On geometric ergodicity of additive and multiplicative transformation-based Markov Chain Monte Carlo in high dimensions

Kushal Kr. Dey[a] **and Sourabh Bhattacharya**[b]

[a]*University of Chicago*
[b]*Indian Statistical Institute*

Abstract. Recently Dutta and Bhattacharya (*Statistical Methodology* **16** (2014) 100–116) introduced a novel Markov Chain Monte Carlo methodology that can simultaneously update all the components of high-dimensional parameters using simple deterministic transformations of a one-dimensional random variable drawn from any arbitrary distribution defined on a relevant support. The methodology, which the authors refer to as transformation-based Markov Chain Monte Carlo (TMCMC), greatly enhances computational speed and acceptance rate in high-dimensional problems. Two significant transformations associated with TMCMC are additive and multiplicative transformations. Combinations of additive and multiplicative transformations are also of much interest. In this work, we investigate geometric ergodicity associated with additive and multiplicative TMCMC, along with their combinations, assuming that the target distribution is multi-dimensional and belongs to the super-exponential family; we also illustrate their efficiency in practice with simulation studies.

1 Introduction

It is well known that in high dimensions traditional Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis–Hastings algorithm, face several challenges, with respect to computational complexity, as well as with convergence issues. Indeed, Bayesian computation often requires inversion of high-dimensional matrices in each MCMC iteration, causing enormous computational burden. Moreover, such high-dimensional problems may converge at an extremely slow rate, because of the complicated posterior dependence among the parameters. This implies the requirement of an extremely large number of iterations, but since even individual iterations may be computationally burdensome, traditional MCMC methods do not seem to be ideally suited for Bayesian analysis of complex, high-dimensional problems.

In an effort to combat the problems, Dutta and Bhattacharya (2014) proposed a novel methodology that can update all the parameters simultaneously in a single

---

block using simple deterministic bijective transformations of a one-dimensional random variable (or any other low-dimensional random variables) drawn from some arbitrary distribution. The idea effectively reduces the high-dimensional random parameter to a one-dimensional parameter, thus dramatically improving computational speed and acceptance rate. Details are provided in Dutta and Bhattacharya (2014).

Among the deterministic, bijective transformations, Dutta and Bhattacharya (2014) recommend the additive and the multiplicative transformations. Here it is important to mention that the multiplicative transformation is designed to update parameters on the real line, not just on $(0, \infty)$, and thus, cannot be represented as the log-additive transformation. In Sections 1.1 and 1.2, we provide brief overviews of additive and multiplicative TMCMC, respectively. In Section 1.3, we briefly explain additive–multiplicative TMCMC, which is a combination of additive and multiplicative TMCMC.

This paper deals with geometric ergodicity (or geometric rate of convergence) of the TMCMC chain (both additive and multiplicative, along with their mixtures of two kinds) to the multi-dimensional stationary distribution. The geometric ergodicity property, apart from theoretically ensuring convergence of the underlying Markov chain to the stationary distribution at a geometric rate, also ensures asymptotic stability of a regular family of stochastic estimates through the application of the central limit theorem (see Meyn and Tweedie (1993), Chapter 17, and Jones and Hobert (2001), Section 5.3). The geometric ergodicity of the Random Walk Metropolis–Hastings (RWMH) chain is already well documented (see Mengersen and Tweedie (1996), Roberts and Tweedie (1996), Jarner and Hansen (2000)). Some extensions of these results to chains with polynomial rates of convergence and specific forms of target densities (for instance, heavy tailed families) are also available in the literature (Jarner and Roberts (2002, 2007)). In this paper, we present conditions that guarantee geometric ergodicity of the TMCMC chain corresponding to both additive and multiplicative moves, when the target distribution is multi-dimensional. Crucially, we assume that the target distribution belongs to the super-exponential family. Note that the super-exponential assumption has also been crucially used by Jarner and Hansen (2000) for proving geometric ergodicity of RWMH for multi-dimensional target distributions.

While dealing with multiplicative TMCMC, we encounter a technical problem, which is bypassed by forming an appropriate mixture of additive and multiplicative moves, to which we refer as "essentially fully" multiplicative TMCMC. We also consider a usual mixture of additive and multiplicative moves. We establish geometric ergodicity of both kinds of mixtures and demonstrate with simulation studies that the usual mixture outperforms RWMH, additive TMCMC, as well as "essentially fully" multiplicative TMCMC.

In Section 2, we give conditions for geometric ergodicity of additive TMCMC. The approach to establishing geometric ergodicity of the TMCMC chains associated with multiplicative TMCMC is more complicated and is covered in detail in

Section 3. In Section 4, we illustrate the practical implications of our theoretical results by conducting simulation studies, where we numerically compare convergence issues of the TMCMC approach with that of RWMH, especially in high dimensions. In Section 5, we discuss extension of our approach to situations where the high-dimensional target densities are not in the super-exponential family but can be dealt with using special techniques, in particular, a diffeomorphism based method developed by Johnson and Geyer (2012), and conduct detailed simulation studies in such set-up, demonstrating that TMCMC very significantly outperforms RWMH that set-up. Concluding remarks are provided in Section 6.

1.1 Additive TMCMC

Suppose that we are simulating from a d dimensional space (usually $\mathbb{R}^d$), and suppose we are currently at a point $x = (x_1, \ldots, x_d)$. Let us define d random variables $b_1, \ldots, b_d$, such that, for $i = 1, \ldots, d$,

$$b_i = \begin{cases} +1 & \text{with probability } p_i; \\ -1 & \text{with probability } 1 - p_i. \end{cases} \tag{1}$$

The additive TMCMC uses moves of the following type:

$$(x_1, \ldots, x_d) \rightarrow (x_1 + b_1\epsilon, \ldots, x_d + b_d\epsilon),$$

where $\epsilon \sim g^{(1)} = q^{(1)}(\cdot)I_{\{\epsilon>0\}}$. Here $q^{(1)}(\cdot)$ is an arbitrary density with support $\mathbb{R}_+$, the positive part of the real line, and for any set A, I_A denotes the indicator function of A. We define $T_b^{(1)}(x, \epsilon) = (x_1 + b_1\epsilon, \ldots, x_d + b_d\epsilon)$ to be the additive transformation of x corresponding to the 'move-type' b. In this work, we shall assume that $p_i = 1/2$ for $i = 1, \ldots, d$. Note that the Jacobian of the additive transformations is one.

