pseudo-RWM process are identical to those of the RWM process, conditioned upon a move in the RWM process being accepted, that is, its jump chain. For $d \geq 1$, let $\hat{\mathbf{X}}_0^d, \hat{\mathbf{X}}_1^d, \dots$ denote the successive states of the pseudo-RWM process, where $\hat{\mathbf{X}}_0^d \sim \pi_d(\cdot)$. The pseudo-RWM process is a Markov process, where for $t \geq 0$, $\hat{\mathbf{X}}_{t+1}^d = \hat{\mathbf{X}}_t^d + \sigma_d \hat{\mathbf{Z}}_t^d$ and given that $\hat{\mathbf{X}}_t^d = \mathbf{x}^d$, $\hat{\mathbf{Z}}_t^d$ has p.d.f.

$$\zeta(\mathbf{z}^d|\mathbf{x}^d) = h_d(\mathbf{z}^d)\alpha(\mathbf{x}^d, \mathbf{x}^d + \sigma_d \mathbf{z}^d)/J_d(\mathbf{x}^d), \qquad \mathbf{z}^d \in (-1, 1)^d.$$

Note that $\zeta(\mathbf{z}^d|\mathbf{x}^d) = 0$ for $\mathbf{z}^d \notin (-1, 1)^d$. Since $\mathbf{X}_0^d, \hat{\mathbf{X}}_0^d \sim \pi_d$, we can couple the two processes to have the same starting value $\mathbf{X}_0^d$. A continued coupling of the two processes is outlined below. Suppose that $\mathbf{X}_t^d = \mathbf{x}^d$. Then for any $s \geq 1$,

$$\mathbb{P}\left(\bigcup_{j=1}^{s} \{\mathbf{X}_{t+j}^d = \mathbf{x}^d\} | \mathbf{X}_t^d = \mathbf{x}^d\right) = \left(1 - J_d(\mathbf{x}^d)\right)^s. \tag{2.4}$$

That is, the number of iterations the RWM algorithm stays at $\mathbf{x}^d$ before moving follows a geometric distribution with "success" probability $J_d(\mathbf{x}^d)$. Therefore for $j \geq 0$, let $M_j(\cdot)$ denote independent geometric random variables, where for $0 < p \leq 1$, $M_j(p)$ denotes a geometric random variable with "success" probability p. For $s \in \mathbb{Z}^+$, let $\hat{M}_s^d = M_s(J(\hat{\mathbf{X}}_s^d))$ and for $t \in \mathbb{Z}^+$, let

$$U_t^d = \sup\left\{s \in \mathbb{Z}^+ : \sum_{j=0}^{s-1} M_j(J_d(\hat{\mathbf{X}}_j^d)) \leq t\right\},$$

where the sum is zero if vacuous. For $s \in \mathbb{Z}^+$, attach $\hat{M}_s^d = M_s(J(\hat{\mathbf{X}}_s^d))$ to $\hat{\mathbf{X}}_s^d$. Thus $\hat{M}_s^d$ denotes the total number of iterations the RWM process spends at $\hat{\mathbf{X}}_s^d$ before moving to $\hat{\mathbf{X}}_{s+1}^d$. Hence, the RWM process can be constructed from $(\hat{\mathbf{X}}_0^d, \hat{M}_0^d), (\hat{\mathbf{X}}_1^d, \hat{M}_1^d), \dots$ by setting $\mathbf{X}_0^d \equiv \hat{\mathbf{X}}_0^d$ and for all $s \geq 1$, $\mathbf{X}_s^d = \hat{\mathbf{X}}_{U_s^d}^d$. Obviously the above process can be reversed by setting $\hat{\mathbf{X}}_t^d$ equal to the tth accepted move in the RWM process.

For each $d \geq 1$, the components of $\mathbf{X}_0^d$ are independent and identically distributed. Therefore we focus attention on the first component as this is indicative of the behavior of the whole process. For $d \geq 1$ and $t \geq 0$, let $V_t^d = X_{[d^2 t],1}^d$ and $\hat{V}_t^d = \hat{X}_{[d^2 t],1}^d$.

THEOREM 2.1. *Fix $l > 0$. For all $d \geq 1$, let $\mathbf{X}_0^d \sim \pi_d$. Then, as $d \to \infty$,*

$$V^d \Rightarrow V$$

in the Skorokhod topology on $D[0, \infty)$, *where V. satisfies the* (*reflected*) *Langevin SDE on* $[0, 1]$

$$dV_t = \sqrt{\phi(l)}\, dB_t + \tfrac{1}{2}\phi(l) g'(V_t)\, dt + dL_t^0(V) - dL_t^1(V) \tag{2.5}$$

with $V_0 \sim f$. Note that B_t is standard Brownian motion,

$$\phi(l) = \frac{l^2}{3}\exp\left(-\frac{f^* l}{2}\right)$$

and $f^* = \lim_{x \downarrow 0}(\frac{f(x)+f(1-x)}{2})$.

Here $\{L_t^y, t \geq 0\}$ denotes the local time of V at y (=0, 1) and the SDE (2.5) corresponds to standard reflection at the boundaries 0 *and* 1 (*see*, *e.g.*, *Chapter* VI *of* [13]).

PROOF. As noted in Section 1, the acceptance probability of the RWM process is inherently random and therefore it is necessary to consider the behavior of the RWM process averaged over $[d^\delta]$ iterations, for $\delta > 0$. Fix $0 < 20\gamma < \beta < \delta < \delta + \gamma < \frac{1}{2}$ and let $\{k_d\}$ be a sequence of positive integers satisfying $[d^\beta] \leq k_d \leq [d^\delta]$. For $s \in \mathbb{Z}^+$, let $\tilde{\mathbf{X}}_s^d = \mathbf{X}_{s[d^\delta]}^d$ and for $t \geq 0$, let $\tilde{V}_t^d = \tilde{X}_{[td^2/[d^\delta]],1}^d$. For all $t \geq 0$, $|X_{t+1,1}^d - X_{t,1}^d| \leq \sigma_d$ and $|[d^2 t] - [d^\delta] \times [d^2 t/[d^\delta]]| \leq [d^\delta]$. Hence, for all $T > 0$,

$$\sup_{0 \leq s \leq T} |\tilde{V}_s^d - V_s^d| \leq [d^\delta]\sigma_d. \tag{2.6}$$

Therefore by [5], Theorem 4.1, $V^d \Rightarrow V$ as $d \to \infty$, if $\tilde{V}^d \Rightarrow V$ as $d \to \infty$. Hence we proceed by showing that

$$\tilde{V}^d \Rightarrow V \qquad \text{as } d \to \infty. \tag{2.7}$$

Let G_d^δ be the (discrete-time) generator of $\tilde{\mathbf{X}}^d$ and let H be an arbitrary test function of the first component only. Thus

$$G_d^\delta H(\mathbf{x}^d) = \frac{d^2}{[d^\delta]}\mathbb{E}[H(\tilde{\mathbf{X}}_1^d) - H(\tilde{\mathbf{X}}_0^d) | \tilde{\mathbf{X}}_0^d = \mathbf{x}^d]. \tag{2.8}$$

The generator G of the (limiting) one-dimensional diffusion V for an arbitrary test function H is given by

$$GH(x) = \phi(l)\left\{\tfrac{1}{2}g'(x)H'(x) + \tfrac{1}{2}H''(x)\right\} \tag{2.9}$$

for all $x \in [0, 1]$ at least for all $H \in \mathcal{D}$, where $\mathcal{D}$ is defined in (2.10) below.

First note that the diffusion defined by (2.9) is regular; see [9], page 366. Therefore by [9], Chapter 8, Corollary 1.2, it is sufficient to restrict attention to functions

$$H \in \mathcal{D} \equiv \{h : h \in \hat{C}([0, 1]) \cap C^2((0, 1)) \cap \mathcal{D}^*, Gh \in \hat{C}([0, 1])\}, \tag{2.10}$$

where $C^2((0, 1))$ is the set of twice differentiable functions upon $(0, 1)$, $\hat{C}[0, 1]$ is the set of bounded continuous functions upon $[0, 1]$ and $\mathcal{D}^*$ is obtained by setting $q_i = 0$ $(i = 0, 1)$ in [9], page 367, (1.11) and is given by

$$\mathcal{D}^* = \{h : h'(0) = h'(1) = 0\}. \tag{2.11}$$

Let $H_1^* = \sup_{0 \le y \le 1} H'(y)$ and $H_2^* = \sup_{0 \le y \le 1} H''(y)$. Then $H \in C^2((0,1))$ combined with $H \in \mathcal{D}^*$ implies that $H_1^* < \infty$. It then follows from g' being bounded on [0, 1] and $GH \in \hat{C}([0,1])$ that $H_2^* < \infty$. These observations will play a key role in Appendix C.

Now (2.7) is proved using [9], Chapter 4, Corollary 8.7, by showing that there exists a sequence of sets $\{\tilde{F}_d\}$ such that for any $T > 0$,

$$\mathbb{P}\left(\bigcup_{j=0}^{[Td^2/[d^\delta]]} \{\mathbf{X}_j^d \notin \tilde{F}_d\}\right) \to 0 \qquad \text{as } d \to \infty \tag{2.12}$$

and

$$\sup_{\mathbf{x}^d \in \tilde{F}_d} |G_d^\delta H(\mathbf{x}^d) - GH(x_1)| \to 0 \qquad \text{as } d \to \infty. \tag{2.13}$$

Let the sets $\{F_d\}$ and $\{\tilde{F}_d\}$ be such that $F_d = \bigcap_{j=1}^4 F_d^j$ and

$$\tilde{F}_d = \left\{\mathbf{x}^d; \mathbb{P}\left(\bigcup_{j=0}^{[d^\delta]} \{\hat{\mathbf{X}}_j^d \notin F_d\} | \hat{\mathbf{X}}_0^d = \mathbf{x}^d\right) \le d^{-3}\right\}, \tag{2.14}$$

where F_d^1, F_d^2, F_d^3 and F_d^4 are defined below. Recall that $b_d^r(\mathbf{x}^d) = \sum_{j=1}^d 1_{\{x_j \in R_d^r\}}$, the total number of components of $\mathbf{x}^d$ in $R_d^r = (0, r/d) \cup (1 - r/d, 1)$. We term R_d^l the rejection region, in that, for any component in R_d^l, there is positive probability of proposing a move outside the hypercube with such moves automatically being rejected. Let

$$F_d^1 = \{\mathbf{x}^d; b_d^l(\mathbf{x}^d) \le \gamma \log d\}, \tag{2.15}$$

$$F_d^2 = \bigcap_{k=[d^\beta]}^{[d^\delta]} \{\mathbf{x}^d; |b_d^{k^{3/4}}(\mathbf{x}^d) - \mathbb{E}[b_d^{k^{3/4}}(\mathbf{X}_0^d)]| \le \sqrt{k}\}, \tag{2.16}$$

$$F_d^3 = \left\{\mathbf{x}^d; \sup_{[d^\beta] \le k_d \le [d^\delta]} \sup_{0 \le r \le l} |\lambda_d(\mathbf{x}^d; r; k_d) - \lambda(r)| \le d^{-\gamma}\right\}, \tag{2.17}$$

$$F_d^4 = \left\{\mathbf{x}^d; \left|\frac{1}{d}\sum_{j=1}^d g'(x_j)^2 - \mathbb{E}_f[g'(X_1)^2]\right| < d^{-1/8}\right\}, \tag{2.18}$$

where $\lambda_d(\mathbf{x}^d; r; k_d) = \mathbb{E}[b_d^r(\mathbf{X}_{k_d}^d) | \mathbf{X}_0^d = \mathbf{x}^d]$ and $\lambda(r) = f^* r(1 + r/2l)$. In Appendix A, we prove (2.12) for the sets $\{\tilde{F}_d\}$ given in (2.14). Note that (2.12) follows immediately from Theorem A.13, (A.74) since $\mathbf{X}_0^d \sim \pi_d$. An outline of the roles played by each F_d^j ($j = 1, 2, 3, 4$) is given below. For $\mathbf{x}^d \in F_d^1$ ($\mathbf{x}^d \in F_d^2$) the total number of components in (*close to*) the rejection region are controlled. For $\mathbf{x}^d \in F_d^3$ after k_d iterations the total number and position of the points $\{\hat{\mathbf{X}}_{k_d}^d | \hat{\mathbf{X}}_0^d = \mathbf{x}^d\}$ in R_d^l

are approximately from the stationary distribution of $\hat{\mathbf{X}}^d$. Finally, for $\mathbf{x}^d \in F_d^4$, $\frac{1}{d}\sum_{j=1}^d g'(x_j)^2 \approx \mathbb{E}_f[g'(X)^2]$; this is the key requirement for the sets $\{F_d\}$ given in [14], cf. [14], page 114, $R_n(x_2, \dots, x_n)$.

The proof of (2.13) splits into two parts and exploits the pseudo-RWM process. Let

$$P_d = \max\left\{K = 0, 1, \dots, [d^\delta - 1]; \frac{1}{[d^\delta]}\sum_{j=0}^{K-1} M_j(J_d(\hat{\mathbf{X}}_j^d)) \le 1\right\} \Big/ [d^\delta], \tag{2.19}$$

the proportion of accepted moves in the first $[d^\delta]$ iterations, where the sum is set equal to zero if vacuous. Then $\tilde{\mathbf{X}}_1^d = \mathbf{X}_{[d^\delta]}^d = \hat{\mathbf{X}}_{[P_d d^\delta]}^d$ and

$$G_d^\delta H(\mathbf{x}^d) = \frac{d^2}{[d^\delta]}\mathbb{E}\big[H\big(\hat{\mathbf{X}}_{[P_d d^\delta]}^d\big) - H(\hat{\mathbf{X}}_0^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d\big]. \tag{2.20}$$

In Appendix B, we show that for all $\mathbf{x}^d \in \tilde{F}_d$, $P_d|\hat{\mathbf{X}}_0^d = \mathbf{x}^d \xrightarrow{p} \exp(-lf^*/2)$ as $d \to \infty$. Consequently, it is useful to introduce $\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d)$ $(0 \le \pi \le 1)$ which is defined for fixed $0 \le \pi \le 1$ as

$$\begin{aligned} \hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) &= \frac{d^2}{[d^\delta]}\mathbb{E}\big[\big(H\big(\hat{\mathbf{X}}_{[\pi d^\delta]}^d\big) - H(\hat{\mathbf{X}}_0^d)\big)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d\big] \\ &= \frac{d^2}{[d^\delta]}\sum_{j=0}^{[\pi d^\delta - 1]} \mathbb{E}[H(\hat{\mathbf{X}}_{j+1}^d) - H(\hat{\mathbf{X}}_j^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d] \\ &= \frac{1}{[d^\delta]}\sum_{j=0}^{[\pi d^\delta - 1]} \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_j^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d], \end{aligned} \tag{2.21}$$

where

$$\hat{G}_d H(\hat{\mathbf{X}}_j^d) = d^2\mathbb{E}[H(\hat{\mathbf{X}}_1^d - \hat{\mathbf{X}}_0^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d]. \tag{2.22}$$

Finally in Appendix C, we prove in Lemma C.6 that

$$\sup_{0 \le \pi \le 1}\sup_{\mathbf{x}^d \in \tilde{F}_d} |\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) - GH(x_1)| \to 0 \qquad \text{as } d \to \infty. \tag{2.23}$$

The triangle inequality is then utilized to prove (2.13) in Lemma C.6 using (2.23) and $P_d|\mathbf{X}_0^d = \mathbf{x}^d \xrightarrow{p} \exp(-lf^*/2)$ as $d \to \infty$. $\square$

It should be noted that in Appendix C, we assume that $\mathbb{E}[g'(X)^2] > 0$, in particular in Lemma C.1. In Appendixes A and B we make no such assumption. However, $\mathbb{E}[g'(X)^2] = 0$ corresponds to $f(x) = 1_{\{0<x<1\}}$ (uniform distribution), and proving Lemma C.6 in this case follows similar but simpler arguments to those given in Appendix C.

A key difference between the diffusion limits for continuous and discontinuous i.i.d. product densities is the dependence of the speed measure $\phi(l)$ upon f. For continuous (suitably differentiable) f, $\phi(l)$ depends upon $I \equiv \mathbb{E}_f[\{f'(X)/f(X)\}^2]$, which is a measure of the "roughness" of f. For discontinuous densities of the form (1.7), $\phi(l)$ depends upon $f^* = \lim_{x\downarrow 0}\{(f(x) + f(1-x))/2\}$, the (mean of the) limit of the density at the boundaries (discontinuities). Discussion of the role of the density f in the behavior of the RWM algorithm is given in Section 3.

The most important consequence of Theorem 2.1 is the following result.

COROLLARY 2.2. *Let* $a(l) = \exp(-f^* l/2)$. *Then*

$$\mathbb{E}_{\pi_d}\mathbb{E}[J_d(\mathbf{X}_0^d)] \to a(l) \qquad \text{as } d \to \infty.$$

$\phi(l)$ *is maximized by* $l = \hat{l} = 4/f^*$ *with*

$$a(\hat{l}) = \exp(-2) = 0.1353.$$

Clearly, if $f(\cdot)$ is known, $\hat{l}$ can be calculated explicitly. However, where MCMC is used, $f(\cdot)$ will often only be known up to the constant of proportionality. This is where Corollary 2.2 has major practical implications, in that, to maximize the speed of the limiting diffusion, and hence, the efficiency of the RWM algorithm, it is sufficient to monitor the average acceptance rate, and to choose l such that the average acceptance rate is approximately e^{-2}. Therefore there is no need to explicitly calculate or estimate the constant of proportionality.

3. Extensions. In this section, we discuss the extent to which the conclusions of Theorem 2.1 extend beyond π_d being an i.i.d. product density upon the d-dimensional hypercube and $\mathbf{X}_0^d \sim \pi_d$. First we present two extensions of Theorem 2.1. The second extension, Theorem 3.3, is an important practical result concerning lower-dimensional updating schema.

Suppose that $f(\cdot)$ is nonzero on the positive half-line. That is,

$$f(x) \propto \exp(g(x)) \qquad (x > 0) \tag{3.1}$$

and $f(x) = 0$ otherwise.

THEOREM 3.1. *Fix* $l > 0$. *For all* $d \geq 1$, *let* $\mathbf{X}_0^d \sim \pi_d$, *given by* (3.1), *with* $\sup_{x\geq 0}|g'(x)| = g^* < \infty$. *Then, as* $d \to \infty$,

$$V^d \Rightarrow V$$

in the Skorokhod topology on $D[0,\infty)$, *where* V. *satisfies the* (*reflected*) *Langevin SDE on* $[0,\infty)$

$$dV_t = \sqrt{\phi(l)}\, dB_t + \tfrac{1}{2}\phi(l) g'(V_t)\, dt + dL_t^0(V)$$

with $V_0 \sim f$, $\phi(l) = \frac{l^2}{3}\exp(-f^\star l/4)$ *and* $f^\star = \lim_{x\downarrow 0} f(x)$.