Thus, a single ϵ is simulated from $q(\cdot)I_{\{\epsilon>0\}}$, which is then either added to, or subtracted from each of the d co-ordinates of x with probability $1/2$. Assuming that the target distribution is proportional to π, the new move $T_b^{(1)}(x, \epsilon)$, corresponding to the move-type b, is accepted with probability

$$\alpha = \min\left\{1, \frac{\pi(T_b^{(1)}(x, \epsilon))}{\pi(x)}\right\}. \tag{2}$$

The path diagram for additive TMCMC that displays the possible regions to which our chain can move to starting from a fixed point, is presented in Figure 1.

In this paper we show, under appropriate and reasonably general assumptions on π, that additive TMCMC with $p_i = 1/2; i = 1, \ldots, d$, is geometrically ergodic for any finite dimension d.
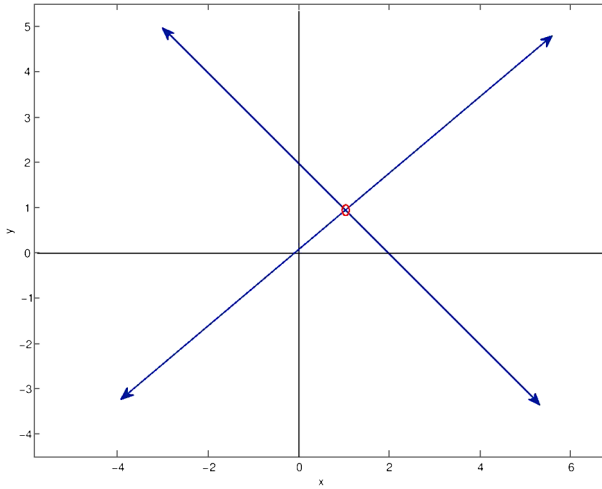
Figure 1 *Path diagram for additive TMCMC in one step from a fixed point denoted by the circle in the middle.*

1.1.1 *Discussion on non-uniform move-type probabilities for additive TMCMC.*
For simplicity of illustration, let us assume that $p_i = p$; $i = 1, \ldots, d$. Also, let $Y = \sum_{i:b_i=1} b_i$. Then $Y \sim \text{Binomial}(d, p)$. The acceptance probability is then given by

$$
\begin{aligned}
\alpha &= \min\left\{1, \left(\frac{p}{1-p}\right)^Y \left(\frac{1-p}{p}\right)^{(d-Y)} \frac{\pi(T_b^{(1)}(x, \epsilon))}{\pi(x)}\right\} \\
&= \min\left\{1, \left(\frac{p}{1-p}\right)^{(2Y-d)} \frac{\pi(T_b^{(1)}(x, \epsilon))}{\pi(x)}\right\}.
\end{aligned}
\tag{3}
$$

Now, as $d \to \infty$, $(2Y - d) \stackrel{\text{a.s.}}{\sim} (2dp - d) = d(2p - 1)$, where, for any two random sequences $\{m_d; d = 1, 2, \ldots\}$ and $\{n_d; d = 1, 2, \ldots\}$, $m_d \stackrel{\text{a.s.}}{\sim} n_d$ indicates $\lim_{d \to \infty} \frac{m_d}{n_d} = 1$, almost surely. Hence, for $p \neq 1/2$, $(\frac{p}{1-p})^{(2Y-d)} \stackrel{\text{a.s.}}{\to} \infty$.

Now note that, for additive TMCMC with a single ϵ, as $d \to \infty$, the ratio $\frac{\pi(T_b^{(1)}(x, \epsilon))}{\pi(x)}$ is expected to converge to zero at a very slow rate. Indeed, it follows from the supplement of Dutta and Bhattacharya (2014) that under the strong log-concavity assumption on π, the acceptance rate with these non-uniform move-type probabilities satisfies the following inequalities as $d \to \infty$:

$$
\left\{2\Phi\left(\sqrt{-\frac{2}{dM_d} \log \frac{1 - \psi_2}{c_d}}\right) - 1\right\} \leq \text{AR}_p \leq \left\{2\Phi\left(\sqrt{-\frac{2}{dM_d} \log \frac{\psi_1}{c_d}}\right) - 1\right\}, \tag{4}
$$

where $0 < \psi_1, \psi_2 < 1$, $M_d = O(d^t)$; $t > 2$, and $c_d = (\frac{p}{1-p})^{d(2p-1)}$. For $p = 1/2$, we obtain the following asymptotic inequality proved in the supplement of Dutta

and Bhattacharya (2014)

$$\left\{2\Phi\left(\sqrt{-\frac{2}{dM_d}\log(1-\psi_2)}\right)-1\right\} \leq \mathrm{AR}_{1/2} \leq \left\{2\Phi\left(\sqrt{-\frac{2}{dM_d}\log\psi_1}\right)-1\right\}, \quad (5)$$

which is a special case of (4). It is shown in the supplement of Dutta and Bhattacharya (2014) that for $p = 1/2$, as $d \to \infty$, the acceptance rate of additive TM-CMC tends to zero at a much slower rate compared to that of the normal random walk Metropolis–Hastings algorithm. In fact, it is easy to see that $\mathrm{AR}_p \to 0$ as $d \to \infty$ for any $p \in (0, 1)$, and quite importantly, it holds that $\frac{\mathrm{AR}_p}{\mathrm{AR}_{1/2}} \to \infty$ as $d \to \infty$. In other words, for high-dimensional target distributions, the additive TMCMC based acceptance rate can be further improved with non-uniform move probabilities.

But an increase in acceptance rate does not necessarily lead to faster convergence of the underlying Markov chain. Hence, although higher acceptance rates are to be expected of additive TMCMC for non-uniform move-type probabilities in high dimensions, faster rates of convergence may not still be achieved. We reserve the investigation of the effects of non-uniform move-type probabilities on convergence rate for our future research.
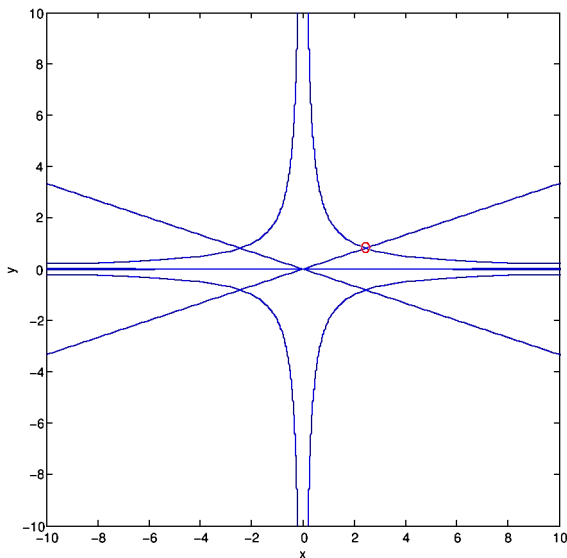
1.2 Multiplicative TMCMC

Again suppose that we are simulating from a d dimensional space (say, $\mathbb{R}^d$), and that we are currently at a point $x = (x_1, \ldots, x_d)$. Let us now modify the definition of the random variables $b_1, \ldots, b_d$, such that, for $i = 1, \ldots, d$,

$$b_i = \begin{cases} +1 & \text{with probability } p_i; \\ 0 & \text{with probability } q_i; \\ -1 & \text{with probability } 1 - p_i - q_i. \end{cases} \quad (6)$$

Let $\epsilon \sim g^{(2)} = q^{(2)}(\cdot)I_{\{|\epsilon|\leq 1\}}$. If $b_i = +1$, then $x_i \to x_i\epsilon$, if $b_i = -1$, then $x_i \to x_i/\epsilon$ and if $b_i = 0$, then $x_i \to x_i$, that is, x_i remains unchanged. Let the transformed coordinate be denoted by x_i^*. Also, let $J(b, \epsilon)$ denote the Jacobian of the transformation $(x, \epsilon) \mapsto (x^*, \epsilon)$. We denote x^* by $T_b^{(2)}(x, \epsilon)$, the multiplicative transformation $(x, \epsilon) \mapsto (x^*, \epsilon)$ associated with the move-type b.

For example, if $d = 2$, then for $b = (1, 1)$, $T_b^{(2)}(x, \epsilon) = (x_1\epsilon, x_2\epsilon)$ and the Jacobian is ϵ^2, for $b = (-1, -1)$, $T_b^{(2)}(x, \epsilon) = (x_1/\epsilon, x_2/\epsilon)$ and $|J(b, \epsilon)| = \epsilon^{-2}$. For $b = (1, -1)$, $b = (-1, 1)$, and $b = (0, 0)$, $T_b^{(2)}(x, \epsilon) = (x_1\epsilon, x_2/\epsilon)$, $(x_1/\epsilon, x_2\epsilon)$, and (x_1, x_2), respectively, and in all these three instances, $|J(b, \epsilon)| = 1$. For $b = (1, 0)$ and $b = (0, 1)$, $T_b^{(2)}(x, \epsilon) = (x_1\epsilon, x_2)$ and $T_b^{(2)}(x, \epsilon) = (x_1, x_2\epsilon)$, respectively, and in both these cases $|J(b, \epsilon)| = |\epsilon|$. For $b = (-1, 0)$ or $b = (0, -1)$, $T_b^{(2)}(x, \epsilon) = (x_1/\epsilon, x_2)$ and $(x_1, x_2/\epsilon)$, respectively, and the Jacobian is $|\epsilon|^{-1}$ in both these cases.

Figure 2 *Path diagram for multiplicative TMCMC in one step from a fixed point denoted by the circle.*

In general, the transformation of the ith coordinate is given by $x_i \epsilon^{b_i}$ and the Jacobian is given by $|\epsilon|^{\sum_{i=1}^{d} b_i}$.

The path diagram for multiplicative TMCMC that displays the possible range of values that our chain can move to starting from a fixed point is presented in Figure 2.

In this paper, for our purpose, we assume that $p_i = q_i = 1/3; i = 1, \ldots, d$. Then assuming that the target distribution is proportional to π, the new move $T_b^{(2)}(x, \epsilon)$ is accepted with probability

$$\alpha = \min\left\{1, \frac{\pi(T_b^{(2)}(x, \epsilon))}{\pi(x)} |J(b, \epsilon)|\right\}. \tag{7}$$

1.2.1 *Discussion on non-uniform move-type probabilities for multiplicative TM-CMC.* For simplicity of illustration, let us assume that $p_i = p$ and $q_i = q$, for $i = 1, \ldots, d$. Also, let $Y = \sum_{i:b_i=1} b_i$, and $Z = \sum_{i:b_i=0} b_i$. Then $Y \sim \text{Binomial}(d, p)$, and $Z \sim \text{Binomial}(d, q)$. The acceptance probability is then given by

$$\alpha = \min\left\{1, \left(\frac{p}{1-p-q}\right)^Y \left(\frac{1-p-q}{p}\right)^{(d-Y-Z)} \frac{\pi(T_b^{(1)}(x, \epsilon))}{\pi(x)}\right\}$$
$$= \min\left\{1, \left(\frac{p}{1-p-q}\right)^{(2Y+Z-d)} \frac{\pi(T_b^{(1)}(x, \epsilon))}{\pi(x)}\right\}. \tag{8}$$

As $d \to \infty$, $(2Y + Z - d) \overset{\text{a.s.}}{\sim} d(2p + q - 1)$. If $2p + q > 1$, then $p > (1 - q)/2$, so that $1 - p - q < (1 - q)/2$. Hence, $p/(1 - p - q) > 1$, implying that $(\frac{p}{1-p-q})^{(2Y+Z-d)} \overset{\text{a.s.}}{\to} \infty$. If, on the other hand $2p + q < 1$, then $p < (1 - q)/2$, and $1 - p - q > (1 - q)/2$, implying $p/(1 - p - q) < 1$. Again, this implies $(\frac{p}{1-p-q})^{(2Y+Z-d)} \overset{\text{a.s.}}{\to} \infty$. In contrast with additive TMCMC, for multiplicative TMCMC the asymptotic form of the acceptance rate is not available yet, even for strongly log-concave target distributions. However, it is clear that for $2p + q \neq 1$, the acceptance rate is asymptotically much higher than that associated with $2p + q = 1$.

Although for convenience of presentation we prove geometric ergodicity of multiplicative TMCMC assuming $p_i = p = 1/3$ and $q_i = q = 1/3$; $i = 1, \ldots, d$, the steps of our proofs remain the same for any other $0 < p, q < 1$, satisfying $2p + q = 1$ (note that for such p, q, $p + q = (1 + q)/2 < 1$ is automatically satisfied). We reserve the cases $2p + q \neq 1$ for our future investigation.

Apart from additive and multiplicative TMCMC we also consider appropriate geometric ergodic mixtures of additive and multiplicative TMCMC, which not only help bypass a somewhat undesirable theoretical assumption regarding the high-dimensional target density π, but as simulation studies demonstrate, appropriate mixtures of additive and multiplicative TMCMC also ensure faster convergence compared to individual additive TMCMC and individual multiplicative TMCMC.