PROOF. The proof of the theorem is virtually identical to the proof of Theorem 2.1, and so, the details are omitted. □

Note that we have assumed that $g'(\cdot)$ is bounded on $[0, \infty)$. This assumption is almost certainly stronger than necessary with $g'(\cdot)$ being Lipschitz and/or satisfying certain moment conditions probably being sufficient; cf. [14].

Theorem 3.1 is unsurprising with the speed of the diffusion depending upon the number of components close to the discontinuity at 0.

COROLLARY 3.2. *Let* $\pi_d(\mathbf{x}^d) = \prod_{i=1}^d f(x_i)$ *where* f *satisfies* (3.1). *Then*

$$\mathbb{E}_{\pi_d}[J_d(\mathbf{X}_0^d)] \to \exp(-f^\star l/4) \equiv a(l) \qquad \text{as } d \to \infty.$$

$\phi(l)$ *is maximized by* $l = \hat{l} = 8/f^\star$ *with*

$$a(\hat{l}) = \exp(-2) = 0.1353.$$

Therefore the conclusions are identical to Corollary 2.2 that in order to maximize the speed of the limiting diffusion it is sufficient to choose l such that the average acceptance rate is e^{-2}.

The second and more important extension of Theorem 2.1 follows on from [10]. In [10], the Metropolis-within-Gibbs algorithm was considered, where only a proportion c $(0 < c \leq 1)$ of the components are updated at each iteration. For given $d \geq 1$, at each iteration $c_d d$ of the components are chosen uniformly at random and new values for these components are proposed using random walk Metropolis with proposal variance $\sigma_{d,c_d}^2 = (l/d)^2$. The remaining $(1 - c_d)d$ components remain fixed at their current values. Finally, it is assumed that $c_d \to c$ as $d \to \infty$.

The following result assumes that $f(\cdot)$ is nonzero on $(0, 1)$ only. The extension to the positive half-line is trivial.

THEOREM 3.3. *Fix* $0 < c \leq 1$ *and* $l > 0$. *For all* $d \geq 1$, *let* $\mathbf{X}_0^d = (X_{0,1}^d, X_{0,2}^d, \dots, X_{0,d}^d)$ *be such that all of its components are distributed according to* $f(\cdot)$. *Then*, *as* $d \to \infty$,

$$V^d \Rightarrow V$$

in the Skorokhod topology, *where* $V_0 \sim f(\cdot)$ *and* V *satisfies the* (*reflected*) *Langevin SDE on* $[0, 1]$

$$dV_t = \sqrt{\phi_c(l)}\, dB_t + \tfrac{1}{2}\phi_c(l) g'(V_t)\, dt + dL_t^0(V) - dL_t^1(V),$$

where B_t *is standard Brownian motion*, $\phi_c(l) = \frac{cl^2}{3}\exp(-cf^* l/2)$ *and* $f^* = \lim_{x \downarrow 0} \frac{f(x)+f(1-x)}{2}$.

Let $a_d^{c_d}(l)$ denote the average acceptance rate of the RWM algorithm in d dimensions where a proportion c_d of the components are updated at each iteration. Let

$$a^c(l) = \exp(-cf^* l/2).$$

We then have the following result which mirrors Corollaries 2.2 and 3.2.

COROLLARY 3.4. *Let $c_d \to c$ as $d \to \infty$. Then*

$$a_d^{c_d}(l) \to a^c(l) \qquad \text{as } d \to \infty.$$

For fixed $0 < c \leq 1$, $\phi_c(l)$ *is maximized by*

$$l = \hat{l}_c = \frac{4}{cf^*}$$

and

$$\phi_c(\hat{l}_c) = \frac{1}{c}\phi_1(\hat{l}_1).$$

Also

$$a(\hat{l}_c) = \exp(-2) = 0.1353.$$

Corollary 3.4 is of fundamental importance from a practical point of view, in that it shows that the optimal speed of the limiting diffusion is inversely proportional to c. Therefore the optimal action is to choose c as close to 0 as possible. Furthermore, we have shown that not only is full-dimensional RWM bad for discontinuous target densities but it is the worst algorithm of all the Metropolis-within-Gibbs RWM algorithms.

We now go beyond i.i.d. product densities with a discontinuity at the boundary and $\mathbf{X}_0^d \sim \pi_d$. We consider general densities on the unit hypercube, discontinuities not at the boundary and $\mathbf{X}_0^d \not\sim \pi_d$. As mentioned in Section 1, for i.i.d. product densities, the speed measure of the limiting one-dimensional diffusion, $\phi(l)$, is equal to the limit, as $d \to \infty$, of the ESJD times d. Therefore we consider the ESJD for the above-mentioned extensions as being indicative of the behavior of the limiting Langevin diffusion. We also highlight an extra criterion which is likely to be required in moving from an ESJD to a Langevin diffusion limit.

Using the proof of Theorem 2.1, it is straightforward to show that

$$\begin{aligned}
\phi(l) &= \frac{l^2}{3}\exp\left(-\frac{lf^*}{2}\right) \\
&= l^2 \mathbb{E}[Z_1^2] \lim_{d\to\infty} \mathbb{E}\left[1_{\{\mathbf{X}_0^d + \sigma_d \mathbf{Z}_1^d \in [0,1]^d\}}\right] \qquad (3.2)\\
&= \frac{l^2}{3} \lim_{d\to\infty} \mathbb{E}\left[\left(\frac{3}{4}\right)^{b_d^l(\mathbf{X}_0^d)}\right].
\end{aligned}$$

The first equality in (3.2) can be proved using Lemma A.6, (A.26), where for $Z_1 \sim U(-1,1)$, $\mathbb{E}[Z_1^2] = 1/3$. The second equality in (3.2) comes from the fact that for $0 < x < \sigma_d$, $f(x) + f(1-x) = 2f^* + O(1/d)$ and for a component $X_{0,i}^d$ uniformly distributed on $(0, l/d)$ or $(1 - l/d, 1)$, $P(X_{0,i}^d + \sigma_d Z_i^d \in [0,1]) = 3/4$. That is, the acceptance probability of a proposed move is dominated by whether or not the proposed move lies inside the d-dimensional unit hypercube. Proposed moves inside the hypercube are accepted with probability $1 - o(d^{-\alpha})$ for any $\alpha < 1/2$; see Lemma A.7. Thus it is the number and behavior of the components at the boundary of the hypercube (the discontinuity) which determine the behavior of the RWM algorithm. This is also seen in Theorems 3.1 and 3.3.

First, we consider discontinuities not at the boundary. Suppose that $\pi_d(\mathbf{x}^d) = \prod_{i=1}^d f(x_i)$, where

$$f(x) \propto 1_{\{x \in [a,b]\}} \exp(g(x)) \qquad (x \in \mathbb{R}) \tag{3.3}$$

for some $a, b \in \mathbb{R}$. Further suppose that $g(\cdot)$ is continuous (twice differentiable) upon $[a,b]$ except at a countable number of points, $\mathcal{P} = \{a_1, a_2, \dots, a_k\}$, say, on (a,b). Set $a_0 = a$ and $a_{k+1} = b$, with $\sigma_d = l/d$. For $j = 0, 1, \dots, k+1$, let $f_j^- = \lim_{x \to a_j -} f(x)$ and $f_j^+ = \lim_{x \to a_j +} f(x)$, with $Y_j^- \sim \mathrm{Po}(l f_j^-/4)$ and $Y_j^+ \sim \mathrm{Po}(l f_j^+/4)$, where $f_0^- = f_{k+1}^+ = Y_0^- = Y_{k+1}^+ = 0$. Then following [12], (4.23), we can show that d times the ESJD

$$d\mathbb{E}\left[\sum_{i=1}^d (X_{1,i}^d - X_{0,i}^d)^2\right] \to \frac{l^2}{3}\mathbb{E}\left[1 \wedge \prod_{j=0}^{k+1}\left(\frac{f_j^-}{f_j^+}\right)^{Y_j^+ - Y_j^-}\right] \qquad \text{as } d \to \infty. \tag{3.4}$$

Thus the optimal scaling of σ_d is again of the form l/d and the acceptance or rejection of a proposed move is determined by the components close to the discontinuities. Furthermore, it is straightforward to show that for each $j = 0, 1, \dots, k+1$, $l^2 (f_j^-/f_j^+)^{Y_j^+ - Y_j^-} \xrightarrow{p} 0$ as $l \to \infty$, implying that the optimal choice of l lies in $(0, \infty)$. Proving a Langevin diffusion for the (normalized) first component of the RWM algorithm should be possible with appropriate local time terms at the discontinuities in f. While (3.4) holds regardless of f_j^- and f_j^+ for a diffusion limit we require that $\min_{1 \le j \le k+1} f_j^-, \min_{0 \le j \le k} f_j^+ > 0$, that is, the density is strictly positive on (a,b). (If this is not the case, the RWM algorithm is reducible in the limit as $d \to \infty$.) Extensions to the case where either $a = -\infty$ and/or $b = \infty$ are straightforward.

Second, we consider general densities which are zero outside the d-dimensional hypercube, $\pi_d(\mathbf{x}^d) \propto 1_{\{\mathbf{x}^d \in [0,1]^d\}} \exp(\mu_d(\mathbf{x}^d))$, where $\mu_d(\cdot)$ is assumed to be continuous and twice differentiable. Let $\sigma_d = l/d$ and assuming that

$$\exp\big(\mu_d(\mathbf{X}_0^d + \sigma_d \mathbf{Z}_1^d) - \mu_d(\mathbf{X}_0^d)\big) \xrightarrow{p} 1 \qquad \text{as } d \to \infty, \tag{3.5}$$

we have that d times the ESJD satisfies

$$d\mathbb{E}\left[\sum_{i=1}^{d}(X_{1,i}^{d}-X_{0,i}^{d})^{2}\right]\to\frac{l^{2}}{3}\lim_{d\to\infty}\mathbb{E}\left[\left(\frac{3}{4}\right)^{b_{d}^{l}(\mathbf{X}_{0}^{d})}\right]\qquad\text{as } d\to\infty. \tag{3.6}$$

Note that (3.5) is a weak condition and should be straightforward to check using a Taylor series expansion of μ_d. For i.i.d. product densities, $b_d^l(\mathbf{X}_0^d)\xrightarrow{D}\mathrm{Po}(2lf^*)$ as $d\to\infty$. More generally, the limiting distribution of $b_d^l(\mathbf{X}_0^d)$ will determine the limit of the right-hand side of (3.6). In particular, so long as there exist $\delta>0$ and $K\in\mathbb{N}$ such that $\mathbb{P}(\lim_{d\to\infty}b_d^l(\mathbf{X}_0^d)\le K)>\delta$, the right-hand side of (3.6) will be nonzero for $l>0$. It is informative to consider what conditions upon π_d are likely to be necessary for a diffusion limit, whether it be one-dimensional or infinite-dimensional as in [7]. Suppose that $b_d^l(\mathbf{X}_0^d)\xrightarrow{D}B$ as $d\to\infty$. For a diffusion limit we will require moment conditions on B, probably requiring that there exists $\varepsilon>0$ such that $\mathbb{E}[\exp(\varepsilon B)]<\infty$. This will be required to control the probability of the RWM algorithm getting "stuck" in the corners of the hypercube. This highlights a key difference between studying the ESJD and a diffusion limit. For the ESJD, we want a positive probability that the total number of components at the boundary of the hypercube is finite in the limit as $d\to\infty$. For the diffusion limit, as seen with the construction of $\{F_d^1\}$ in Theorem 2.1, we want that the probability of there being a large number of components ($O(\log d)$) at the boundary is very small ($o(d^{-2})$).

Third, suppose that $\mathbf{X}_0^d\not\sim\pi_d$. There are very bad starting points in the "corners" of the hypercube. For example, if $\mathbf{X}_0^d=(\exp(-d),\exp(-d),\dots,\exp(-d))$, $J_d(\mathbf{X}_0^d)\approx(0.5+\exp(-d))^d$ which even for $d=100$ is less than 1×10^{-30}. Thus the RWM process is likely to be "stuck" at its starting point for a very long period of time. This is rather pathological and a more interesting question is the situation when $\mathbf{X}_0^d=\mathbf{S}^d$, where the components of $\mathbf{S}^d$ are i.i.d. In particular, suppose that $S_1^d\sim U[0,1]$, so that $\mathbf{X}_0^d$ is chosen uniformly at random over the hypercube. Note that, if $\mathbf{S}^d$ is the uniform distribution,

$$d\mathbb{E}\left[\sum_{i=1}^{d}(X_{1,i}^{d}-X_{0,i}^{d})^{2}|\mathbf{X}_0^d\stackrel{D}{=}\mathbf{S}^d\right]\to\frac{l^{2}}{3}\exp\left(-\frac{l}{2}\right)\qquad\text{as } d\to\infty \tag{3.7}$$

with the right-hand side maximized by taking $\hat{l}=4$ compared with $\hat{l}=4/f^*$ for $\mathbf{X}_0^d\sim\pi_d$. We expect to see similar behavior to [8], in that the optimal σ_d (in terms of the ESJD) will vary as the algorithm converges to the stationary distribution but will be of the form $\sigma_d=l/d$ throughout. The RWM algorithm is unlikely to get "stuck" with it conjectured that for any $T>0$ and $\gamma>0$,

$$\mathbb{P}\left(\bigcup_{t=0}^{[Td^2]}\{b_d^l(\mathbf{X}_t^d)\ge\gamma\log d\}|\mathbf{X}_0^d\stackrel{D}{=}\mathbf{S}^d\right)\to0\qquad\text{as } d\to\infty.$$

Simulations with $f(x) \propto 1_{\{0<x<1\}} \exp(-2x)$ and $f(x) \propto 1_{\{0<x<1\}} \exp(-(x-0.5)^2/2)$ and $d = 10, 20, \dots, 200$ suggest that convergence occurs in $O(d^2)$ iteration. For convergence, we monitor the mean of $\mathbf{X}_t^d$ for $f(x) \propto 1_{\{0<x<1\}} \exp(-2x)$ and the variance of $\mathbf{X}_t^d$ for $f(x) \propto 1_{\{0<x<1\}} \exp(-(x-0.5)^2/2)$.

APPENDIX A: CONSTRUCTION OF THE SETS $\{F_d\}$ AND $\{\tilde{F}_d\}$

The sets F_d consist of the intersection of four sets F_d^i $(i = 1, 2, 3, 4)$. For $i = 1, 2, 3, 4$, we will define F_d^i and discuss the role that it plays in the proof of Theorem 2.1, one at a time. Furthermore, we show that in stationarity it is highly unlikely that $\mathbf{X}_t^d$ does not belong to F_d. Since we rely upon a homogenization argument, it is necessary to go further than the sets F_d to the sets $\tilde{F}_d \subset F_d$. In particular, if $\hat{\mathbf{X}}_0^d \in \tilde{F}_d$, then it is highly unlikely that any of $\hat{\mathbf{X}}_1^d, \hat{\mathbf{X}}_2^d, \dots, \hat{\mathbf{X}}_{[d^\delta]}^d$ do not belong to F_d. The above statement is made precise in Theorem A.13 below, where the constructions of $\{F_d\}$ and $\{\tilde{F}_d\}$ are drawn together.

It is possible that all d components of $\mathbf{X}_0^d$ are in R_d^l. However, this is highly unlikely and we show in Lemma A.1 that with high probability, there are at most $\gamma \log d$ components in the rejection region. Let $F_d^1 = \{\mathbf{x}^d; b_d^l(\mathbf{x}^d) \le \gamma \log d\}$.

LEMMA A.1. *For any $\kappa > 0$,*

$$d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d^1) \to 0 \qquad \text{as } d \to \infty.$$

PROOF. Fix $\kappa > 0$. Note that $\mathbf{X}_0^d \notin F_d^1$ if and only if $b_d^l(\mathbf{X}_0^d) > \gamma \log d$. However,

$$b_d^l(\mathbf{X}_0^d) \sim \mathrm{Bin}\left(d, \int_0^{l/d} \{f(x) + f(1-x)\}\, dx\right)$$

with

$$d \int_0^{l/d} \{f(x) + f(1-x)\}\, dx \to 2f^* l \qquad \text{as } d \to \infty. \tag{A.1}$$

Fix $\rho > \kappa/\gamma$. By Markov's inequality and using independence of the components of $\mathbf{X}_0^d$,

$$\begin{aligned}
&d^\kappa \mathbb{P}\big(b_d^l(\mathbf{X}_0^d) > \gamma \log d\big) \\
&\quad \le d^\kappa \mathbb{E}[\exp(\rho b_d^l(\mathbf{X}_0^d))]/\exp(\rho\gamma \log d) \\
&\quad = d^\kappa \mathbb{E}\big[\exp\big(\rho 1_{\{X_{0,1}^d \in R_d^l\}}\big)\big]^d / d^{\rho\gamma} \qquad \text{(A.2)}\\
&\quad = d^\kappa \left(1 + (e^\rho - 1) \int_0^{l/d} \{f(x) + f(1-x)\}\, dx\right)^d \Big/ d^{\rho\gamma} \\
&\quad \le d^{\kappa - \rho\gamma} \exp\left((e^\rho - 1) d \int_0^{l/d} \{f(x) + f(1-x)\}\, dx\right).
\end{aligned}$$

The lemma follows since (A.1) implies that the right-hand side of (A.2) converges to 0 as $d \to \infty$. $\square$

For $\mathbf{x}^d \in F_d^1$, it follows from (2.3) that

$$J_d(\mathbf{x}^d) \geq \exp(-lg^*)2^{-b_d^l(\mathbf{x}^d)} \geq \exp(-lg^*)2^{-\gamma \log d} \geq \exp(-lg^*)d^{-\gamma}. \tag{A.3}$$

This is a useful lower bound for the acceptance probability and as a result the random walk Metropolis algorithm does not get "stuck" at values of $\mathbf{x}^d \in F_d^1$. To assist with the homogenizing arguments, we define $\{\tilde{F}_d^1\}$ by

$$\tilde{F}_d^1 = \left\{ \mathbf{x}^d ; \mathbb{P}\left(\bigcup_{j=0}^{[d^\delta]} \hat{\mathbf{X}}_j^d \notin F_d^1 | \hat{\mathbf{X}}_0^d = \mathbf{x}^d \right) \leq d^{-3} \right\}. \tag{A.4}$$

That is, by starting in $\tilde{F}_d^1$ it is highly unlikely that the pseudo-RWM algorithm leaves F_d^1 in $[d^\delta]$ iterations. To study $\tilde{F}_d^1$ and later $\tilde{F}_d$ we require the following lemmas.