1.3 Additive–multiplicative TMCMC

Dutta and Bhattacharya (2014) described another TMCMC algorithm that uses the additive transformation for some co-ordinates of x and the multiplicative transformation for the remaining coordinates. Dutta and Bhattacharya (2014) refer to this as additive–multiplicative TMCMC. Let the target density π be supported on $\mathbb{R}^d$. Then, if the additive transformation is used for the ith coordinate, we update x_i to $x_i + b_i\epsilon_1$, where b_i is defined by (1), and $\epsilon \sim g^{(1)}$. On the other hand, if for any co-ordinate x_j, the multiplicative transformation is used, then we simulate b_j following (6), simulate $\epsilon_2 \sim g^{(2)}$, and update x_j to either $x_j\epsilon_2$ or x_j/ϵ_2 accordingly as $b_j = +1$ or -1. If $b_j = 0$, then we leave x_j unchanged. The new proposal is accepted with probability having the same form as (7). Note that unlike the cases of additive TMCMC and multiplicative TMCMC, which use a single ϵ to update all the d coordinates of x, here we need two ϵ's: ϵ_1 and ϵ_2, to update the d co-ordinates.

The proof of geometric ergodicity of additive–multiplicative TMCMC is almost the same as that of multiplicative TMCMC, and hence we omit it from this paper.

1.4 Geometric ergodicity

Let P be the transition kernel of a ψ-irreducible, aperiodic, positive Harris recurrent Markov chain with the stationary distribution π. Then the chain is geometri-

cally ergodic if there exist a function $V \geq 1$ which is finite at least at one point, and constants $0 < \rho < 1$ and $M < \infty$ satisfying

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq M V(x) \rho^n \qquad \forall n \geq 1, \tag{9}$$

where $\|\mu\|_{\mathrm{TV}} = \sup_{g:|g| \leq V} \mu(g)$ denotes the *total variation norm*. A standard way of checking geometric ergodicity is a result that involves small sets and the 'geometric drift condition'. A set E is called small if there exists $m > 0$, $\delta > 0$ and a probability measure ν such that for $x \in E$,

$$P^m(x, \cdot) \geq \delta \nu(\cdot). \tag{10}$$

P is said to have geometric drift to a small set E if there is a function $V \geq 1$, finite for at least one point, and constants $\lambda < 1$ and $\zeta < \infty$ so that

$$PV(x) \leq \lambda V(x) + \zeta I_E(x), \tag{11}$$

where $PV(x) = \int V(y) P(x, y) \, dy$ is the expectation of V after one transition given that one starts at the point x, and $I_E(x) = 1$ if $x \in E$ and 0 otherwise, is the indicator function. Theorems 14.0.1 and 15.0.1 in Meyn and Tweedie (1993) establish the fact that if P has a geometric drift to a small set E, then under certain regularity conditions, P is π-almost everywhere geometrically ergodic and the converse is also true.

We now provide necessary and sufficient conditions in favour of (11); the result can be thought of as an adaptation of Lemma 3.5 of Jarner and Hansen (2000).

Lemma 1.1. *Assume that the Markov transition kernel P is associated with additive, multiplicative, or additive–multiplicative TMCMC. If there exists V such that $V \geq 1$ and finite on bounded support, such that the following hold*

$$\limsup_{\|x\| \to \infty} \frac{PV(x)}{V(x)} < 1 \quad and \tag{12}$$

$$\frac{PV(x)}{V(x)} < \infty \qquad \forall x, \tag{13}$$

then V satisfies the geometric drift condition (11) and hence the chain must be geometrically ergodic. Also, if for some V finite, the geometric drift condition is satisfied, then the above conditions must also hold true.

Proof. Assume that for some V finite and $V \geq 1$, the geometric drift condition (11) is satisfied. Now, dividing both sides by $V(x)$, we get

$$\frac{PV(x)}{V(x)} \leq \lambda + \zeta \frac{I_E(x)}{V(x)}.$$

Since V is finite, then given that $V \geq 1$, we have

$$\frac{PV(x)}{V(x)} \leq \lambda + \zeta < \infty.$$

Also if $\|x\| \to \infty$ then as E is a bounded small set, $I_E(x) \to 0$, and hence

$$\limsup_{\|x\|\to\infty} \frac{PV(x)}{V(x)} \leq \lambda < 1.$$

For the converse, let us fix a value $\gamma < 1$. Let R be particularly large so that if $\|x\| > R$, then

$$\frac{PV(x)}{V(x)} < \gamma \qquad \text{if } \|x\| > R \implies PV(x) < \gamma V(x) \qquad \text{if } \|x\| > R.$$

Since

$$PV(x) \leq \frac{PV(x)}{V(x)} V(x),$$

and since $\frac{PV(x)}{V(x)}$ is finite by hypothesis (13) and the function V is also finite on any bounded set, this implies that $PV(x)$ is finite on $E = \{x : \|x\| \leq R\}$, which is closed and bounded.

Take ζ to be the maximum value (which must be finite) that $PV(x)$ can attain on the set E. In the supplement of Dutta and Bhattacharya (2014) it is shown that for additive TMCMC, sets of the form $E = \{x : \|x\| \leq R\}$ are small. Defining

$$\mathcal{V} = \{(v_1, \ldots, v_d) \in \mathbb{R}^d : v_i = 0 \text{ for at least one } i \in \{1, \ldots, d\}\}, \qquad (14)$$

in the Appendix we will show that compact subsets of $\mathbb{R}^d \backslash \mathcal{V}$, which we denote by E^*, are small for multiplicative TMCMC; the same result also holds for additive–multiplicative TMCMC. Hence, for all x, if $\mathbb{E}$ is either E or E^*,

$$PV(x) \leq \gamma V(x) + \zeta I_{\mathbb{E}}(x).$$

This proves the lemma. $\square$

So, in order to check geometric ergodicity, it is enough to prove (12) and (13) for the given chain.

2 Geometric ergodicity of additive TMCMC

We shall now provide necessary and sufficient conditions for geometric ergodicity for additive TMCMC for a broad class of distributions. This proof follows on the lines of Jarner and Hansen (2000) and has been suitably modified for our additive TMCMC case. First, we define the notion of super-exponential densities.

A density π is said to be super-exponential if it is positive with continuous first derivative and satisfies

$$\lim_{\|x\|\to\infty} n(x)' \nabla \log \pi(x) = -\infty, \qquad (15)$$

where $n(x)$ denotes the unit vector $\frac{x}{\|x\|}$. This would imply that for any $K > 0$, $\exists\, R > 0$ such that

$$\frac{\pi(x + cn(x))}{\pi(x)} \leq e^{-cK}; \qquad \|x\| \geq R, c \geq 0. \tag{16}$$

In words, the above definition entails that π is decaying at a rate faster than exponential along any direction. It is very easy to check that the Gaussian (univariate as well as multivariate for any variance covariance matrix) or the Gamma distributions (univariate or independent multivariate) indeed satisfy these conditions.

Let the acceptance region and the (potential) rejection region corresponding to the move-type b be defined by $A^{(1)}(b, x) = \{\epsilon : \pi(T_b^{(1)}(x, \epsilon)) \geq \pi(x)\}$ and $R^{(1)}(b, x) = \{\epsilon : \pi(T_b^{(1)}(x, \epsilon)) < \pi(x)\}$, respectively. Also, let $A^{(1)}(x) = \bigcup_{b_1,\ldots,b_d} A^{(1)}(b, x)$ and $R^{(1)}(x) = \bigcap_{b_1,\ldots,b_d} R^{(1)}(b, x)$ denote the overall acceptance region and the overall potential rejection region, respectively.

Let $Q^{(1)}(x, B)$ denote the probability corresponding to the additive TMCMC proposal of reaching the Borel set B from x in one step. Let $P^{(1)}$ denote the Markov transition kernel associated with additive TMCMC. Then the following theorem establishes geometric ergodicity of additive TMCMC in the super-exponential set-up.

Theorem 2.1. *If the target density π is super-exponential and has contours that are nowhere piecewise parallel to $\{x : |x_1| = |x_2| = \cdots = |x_d|\}$, then the additive TMCMC chain satisfies geometric drift if and only if*

$$\liminf_{\|x\| \to \infty} Q^{(1)}(x, A^{(1)}(x)) > 0. \tag{17}$$

Proof.

Following the notation of Jarner and Hansen (2000), let $C_{\pi(x)}$ be the contour of the density π corresponding to the value $\pi(x)$. We define the radial cone $C_{\pi(x)}(\delta)$ around $C_{\pi(x)}$ to be

$$C_{\pi(x)}(\delta) = \{y + sn(y) : y \in C_{\pi(x)}, -\delta < s < \delta\}. \tag{18}$$

See Figure 1 of Jarner and Hansen (2000) for visualizing these regions in two dimensions.

By (17), there exists a $\eta > 0$ such that

$$\limsup_{\|x\| \to \infty} Q^{(1)}(x, R^{(1)}(x)) \leq 1 - 2\eta^{1/2}. \tag{19}$$

Take the belt length δ such that the probability that a move from x, the starting point, falls within this δ belt is less than η. That it is possible can be seen as follows. Note that there exists a compact set E such that

$$Q^{(1)}(x, E^c) < \frac{\eta}{2}. \tag{20}$$

So, for given δ, if we can ensure that our proposal distribution satisfies

$$Q^{(1)}(x, C_{\pi(x)}(\delta) \cap E) < \frac{\eta}{2}, \tag{21}$$

then we are done. Note that for any point on the contour, the probability that the additive TMCMC moves result in a value within $C_{\pi(x)}(\delta)$ is bounded above by $2c\delta$, for some finite c (since this probability is $2 \int_0^\delta g(\epsilon)\,d\epsilon \leq 2c\delta$, as $g(\epsilon) \leq c$ on $(0, \delta)$, for $0 < c < \infty$) and thus can be made as small as desired by choosing δ sufficiently small. The above argument is easy to visualize in two dimensions as depicted in Figure 1 and Figure 1 of Jarner and Hansen (2000)—for any point in the first quadrant part of the contour, the probability that the outer and inner TMCMC moves given, respectively, by $(+\epsilon, +\epsilon)$ and $(-\epsilon, -\epsilon)$, land within $C_{\pi(x)}(\delta)$, is bounded above by $2c\delta$. The same argument applies to the other three quadrants. For the other moves, note that since the contours (intersected with E) are nowhere piecewise parallel to $\{x : |x_1| = \cdots = |x_d|\}$, the moves can fall in only finite number of regions of $C_{\pi(x)}(\delta) \cap E$. Infinitely many regions can be ruled out because of the intersection with E, which is compact. If that was the case, then this infinite collection of interesting points would have a limit point in E, which is not possible as the points are isolated.

Now, there exists R_η so that for any point y outside the δ bound around x and in the rejection region, it holds that

$$\frac{\pi(y)}{\pi(x)} < \eta; \qquad \|x\| > R_\eta. \tag{22}$$

This can be seen by taking the shortest line from y to the origin; suppose it intersects (after extending if needed) the contour $C_{\pi(x)}$ at z. There will be two such values of z, and we choose the one that is nearest to x. Then, by (16) and the fact that $\pi(x)$ is the same as $\pi(z)$ (since x and z are on the same contour), we obtain (22). To ensure that this z indeed satisfies $\|z\| > R_\eta$, consider the set E, which is the set where effectively all the moves fall. Join each point in E to the origin by a straight line and extend it if needed to intersect the contour; consider those points of intersections which are closest to x. The points of intersections yield a segment, $D(x)$, of the contour, which contains x and is bounded and closed. Now since this set is bounded, we can always choose x with large enough norm so that all the points in E associated with $D(x)$ have norms greater than R_η. Since z is one of such points, we are done.

On the other hand, if y is outside the δ bound around x but falls in the acceptance region, then by the same arguments it holds that

$$\frac{\pi(x)}{\pi(y)} < \eta. \tag{23}$$

Now, with $V(x) = \frac{c}{\sqrt{\pi(x)}}$ for some $c > 0$ chosen appropriately, it holds that

$$
\frac{P^{(1)}V(x)}{V(x)} = \frac{1}{2^d} \sum_{b_1,\ldots,b_d} \int_{A^{(1)}(x)} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1 + b_1\epsilon,\ldots,x_d + b_d\epsilon)} \right]^{1/2} g^{(1)}(\epsilon)\, d\epsilon
$$

$$
+ \frac{1}{2^d} \sum_{b_1,\ldots,b_d} \int_{R^{(1)}(x)} \left[1 - \frac{\pi(x_1 + b_1\epsilon,\ldots,x_d + b_d\epsilon)}{\pi(x_1,\ldots,x_d)} \right. \tag{24}
$$

$$
+ \left. \left\{ \frac{\pi(x_1 + b_1\epsilon,\ldots,x_d + b_d\epsilon)}{\pi(x_1,\ldots,x_d)} \right\}^{1/2} \right] g^{(1)}(\epsilon)\, d\epsilon.
$$

We split the integrals over $A^{(1)}(x)$ and that over $R^{(1)}(x)$ into two parts—within $C_{\pi(x)}(\delta)$ and outside $C_{\pi(x)}(\delta)$. Since $\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)} < 1$ on $A^{(1)}(x)$, it follows from (20) and (21) that

$$
\int_{A^{(1)}(x) \cap C_{\pi(x)}(\delta)} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1 + b_1\epsilon,\ldots,x_d + b_d\epsilon)} \right]^{1/2} g^{(1)}(\epsilon)\, d\epsilon < \frac{\eta}{2}. \tag{25}
$$