LEMMA A.2. *For a random variable X, suppose that there exist $\delta, \varepsilon > 0$ such that*

$$\mathbb{P}(X \in A | X \in B) \leq \delta\varepsilon \tag{A.5}$$

and for all $x \in D^C$, $\mathbb{P}(X \in A | X = x) \geq \varepsilon$. Then

$$\mathbb{P}(X \notin D | X \in B) \leq \delta. \tag{A.6}$$

PROOF. First note that

$$\begin{aligned} \mathbb{P}(X \in A | X \in B) &\geq \mathbb{P}(X \in A \cap X \in D^C | X \in B) \\ &= \mathbb{P}(X \in A | X \in D^C, X \in B)\mathbb{P}(X \notin D | X \in B). \end{aligned} \tag{A.7}$$

The lemma follows from rearranging (A.7) and using (A.5) and $\mathbb{P}(X \in A | X \in D^C, X \in B) \geq \varepsilon$. $\square$

LEMMA A.3. *Suppose that a sequence of sets $\{F_d^\star\}$ is such that there exists $\kappa > 0$ such that*

$$d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d^\star) \to 0 \qquad \text{as } d \to \infty. \tag{A.8}$$

Fix $\varepsilon > 0$ and let

$$\tilde{F}_d^\star = \left\{ \mathbf{x}^d ; \mathbb{P}\left(\bigcup_{i=0}^{[d^\delta]} \{\hat{\mathbf{X}}_i^d \notin F_d^\star \cap F_d^1\} | \hat{\mathbf{X}}_0^d = \mathbf{x}^d \right) \leq d^{-\varepsilon} \right\}. \tag{A.9}$$

Then

$$d^{\kappa-(2+\delta+\gamma+\varepsilon)} \mathbb{P}(\hat{\mathbf{X}}_0^d \notin \tilde{F}_d^\star) \to 0 \qquad \text{as } d \to \infty. \tag{A.10}$$

PROOF. Since $\mathbf{X}_i^d \sim \pi_d$,

$$\mathbb{P}\left(\bigcup_{i=0}^{[d^{2+\delta+\gamma}]} \{\mathbf{X}_i^d \notin F_d^\star \cap F_d^1\}\right) \le d^{2+\delta+\gamma}\mathbb{P}(\mathbf{X}_0^d \notin F_d^* \cap F_d^1). \tag{A.11}$$

Therefore for all sufficiently large d,

$$\mathbb{P}(\mathbf{X}_0^d \in F_d^* \cap F_d^1) \ge 1 - \mathbb{P}(\mathbf{X}_0^d \notin F_d^*) - \mathbb{P}(\mathbf{X}_0^d \notin F_d^1) \ge \tfrac{1}{2}. \tag{A.12}$$

By Bayes's theorem, $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) \le \mathbb{P}(A)/\mathbb{P}(B)$. Therefore taking $A = \bigcup_{i=0}^{[d^{2+\delta+\gamma}]}\{\mathbf{X}_i^d \notin F_d^* \cap F_d^1\}$ and $B = \{\mathbf{X}_0^d \in F_d^* \cap F_d^1\}$, it follows from (A.11) and (A.12) that

$$\mathbb{P}\left(\bigcup_{i=0}^{[d^{2+\delta+\gamma}]} \{\mathbf{X}_i^d \notin F_d^\star \cap F_d^1\} | \mathbf{X}_0^d \in F_d^\star \cap F_d^1\right) \le \frac{d^{2+\delta+\gamma}\mathbb{P}(\mathbf{X}_0^d \notin F_d^\star \cap F_d^1)}{1/2}. \tag{A.13}$$

Let

$$\hat{F}_d^\star = \left\{\mathbf{x}^d; \mathbb{P}\left(\bigcup_{i=0}^{[d^{2+\delta+\gamma}]} \{\mathbf{X}_i^d \notin F_d^\star \cap F_d^1\} | \mathbf{X}_0^d = \mathbf{x}^d\right) \le d^{-\varepsilon}\right\}. \tag{A.14}$$

It follows from Lemmas A.1 and A.2 that

$$d^{\kappa-(2+\delta+\gamma+\varepsilon)}\mathbb{P}(\mathbf{X}_0^d \notin \hat{F}_d^\star | \mathbf{X}_0^d \in F_d^\star \cap F_d^1) \to 0 \qquad \text{as } d \to \infty. \tag{A.15}$$

Since $d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d^\star \cap F_d^1) \to 0$ as $d \to \infty$, it follows from (A.15) that

$$d^{\kappa-(2+\delta+\gamma+\varepsilon)}\mathbb{P}(\mathbf{X}_0^d \notin \hat{F}_d^\star) \to 0 \qquad \text{as } d \to \infty. \tag{A.16}$$

For $d \ge 1$ and $i = 0, 1, 2, \dots,$ let $\{\theta_i^d\}$ be independent and identically distributed Bernoulli random variables with $\mathbb{P}(\theta_0^d = 1) = \exp(-lg^*)2^{-\gamma \log d}$ where $g^* = \max_{\{0 \le x \le 1\}} |g'(x)|$. It is straightforward using Hoeffding's inequality to show that

$$d^\kappa \mathbb{P}\left(\sum_{i=1}^{[d^{2+\delta+\gamma}]} \theta_i^d < d^\delta\right) \to 0 \qquad \text{as } d \to \infty. \tag{A.17}$$

Now $\{\theta_j^d\}$ and $\{\mathbf{X}_j^d\}$ can be constructed upon a common probability space such that if $\theta_j^d = 1$ and $\mathbf{X}_j^d \in F_d^1$, $\mathbf{X}_{j+1}^d \ne \mathbf{X}_j^d$. For $k, n \ge 0$, consider $\hat{\mathbf{X}}_k^d$, if $\sum_{i=1}^n \theta_i^d \ge k$ and $\bigcap_{j=0}^n \{\mathbf{X}_j^d \in F_d^* \cap F_d^1\}$, a coupling exists such that there exists $J_k \in \{k, k+1, \dots, n\}$ such that $\hat{\mathbf{X}}_k^d = \mathbf{X}_{J_k}^d \in F_d^* \cap F_d^1$. Exploiting the above coupling, $\bigcap_{j=0}^{[d^{2+\delta+\gamma}]}\{\mathbf{X}_j^d \in F_d^\star \cap F_d^1\}$ and $\sum_{i=1}^{[d^{2+\delta+\gamma}]} \theta_i^d \ge d^\delta$ together imply that $\bigcap_{j=0}^{[d^\delta]}\{\hat{\mathbf{X}}_j^d \in F_d^\star \cap F_d^1\}$. Thus

$$\mathbb{P}(\hat{\mathbf{X}}_0^d \notin \tilde{F}_d^\star) \le \mathbb{P}(\mathbf{X}_0^d \notin \hat{F}_d^\star) + \mathbb{P}\left(\sum_{i=1}^{[d^{2+\delta+\gamma}]} \theta_i^d < d^\delta\right), \tag{A.18}$$

and (A.10) follows from (A.16), (A.17) and (A.18). □

As noted in Section 2, we follow [11] by considering the behavior of the random walk Metropolis algorithm over steps of size $[d^\delta]$ iterations. We find that a single component moves only a small distance in $[d^\delta]$ iterations, while over $[d^\delta]$ iterations the acceptance probability, which is dominated by the number and position of components in R_d^l, "forgets" its starting value. Moreover, we show that approximately $\exp(-f^* l/2)[d^\delta]$ of the proposed moves are accepted. However, we need to control the number of components which are *close to* the rejection region (F_d^2) and the distribution of the position of the components in the rejection region after $[d^\beta]$ iterations (F_d^3), where $0 < \beta < \delta$.

For any $k \geq 1$, let

$$\hat{F}_d^2(k) = \{\mathbf{x}^d : |b_d^{k^{3/4}}(\mathbf{x}^d) - \mathbb{E}[b_d^{k^{3/4}}(\mathbf{X}_0^d)]| \leq \sqrt{k}\}$$

and let

$$F_d^2 = \bigcap_{k=[d^\beta]}^{[d^\delta]} \hat{F}_d^2(k). \tag{A.19}$$

Before studying F_d^2, we state a simple, useful result concerning the central moments of a sequence of binomial random variables.

LEMMA A.4. *Let $B_d \sim \mathrm{Bin}(d, p_d)$. Suppose that $p_d \to 0$ and $dp_d \to \infty$ as $d \to \infty$; then for any $m \in \mathbb{N}$,*

$$\mathbb{E}\big[(B_d - \mathbb{E}[B_d])^{2m}\big]/(dp_d)^m \to \prod_{j=1}^m (2j-1) \qquad \text{as } d \to \infty. \tag{A.20}$$

LEMMA A.5. *For any $\kappa > 0$ and sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \leq k_d \leq [d^\delta]$,*

$$d^\kappa \mathbb{P}\big(\mathbf{X}_0^d \notin \hat{F}_d^2(k_d)\big) \to 0 \qquad \text{as } d \to \infty. \tag{A.21}$$

Consequently, for any $\kappa > 0$, $d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d^2) \to 0$ as $d \to \infty$.

PROOF. Fix $\kappa > 0$. By stationarity and Markov's inequality, for all $m \in \mathbb{N}$,

$$\begin{aligned} d^\kappa \mathbb{P}\big(\mathbf{X}_0^d \notin \hat{F}_d^2(k_d)\big) &= d^\kappa \mathbb{P}\big(|b_d^{k_d^{3/4}}(\mathbf{X}_0^d) - \mathbb{E}[b_d^{k_d^{3/4}}(\mathbf{X}_0^d)]| \geq \sqrt{k_d}\big) \\ &\leq \frac{d^\kappa}{k_d^m} \mathbb{E}\big[\big(b_d^{k_d^{3/4}}(\mathbf{X}_0^d) - \mathbb{E}[b_d^{k_d^{3/4}}(\mathbf{X}_0^d)]\big)^{2m}\big]. \end{aligned} \tag{A.22}$$

However, $b_d^{k_d^{3/4}}(\mathbf{X}_0^d) \sim \mathrm{Bin}(d, \int_0^{k_d^{3/4}/d} \{f(x) + f(1-x)\}\,dx)$, so by Lemma A.4 for any $m \in \mathbb{N}$, for all sufficiently large d,

$$\mathbb{E}\big[\big(b_d^{k_d^{3/4}}(\mathbf{X}_0^d) - \mathbb{E}[b_d^{k_d^{3/4}}(\mathbf{X}_0^d)]\big)^{2m}\big] \leq K_m k_d^{3m/4},$$

where $K_m = \prod_{j=1}^{m}(2j-1)+1$. Since $k_d \geq [d^\beta]$, the right-hand side of (A.22) converges to 0 as $d \to \infty$ by taking $m > 4\kappa/\beta$, proving (A.21).

Note that

$$d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d^2) \leq d^\kappa \sum_{k=[d^\beta]}^{[d^\delta]} \mathbb{P}(\mathbf{X}_0^d \notin \hat{F}_d^2(k)). \tag{A.23}$$

The right-hand side of (A.23) converges to 0 as $d \to \infty$ since (A.21) holds with κ replaced by $\kappa + \delta$. $\square$

Before considering F_d^3, the distribution of the position of the components in the rejection region after $[d^\beta]$ iterations, we introduce a simple random walk on the hypercube (RWH). The biggest problem in analyzing the RWM or pseudo-RWM algorithm is the dependence between the components. However, the dependence is weak and whether or not a proposed move is accepted is dominated by whether or not the proposed moves lies inside or outside the hypercube. Therefore we couple the RWM algorithm to the simpler RWH algorithm.

For $d \geq 1$, define the RWH algorithm as follows. Let $\mathbf{W}_k^d$ denote the position of the RWH algorithm after k iterations. Then

$$\mathbf{W}_{k+1}^d = \begin{cases} \mathbf{W}_k^d + \sigma_d \mathbf{Z}_{k+1}^d, & \text{if } \mathbf{W}_k^d + \sigma_d \mathbf{Z}_{k+1}^d \in [0,1]^d, \\ \mathbf{W}_k^d, & \text{otherwise.} \end{cases} \tag{A.24}$$

That is, the RWH algorithm simply accepts all proposed moves which remain inside the hypercube and rejects all proposed moves outside the hypercube. Define the pseudo-RWH algorithm in the obvious fashion with $\hat{\mathbf{W}}_k^d = (\hat{W}_{k,1}^d, \hat{W}_{k,2}^d, \ldots, \hat{W}_{k,d}^d)$ denoting the position of the pseudo-RWH algorithm at iteration k. Then for $1 \leq i \leq d$, $\hat{W}_{k+1,i}^d = \hat{W}_{k,i}^d + \sigma_d \hat{Z}_{k+1,i}^d$, where $\hat{Z}_{k+1,i}^d \sim U[(-\hat{W}_{k,i}^d/\sigma_d) \vee -1, (\hat{W}_{k,i}^d/\sigma_d) \wedge 1]$.

For our purposes it will suffice to consider the coupling of the pseudo-RWM and pseudo-RWH algorithms over $[d^\delta]$ iterations and study how the pseudo-RWH algorithm evolves over $[d^\delta]$ iterations. Note that the RWH algorithm coincides with the RWM algorithm with a uniform target density over the d-dimensional cube, so in this case the coupling is exact.

The components of the pseudo-RWH algorithm behave independently. For $x \in (0,1)$, let $\omega_d(x) = \mathbb{P}(0 < x + \sigma_d Z < 1)$ and for $\mathbf{x}^d \in (0,1)^d$, let $\Omega_d(\mathbf{x}^d) = \prod_{j=1}^d \omega_d(x_j)$. Then $\Omega_d(\mathbf{x}^d)$ is the probability that a proposed move from $\mathbf{x}^d$ is accepted in the RWH algorithm.

LEMMA A.6. *For any $\alpha < \frac{1}{2}$ and $\mathbf{x}^d \in [0,1]^d$, there exists a coupling such that*

$$d^\alpha \mathbb{P}(\mathbf{X}_1^d \neq \mathbf{W}_1^d | \mathbf{X}_0^d \equiv \mathbf{W}_0^d = \mathbf{x}^d) \to 0 \qquad \textit{as } d \to \infty. \tag{A.25}$$

PROOF. Let $U \sim U[0,1]$; then we can couple $\mathbf{X}_1^d$ and $\mathbf{W}_1^d$ using $\mathbf{Z}_1^d$ and U as follows. Let

$$\mathbf{W}_1^d = \begin{cases} \mathbf{x}^d + \sigma_d \mathbf{Z}_1^d, & \text{if } \mathbf{x}^d + \sigma_d \mathbf{Z}_1^d \in [0,1]^d, \\ \mathbf{x}^d, & \text{otherwise,} \end{cases}$$

$$\mathbf{X}_1^d = \begin{cases} \mathbf{x}^d + \sigma_d \mathbf{Z}_1^d, & \text{if } \mathbf{x}^d + \sigma_d \mathbf{Z}_1^d \in [0,1]^d \\ & \text{and } U \le 1 \wedge \exp\left(\sum_{j=1}^d \{g(x_j + \sigma_d Z_{1,j}) - g(x_j)\}\right), \\ \mathbf{x}^d, & \text{otherwise.} \end{cases}$$

Therefore, $\mathbf{X}_1^d \neq \mathbf{W}_1^d$ if $\mathbf{x}^d + \sigma_d \mathbf{Z}_1^d \in [0,1]^d$ and $U > 1 \wedge \exp(\sum_{j=1}^d \{g(x_j + \sigma_d Z_{1,j}) - g(x_j)\})$. Thus

$$\begin{aligned} d^\alpha \mathbb{P}(\mathbf{X}_1^d \neq \mathbf{W}_1^d | \mathbf{X}_0^d \equiv \mathbf{W}_0^d = \mathbf{x}^d) &= d^\alpha \mathbb{P}\Bigg(\mathbf{x}^d + \sigma_d \mathbf{Z}_1^d \in [0,1]^d, \\ &\qquad U > 1 \wedge \exp\left(\sum_{j=1}^d \{g(x_j + \sigma_d Z_{1,j}) - g(x_j)\}\right)\Bigg) \\ &= d^\alpha \mathbb{E}\Bigg[\prod_{j=1}^d 1_{\{0 < x_j + \sigma_d Z_{1,j} < 1\}} \\ &\qquad \times \left\{1 - 1 \wedge \exp\left(\sum_{j=1}^d \{g(x_j + \sigma_d Z_{1,j}) - g(x_j)\}\right)\right\}\Bigg] \\ &\le d^\alpha \mathbb{E}\left[\left|\sum_{j=1}^d \{g(x_j + \sigma_d Z_{1,j}) - g(x_j)\}\right|\right], \end{aligned} \tag{A.26}$$

since for all $y \in \mathbb{R}$, $|1 - \{1 \wedge \exp(y)\}| \le |y|$.

By Taylor's theorem, for $1 \le j \le d$, there exists ξ_j^d lying between 0 and $\sigma_d Z_{1,j}$ such that

$$g(x_j + \sigma_d Z_{1,j}) - g(x_j) = g'(x_j)\sigma_d Z_{1,j} + \frac{g''(x_j + \xi_j^d)}{2}(\sigma_d Z_{1,j})^2. \tag{A.27}$$

Since $g(\cdot)$ is continuously twice differentiable on $(0,1)$, there exists $K < \infty$ such that

$$\left|\sum_{j=1}^d \{g(x_j + \sigma_d Z_{1,j}) - g(x_j)\}\right| \le \left|\frac{l}{d}\sum_{j=1}^d g'(x_j) Z_{1,j}\right| + \frac{Kl^2}{2d}. \tag{A.28}$$

Since the components of $\mathbf{Z}_1^d$ are independent, by Jensen's inequality, (A.28) and $\mathbb{E}[(X+c)^2] \le 2\mathbb{E}[X^2] + 2c^2$, for any random variable X and constant c, we have that

$$\begin{aligned}
&d^\alpha \mathbb{P}(\mathbf{X}_1^d \neq \mathbf{W}_1^d | \mathbf{X}_0^d \equiv \mathbf{W}_0^d = \mathbf{x}^d) \\
&\quad \le d^\alpha \left(2\left\{ \frac{l^2}{3d^2} \sum_{j=1}^d g'(x_j)^2 + \frac{K^2 l^4}{4d^2} \right\} \right)^{1/2} \qquad \text{(A.29)} \\
&\quad \to 0 \qquad \text{as } d \to \infty,
\end{aligned}$$

and the lemma is proved. □

COROLLARY A.7. *Fix* $0 < \alpha < \frac{1}{2} - \delta$.
For any $\mathbf{x}^d \in [0,1]^d$, *there exists a coupling such that*

$$d^\alpha \mathbb{P}\left(\bigcup_{j=0}^{[d^\delta]} \{\mathbf{X}_j^d \neq \mathbf{W}_j^d\} | \mathbf{X}_0^d \equiv \mathbf{W}_0^d = \mathbf{x}^d \right) \to 0 \qquad \textit{as } d \to \infty. \tag{A.30}$$

Moreover, if $\mathbf{x}^d \in \tilde{F}_d^1$ *and* $\alpha + \delta + \gamma < \frac{1}{2}$, *there exists a coupling such that*

$$d^\alpha \mathbb{P}\left(\bigcup_{j=0}^{[d^\delta]} \{\hat{\mathbf{X}}_j^d \neq \hat{\mathbf{W}}_j^d\} | \mathbf{X}_0^d \equiv \mathbf{W}_0^d = \mathbf{x}^d \right) \to 0 \qquad \textit{as } d \to \infty. \tag{A.31}$$

For $r \ge 0$ and $k = 0, 1, 2, \dots,$ let

$$\chi_j^d(x_j; r; k) = \begin{cases} 1, & \text{if } \hat{W}_{k,j}^d \in R_d^r \text{ given that } \hat{W}_{0,j}^d = x_j, \\ 0, & \text{otherwise.} \end{cases} \tag{A.32}$$

Let $q^d(x; r; k) = \mathbb{E}[\chi_1^d(x; r; k)]$ and let $\lambda_d(\mathbf{x}^d; r; k) = \sum_{j=1}^d q^d(x_j; r; k)$. Note that the movement of the components of the pseudo-RWH algorithm are independent.

The next stage in the proof is to show that, if $\hat{\mathbf{X}}_0^d$ is started in F_d^3, then after k_d iterations of the pseudo-RWM algorithm has forgotten its starting value in terms of the total number and position of the components in R_d^l (the rejection region). Moreover, the total number and position of the components in R_d^l after k_d iterations of the pseudo-RWM algorithm are approximately from the stationary distribution of $\hat{\mathbf{X}}_\cdot^d$. Before defining and studying $\{F_d^3\}$, we require the following lemma and associated corollary concerning the distribution of the components in the rejection region after k_d steps.

LEMMA A.8. *Let* $\{k_d\}$ *be any sequence of positive integers satisfying* $[d^\beta] \le k_d \le [d^\delta]$.

For any sequence of $\{\mathbf{x}^d\}$ *such that* $\mathbf{x}^d \in F_d^1 \cap F_d^2$,

$$d^{2\gamma} \sum_{i=1}^{d} q^d(x_i^d; l; k_d)^2 \to 0 \qquad \text{as } d \to \infty. \tag{A.33}$$

Also for all $0 < x < 1$,

$$d^{2\gamma} q^d(x; l; k_d) \to 0 \qquad \text{as } d \to \infty. \tag{A.34}$$

PROOF. Fix $\mathbf{x}^d \in F_d^1 \cap F_d^2$ and set $\hat{\mathbf{W}}_0^d = \mathbf{x}^d$.

To prove (A.33) and (A.34) we couple the components of $\hat{\mathbf{W}}_t^d$ to a simple reflected random walk process $\{S_t^d; t \geq 0\}$. Set $S_0^d = x$ for some $0 < x < 1$. Let $\tilde{Z}_1, \tilde{Z}_2, \dots$ be i.i.d. according to $U[-1, 1]$. For $t \geq 1$, set $S_{t+1}^d = S_t^d + \sigma_d \tilde{Z}_{t+1}$ with reflection at the boundaries 0 and 1 so that $S_t^d \in (0, 1)$. For $x \in (0, 1)$, let $p^d(x; l, k_d) = \mathbb{P}(S_{k_d}^d \in R_d^r | S_0^d = x)$.

Consider $\hat{W}_{t,1}^d$ with identical arguments applying for the other components of $\hat{\mathbf{W}}_t^d$. Since $k_d \sigma_d \to 0$ as $d \to \infty$, we assume that d is such that $(k_d + 1)\sigma_d < \frac{1}{2}$. Then

$$\mathbb{P}\big(\hat{W}_{k_d,1}^d \in R_d^r | \hat{W}_{0,1}^d \in \big((k_d + 1)\sigma_d, 1 - (k_d + 1)\sigma_d\big)\big) = 0. \tag{A.35}$$

For $x \in (0, (k_d + 1)\sigma_d) \cup (1 - (k_d + 1)\sigma_d, 1)$ we can couple S_t^d and $\hat{W}_{t,1}^d$ such that

$$q^d(x; l; k_d) \leq p^d(x; l; k_d). \tag{A.36}$$

For $\sigma_d < y < 1 - \sigma_d$, if $S_t^d = \hat{W}_{t,1}^d = y$, then set $S_{t+1}^d = \hat{W}_{t+1,1}^d = y + \sigma_d \tilde{Z}_{t+1}$. Now if $y < \sigma_d$ $(y > 1 - \sigma_d)$, $\tilde{Z}_{t+1}$ and $\hat{Z}_{t+1,1}^d$ can be coupled such that, if $S_t^d = \hat{W}_{t,1}^d = y$, then $S_{t+1}^d \leq \hat{W}_{t+1,1}^d$ $(S_{t+1}^d \geq \hat{W}_{t+1,1}^d)$. Furthermore, for $y_1 < y_2 < 1/2$ $(y_1 > y_2 > 1/2)$, the above coupling can be extended to give, if $S_t^d = y_1$ and $\hat{W}_{t,1}^d = y_2$, then $S_{t+1}^d < W_{t+1}^d$ $(S_{t+1}^d > W_{t+1}^d)$. Since in k_d iterations either process can move at most a distance $k_d \sigma_d$, (A.36) follows from the above coupling.

Without loss of generality, we assume that $0 < x < (k_d + 1)\sigma_d$ [symmetry arguments apply for $1 - (k_d + 1)\sigma_d < x < 1$]. By the reflection principle,

$$\begin{aligned} p^d(x; l; k_d) &= \mathbb{P}\left(-\sigma_d < x + \sigma_d \sum_{i=1}^{k_d} \tilde{Z}_i < \sigma_d\right) \\ &= \mathbb{P}\left(-1 < \frac{x}{\sigma_d} + \sum_{i=1}^{k_d} \tilde{Z}_i < 1\right). \end{aligned} \tag{A.37}$$

By the Berry–Esséen theorem, there exists a positive constant, $K_1 < \infty$ say, such that for all $z \in \mathbb{R}$,

$$\left| \mathbb{P}\left(\sqrt{\frac{3}{k_d}} \sum_{i=0}^{k_d - 1} Z_i \leq z\right) - \Phi(z) \right| \leq \frac{K_1}{\sqrt{k_d}}, \tag{A.38}$$

where $\Phi(\cdot)$ denotes the c.d.f. of a standard normal. Therefore it follows from (A.37) and (A.38) that there exists a positive constant, $K_2 < \infty$ say, such that for all $x \in (0, 1)$,

$$p^d(x; l; k_d) \leq \frac{K_2}{\sqrt{k_d}}. \tag{A.39}$$

By Hoeffding's inequality, for any $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{i=1}^{k_d} \tilde{Z}_i\right| > \varepsilon k_d^{3/4}\right) &\leq 2\exp\left(-\frac{2(\varepsilon k_d^{3/4})^2}{2^2 k_d}\right) \\ &= 2\exp(-\varepsilon^2\sqrt{k_d}/2). \end{aligned} \tag{A.40}$$

Hence for $k_d^{3/4}/d < x < (k_d + 1)/d$, by taking $\varepsilon = 1/2l$ in (A.40), we have that

$$\begin{aligned} dp^d(x; l; k_d) &= d\mathbb{P}\left(\left|x + \sigma_d \sum_{i=1}^{k_d} \tilde{Z}_i\right| < \sigma_d\right) \\ &\leq d\mathbb{P}\left(\left|\sigma_d \sum_{i=1}^{k_d} \tilde{Z}_i\right| > \frac{k_d^{3/4}}{2d}\right) \\ &\leq 2d\exp\left(-\frac{\sqrt{k_d}}{8l^2}\right) \to 0 \qquad \text{as } d \to \infty. \end{aligned} \tag{A.41}$$

Furthermore, note that for $(k_d + 1)\sigma_d < x < 1 - (k_d + 1)\sigma_d$, $p(x; l; k_d) = 0$.

Then (A.34) follows immediately from (A.36) and the above bounds for $p(x; l; k_d)$ since $d^{2\gamma}/\sqrt{k_d} \to 0$ as $d \to \infty$.

Finally, for $\mathbf{x}^d \in F_d^1 \cap F_d^2$, it follows from (A.36), (A.38) and (A.39) that there exists $K_3 < \infty$ such that

$$\begin{aligned} d^{2\gamma} \sum_{i=1}^{d} q^d(x_i^d; l; k_d)^2 &\leq d^{2\gamma} \sum_{i=1}^{d} p^d(x_i^d; l; k_d)^2 \\ &\leq d^{2\gamma}\left\{K_3 k_d^{3/4}\left(\frac{K_2}{\sqrt{k_d}}\right)^2 + 2d\exp\left(-\frac{\sqrt{k_d}}{8l^2}\right)\right\} \end{aligned} \tag{A.42}$$

with the right-hand side of (A.42) converging to 0 as $d \to \infty$. □

COROLLARY A.9. *For any $m \geq 2$, any sequence $\{r_d\}$ satisfying $0 \leq r_d \leq l$ and any sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \leq k_d \leq [d^\delta]$, there exists $K < \infty$, such that for all $d \geq 1$,*

$$\mathbb{E}[q^d(X_{0,1}^d; r_d; k_d)^m] \leq K d^{-(1+\beta m/8)}. \tag{A.43}$$

PROOF. Fix $m \geq 2$. Note that

$$\begin{aligned}\mathbb{E}[q^d(X_{0,1}^d; r_d; k_d)^m] &\leq \mathbb{E}[q^d(X_{0,1}^d; l; k_d)^m] \\ &= \int_0^1 q^d(x; l; k_d)^m f(x)\,dx \\ &= \int_{R_d^{k_d^{3/4}}} q^d(x; l; k_d)^m f(x)\,dx \\ &\quad + \int_{(R_d^{k_d^{3/4}})^C} q^d(x; l; k_d)^m f(x)\,dx.\end{aligned} \tag{A.44}$$

The two terms on the right-hand side of (A.44) are bounded using (A.39) and (A.41), respectively. Thus it follows from the proof of Lemma A.8 that there exist constants $K_1, K_2 < \infty$ such that, for all $d \geq 1$,

$$\begin{aligned}\mathbb{E}[q^d(X_{0,1}^d; r_d; k_d)^m] &\leq \int_{R_d^{k_d^{3/4}}} \left(\frac{K_1}{\sqrt{k_d}}\right)^m f(x)\,dx \\ &\quad + \int_{(R_d^{k_d^{3/4}})^C} \left\{2\exp\left(-\frac{\sqrt{k_d}}{8l^2}\right)\right\}^m f(x)\,dx \\ &\leq \mathbb{P}(X_{0,1}^d \in R_d^{k_d^{3/4}})\left(\frac{K_1}{\sqrt{k_d}}\right)^m \\ &\quad + \mathbb{P}(X_{0,1}^d \notin R_d^{k_d^{3/4}}) \times 2\exp\left(-\frac{\sqrt{k_d}}{8l^2}\right) \\ &\leq K_2 \frac{k_d^{3/4}}{d} k_d^{-m/2} + 2\exp\left(-\frac{\sqrt{k_d}}{8l^2}\right).\end{aligned} \tag{A.45}$$

The corollary follows from (A.45) since $m \geq 2$ and $k_d \geq [d^\beta]$. $\square$

We are now in position to define $\{F_d^3\}$. For any $0 \leq r \leq l$ and $k \in \mathbb{Z}^+$, let

$$\hat{F}_d^3(r; k) = \{\mathbf{x}^d : |\lambda_d(\mathbf{x}^d; r; k) - \lambda(r)| < d^{-\gamma}/8\}, \tag{A.46}$$

where $\lambda(r) = f^* r(1 + r/2l)$. Let

$$F_d^3 = \left\{\mathbf{x}^d : \sup_{[d^\beta] \leq k_d \leq [d^\delta]} \sup_{0 \leq r \leq l} |\lambda_d(\mathbf{x}^d; r; k_d) - \lambda(r)| < d^{-\gamma}\right\}. \tag{A.47}$$

We study $\{\hat{F}_d^3(r_d, k_d)\}$ as a prelude to analyzing $\{F_d^3\}$ where r_d and k_d are defined in Lemma A.10 below.

LEMMA A.10. *For any sequence $\{r_d\}$ satisfying $0 \leq r_d \leq l$, any sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \leq k_d \leq [d^\delta]$ and $\kappa > 0$,*

$$d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin \hat{F}_d^3(r_d, k_d)) \to 0 \qquad \text{as } d \to \infty. \tag{A.48}$$

PROOF. By the triangle inequality,

$$
\begin{aligned}
&d^\kappa \mathbb{P}\big(\mathbf{X}_0^d \notin \hat{F}_d^3(r_d; k_d)\big) \\
&\quad \le d^\kappa \mathbb{P}\big(|\lambda_d(\mathbf{X}_0^d; r_d; k_d) - \mathbb{E}[\lambda_d(\mathbf{X}_0^d; r_d; k_d)]| > d^{-\gamma}/16\big) \\
&\qquad + d^\kappa \mathbb{P}\big(|\mathbb{E}[\lambda_d(\mathbf{X}_0^d; r_d; k_d)] - \lambda(r_d)| > d^{-\gamma}/16\big).
\end{aligned} \tag{A.49}
$$

In turn we show that the two terms on the right-hand side of (A.49) converge to 0 as $d \to \infty$.

By Markov's inequality, we have that for any $m \in \mathbb{N}$,

$$
\begin{aligned}
&d^\kappa \mathbb{P}\big(|\lambda_d(\mathbf{X}_0^d; r_d; k_d) - \mathbb{E}[\lambda_d(\mathbf{X}_0^d; r_d; k_d)]| > d^{-\gamma}/16\big) \\
&\quad \le 16^m d^{\kappa + m\gamma} \mathbb{E}\left[\left(\sum_{j=1}^d \{q^d(X_{0,j}; r_d; k_d) - \mathbb{E}[q^d(X_{0,j}; r_d; k_d)]\}\right)^m\right] \\
&\quad = 16^m d^{\kappa + m\gamma} \sum_{i_1=1}^d \cdots \sum_{i_m=1}^d \mathbb{E}\left[\prod_{j=1}^m \{q^d(X_{0,i_j}; r_d; k_d) \right. \\
&\qquad\qquad \left. - \mathbb{E}[q^d(X_{0,i_j}; r_d; k_d)]\}\right].
\end{aligned} \tag{A.50}
$$

Since the components of $\mathbf{X}_0^d$ are independent and identically distributed, we have for any $\{i_1, i_2, \dots, i_m\}$, there exists $1 \le J \le m$ and $l_1, l_2, \dots, l_J \ge 1$ with $l_1 + l_2 + \cdots + l_J = m$ such that

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{j=1}^m \{q^d(X_{0,i_j}; r_d; k_d) - \mathbb{E}[q^d(X_{0,i_j}; r_d; k_d)]\}\right] \\
&\qquad = \prod_{j=1}^J \mathbb{E}\big[\{q^d(X_{0,1}; r_d; k_d) - \mathbb{E}[q^d(X_{0,1}; r_d; k_d)]\}^{l_j}\big].
\end{aligned} \tag{A.51}
$$

Note that if any $l_j = 1$, then the right-hand side of (A.51) is equal to 0. By Corollary A.9, if $l_1, l_2, \dots, l_J \ge 2$, there exists $K_1 < \infty$ such that the right-hand side of (A.51) is less than or equal to $\prod_{j=1}^J \{K_1 d^{-(1 + l_j \beta/8)}\} = K_1^J d^{-J} d^{-m\beta/8}$. Furthermore, there exists $K_2 < \infty$ such that for any $1 \le J \le m$ and $l_1, l_2, \dots, l_J \ge 2$, there are at most $K_2 d^J$ configurations of $\{i_1, i_2, \dots, i_m\}$ such that for $j = 1, 2, \dots, J$, l_j of the components are the same. Therefore there exists $K < \infty$ such that

$$
\begin{aligned}
&\sum_{i_1=1}^d \cdots \sum_{i_m=1}^d \mathbb{E}\left[\prod_{j=1}^m \{q^d(X_{0,i_j}; r_d; k_d) - \mathbb{E}[q^d(X_{0,i_j}; r_d; k_d)]\}\right] \\
&\qquad \le K d^{-m\beta/8}.
\end{aligned} \tag{A.52}
$$

Taking $m > \kappa/(\beta/8 - \gamma)$, it follows from (A.52) that the right-hand side of (A.50) converges to 0 as $d \to \infty$.