Now note that

$$
\left| \frac{\|x_n + b\epsilon\|^2}{\|x_n\|^2} - 1 \right| \leq \frac{2\epsilon|b'x_n|}{\|x_n\|^2} + \frac{d\epsilon^2}{\|x_n\|^2}
$$

$$
\leq \frac{2\epsilon}{\|x_n\|} + \frac{d\epsilon^2}{\|x_n\|^2} \quad \left(\text{since } |b'x_n| \leq \sqrt{\|b\|^2 \|x_n\|^2} = \|x_n\|\right). \tag{26}
$$

Let $\mathcal{N}_\epsilon$ denote a null set associated with the probability distribution of ϵ. Then for all $\omega \in \mathcal{N}_\epsilon^c$ such that $\epsilon(\omega) \in E$, for any compact set E, (26) goes to zero. That is, for $\omega \in \mathcal{N}_\epsilon^c \cap \epsilon^{-1}(E)$, $\frac{\|x_n + b\epsilon\|}{\|x_n\|} \to 1$. Thus, for $n > N_0(\eta_2(\omega))$ for some $N_0(\eta_2(\omega))$ depending upon $\eta_2(\omega)$ such that $\|x_n\| > \frac{R_\eta}{1 - \eta_2(\omega)} > R_\eta$, since $1 + \eta_2(\omega) > \frac{\|x_n + b\epsilon\|}{\|x_n\|} > 1 - \eta_2(\omega)$ for $\omega \in \mathcal{N}_\epsilon^c \cap \epsilon^{-1}(E)$, we have

$$
\|x_n + b\epsilon\| > (1 - \eta_2(\omega))\|x_n\| > R_\eta. \tag{27}
$$

Note that for any given $\zeta > 0$, we can choose E_ζ such that $Q^{(1)}(x, E_{\zeta,x,b}^c) < \zeta$, for any x and b, where $E_{\zeta,x,b}^c = \{x + b\epsilon : \epsilon \in E_\zeta^c\}$. Thus, we can choose $\zeta > 0$ such that

$$
Q^{(1)}\left(x, A^{(1)}(x) \cap C_{\pi(x)}^c(\delta) \cap E_{\zeta,x,b}^c\right) < \frac{\eta^{1/2}}{2} Q^{(1)}\left(x, A^{(1)}(x)\right). \tag{28}
$$

Now, it follows from (27) and (23) that for given $\eta > 0$, we can choose R_η such that for $\|x\| > R_\eta$, $\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)} < \frac{\eta}{4}$. Hence,

$$
\int_{A^{(1)}(x) \cap C_{\pi(x)}^c(\delta) \cap E_{\zeta,x,b}} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1 + b_1\epsilon,\ldots,x_d + b_d\epsilon)} \right]^{1/2} g^{(1)}(\epsilon)\, d\epsilon
$$

$$
< \frac{\eta^{1/2}}{2} Q^{(1)}\left(x, A^{(1)}(x)\right). \tag{29}
$$

Also, since $\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)} < 1$ on $A^{(1)}(x)$, it follows from (28) that

$$\int_{A^{(1)}(x)\cap C^c_{\pi(x)}(\delta)\cap E^c_{\zeta,x,b}} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)}\right]^{1/2} g^{(1)}(\epsilon)\,d\epsilon$$
$$< \frac{\eta^{1/2}}{2} Q^{(1)}(x, A^{(1)}(x)). \tag{30}$$

Thus, for $\|x\| > R_\eta$, it follows from (29) and (30) that

$$\int_{A^{(1)}(x)\cap C^c_{\pi(x)}(\delta)} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)}\right]^{1/2} g^{(1)}(\epsilon)\,d\epsilon$$
$$< \eta^{1/2} Q^{(1)}(x, A^{(1)}(x)). \tag{31}$$

Now note that on $R^{(1)}(x)$, $1 - \frac{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)}{\pi(x_1,\ldots,x_d)} < 1$, so that

$$\int_{R^{(1)}(x)} \left[1 - \frac{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)}{\pi(x_1,\ldots,x_d)}\right] g^{(1)}(\epsilon)\,d\epsilon < Q^{(1)}(x, R^{(1)}(x)). \tag{32}$$

For the integral $\int_{R^{(1)}(x)} \frac{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)}{\pi(x_1,\ldots,x_d)} g^{(1)}(\epsilon)\,d\epsilon$, breaking up $R^{(1)}(x)$ into $R^{(1)}(x) \cap C_{\pi(x)}(\delta)$ and $R^{(1)}(x) \cap C^c_{\pi(x)}(\delta)$ we obtain, in exactly the same way as (25) and (31), the following:

$$\int_{R^{(1)}(x)} \left[\frac{\pi(x_1+b_1\epsilon,\ldots,x_d+b_d\epsilon)}{\pi(x_1,\ldots,x_d)}\right]^{1/2} g^{(1)}(\epsilon)\,d\epsilon$$
$$< \frac{\eta}{2} + \eta^{1/2} Q^{(1)}(x, R^{(1)}(x)), \tag{33}$$

for $\|x\| > R_\eta$. Combining (25), (31), (32) and (33), we obtain

$$\frac{P^{(1)}V(x)}{V(x)} < \eta + \eta^{1/2} Q^{(1)}(x, A^{(1)}(x)) + (1+\eta^{1/2})Q^{(1)}(x, R^{(1)}(x))$$
$$= \eta + \eta^{1/2} + Q^{(1)}(x, R^{(1)}(x)). \tag{34}$$

Using (19), we obtain

$$\limsup_{\|x\|\to\infty} \frac{PV(x)}{V(x)} < \eta + \eta^{1/2} + \limsup_{\|x\|\to\infty} Q^{(1)}(x, R^{(1)}(x))$$
$$= 1 - \eta^{1/2} + \eta$$
$$< 1.$$

Thus, (12) is satisfied. Since all the ratios in the integrals of (24) are less than 1, it is clear that $P^{(1)}V(x)/V(x) < \infty$ for all x, satisfying (13). This proves geometric ergodicity of additive TMCMC.

Now we prove that if additive TMCMC is geometrically ergodic, then (17) is satisfied. In fact, we prove that if condition (17) is not satisfied, that is, if $\limsup_{\|x\|\to\infty} Q^{(1)}(x, R^{(1)}(x)) = 1$, then $\limsup_{\|x\|\to\infty} P^{(1)}(x, \{x\}) = 1$. Indeed, it follows from Theorem 5.1 of Roberts and Tweedie (1996) that the latter condition implies that $P^{(1)}$ is not geometrically ergodic.