The lemma follows by showing that for all sufficiently large d,

$$|\mathbb{E}[\lambda_d(\mathbf{X}_0^d; r_d; k_d)] - \lambda(r_d)| \leq d^{-\gamma}/16. \tag{A.53}$$

Note that

$$\begin{aligned}\mathbb{E}[\lambda_d(\mathbf{X}_0^d; r_d; k_d)] &= d\mathbb{E}[q^d(X_{0,1}; r; k_d)] \\ &= d\int_0^{k_d^{3/4}/d} q^d(x; r_d; k_d) f(x)\,dx \\ &\quad + d\int_{k_d^{3/4}/d}^{1-k_d^{3/4}/d} q^d(x; r_d; k_d) f(x)\,dx \\ &\quad + d\int_{1-k_d^{3/4}/d}^{1} q^d(x; r_d; k_d) f(x)\,dx.\end{aligned} \tag{A.54}$$

By (A.41), the second integral on the right-hand side of (A.54) is bounded above by $d \times 2\exp(-\sqrt{k_d}/8l^2) \to 0$ as $d \to \infty$. Let $f_\star = \sup_{0 \leq x \leq 1} |f'(x)|$. Then by Taylor's theorem, for $0 \leq x \leq k_d^{3/4}/d$,

$$|f(x) - f(0)| \leq x \sup_{0 \leq y \leq x} f'(y) \leq f_\star k_d^{3/4}/d. \tag{A.55}$$

Thus

$$\begin{aligned}&\left| d\int_0^{k_d^{3/4}/d} q^d(x; r_d; k_d) f(x)\,dx - f(0) d\int_0^{k_d^{3/4}/d} q^d(x; r_d; k_d)\,dx \right| \\ &\qquad \leq d \times f_\star \frac{k_d^{3/4}}{d} \times \int_0^{k_d^{3/4}/d} q^d(x; r_d; k_d)\,dx.\end{aligned} \tag{A.56}$$

Similarly, we have that

$$\begin{aligned}&\left| d\int_{1-k_d^{3/4}/d}^{1} q^d(x; r_d; k_d) f(x)\,dx - f(1) d\int_{1-k_d^{3/4}/d}^{1} q^d(x; r_d; k_d)\,dx \right| \\ &\qquad \leq d \times f_\star \frac{k_d^{3/4}}{d} \times \int_{1-k_d^{3/4}/d}^{1} q^d(x; r_d; k_d)\,dx.\end{aligned} \tag{A.57}$$

By symmetry, $q^d(1 - x; r_d; k_d) = q^d(x; r_d; k_d)$, so

$$d^\gamma \left| \mathbb{E}[\lambda_d(\mathbf{X}_0^d; r_d; k_d)] - 2f^* d\int_0^1 q^d(x; r_d; k_d)\,dx \right| \to 0 \qquad \text{as } d \to \infty. \tag{A.58}$$

Since $\int_0^1 \omega_d(y)\,dy \geq 1 - 2\sigma_d$, using Lemma A.8, (A.34), we have that, for all sufficiently large d,

$$\begin{aligned}
&d^{\gamma}\left|d\int_0^1 q^d(x;r_d;k_d)\,dx - d\int_0^1 q(x;r_d;k_d)\frac{\omega_d(x)}{\int_0^1 \omega_d(y)\,dy}\,dx\right| \\
&\quad\leq 4d^{1+\gamma}\int_0^{\sigma_d} q^d(x;r_d;k_d)\,dx \\
&\qquad + d\int_{\sigma_d}^{1-\sigma_d}\left\{\frac{1}{\int_0^1 \omega_d(y)\,dy} - 1\right\}q^d(x;r_d;k_d)\,dx \\
&\quad\leq 4d^{1+\gamma}\sigma_d d^{-2\gamma} + d^{1+\gamma}\int_0^1 \frac{2\sigma_d}{\int_0^1 \omega_d(y)\,dy}q^d(x;r_d;k_d)\,dx.
\end{aligned} \tag{A.59}$$

Let $p^d(x;r_d;k_d)$ be defined as in Lemma A.8. Note that $U[0,1]$ is the stationary distribution of a reflected random walk on $(0,1)$. Therefore for any $k \geq 1$,

$$\int_0^1 p^d(x;r_d;k)\,dx = \int_0^1 p^d(x;r_d;0)\,dx = \frac{2r_d}{d}. \tag{A.60}$$

Therefore, it follows from Lemma A.8, (A.36), that

$$d\int_0^1 q^d(x;r_d;k_d)\,dx \leq d\int_0^1 p^d(x;r_d;k_d)\,dx = 2r_d. \tag{A.61}$$

Hence the right-hand side of (A.59) converges to 0 as $d \to \infty$.

Note that the stationary distribution of a single component of the pseudo-RWH algorithm has p.d.f. $\omega_d(x)/\int_0^1 \omega_d(y)\,dy$ $(0 < x < 1)$. Therefore

$$\begin{aligned}
d\int_0^1 q^d(x;r_d;k_d)\frac{\omega_d(x)}{\int_0^1 \omega_d(y)\,dy}\,dx &= d\int_0^1 q^d(x;r_d;0)\frac{\omega_d(x)}{\int_0^1 \omega_d(y)\,dy}\,dx \\
&= \frac{r_d}{2}\left(1+\frac{r_d}{2l}\right)\Big/\left(1-\frac{l}{2d}\right).
\end{aligned} \tag{A.62}$$

Finally, combining (A.58), (A.59) and (A.62), we have that (A.53) holds and the lemma is proved. □

LEMMA A.11. *For any* $\kappa > 0$,

$$d^{\kappa}\mathbb{P}(\mathbf{X}_0^d \notin F_d^3) \to 0 \qquad \text{as } d \to \infty. \tag{A.63}$$

PROOF. Fix $\kappa > 0$. Fix a sequence of positive integers $\{k_d\}$ such that $[d^{\beta}] \leq k_d \leq [d^{\delta}]$. Fix $\theta > \gamma$ and let $\mathcal{S}_d = \{0, d^{-\theta}, 2d^{-\theta}, \ldots, [ld^{\theta}]d^{-\theta}, l\}$. Thus the elements of $\mathcal{S}_d$ are separated by a distance of at most $d^{-\theta}$.

For any $0 \le r \le l$ and $d \ge 1$, there exist $\tilde{r}_d, \hat{r}_d \in \mathcal{S}_d$ such that $\tilde{r}_d \le r < \hat{r}_d$ with $\hat{r}_d - \tilde{r}_d \le d^{-\theta}$. By the triangle inequality,

$$\begin{aligned}
&|\lambda_d(\mathbf{X}_0^d; r; k_d) - \lambda(r)| \\
&\quad \le \lambda_d(\mathbf{X}_0^d; r; k_d) - \lambda_d(\mathbf{X}_0^d; \tilde{r}_d; k_d) + |\lambda_d(\mathbf{X}_0^d; \tilde{r}_d; k_d) - \lambda(\tilde{r}_d)| \\
&\qquad + \lambda(r) - \lambda(\tilde{r}_d) \\
&\quad \le \lambda_d(\mathbf{X}_0^d; \hat{r}_d; k_d) - \lambda_d(\mathbf{X}_0^d; \tilde{r}_d; k_d) + |\lambda_d(\mathbf{X}_0^d; \tilde{r}_d; k_d) - \lambda(\tilde{r}_d)| \\
&\qquad + \lambda(\hat{r}_d) - \lambda(\tilde{r}_d) \\
&\quad \le |\lambda_d(\mathbf{X}_0^d; \hat{r}_d; k_d) - \mathbb{E}[\lambda_d(\mathbf{X}_0^d; \hat{r}_d; k_d)]| \\
&\qquad + 2|\lambda_d(\mathbf{X}_0^d; \tilde{r}_d; k_d) - \mathbb{E}[\lambda_d(\mathbf{X}_0^d; \tilde{r}_d; k_d)]| \\
&\qquad + 2|\lambda(\hat{r}_d) - \lambda(\tilde{r}_d)|.
\end{aligned} \tag{A.64}$$

By Lemma A.10, for any sequence $\{r_d\}$ satisfying $0 \le r_d \le l$,

$$d^{\kappa+\theta+\delta}\mathbb{P}\biggl(|\lambda_d(\mathbf{X}_0^d; r_d; k_d) - \lambda(r_d)| > \frac{d^{-\gamma}}{8}\biggr) \to 0 \qquad \text{as } d \to \infty. \tag{A.65}$$

Hence

$$d^{\kappa+\delta}\mathbb{P}\biggl(\max_{r \in \mathcal{S}_d} |\lambda_d(\mathbf{X}_0^d; r; k_d) - \lambda(r)| > \frac{d^{-\gamma}}{8}\biggr) \to 0 \qquad \text{as } d \to \infty. \tag{A.66}$$

For all sufficiently large d,

$$\sup_{0 \le r_d, s_d \le l, |r_d - s_d| \le d^{-\theta}} |\lambda(r_d) - \lambda(s_d)| \le \frac{d^{-\gamma}}{16}. \tag{A.67}$$

Therefore it follows from (A.64), (A.66) and (A.67) that

$$d^{\kappa+\delta}\mathbb{P}\Bigl(\sup_{0 \le r \le l} |\lambda_d(\mathbf{X}_0^d; r; k_d) - \lambda(r)| > d^{-\gamma}\Bigr) \to 0 \qquad \text{as } d \to \infty. \tag{A.68}$$

Since (A.68) holds for any sequence $\{k_d\}$ satisfying $[d^\beta] \le k_d \le [d^\delta]$, the lemma follows since

$$\begin{aligned}
&d^\kappa \mathbb{P}\Bigl(\sup_{[d^\beta] \le k \le [d^\delta]} \sup_{0 \le r \le l} |\lambda_d(\mathbf{X}_0^d; r; k) - \lambda(r)| > d^{-\gamma}\Bigr) \\
&\quad \le d^\kappa \sum_{k=[d^\beta]}^{[d^\delta]} \mathbb{P}\Bigl(\sup_{0 \le r \le l} |\lambda_d(\mathbf{X}_0^d; r; k) - \lambda(r)| > d^{-\gamma}\Bigr).
\end{aligned} \tag{A.69}$$

□

Finally, we consider

$$F_d^4 = \biggl\{\mathbf{x}^d; \biggl|\frac{1}{d}\sum_{j=1}^d g'(x_j)^2 - \mathbb{E}[g'(X_1)^2]\biggr| < d^{-1/8}\biggr\}. \tag{A.70}$$

The sets $\{F_d^4\}$ mirror the sets $\{F_n\}$ in [14] and are used when considering $G_d^\delta H(\mathbf{x}^d)$ and $\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d)$ but play no role in analyzing P_d.

LEMMA A.12. *For any* $\kappa > 0$,

$$d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d^4) \to 0 \qquad \textit{as } d \to \infty. \tag{A.71}$$

PROOF. Let $g^* = \sup_{0 \le y \le 1} |g'(y)|$ and fix $\kappa > 0$. Then by Hoeffding's inequality,

$$\begin{aligned} d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d^4) &= d^\kappa \mathbb{P}\left(\left|\sum_{j=1}^d g'(X_{0,j})^2 - d\mathbb{E}[g'(X_{0,1})^2]\right| > d^{7/8}\right) \\ &\le d^\kappa \times 2\exp\left(-\frac{2d^{7/4}}{d(g^*)^4}\right) \to 0 \qquad \text{as } d \to \infty. \end{aligned} \tag{A.72}$$

□

Finally we are in position to consider $\{F_d\}$ and $\{\tilde{F}_d\}$. Recall that, for $d \ge 1$, $F_d = F_d^1 \cap F_d^2 \cap F_d^3 \cap F_d^4$ and

$$\tilde{F}_d = \left\{\mathbf{x}^d; \mathbb{P}\left(\bigcup_{j=0}^{[d^\delta]} \hat{\mathbf{X}}_j^d \notin F_d | \hat{\mathbf{X}}_0^d = \mathbf{x}^d\right) \le d^{-3}\right\}.$$

Combining Lemmas A.1, A.5, A.11 and A.12, we have the following theorem.

THEOREM A.13. *For any* $\kappa > 0$,

$$d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin F_d) \to 0 \qquad \textit{as } d \to \infty. \tag{A.73}$$

Hence, by Lemma A.3, *for any* $\kappa > 0$,

$$d^\kappa \mathbb{P}(\mathbf{X}_0^d \notin \tilde{F}_d) \to 0 \qquad \textit{as } d \to \infty. \tag{A.74}$$

Also using the couplings outlined above, we have that

$$\mathbb{P}\left(\bigcup_{j=0}^{[d^\delta]} \{\hat{\mathbf{W}}_j^d \notin F_d\} | \hat{\mathbf{W}}_0^d \in \tilde{F}_d\right) \to 0 \qquad \textit{as } d \to \infty. \tag{A.75}$$

APPENDIX B: PROOF OF $P_d | \mathbf{X}_0^d = \mathbf{x}^d \xrightarrow{p} \exp(-lf^*/2)$

We show that for any sequence $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d \in \tilde{F}_d$,

$$P_d | \mathbf{X}_0^d = \mathbf{x}^d \xrightarrow{p} \exp(-lf^*/2) \qquad \text{as } d \to \infty. \tag{B.1}$$

The key result is Lemma B.1 which states that after k_d iterations, the configuration of the components in the rejection region R_d^l resemble the configuration of the points of a Poisson point process with rate $\lambda(r) = f^* r(1 + r/2l)$ on the interval $[0, l]$.

For any $n \in \mathbb{N}$ and $1 \le i \le n$, let

$$S_n^d(\mathbf{x}^d; i; k) = \sum_{j=1}^{d} \{\chi_i^d(x_j; il/n; k) - \chi_j^d(x_j; (i-1)l/n; k)\}$$

with

$$\mathbf{S}_n^d(\mathbf{x}^d; k) = (S_n^d(\mathbf{x}^d; 1; k), S_n^d(\mathbf{x}^d; 2; k), \dots, S_n^d(\mathbf{x}^d; n; k)).$$

Let $\mathbf{S}_n = (S_n^1, S_n^2, \dots, S_n^n)$ where the components of $\mathbf{S}_n$ are independent Poisson random variables with $S_n^i \sim \mathrm{Po}(\lambda_{n,i})$ and

$$\lambda_{n,i} = \lambda(il/n) - \lambda((i-1)l/n) \qquad (1 \le i \le n).$$

LEMMA B.1. *For any $n \in \mathbb{N}$, any sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \le k_d \le [d^\delta]$ and $\mathbf{x}^d \in F_d$,*

$$\mathbf{S}_n^d(\mathbf{x}^d; k_d) \xrightarrow{D} \mathbf{S}_n \qquad \text{as } d \to \infty.$$

PROOF. Fix $n \in \mathbb{N}$ and $\mathbf{x}^d \in F_d$. Let

$$\check{\mathbf{S}}_n^d(\mathbf{x}^d; k_d) = (\check{S}_n^d(\mathbf{x}^d; 1; k_d), \check{S}_n^d(\mathbf{x}^d; 2; k_d), \dots, \check{S}_n^d(\mathbf{x}^d; n; k_d)),$$

where for $1 \le i \le n$, $\check{S}_n^d(\mathbf{x}^d; i; k_d)$ are independent Poisson random variables with means

$$\lambda_{n,i}^d(\mathbf{x}^d; k_d) = \lambda_d(\mathbf{x}^d; il/n; k_d) - \lambda_d(\mathbf{x}^d; (i-1)l/n; k_d).$$

The lemma is proved by showing that

$$\begin{aligned} d_{TV}(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d; k_d), \mathbf{S}_n) &\le d_{TV}(\mathbf{S}_n^d(\mathbf{x}^d; k_d), \check{\mathbf{S}}_n^d(\mathbf{x}^d; k_d)) \\ &\quad + d_{TV}(\check{\mathbf{S}}_n^d(\mathbf{x}^d; k_d), \mathbf{S}_n) \\ &\to 0 \qquad \text{as } d \to \infty. \end{aligned} \tag{B.2}$$

By [1], Theorem 1,

$$d_{TV}(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d; k_d), \check{\mathbf{S}}_n^d(\mathbf{x}^d; k_d)) \le \sum_{i=1}^{d} q^d(x_i; l; k_d)^2. \tag{B.3}$$

By Lemma A.8, (A.33) the right-hand side of (B.3) converges to 0 as $d \to \infty$.

For the second term on the right-hand side of (B.2), it suffices to show that

$$\check{\mathbf{S}}_n^d(\mathbf{x}^d; k_d) \xrightarrow{D} \mathbf{S}_n \qquad \text{as } d \to \infty.$$

(For discrete random variables convergence in distribution and convergence in total variation distance are equivalent; see [2], page 254.)

The components of $\check{\mathbf{S}}_n^d(\mathbf{x}^d; k_d)$ and $\mathbf{S}_n$ are independent, and therefore it is sufficient to show that, for all $1 \le i \le n$,

$$\check{S}_n^d(\mathbf{x}^d; i; k_d) \xrightarrow{D} S_{n,i} \qquad \text{as } d \to \infty. \tag{B.4}$$

For all $1 \le i \le n$, (B.4) holds, if

$$\lambda_{n,i}^d(\mathbf{x}^d; k_d) \to \lambda_{n,i} \qquad \text{as } d \to \infty. \tag{B.5}$$

Therefore the lemma follows from (B.5) since $[d^\beta] \le k_d \le [d^\delta]$ and $\mathbf{x}^d \in F_d^3$. [See (2.17) for the construction of $\{F_d^3\}$.] $\square$

Lemma B.1 is the key result stating that if the pseudo-RWH process is started from the set F_d, then after $[d^\beta]$ iterations the distribution of the components in the rejection region are approximately given by $\mathbf{S}_n$. We show that studying the pseudo-RWH algorithm over $[d^\delta]$ iterations suffices in analyzing $T_d(\pi) = \frac{1}{[d^\delta]} \sum_{j=0}^{[\pi d^\delta - 1]} M_j(J_d(\hat{\mathbf{X}}_j^d))$. Note that P_d satisfies

$$T_d(P_d) \le 1 < T_d(P_d + 1/[d^\delta]). \tag{B.6}$$

Let $\hat{T}_d(\pi) = \frac{1}{[d^\delta]} \sum_{j=0}^{[\pi d^\delta - 1]} M_j(\Omega_d(\hat{\mathbf{W}}_j^d))$. Before establishing a coupling between $T_d(\pi)$ and $\hat{T}_d(\pi)$, we give a simple coupling for geometric random variables.