We can choose a compact set E such that $Q^{(1)}(x, E^c) < \eta$ and can choose δ small enough such that $\limsup_{\|x\|\to\infty} Q^{(1)}(x, C_{\pi(x)}(\delta) \cap E) \le \eta$. This and the fact (22) imply that

$$\limsup_{\|x\|\to\infty} P^{(1)}(x, \{x\})$$

$$\ge \limsup_{\|x\|\to\infty} \frac{1}{2^d} \sum_{b_1,\ldots,b_d} \int_{R^{(1)}(x)} \left[1 - \frac{\pi(x_1 + b_1\epsilon, \ldots, x_d + b_d\epsilon)}{\pi(x_1, \ldots, x_d)} \right] g^{(1)}(\epsilon) \, d\epsilon$$

$$\ge \limsup_{\|x\|\to\infty} \frac{1}{2^d} \sum_{b_1,\ldots,b_d} \int_{R^{(1)}(x) \cap E \cap [C_{\pi(x)}(\delta)]^c} \left[1 - \frac{\pi(x_1 + b_1\epsilon, \ldots, x_d + b_d\epsilon)}{\pi(x_1, \ldots, x_d)} \right]$$

$$\times g^{(1)}(\epsilon) \, d\epsilon$$

$$\ge (1 - \eta) \limsup_{\|x\|\to\infty} Q\left(x, R^{(1)}(x) \cap E \cap \left[C_{\pi(x)}(\delta)\right]^c\right)$$

$$\ge (1 - \eta)(1 - 2\eta).$$

Since $\eta > 0$ is arbitrary, the proof is complete. $\qquad\square$

Note that for spherically symmetric super-exponential distributions (for example, standard Gaussian), the conditions of Theorem 2.1 naturally hold. For instance, the fact that no part of the contour is parallel to $\{x : |x_1| = |x_2| = \cdots = |x_d|\}$ is quite obvious. To check that $\liminf_{\|x\|\to\infty} Q^{(1)}(x, A^{(1)}(x)) > 0$, first perceive (see Figure 3) that at any point in the first quadrant, the inward direction stays in the acceptance region if the magnitude of the inward direction does not exceed the
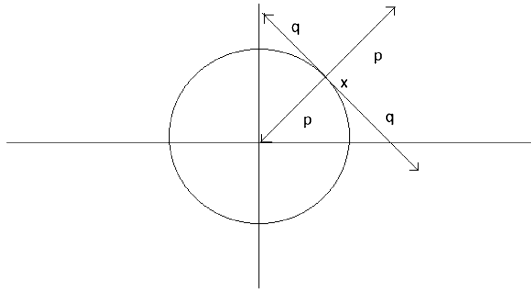
Figure 3 *A contour of a spherically symmetric distribution. Here x is the current state lying on the contour (first quadrant), and the four directions that can be taken by the next move of additive TMCMC, are displayed. Here $p = q = 1/2$ are the move-type probabilities.*

diameter of the contour containing x. However, the inward direction can land in the rejection region on the other side of the contour if the magnitude of the inward direction exceeds the diameter of the contour $C_{\pi(x)}$. Since $\|x\|$ is the radius of $C_{\pi(x)}$, in order to ensure that the inward move falls in $A^{(1)}(x)$ with high probability when $\|x\|$ is large, we must choose the proposal density $g^{(1)}(\epsilon)$ in such a way that too large step sizes compared to $\|x\|$, have small probabilities. Thus, for our purpose, first let M_η be such that $\int_0^{M_\eta} g^{(1)}(\epsilon)\, d\epsilon > 1 - \eta$. Now choose x such that $\|x\| > 3M_\eta$ (radius of $C_{\pi(x)}$ is greater than $3M_\eta$). Then $Q^{(1)}(x, A^{(1)}(x)) > \frac{1-\eta}{4} > 0$. Now consider any sequence x_n with $\|x_n\| \to \infty$, where x_n has norm greater than $3M_\eta$ for all but finite n. Then along this sequence, the limit of $Q^{(1)}(x, A^{(1)}(x))$ is greater than $\frac{1-\eta}{4}$. Thus, $\liminf_{\|x\| \to \infty} Q^{(1)}(x, A^{(1)}(x)) > 0$ condition is satisfied.

Note that the constraint that no part of the contour can be piecewise parallel to $\{x : |x_1| = \cdots = |x_d|\}$ does not really cause too much of a problem because the only common distribution that satisfies this property is the Laplace distribution and it is not super-exponential.

3 Geometric ergodicity of multiplicative TMCMC

In the one-dimensional case, geometric ergodicity of multiplicative TMCMC has been established by Dutta (2012), assuming that the target density is regularly varying in an appropriate sense. Here we extend the result to arbitrary dimensions, of course without the aid of the regularly varying assumption, since such an assumption is not well-defined in high dimensions. Note however, that since vectors $v \in \mathcal{V}$ (where $\mathcal{V}$ is defined in (14)), cannot belong to small sets associated with multiplicative TMCMC, to prove geometric ergodicity we also need to show that $\limsup_{\|x-v\| \to 0} P^{(2)} V(x)/V(x) < 1$ for all $v \in \mathcal{V}$. This seems to be too demanding a requirement. In the one-dimensional case, 0 is the only point which cannot belong to small sets, and the proof of geometric ergodicity in this case requires showing $\limsup_{|x| \to 0} P^{(2)} V(x)/V(x) < 1$. This has been established by Dutta (2012), however, the technique of his proof could not assist us in our complicated, high-dimensional case.

If one has the liberty to assume, in the high-dimensional case, that there is an arbitrarily small, compact neighborhood $\mathbb{N}_0$, of $\mathbf{0} = (0, 0, \ldots, 0)'$, which has zero probability under the target density π, then the proof of $\limsup_{\|x-v\| \to 0} P^{(2)} V(x)/V(x) < 1$ for all $v \in \mathcal{V}$ is not required. Although for practical purposes this is not a very stringent assumption, from the theoretical standpoint this is somewhat disconcerting. In the next subsections, we introduce two different kinds of geometric ergodic mixtures of additive and multiplicative TMCMC kernels that do not require the undesirable assumption $\pi(\mathbb{N}_0) = 0$. The first mixture we introduce is essentially multiplicative TMCMC in a sense to be made precise subsequently, whereas the second mixture is a straightforward convex combination of additive and multiplicative TMCMC kernels.