LEMMA B.2. *Suppose that $0 \le q < p \le 1$ and that X and Y are independent geometric random variables with success probabilities p and q, respectively, that is, $X \sim M(p)$ and $Y \sim M(q)$. Let A be a Bernoulli random variable with $\mathbb{P}(A = 1) = q/p$ and $Z \sim M(q)$. Then if A, X, Y and Z are mutually independent,*

$$Y \stackrel{D}{=} X + (1 - A)Z. \tag{B.7}$$

Therefore there exists a coupling of X and Y such that

$$\mathbb{P}(X \ne Y) = P(A = 0) = \frac{p - q}{p}. \tag{B.8}$$

LEMMA B.3. *For any $0 < \pi \le 1$ and $\mathbf{x}^d \in \tilde{F}_d$, there exists a coupling of $T_d(\pi)$ and $\hat{T}_d(\pi)$ such that*

$$\mathbb{P}\big(T_d(\pi) \ne \hat{T}_d(\pi) | \hat{\mathbf{X}}_0^d \equiv \hat{\mathbf{W}}_0^d = \mathbf{x}^d\big) \to 0 \qquad \textit{as } d \to \infty. \tag{B.9}$$

PROOF. For $\mathbf{x}^d \in \tilde{F}_d$, by Corollary A.7, we have that

$$\mathbb{P}\left(\bigcup_{j=0}^{[d^\delta]} \{\hat{\mathbf{X}}_j^d \ne \hat{\mathbf{W}}_j^d\} | \hat{\mathbf{X}}_0^d \equiv \hat{\mathbf{W}}_0^d = \mathbf{x}^d \right) \to 0 \qquad \text{as } d \to \infty. \tag{B.10}$$

Suppose that for $j = 0, 1, \dots, [d^\delta]$, $\hat{\mathbf{W}}_j^d = \hat{\mathbf{X}}_j^d \in F_d^1$. Then using Lemma B.2, (B.8), $M_j(J_d(\hat{\mathbf{X}}_j^d))$ and $M_j(\Omega_d(\hat{\mathbf{W}}_j^d))$ can be coupled such that

$$\mathbb{P}\big(M_j(J_d(\hat{\mathbf{X}}_j^d)) \neq M_j(\Omega_d(\hat{\mathbf{W}}_j^d)) | \hat{\mathbf{W}}_j^d = \hat{\mathbf{X}}_j^d \in F_d^1\big) \leq \frac{\Omega_d(\hat{\mathbf{X}}_j^d) - J_d(\hat{\mathbf{X}}_j^d)}{\Omega_d(\hat{\mathbf{X}}_j^d)}. \tag{B.11}$$

Since $\hat{\mathbf{X}}_j^d \in F_d^1$, $\Omega_d(\hat{\mathbf{X}}_j^d) \geq 2^{-\gamma \log d} \geq d^{-\gamma}$, the right-hand side of (B.11) is less than $d^\gamma \{\Omega_d(\hat{\mathbf{X}}_j^d) - J_d(\hat{\mathbf{X}}_j^d)\}$. Note that

$$\mathbb{P}(\hat{\mathbf{W}}_{j+1}^d \neq \hat{\mathbf{X}}_{j+1}^d | \hat{\mathbf{W}}_j^d = \hat{\mathbf{X}}_j^d \in F_d^1) = \Omega_d(\hat{\mathbf{X}}_j^d) - J_d(\hat{\mathbf{X}}_j^d),$$

so by Lemma A.6 for any $\alpha < \frac{1}{2}$, $d^{\alpha-\gamma}$ times the right-hand side of (B.11) converges to 0 as $d \to \infty$. Taking α such that $\delta + \gamma < \alpha < \frac{1}{2}$,

$$\begin{aligned}&\sum_{j=0}^{[d^\delta]} \mathbb{P}\big(M_j(J_d(\hat{\mathbf{X}}_j^d)) \neq M_j(\Omega_d(\hat{\mathbf{W}}_j^d)) | \hat{\mathbf{W}}_j^d = \hat{\mathbf{X}}_j^d \in F_d^1\big) \\ &\qquad \to 0 \qquad \text{as } d \to \infty.\end{aligned} \tag{B.12}$$

The lemma then follows from (B.10) and (B.12). $\square$

We show that it suffices to study $\tilde{T}_d(\pi) = \frac{1}{[d^\delta]} \sum_{j=0}^{[\pi d^\delta - 1]} \Omega_d(\hat{\mathbf{W}}_j^d)^{-1}$. In other words, replace the mean of the geometric random variables $\{M(\Omega_d(\hat{\mathbf{W}}_0^d)), M(\Omega_d(\hat{\mathbf{W}}_1^d)), \dots, M(\Omega_d(\hat{\mathbf{W}}_{[\pi d^\delta - 1]}^d))\}$ by the mean of the means of the geometric random variables.

LEMMA B.4. *For any* $0 < \pi \leq 1$ *and for any sequence of* $\{\mathbf{x}^d\}$ *such that* $\mathbf{x}^d \in \tilde{F}_d$, $\hat{T}_d(\pi) | \hat{\mathbf{W}}_0^d = \mathbf{x}^d \xrightarrow{D} \pi \exp(f^* l/2)$ *if* $\tilde{T}_d(\pi) | \hat{\mathbf{W}}_0^d = \mathbf{x}^d \xrightarrow{D} \pi \exp(f^* l/2)$ *as* $d \to \infty$.

PROOF. Let $A_d = \bigcup_{j=0}^{[d^\delta]} \{\hat{\mathbf{W}}_j^d \notin F_d\}$. Then for any $\mathbf{x}^d \in \tilde{F}_d$, $\mathbb{P}(A_d | \hat{\mathbf{W}}_0^d = \mathbf{x}^d) \to 0$ as $d \to \infty$.

For any $\tau \in \mathbb{R}$ with $i = \sqrt{-1}$, the characteristic function of $\hat{T}_d(\pi)$ conditional upon A_d^C and $\hat{\mathbf{W}}_0^d = \mathbf{x}^d$ is given by

$$\begin{aligned}&\mathbb{E}[\exp(i\tau \hat{T}_d(\pi)) | A_d^C, \hat{\mathbf{W}}_0^d = \mathbf{x}^d] \\ &\quad = \mathbb{E}\left[\prod_{j=0}^{[\pi d^\delta - 1]} \mathbb{E}\left[\exp\left(\frac{i\tau}{[d^\delta]} M_j(\Omega_d(\hat{\mathbf{W}}_j^d))\right) \middle| A_d^C, \{\hat{\mathbf{W}}^d\}\right] \middle| A_d^C, \hat{\mathbf{W}}_0^d = \mathbf{x}^d\right] \\ &\quad = \mathbb{E}\left[\prod_{j=0}^{[\pi d^\delta - 1]} \frac{\exp(i\tau/[d^\delta]) \Omega_d(\hat{\mathbf{W}}_j^d)}{1 - (1 - \Omega_d(\hat{\mathbf{W}}_j^d)) \exp(i\tau/[d^\delta])} \middle| A_d^C, \hat{\mathbf{W}}_0^d = \mathbf{x}^d\right].\end{aligned} \tag{B.13}$$

Conditional upon A_d^C, $\Omega_d(\hat{\mathbf{W}}_j^d)^{-1} \leq 2^{\gamma \log d} \leq d^\gamma$. Hence, for all $0 \leq j \leq [\pi d^\delta - 1]$,

$$\frac{\exp(i\tau/[d^\delta])\Omega_d(\hat{\mathbf{W}}_j^d)}{1-(1-\Omega_d(\hat{\mathbf{W}}_j^d))\exp(i\tau/[d^\delta])} = 1 + \frac{i\tau}{[d^\delta]}\Omega_d(\hat{\mathbf{W}}_j^d)^{-1} + o\left(\frac{1}{[d^\delta]}\right). \tag{B.14}$$

Thus $\mathbb{E}[\exp(i\tau\hat{T}_d(\pi))|A_d^C, \hat{\mathbf{W}}_0^d = \mathbf{x}^d]$ has the same limit as $d \to \infty$ (should one exist) as

$$\mathbb{E}\left[\prod_{j=0}^{[\pi d^\delta - 1]}\left(1 + \frac{i\tau}{[d^\delta]}\Omega_d(\hat{\mathbf{W}}_j^d)^{-1}\right)\Big| A_d^C, \hat{\mathbf{W}}_0^d = \mathbf{x}^d\right], \tag{B.15}$$

which in turn has the same limit as $d \to \infty$ as

$$\begin{aligned} &\mathbb{E}\left[\prod_{j=0}^{[\pi d^\delta - 1]}\exp\left(\frac{i\tau}{[d^\delta]}\Omega_d(\hat{\mathbf{W}}_j^d)^{-1}\right)\Big| A_d^C, \hat{\mathbf{W}}_0^d = \mathbf{x}^d\right] \\ &\quad = \mathbb{E}[\exp(i\tau\tilde{T}_d(\pi))|A_d^C, \hat{\mathbf{W}}_0^d = \mathbf{x}^d]. \end{aligned} \tag{B.16}$$

The lemma follows since $\mathbb{P}(A_d^C|\hat{\mathbf{W}}_0^d = \mathbf{x}^d) \to 1$ as $d \to \infty$. □

We shall show that $\tilde{T}_d(\pi) \xrightarrow{p} \exp(lf^*/2)$ as $d \to \infty$ using Chebyshev's inequality in Lemma B.9. We require preliminary results concerning $\operatorname{cov}(\Omega_d(\hat{\mathbf{W}}_j^d)^{-1}, \Omega_d(\hat{\mathbf{W}}_{j+k}^d)^{-1}|\hat{\mathbf{W}}_0^d = \mathbf{x}^d)$ with the key results given in Lemma B.8. First, however, we introduce useful upper and lower bounds for $\Omega_d(\mathbf{x}^d)^{-1}$ which allow us to exploit Lemma B.1 and prove uniform integrability $\{\tilde{T}_d(\pi)\}$.

For $n \in \mathbb{N}$, $1 \leq i \leq n$ and $\mathbf{x}^d \in (0,1)^d$, let $\tilde{b}_d^{n,i}(\mathbf{x}^d) = b_d^{il/n}(\mathbf{x}^d) - b_d^{(i-1)l/n}(\mathbf{x}^d)$ with $\tilde{\mathbf{b}}_d^n(\mathbf{x}^d) = (\tilde{b}_d^{n,1}(\mathbf{x}^d), \tilde{b}_d^{n,2}(\mathbf{x}^d), \dots, \tilde{b}_d^{n,n}(\mathbf{x}^d))$. For $n \in \mathbb{N}$ and $\mathbf{s} = (s_1, s_2, \dots, s_n) \in \mathbb{R}^n$, let

$$\check{\nu}_n(\mathbf{s}) = \prod_{j=1}^n \left(\frac{1}{2} + \frac{j-1}{2n}\right)^{-s_j}, \tag{B.17}$$

$$\hat{\nu}_n(\mathbf{s}) = \prod_{j=1}^n \left(\frac{1}{2} + \frac{j}{2n}\right)^{-s_j}. \tag{B.18}$$

Then for all $\mathbf{x}^d \in (0,1)^d$,

$$\hat{\nu}_n(\tilde{\mathbf{b}}_d^n(\mathbf{x}^d)) \leq \Omega_d(\mathbf{x}^d)^{-1} \leq \check{\nu}_n(\tilde{\mathbf{b}}_d^n(\mathbf{x}^d)) \leq 2^{b_d^l(\mathbf{x}^d)}. \tag{B.19}$$

LEMMA B.5. *For any $m \in \mathbb{N}$, any sequence of $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d \in \tilde{F}_d$ and any sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \leq k_d \leq [d^\delta]$,*

$$\mathbb{E}\big[\big(2^{b_d^l(\hat{\mathbf{W}}_{k_d}^d)}\big)^m|\hat{\mathbf{W}}_0^d = \mathbf{x}^d\big] \to \exp\big((2^m - 1)\lambda(l)\big) \qquad \text{as } d \to \infty. \tag{B.20}$$

PROOF. Note that $\{b_d^l(\hat{\mathbf{W}}_{k_d}^d)|\hat{\mathbf{W}}_0^d = \mathbf{x}^d\} = \sum_{j=1}^d \chi_j(x_j; l; k_d)$. Then since the $\{\chi_j(x_j; l; k_d)\}$ are independent Bernoulli random variables,

$$\begin{aligned}\mathbb{E}\big[\big(2^{b_d^l(\hat{\mathbf{W}}_{k_d}^d)}\big)^m|\hat{\mathbf{W}}_0^d = \mathbf{x}^d\big] &= \prod_{j=1}^d \mathbb{E}\big[(2^m)^{\chi_j(x_j;l;k_d)}|\hat{\mathbf{W}}_0^d = \mathbf{x}^d\big] \\ &= \prod_{j=1}^d \big\{\big(1 - q^d(x_j; l; k_d)\big) + 2^m q^d(x_j; l; k_d)\big\}.\end{aligned} \tag{B.21}$$

By Lemma A.8, (A.33), for $\mathbf{x}^d \in \tilde{F}_d$, $\sum_{j=1}^d q^d(x_j; l; k_d)^2 \to 0$ as $d \to \infty$, so the right-hand side of (B.21) has the same limit as $d \to \infty$ as

$$\prod_{j=1}^d \exp\big((2^m - 1)q^d(x_j; l; k_d)\big) = \exp\big((2^m - 1)\lambda_d(\mathbf{x}^d; l; k_d)\big). \tag{B.22}$$

The lemma follows since for any $\mathbf{x}^d \in \tilde{F}_d$, $\lambda_d(\mathbf{x}^d; l; k_d) \to \lambda(l)$ as $d \to \infty$. □

LEMMA B.6. *Fix $m, n \in \mathbb{N}$. For any sequence $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d \in F_d$, and any sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \le k_d \le [d^\delta]$, we have that*

$$\mathbb{E}[\check{\nu}_n(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d; k_d))^m] \to \mathbb{E}[\check{\nu}_n(\mathbf{S}_n)^m] \qquad \text{as } d \to \infty,$$
$$\mathbb{E}[\hat{\nu}_n(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d; k_d))^m] \to \mathbb{E}[\hat{\nu}_n^m(\mathbf{S}_n)^m] \qquad \text{as } d \to \infty.$$

PROOF. By [6], Theorem 29.2, and Lemma B.1

$$\check{\nu}_n(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d; k_d))^m \xrightarrow{D} \check{\nu}_n(\mathbf{S}_n)^m \qquad \text{as } d \to \infty, \tag{B.23}$$
$$\hat{\nu}_n(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d; k_d))^m \xrightarrow{D} \hat{\nu}_n(\mathbf{S}_n)^m \qquad \text{as } d \to \infty. \tag{B.24}$$

The lemma follows since (B.19) and Lemma B.5 ensure the uniform integrability of the left-hand sides of (B.23) and (B.24). □

LEMMA B.7. *For any sequence $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d \in F_d$ and sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \le k_d \le [d^\delta]$,*

$$\mathbb{E}[\Omega_d(\hat{\mathbf{W}}_{k_d}^d)^{-1}|\hat{\mathbf{W}}_0^d = \mathbf{x}^d] \to \exp(f^* l/2) \qquad \text{as } d \to \infty. \tag{B.25}$$

For any $\mathbf{x}^d \in \tilde{F}_d$ and sequences of positive integers $\{i_d\}$ and $\{k_d\}$ satisfying $[d^\beta] \le k_d \le [d^\delta]$ and $i_d + k_d \le [d^\delta]$,

$$\mathbb{E}[\Omega_d(\hat{\mathbf{W}}_{i_d+k_d}^d)^{-1}|\hat{\mathbf{W}}_{i_d}^d, \hat{\mathbf{W}}_0^d = \mathbf{x}^d] \xrightarrow{p} \exp(f^* l/2) \qquad \text{as } d \to \infty. \tag{B.26}$$

PROOF. An immediate consequence of Lemma B.6 is that

$$\lim_{d\to\infty} \mathbb{E}[\check{\nu}_n(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d;k_d))], \lim_{d\to\infty} \mathbb{E}[\hat{\nu}_n(\tilde{\mathbf{S}}_n^d(\mathbf{x}^d;k_d))] \to \exp(f^*l/2) \qquad \text{as } n\to\infty,$$

from which (B.25) follows by (B.19).

By Theorem A.13, (A.75), $\mathbb{P}(\hat{\mathbf{W}}_{i_d}^d \in F_d|\hat{\mathbf{W}}_0^d \in \tilde{F}_d) \to 1$ as $d\to\infty$, so (B.26) follows from (B.25). □

LEMMA B.8. *For any sequence* $\{\mathbf{x}^d\}$ *such that* $\mathbf{x}^d \in \tilde{F}_d$ *and any sequences of positive integers* $\{i_d\}$ *and* $\{k_d\}$ *satisfying* $[d^\beta] \le i_d, k_d \le [d^\delta]$,

$$\operatorname{cov}\bigl(\Omega_d(\hat{\mathbf{W}}_{i_d}^d)^{-1}, \Omega_d(\hat{\mathbf{W}}_{i_d+k_d}^d)^{-1}|\hat{\mathbf{W}}_0^d=\mathbf{x}^d\bigr)\to 0 \qquad \text{as } d\to\infty \tag{B.27}$$

and

$$\begin{aligned}&\operatorname{var}\bigl(\Omega_d(\hat{\mathbf{W}}_{k_d}^d)^{-1}|\hat{\mathbf{W}}_0^d=\mathbf{x}^d\bigr)\\ &\quad\to \exp\bigl(f^*l\{4\log 2-\tfrac{3}{2}\}\bigr)-\exp(f^*l) \qquad \text{as } d\to\infty.\end{aligned} \tag{B.28}$$

PROOF. Using (B.19), Lemma B.5 and Markov's inequality, it is straightforward to show that for any $\delta>0$, there exists $K<\infty$ such that

$$\mathbb{P}\bigl(\Omega_d(\hat{\mathbf{W}}_{j_d}^d)^{-1} > K|\hat{\mathbf{W}}_0^d=\mathbf{x}^d\bigr) \le \mathbb{P}\bigl(2^{b_d^l(\hat{\mathbf{W}}_{j_d}^d)} > K|\hat{\mathbf{W}}_0^d=\mathbf{x}^d\bigr)\le\delta. \tag{B.29}$$

Therefore it follows from Lemma B.7 that, for any sequence $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d \in \tilde{F}_d$,

$$\begin{aligned}&\Omega_d(\hat{\mathbf{W}}_{j_d}^d)^{-1}\{\mathbb{E}[\Omega_d(\hat{\mathbf{W}}_{j_d+k_d}^d)^{-1}|\hat{\mathbf{W}}_{j_d}^d, \hat{\mathbf{W}}_0^d=\mathbf{x}^d]\\ &\qquad\qquad - \mathbb{E}[\Omega_d(\hat{\mathbf{W}}_{j_d+k_d}^d)^{-1}|\hat{\mathbf{W}}_0^d=\mathbf{x}^d]\}|\hat{\mathbf{W}}_0^d=\mathbf{x}^d\\ &\quad\xrightarrow{p} 0 \qquad \text{as } d\to\infty.\end{aligned} \tag{B.30}$$

The uniform integrability of the left-hand side of (B.30) follows from (B.19) and Lemma B.5. Hence (B.27) follows.

It is straightforward to show that $\mathbb{E}[\check{\nu}_n(\mathbf{S}_n)^2], \mathbb{E}[\hat{\nu}_n(\mathbf{S}_n)^2]\to\exp(f^*l(4\log 2 - 3/2))$ as $n\to\infty$. Therefore from (B.19) and Lemma B.5, we have that

$$\mathbb{E}[\Omega_d(\hat{\mathbf{W}}_{k_d}^d)^{-2}|\hat{\mathbf{W}}_0^d=\mathbf{x}^d]\to\exp\bigl(f^*l(4\log 2-3/2)\bigr) \qquad \text{as } d\to\infty. \tag{B.31}$$

Then (B.28) follows immediately. □

We are now in position to prove Lemma B.9, which is the final step in proving that for any sequence $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d\in\tilde{F}_d$, $P_d|\mathbf{X}_0^d=\mathbf{x}^d \xrightarrow{p} \exp(-f^*l/2)$ as $d\to\infty$.

LEMMA B.9. *For any* $0<\pi\le 1$ *and any sequence* $\{\mathbf{x}^d\}$ *such that* $\mathbf{x}^d\in\tilde{F}_d$,

$$\tilde{T}_d(\pi)|\hat{\mathbf{W}}_0^d=\mathbf{x}^d \xrightarrow{p} \pi\exp(f^*l/2) \qquad \text{as } d\to\infty. \tag{B.32}$$

PROOF. Fix a sequence $\{\mathbf{x}^d\}$. Let $\tilde{T}_d^1(\pi) = \frac{1}{[d^\delta]} \sum_{j=0}^{[d^\beta-1]} \Omega_d(\hat{\mathbf{W}}_j^d)^{-1}$ and let $\tilde{T}_d^2(\pi) = \frac{1}{[d^\delta]} \sum_{j=[d^\beta]}^{[\pi d^\delta-1]} \Omega_d(\hat{\mathbf{W}}_j^d)^{-1}$. Thus $\tilde{T}_d(\pi) = \tilde{T}_d^1(\pi) + \tilde{T}_d^2(\pi)$.

Let $A_d = \sum_{j=0}^{[d^\delta]} \{\hat{\mathbf{W}}_j^d \notin F_d^1\}$. By Theorem A.13, (A.75), $\mathbb{P}(A_d | \hat{\mathbf{W}}_0^d = \mathbf{x}^d) \to 0$ as $d \to \infty$ and conditional upon A_d^C, $\tilde{T}_d^1(\pi) \leq \frac{[d^\beta] d^\gamma}{[d^\delta]}$. Hence $\tilde{T}_d^1(\pi) | \hat{\mathbf{W}}_0^d = \mathbf{x}^d \xrightarrow{p} 0$ as $d \to \infty$.

By Lemma B.7, (B.25),

$$\mathbb{E}[\tilde{T}_d^2(\pi) | \hat{\mathbf{W}}_0^d = \mathbf{x}^d] = \frac{1}{[d^\delta]} \sum_{j=[d^\beta]}^{[\pi d^\delta-1]} \mathbb{E}[\Omega_d(\hat{\mathbf{W}}_j^d)^{-1} | \hat{\mathbf{W}}_0^d = \mathbf{x}^d] \tag{B.33}$$
$$\to \pi \exp(f^* l/2).$$

By Chebyshev's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}(|\tilde{T}_d^2(\pi) - \mathbb{E}[\tilde{T}_d^2(\pi) | \hat{\mathbf{W}}_0^d = \mathbf{x}^d]| > \varepsilon | \hat{\mathbf{W}}_0^d = \mathbf{x}^d) \tag{B.34}$$
$$\leq \frac{1}{\varepsilon^2 [d^\delta]^2} \sum_{j=[d^\beta]}^{[\pi d^\delta-1]} \sum_{l=[d^\beta]}^{[\pi d^\delta-1]} \operatorname{cov}(\Omega_d(\hat{\mathbf{W}}_j^d)^{-1}, \Omega_d(\hat{\mathbf{W}}_l^d)^{-1} | \hat{\mathbf{W}}_0^d = \mathbf{x}^d).$$

Since for all j, l,

$$\operatorname{cov}(\Omega_d(\hat{\mathbf{W}}_j^d)^{-1}, \Omega_d(\hat{\mathbf{W}}_l^d)^{-1} | \hat{\mathbf{W}}_0^d = \mathbf{x}^d) \tag{B.35}$$
$$\leq \operatorname{var}(\Omega_d(\hat{\mathbf{W}}_j^d)^{-1} | \hat{\mathbf{W}}_0^d = \mathbf{x}^d)^{1/2} \operatorname{var}(\Omega_d(\hat{\mathbf{W}}_l^d)^{-1} | \hat{\mathbf{W}}_0^d = \mathbf{x}^d)^{1/2},$$

it is straightforward to show, using Lemma B.8, that the right-hand side of (B.34) converges to 0 as $d \to \infty$. Thus $\tilde{T}_d^2(\pi) | \hat{\mathbf{W}}_0^d = \mathbf{x}^d \xrightarrow{p} \pi \exp(f^* l/2)$ as $d \to \infty$ and the lemma follows immediately. $\square$

THEOREM B.10. *For any sequence $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d \in \tilde{F}_d$,*

$$P_d | \mathbf{X}_0^d = \mathbf{x}^d \xrightarrow{p} \exp(-f^* l/2) \qquad \text{as } d \to \infty. \tag{B.36}$$

PROOF. For any $0 < \pi \leq 1$, by Lemmas B.3, B.4 and B.9,

$$T_d(\pi) | \hat{\mathbf{X}}_0^d = \mathbf{x}^d \xrightarrow{p} \pi \exp(f^* l/2). \tag{B.37}$$

Since P_d satisfies $T_d(P_d) \leq 1 < T_d(P_d + 1/[d^\delta])$, for any $\varepsilon > 0$,

$$\mathbb{P}(|P_d - \exp(-f^* l/2)| > \varepsilon | \mathbf{X}_0^d = \mathbf{x}^d)$$
$$\leq \mathbb{P}(T_d(\exp(-f^* l/2) - \varepsilon/2) > 1 | \hat{\mathbf{X}}_0^d = \mathbf{x}^d) \tag{B.38}$$
$$+ \mathbb{P}(T_d(\exp(-f^* l/2) + \varepsilon/2) \leq 1 | \hat{\mathbf{X}}_0^d = \mathbf{x}^d)$$

for all sufficiently large d. The lemma follows, since (B.37) ensures that the right-hand side of (B.38) converges to 0 as $d \to \infty$. $\square$

APPENDIX C: PROOF OF (2.13)

From Appendix B, we have that for any sequence $\{\mathbf{x}^d\}$, such that $\mathbf{x}^d \in \tilde{F}_d$, $P_d|\mathbf{X}_0^d = \mathbf{x}^d \xrightarrow{p} \exp(-lf^*/2)$ as $d \to \infty$. Therefore we proceed by showing that, for any $0 \le \pi \le 1$,

$$\sup_{\mathbf{x}^d \in F_d} |\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) - \pi \hat{G} H(x_1)| \to 0 \qquad \text{as } d \to \infty, \tag{C.1}$$

where $\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) = \frac{d^2}{[d^\delta]} \mathbb{E}[(H(\hat{\mathbf{X}}^d_{[\pi d^\delta]}) - H(\hat{\mathbf{X}}_0^d))|\hat{\mathbf{X}}_0^d = \mathbf{x}^d]$ is defined in (2.21) and

$$\hat{G} H(x) = \frac{l^2}{3}\left\{\frac{1}{2} g'(x) H'(x) + \frac{1}{2} H''(x)\right\}. \tag{C.2}$$

Equation (2.13) will then be proved using the triangle inequality.

We analyze $\hat{G}_d H(\hat{\mathbf{X}}_j^d) = d^2 \mathbb{E}[H(\hat{\mathbf{X}}_1^d - \hat{\mathbf{X}}_0^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d]$, which is defined in (2.22), before using (2.21) to study $\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d)$. However, first we require some definitions and preliminary results. Throughout we will utilize the following key facts noted in Section 2: $H'(0) = H'(1) = 0$ and that $H_1^*, H_2^* < \infty$, where $H_1^* = \sup_{0 \le y \le 1} |H'(y)|$ and $H_2^* = \sup_{0 \le y \le 1} |H''(y)|$.

We follow [7] and [10] in noting that, for any function h which is a twice differentiable function on $\mathbb{R}$, the function $z \mapsto 1 \wedge e^{h(z)}$ is also twice differentiable, except at a countable number of points, with first derivative given Lebesgue almost everywhere by the function

$$\frac{d}{dz} 1 \wedge e^{h(z)} = \begin{cases} h'(z) e^{h(z)}, & \text{if } h(z) < 0, \\ 0, & \text{if } h(z) \ge 0. \end{cases}$$

The second derivative can similarly be obtained but will not be explicitly required for our calculations.

For $-1 \le z \le 1$, let $J_d^z(\mathbf{x}^d)$ denote the probability of accepting a move in the RWM algorithm given that $Z_{1,1} = z$ and let

$$\begin{aligned} \tilde{J}_d^0(\mathbf{x}^d) = \mathbb{E}\Bigg[&\exp\left(\sum_{j=2}^d \{g(x_j + \sigma_d Z_{1,j}) - g(x_j)\}\right) \\ &\times 1_{\{\sum_{j=2}^d (g(x_j + \sigma_d Z_{1,j}) - g(x_j)) < 0\}} \prod_{j=2}^d 1_{\{0 < x_j + \sigma_d Z_{1,j} < 1\}} \Bigg]. \end{aligned} \tag{C.3}$$

Then for all $\mathbf{x}^d$, using Taylor's theorem,

$$J_d^z(\mathbf{x}^d) = 1_{\{0 < x_1 + \sigma_d z < 1\}} \{J_d^0(\mathbf{x}^d) + \sigma_d g'(x_1) z \tilde{J}_d^0(\mathbf{x}^d) + O(\sigma_d^2)\}. \tag{C.4}$$

Therefore for $x_1 \in (\sigma_d, 1 - \sigma_d)$,

$$J_d(\mathbf{x}^d) = J_d^0(\mathbf{x}^d) + O(\sigma_d^2). \tag{C.5}$$

LEMMA C.1.

$$\sup_{\mathbf{x}^d \in F_d} \left| \frac{\tilde{J}_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d)} - \frac{1}{2} \right| \to 0 \qquad as\ d \to \infty. \tag{C.6}$$

PROOF. Let $\tilde{\Omega}_d^0(\mathbf{x}^d) = \mathbb{E}[\prod_{j=2}^d 1_{\{0<x_j+\sigma_d Z_{1,j}<1\}} 1_{\{\sum_{j=2}^d (g(x_j+\sigma_d Z_{1,j})-g(x_j))<0\}}]$ and let $\Omega_d^0(\mathbf{x}^d) = \mathbb{E}[\prod_{j=2}^d 1_{\{0<x_j+\sigma_d Z_{1,j}<1\}}]$, the probability a proposed move stays inside the unit cube given that the first component does not move. The proof of (A.26) can be adapted to show that, for any $\alpha < \frac{1}{2}$, $d^\alpha |\Omega_d^0(\mathbf{x}^d) - J_d^0(\mathbf{x}^d)|, d^\alpha |\tilde{\Omega}_d^0(\mathbf{x}^d) - \tilde{J}_d^0(\mathbf{x}^d)| \to 0$ as $d \to \infty$. Therefore since for $\mathbf{x}^d \in F_d$, $J_d^0(\mathbf{x}^d), \Omega_d^0(\mathbf{x}^d) \geq \exp(-lg^*)d^{-\gamma}$, (A.3), we have that

$$\sup_{\mathbf{x}^d \in F_d} \left| \frac{\tilde{J}_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d)} - \frac{\tilde{\Omega}_d^0(\mathbf{x}^d)}{\Omega_d^0(\mathbf{x}^d)} \right| \to 0 \qquad \text{as } d \to \infty. \tag{C.7}$$

Let $\mathcal{B}_d(\mathbf{x}^d) = \{2 \leq j \leq d; x_j \in R_d^l\}$ and let $I_d(\mathbf{x}^d) = \sum_{j \notin \mathcal{B}_d(\mathbf{x}^d)} \sigma_d g'(x_j) Z_{1,j}$. Since $|\mathcal{B}_d(\mathbf{x}^d)| \leq \gamma \log d$, we have that

$$\left| \sum_{j \in \mathcal{B}_d(\mathbf{x}^d)} \big(g(x_j + \sigma_d Z_{1,j}) - g(x_j)\big) \right| \leq (\gamma \log d) \sigma_d g^*. \tag{C.8}$$

Then using a Taylor series expansion, there exists $K < \infty$ such that, for all $\mathbf{x}^d \in F_d$,

$$I_d(\mathbf{x}^d) - \frac{K \log d}{d} \leq \sum_{j=2}^d \big(g(x_j + \sigma_d Z_j) - g(x_j)\big) \leq I_d(\mathbf{x}^d) + \frac{K \log d}{d}. \tag{C.9}$$

Since $Z_{1,1}, Z_{1,2}, \ldots,$ are independent, and whether or not a proposed move from $\mathbf{x}^d$ stays inside the hypercube depends only upon $\mathcal{B}^d(\mathbf{x}^d)$,

$$\begin{aligned} &\Omega_d^0(\mathbf{x}^d) \mathbb{P}\big(I_d(\mathbf{x}^d) < -K \log d/d\big) \\ &\quad \leq \tilde{\Omega}_d^0(\mathbf{x}^d) \leq \Omega_d^0(\mathbf{x}^d) \mathbb{P}\big(I_d(\mathbf{x}^d) < K \log d/d\big). \end{aligned} \tag{C.10}$$

For all $\mathbf{x}^d \in F_d$, $\frac{1}{d} \sum_{j=1}^d g'(x_j)^2 \to \mathbb{E}[g'(X_1)^2]$, so

$$\sqrt{d} I_d(\mathbf{x}^d) \xrightarrow{D} N(0, \mathbb{E}[g'(X_1)^2]) \qquad \text{as } d \to \infty.$$

Therefore it follows that

$$\sup_{\mathbf{x}^d \in F_d} \left| \frac{\tilde{\Omega}_d^0(\mathbf{x}^d)}{\Omega_d^0(\mathbf{x}^d)} - \frac{1}{2} \right| \to 0 \qquad \text{as } d \to \infty \tag{C.11}$$

with the lemma following from (C.7) and (C.11) by the triangle inequality. □

LEMMA C.2. *For* $x_1 \in (\sigma_d, 1-\sigma_d)$ *and* $\mathbf{x}^d \in F_d$,

$$\hat{G}_d H(\mathbf{x}^d) = \frac{l^2}{3}\left\{\frac{1}{2}H''(x_1) + \frac{\tilde{J}_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d)} g'(x_1)H'(x_1)\right\} + \varepsilon_d, \tag{C.12}$$

where $\varepsilon_d \to 0$ *as* $d \to \infty$.

For $x_1 \in R_d^l$,

$$|\hat{G}_d H(\mathbf{x}^d)| \le \tfrac{3}{2} H_2^* l^2. \tag{C.13}$$

PROOF. For $d \ge 1$, fix $\mathbf{x}^d \in F_d$ and suppose that $x_1 \in (\sigma_d, 1-\sigma_d)$. Then

$$\begin{aligned} \hat{G}_d H(\mathbf{x}^d) &= d^2 \mathbb{E}[H(\hat{\mathbf{X}}_1^d) - H(\hat{\mathbf{X}}_0^d) | \hat{\mathbf{X}}_0^d = \mathbf{x}^d] \\ &= \frac{d^2}{J_d(\mathbf{x}^d)} \mathbb{E}\left[\left(H(\mathbf{x}^d + \sigma_d \mathbf{Z}^d) - H(\mathbf{x}^d)\right)\left\{1 \wedge \frac{\pi_d(\mathbf{x}^d + \sigma_d \mathbf{Z}^d)}{\pi_d(\mathbf{x}^d)}\right\}\right]. \end{aligned} \tag{C.14}$$

The right-hand side of (C.14) is familiar in that it is the generator of the RWM-algorithm divided by the acceptance probability; see, for example, [14], page 113.

First, note that

$$\begin{aligned} H(x_1 + \sigma_d Z_1) - H(x_1) = \sigma_d Z_1 H'(x_1) &+ \frac{\sigma_d^2}{2} Z_1^2 H''(x_1) \\ &+ \frac{\sigma_d^2}{2} Z_1^2 \{H''(x_1 + \psi_1^d) - H''(x_1)\}. \end{aligned}$$

Using (C.4), (C.5) and noting that $0 < x_1 + \sigma_d Z_1 < 1$, we have that

$$\begin{aligned} \hat{G}_d H(\mathbf{x}^d) = {} & \frac{d^2}{J_d^0(\mathbf{x}^d) + O(\sigma_d^2)} \\ & \times \mathbb{E}\Bigg[\Bigg\{\sigma_d Z_1 H'(x_1) + \frac{\sigma_d^2}{2} Z_1^2 H''(x_1) \\ & \qquad + \frac{\sigma_d^2}{2} Z_1^2 \{H''(x_1 + \psi_1^d) - H''(x_1)\}\Bigg\} \\ & \qquad \times \{J_d^0(\mathbf{x}^d) + \tilde{J}_d^0(\mathbf{x}^d)\sigma_d g'(x_1) Z_1 + O(\sigma_d^2)\} 1_{\{0 < x_1 + \sigma_d Z_1 < 1\}}\Bigg] \\ = {} & \frac{d^2 J_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d) + O(\sigma_d^2)} \sigma_d \mathbb{E}[Z_1] H'(x_1) \\ & + \frac{d^2 J_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d) + O(\sigma_d^2)} \frac{\sigma_d^2}{2} \mathbb{E}[Z_1^2] H''(x_1) \\ & + \frac{d^2 J_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d) + O(\sigma_d^2)} \frac{\sigma_d^2}{2} \mathbb{E}[Z_1^2\{H''(x_1 + \psi_1^d) - H''(x_1)\}] \end{aligned} \tag{C.15}$$

$$+ \frac{d^2 \tilde{J}_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d) + O(\sigma_d^2)} \sigma_d^2 g'(x_1) H'(x_1) \mathbb{E}[Z_1^2]$$
$$+ \frac{d^2}{J_d^0(\mathbf{x}^d) + O(\sigma_d^2)} O(\sigma_d^3).$$

The first term on the right-hand side of (C.15) is 0. Since $H_2^* < \infty$, by the continuous mapping theorem, $\{H''(x_1 + \psi_1^d) - H''(x_1)\} \xrightarrow{p} 0$ as $d \to \infty$ and then since Z_1 is bounded the third term on the right-hand side of (C.15) converges to 0 as $d \to \infty$. For $\mathbf{x}^d \in F_d$, $J_d^0(\mathbf{x}^d) \geq e^{-lg^*} d^{-\gamma}$, and so, the right-hand side of (C.15) equals

$$\frac{l^2}{3} \left\{ \frac{1}{2} H''(x_1) + \frac{\tilde{J}_d^0(\mathbf{x}^d)}{J_d^0(\mathbf{x}^d)} g'(x_1) H'(x_1) \right\} + \varepsilon_d,$$

where $\varepsilon_d \to 0$ as $d \to \infty$. Thus (C.12) is proved.

The proof of (C.13) follows straightforwardly using Taylor series expansions since $H'(0) = H'(1) = 0$. □

Since $g^* = \sup_{0 \leq y \leq 1} |g'(y)|$, $H_1^*, H_2^* < \infty$, an immediate consequence of Lemma C.2 is that, there exists $K^* < \infty$ such that

$$\sup_d \sup_{\mathbf{x}^d \in F_d} |\hat{G}_d H(\mathbf{x}^d)| \leq K^*. \tag{C.16}$$

LEMMA C.3. *For any sequence of positive integers $\{k_d\}$ satisfying $[d^\beta] \leq k_d \leq [d^\delta]$,*

$$\sup_{\mathbf{x}^d \in \tilde{F}_d} \left| \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_{k_d}^d) | \hat{\mathbf{X}}_0^d = \mathbf{x}^d] - \hat{G} H(x_1) \right| \to 0 \qquad \text{as } d \to \infty. \tag{C.17}$$

PROOF. Fix $\{k_d\}$ and note that

$$\begin{aligned} &\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_{k_d}^d) | \hat{\mathbf{X}}_0^d = \mathbf{x}^d] \\ &\quad = \mathbb{P}(\hat{\mathbf{X}}_{k_d}^d \in F_d | \hat{\mathbf{X}}_0^d = \mathbf{x}^d) \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_{k_d}^d) | \hat{\mathbf{X}}_0^d = \mathbf{x}^d, \hat{\mathbf{X}}_{k_d}^d \in F_d] \\ &\qquad + \mathbb{P}(\hat{\mathbf{X}}_{k_d}^d \notin F_d | \hat{\mathbf{X}}_0^d = \mathbf{x}^d) \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_{k_d}^d) | \hat{\mathbf{X}}_0^d = \mathbf{x}^d, \hat{\mathbf{X}}_{k_d}^d \notin F_d]. \end{aligned} \tag{C.18}$$

Since $H \in \mathcal{D}$, $H_0^* = \sup_{0 \leq y \leq 1} |H(y)| < \infty$. Therefore, for all $\mathbf{y}^d \in [0,1]^d$, $\hat{G}_d H(\mathbf{y}^d) \leq 2d^2 H_0^*$. By (2.14), $\sup_{\mathbf{x}^d \in \tilde{F}_d} d^2 \mathbb{P}(\hat{\mathbf{X}}_{k_d}^d \notin F_d | \hat{\mathbf{X}}_0^d = \mathbf{x}^d) \to 0$ as $d \to \infty$. Thus the latter term on the right-hand side of (C.18) converges to 0 as $d \to \infty$.

Now

$$\begin{aligned}
&\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d})|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d] \\
&\quad = \mathbb{P}(\hat{X}^d_{k_d,1} \notin R^l_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d) \\
&\qquad \times \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d})|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \notin R^l_d] \\
&\qquad + \mathbb{P}(\hat{X}^d_{k_d,1} \in R^l_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d) \\
&\qquad\quad \times \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d})|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \in R^l_d].
\end{aligned} \tag{C.19}$$

Consider first the latter term on the right-hand side of (C.19). By Lemma C.2, (C.13),

$$\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d})|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \in R^l_d] \le \tfrac{3}{2} l^2 H^*_2. \tag{C.20}$$

Note that

$$\begin{aligned}
\mathbb{P}(\hat{X}^d_{k_d,1} \in R^l_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d) &= \frac{\mathbb{P}(\hat{X}^d_{k_d,1} \in R^l_d, \hat{\mathbf{X}}^d_{k_d} \in F_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d)}{\mathbb{P}(\hat{\mathbf{X}}^d_{k_d} \in F_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d)} \\
&\le \frac{\mathbb{P}(\hat{X}^d_{k_d,1} \in R^l_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d)}{\mathbb{P}(\hat{\mathbf{X}}^d_{k_d} \in F_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d)}.
\end{aligned} \tag{C.21}$$

By (2.14), for $\mathbf{x}^d \in \tilde{F}_d$, $\mathbb{P}(\hat{\mathbf{X}}^d_{k_d} \in F_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d) \to 1$ as $d \to \infty$. Use Corollary A.7 and Lemma A.8 to show that $\mathbb{P}(\hat{X}^d_{k_d,1} \in R^l_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d) \to 0$ as $d \to \infty$. Hence, the right-hand side of (C.21) converges to 0 as $d \to \infty$ and consequently the latter term on the right-hand side of (C.19) converges to 0 as $d \to \infty$.

It follows from the above arguments that

$$\min_{\mathbf{x}^d \in \tilde{F}_d} \mathbb{P}(\hat{X}^d_{k_d,1} \notin R^l_d, \hat{\mathbf{X}}^d_{k_d} \in F_d|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d) \to 1 \qquad \text{as } d \to \infty. \tag{C.22}$$

Also it follows from (C.16) that there exists $K < \infty$ such that

$$\sup_d \sup_{\mathbf{x}^d \in \tilde{F}_d} \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d})|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \notin R^l_d] \le K. \tag{C.23}$$

Therefore, it is straightforward using (C.18), (C.19) and the triangle inequality to show that

$$\begin{aligned}
&\sup_{\mathbf{x}^d \in \tilde{F}_d} \big|\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d})|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d] \\
&\qquad - \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d})|\hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \notin R^l_d]\big| \\
&\qquad \to 0 \qquad \text{as } d \to \infty.
\end{aligned} \tag{C.24}$$

By Lemma C.2, (C.12), there exists $\varepsilon_d^1 \to 0$ as $d \to \infty$, such that

$$
\begin{aligned}
&\sup_{\mathbf{x}^d \in \tilde{F}_d} |\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_{k_d}) - \hat{G}H(\hat{X}^d_{k_d,1}) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \notin R^l_d]| \\
&\qquad \le \frac{l^2}{3} \sup_{0 \le y \le 1} |g'(y)H'(y)| \\
&\qquad\qquad \times \sup_{\mathbf{x}^d \in \tilde{F}_d} \mathbb{E}\biggl[\biggl|\frac{\tilde{J}^0_d(\hat{\mathbf{X}}^d_{k_d})}{J^0_d(\hat{\mathbf{X}}^d_{k_d})} - \frac{1}{2}\biggr| \bigg| \hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \notin R^l_d\biggr] + \varepsilon^1_d \\
&\qquad \le \frac{l^2}{3} g^* H^*_1 \sup_{\mathbf{y}^d \in F_d} \biggl|\frac{\tilde{J}^0_d(\mathbf{y}^d)}{J^0_d(\mathbf{y}^d)} - \frac{1}{2}\biggr| + \varepsilon^1_d.
\end{aligned}
\tag{C.25}
$$

By Lemma C.1, the right-hand side of (C.25) converges to 0 as $d \to \infty$.

Using the triangle inequality, the lemma follows by showing that

$$
\begin{aligned}
&\sup_{\mathbf{x}^d \in \tilde{F}_d} |\mathbb{E}[\hat{G}H(\hat{X}^d_{k_d,1}) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d, \hat{\mathbf{X}}^d_{k_d} \in F_d, \hat{X}^d_{k_d,1} \notin R^l_d] - \hat{G}H(x_1)| \\
&\qquad \to 0 \qquad \text{as } d \to \infty.
\end{aligned}
\tag{C.26}
$$

Note that $|\hat{X}^d_{k_d,1} - x_1| \le k_d \sigma_d$, and so, (C.26) follows since $\hat{G}H(\cdot)$ is continuous. □

We are in position to prove (C.1).

LEMMA C.4. *For any* $0 \le \pi \le 1$,

$$
\sup_{\mathbf{x}^d \in \tilde{F}_d} |\hat{G}^{\delta,\pi}_d(\mathbf{x}^d) - \pi \hat{G}H(x_1)| \to 0 \qquad \textit{as } d \to \infty.
\tag{C.27}
$$

PROOF. Since (C.27) trivially holds for $\pi = 0$, we assume that $\pi > 0$. For all sufficiently large d, by the triangle inequality,

$$
\begin{aligned}
&|\hat{G}^{\delta,\pi}_d(\mathbf{x}^d) - \pi \hat{G}H(x_1)| \\
&\qquad = \Biggl|\frac{1}{[d^\delta]} \sum_{j=0}^{[\pi d^\delta - 1]} \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_j) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d] - \pi \hat{G}H(x_1)\Biggr| \\
&\qquad \le \Biggl|\frac{1}{[d^\delta]} \sum_{j=0}^{[d^\beta]-1} \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_j) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d]\Biggr| \\
&\qquad\qquad + \frac{1}{[d^\delta]} \sum_{j=[d^\beta]}^{[\pi d^\delta - 1]} |\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}^d_j) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d] - \hat{G}H(x_1)| \\
&\qquad\qquad + \biggl(\pi - \frac{[\pi d^\delta] - [d^\beta]}{[d^\delta]}\biggr) \hat{G}H(x_1).
\end{aligned}
\tag{C.28}
$$

Since

$$\begin{aligned}
&\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_j^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d] \\
&\quad = \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_j^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d, \hat{\mathbf{X}}_j^d \in F_d]\mathbb{P}(\hat{\mathbf{X}}_j^d \in F_d|\hat{\mathbf{X}}_0^d = \mathbf{x}^d) \\
&\qquad + \mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_j^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d, \hat{\mathbf{X}}_j^d \notin F_d]\mathbb{P}(\hat{\mathbf{X}}_j^d \notin F_d|\hat{\mathbf{X}}_0^d = \mathbf{x}^d),
\end{aligned} \tag{C.29}$$

it is straightforward, following a similar argument to the proof of Lemma C.3, (C.23), to show that there exists $\tilde{K} < \infty$ such that, for all $0 \leq j \leq [d^\delta]$,

$$\sup_{\mathbf{x}^d \in \tilde{F}_d} \left|\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_j^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d]\right| \leq \tilde{K}. \tag{C.30}$$

Therefore the first term on the right-hand side of (C.29) is bounded by $[d^\beta]\tilde{K}/[d^\delta]$. By Lemma C.3 the supremum over $\mathbf{x}^d \in \tilde{F}_d$ of the second term on the right-hand side of (C.28) converges to 0 as $d \to \infty$ and the lemma follows. □

COROLLARY C.5.

$$\sup_{0 \leq \pi \leq 1} \sup_{\mathbf{x}^d \in \tilde{F}_d} |\hat{G}_d^{\delta,\pi}(\mathbf{x}^d) - \pi \hat{G} H(x_1)| \to 0 \qquad \text{as } d \to \infty. \tag{C.31}$$

PROOF. Fix $\varepsilon > 0$ and let $\Pi_\varepsilon = \{0, \varepsilon, 2\varepsilon, \ldots, [1/\varepsilon]\varepsilon, 1\}$. It follows from Lemma C.4 that, for all sufficiently large d,

$$\max_{\pi \in \Pi_\varepsilon} \sup_{\mathbf{x}^d \in \tilde{F}_d} |\hat{G}_d^{\delta,\pi}(\mathbf{x}^d) - \pi \hat{G} H(x_1)| \leq \varepsilon. \tag{C.32}$$

Consider any $0 \leq \pi \leq 1$. There exists $\tilde{\pi} \in \Pi_\varepsilon$ such that $\tilde{\pi} \leq \pi < \tilde{\pi} + \varepsilon$. By the triangle inequality,

$$\begin{aligned}
&|\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) - \pi \hat{G} H(x_1)| \\
&\quad \leq |\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) - \hat{G}_d^{\delta,\tilde{\pi}} H(\mathbf{x}^d)| + |\hat{G}_d^{\delta,\tilde{\pi}} H(\mathbf{x}^d) - \tilde{\pi} \hat{G} H(x_1)| \\
&\qquad + (\pi - \tilde{\pi})|\hat{G} H(x_1)|.
\end{aligned} \tag{C.33}$$

Again by the triangle inequality,

$$\begin{aligned}
&\sup_{\mathbf{x}^d \in \tilde{F}_d} |\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) - \hat{G}_d^{\delta,\tilde{\pi}} H(\mathbf{x}^d)| \\
&\qquad \leq \frac{1}{[d^\delta]} \sum_{j=[\tilde{\pi} d^\delta]}^{[\pi d^\delta - 1]} \sup_{\mathbf{x}^d \in \tilde{F}_d} \left|\mathbb{E}[\hat{G}_d H(\hat{\mathbf{X}}_j^d)|\hat{\mathbf{X}}_0^d = \mathbf{x}^d]\right|.
\end{aligned} \tag{C.34}$$

Since for all sufficiently large d, $([\pi d^\delta - 1] - [\tilde{\pi} d^\delta])/[d^\delta] \leq 2\varepsilon$, it follows from (C.30) that the right-hand side of (C.34) is bounded by $2\tilde{K}\varepsilon$, where $\tilde{K}$ is defined in Lemma C.4.

Let $\hat{K} = 2\tilde{K} + 1 + \sup_{0 \le y \le 1} |\hat{G}H(y)|$. Note that since $g^*, H_1^*, H_2^* < \infty$, we have that $\hat{K} < \infty$. Therefore it follows from (C.33) that for all sufficiently large d,

$$\sup_{\mathbf{x}^d \in \tilde{F}_d} |\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) - \pi \hat{G}H(x_1)| \le \hat{K}\varepsilon. \tag{C.35}$$

Since (C.35) holds for all $0 \le \pi \le 1$ and $\varepsilon > 0$, the lemma follows. □

Finally we are in position to prove (2.13), and hence complete the proof of Theorem 2.1.

LEMMA C.6.

$$\sup_{\mathbf{x}^d \in \tilde{F}_d} |G_d^{\delta} H(\mathbf{x}^d) - GH(x_1)| \to 0 \qquad \textit{as } d \to \infty. \tag{C.36}$$

PROOF. Note that $G_d^{\delta} H(\mathbf{x}^d)$ is given by (2.8) and $GH(x_1) = \exp(-lf^*/2) \times \hat{G}H(x_1)$. Therefore by the triangle inequality,

$$\begin{aligned}
&\sup_{\mathbf{x}^d \in \tilde{F}_d} |G_d^{\delta} H(\mathbf{x}^d) - GH(x_1)| \\
&\quad = \sup_{\mathbf{x}^d \in \tilde{F}_d} \left| \frac{d^2}{[d^\delta]} \mathbb{E}[H(\hat{\mathbf{X}}^d_{[P_d d^\delta]}) - H(\hat{\mathbf{X}}^d_0) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d] - \exp(-lf^*/2)\hat{G}H(x_1) \right| \\
&\quad \le \sup_{\mathbf{x}^d \in \tilde{F}_d} \left| \mathbb{E}\left[\frac{d^2}{[d^\delta]} \big(H(\hat{\mathbf{X}}^d_{[P_d d^\delta]}) - H(\hat{\mathbf{X}}^d_0)\big) - P_d \hat{G}H(x_1) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d \right] \right| \\
&\qquad + \sup_{\mathbf{x}^d \in \tilde{F}_d} |\mathbb{E}[P_d \hat{G}H(x_1) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d] - \exp(-lf^*/2)\hat{G}H(x_1)| \\
&\quad \le \sup_{0 \le \pi \le 1} \sup_{\mathbf{x}^d \in \tilde{F}_d} \left| \mathbb{E}\left[\frac{d^2}{[d^\delta]} \big(H(\hat{\mathbf{X}}^d_{[\pi d^\delta]}) - H(\hat{\mathbf{X}}^d_0)\big) - \pi \hat{G}H(x_1) | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d \right] \right| \\
&\qquad + \sup_{\mathbf{x}^d \in \tilde{F}_d} |\mathbb{E}[P_d | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d] - \exp(-lf^*/2)| \sup_{0 \le y \le 1} |\hat{G}H(y)| \\
&\quad \le \sup_{0 \le \pi \le 1} \sup_{\mathbf{x}^d \in \tilde{F}_d} |\hat{G}_d^{\delta,\pi} H(\mathbf{x}^d) - \pi \hat{G}H(x_1)| \\
&\qquad + \sup_{\mathbf{x}^d \in \tilde{F}_d} |\mathbb{E}[P_d | \hat{\mathbf{X}}^d_0 = \mathbf{x}^d] - \exp(-lf^*/2)| \sup_{0 \le y \le 1} |\hat{G}H(y)|.
\end{aligned} \tag{C.37}$$

By Corollary C.5, the first term on the right-hand side of (C.37) converges to 0 as $d \to \infty$. By Theorem B.10, for any sequence $\{\mathbf{x}^d\}$ such that $\mathbf{x}^d \in \tilde{F}_d$,

$P_d|\mathbf{X}_0^d = \mathbf{x}^d \xrightarrow{p} \exp(-lf^*/2)$ as $d \to \infty$. Hence the latter term on the right-hand side of (C.37) converges to 0 as $d \to \infty$, since $g^*, H_1^*, H_2^* < \infty$ implies that $\sup_{0 \le y \le 1} |\hat{G}H(y)| < \infty$. □

Acknowledgments. We thank the anonymous referees for their helpful comments which have improved the presentation of the paper.

REFERENCES

[1] BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A** 175–184. MR0974580

[2] BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. Oxford Univ. Press, Oxford. MR1163825

[3] BÉDARD, M. (2007). Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Probab.* **17** 1222–1244. MR2344305

[4] BÉDARD, M. (2008). Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Process. Appl.* **118** 2198–2222. MR2474348

[5] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[6] BILLINGSLEY, P. (1979). *Probability and Measure*. Wiley, New York. MR0534323

[7] BREYER, L. A. and ROBERTS, G. O. (2000). From Metropolis to diffusions: Gibbs states and optimal scaling. *Stochastic Process. Appl.* **90** 181–206. MR1794535

[8] 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

[9] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes, Characterization and Convergence*. Wiley, New York. MR0838085

[10] NEAL, P. and ROBERTS, G. (2006). Optimal scaling for partially updating MCMC algorithms. *Ann. Appl. Probab.* **16** 475–515. MR2244423

[11] NEAL, P. and ROBERTS, G. (2008). Optimal scaling for random walk Metropolis on spherically constrained target densities. *Methodol. Comput. Appl. Probab.* **10** 277–297. MR2399684

[12] NEAL, P. J. and ROBERTS, G. O. (2011). Optimal scaling of random walk Metropolis algorithms with non-Gaussian proposals. *Methodol. Comput. Appl. Probab.* **13** 583–601.

[13] REVUZ, D. and YOR, M. (1998). *Continuous Martingales and Brownian Motion*. Springer, Berlin.

[14] 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

[15] ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450

[16] SHERLOCK, C. (2006). Methodology for inference on the Markov modulated Poisson process and theory for optimal scaling of the random walk Metropolis. Ph.D. thesis, Lancaster Univ.

[17] SHERLOCK, C. and ROBERTS, G. (2009). Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli* **15** 774–798. MR2555199

P. NEAL
SCHOOL OF MATHEMATICS
UNIVERSITY OF MANCHESTER
SACKVILLE STREET
MANCHESTER
M60 1QD
UNITED KINGDOM
E-MAIL: P.Neal-2@manchester.ac.uk

G. ROBERTS
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY
CV4 7AL
UNITED KINGDOM
E-MAIL: Gareth.O.Roberts@warwick.ac.uk

W. K. YUEN
DEPARTMENT OF MATHEMATICS
BROCK UNIVERSITY
ST. CATHARINES
ON L2S 3A1
CANADA
E-MAIL: jyuen@brocku.ca