3.1 A new mixture-based Markov transition kernel with "essentially full" weight on multiplicative TMCMC

We break up π into a mixture of two densities: π_1, supported on $\mathbb{N}_0$, and π_2, supported on $\mathbb{N}_0^c$. That is, we write

$$
\begin{aligned}
\pi(x) &= \pi(\mathbb{N}_0)\frac{\pi(x)}{\pi(\mathbb{N}_0)}I\{x \in \mathbb{N}_0\} + \pi(\mathbb{N}_0^c)\frac{\pi(x)}{\pi(\mathbb{N}_0^c)}I\{x \in \mathbb{N}_0^c\} \\
&= \pi(\mathbb{N}_0)\pi_1(x) + \pi(\mathbb{N}_0^c)\pi_2(x),
\end{aligned}
\tag{35}
$$

where

$$
\pi_1(x) = \frac{\pi(x)}{\pi(\mathbb{N}_0)}I\{x \in \mathbb{N}_0\} \quad \text{and} \quad \pi_2(x) = \frac{\pi(x)}{\pi(\mathbb{N}_0^c)}I\{x \in \mathbb{N}_0^c\}.
\tag{36}
$$

Clearly, $\pi_2(\mathbb{N}_0) = 0$. In fact, as we elaborate below, the above mixture representation transfers the requirement $\pi(\mathbb{N}_0) = 0$ to $\pi_2(\mathbb{N}_0) = 0$.

Now consider the following Markov chain: for any $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$, with $\mathcal{B}(\mathbb{R}^d)$ being the Borel σ-field of $\mathbb{R}^d$,

$$
P(x, A) = \pi(\mathbb{N}_0)P^{(1)}(x, A) + \pi(\mathbb{N}_0^c)P^{(2)}(x, A),
\tag{37}
$$

where $P^{(1)}(x, \cdot)$ and $P^{(2)}(x, \cdot)$ are Markov transition kernels corresponding to additive TMCMC converging to π_1 and multiplicative TMCMC converging to π_2, respectively. We choose the proposal density $g^{(2)}$ for multiplicative TMCMC such that there is a one-dimensional, arbitrarily small neighborhood of 0 which receives zero probability under $g^{(2)}$. We denote the one-dimensional neighborhood of 0 by $\mathcal{N}_0$. We also assume that there exist arbitrarily small neighborhoods $\mathcal{N}_{+1}$ and $\mathcal{N}_{-1}$ of $+1$ and -1 respectively, which receive zero probability under $g^{(2)}$.

In order to implement the mixture kernel P, we can separately run two chains—one is additive TMCMC converging to π_1 and another is multiplicative TMCMC converging to π_2, both chains starting at the same initial value x_0. Since both the chains are positive Harris recurrent on $\mathbb{R}^d\setminus\mathcal{V}$ ($P^{(1)}$ is positive Harris recurrent on $\mathbb{R}^d$ and $P^{(2)}$ is positive Harris recurrent on $\mathbb{R}^d\setminus\mathcal{V}$) convergence to both π_1 and π_2 occurs for the initial value x_0 ($\neq \mathbf{0}$), even though the supports of π_1 and π_2 are disjoint. In practice, it will be convenient to choose x_0 from the boundary between $\mathbb{N}_0$ and $\mathbb{N}_0^c$. Thus, for any initial value x_0, we will have an additive TMCMC chain $\{x_1^{(k)}; k = 0, 1, 2, \ldots\}$ converging to π_1 and another multiplicative TMCMC chain $\{x_2^{(k)}; k = 0, 1, 2, \ldots\}$ converging to π_2, with $x_1^{(0)} = x_2^{(0)} = x_0$.

Finally, for each $k = 1, 2, \ldots$, we select and store $x_1^{(k)}$ with probability $\pi(\mathbb{N}_0)$ and $x_2^{(k)}$ with probability $[1 - \pi(\mathbb{N}_0)]$. Thus, for $k > 1$, the chain $\{P^{(1)}\}^k$ depends only on $x_1^{(k-1)}$, and not on $x_2^{(k-1)}$. Similarly, $\{P^{(2)}\}^k$ depends only on $x_2^{(k-1)}$ and not on $x_1^{(k-1)}$.

Thus, the mixture P uses additive TMCMC to simulate only from π_1, and uses multiplicative TMCMC to simulate only from π_2. Since $\pi(\mathbb{N}_0)$ is negligibly small, the mixture P gives "essentially full" weight to multiplicative TMCMC.

If we can prove that $P^{(2)}$ is geometrically ergodic for π_2, then because $P^{(1)}$ is geometrically ergodic for π_1 (in fact, uniformly ergodic for π_1 since the support $\mathbb{N}_0$ of π_1 is compact), it will follow that P itself is geometrically ergodic. See Appendix B for a proof of this statement.

Note that $\pi(\mathbb{N}_0)$ is unknown and needs to be estimated for implementing P. In Appendix C, we present an importance sampling based idea regarding this, also demonstrating why the estimated probability is expected to yield the same TM-CMC samples as the exact value of $\pi(\mathbb{N}_0)$.

The mixture kernel P given by (37) is designed to give almost full weight to multiplicative TMCMC. It is also possible to consider a more conventional mixture of additive and multiplicative TMCMC, which is also geometrically ergodic but combines the good features of both the algorithms to yield a more efficient TM-CMC sampler, and does not require estimation of $\pi(\mathbb{N}_0)$. In the next subsection, we discuss this in detail, also elucidating how (37) differs from the combination of additive and multiplicative TMCMC in a traditional mixture set-up.

3.2 Combination of additive and multiplicative TMCMC in a traditional mixture set-up

Instead of (37), we could define a mixture of the form

$$P^*(x, A) = p P^{(1)}(x, A) + (1 - p) P^{(2)}(x, A), \tag{38}$$

where $0 < p < 1$ is any choice of mixing probability. The transition kernels $P^{(1)}(x, \cdot)$ and $P^{(2)}(x, \cdot)$, as before, are additive and multiplicative TMCMC, respectively, but here each of them converges to the target density π, unlike the case of (37) where $P^{(1)}(x, \cdot)$ converged to π_1 and $P^{(2)}(x, \cdot)$ converged to π_2.

For the implementation of P^*, one can first simulate $u \sim U(0, 1)$; if $u < p$, then additive TMCMC will be implemented, else multiplicative TMCMC should be used. Thus, unlike the case of (37), we have a single chain $\{x^{(k)}; k = 0, 1, 2, \ldots\}$ converging to π. Note also, that P^* implements both additive and multiplicative TMCMC on the entire support of π. In contrast, P, given by (37), implements additive TMCMC only for π_1, which is supported on $\mathbb{N}_0$ and implements multiplicative TMCMC only for π_2, which is supported on $\mathbb{N}_0^c$.

Recall that, in Section 2, we have already shown, for $V(x) = c/\sqrt{\pi(x)}$, that $\limsup_{\|x\| \to \infty} \frac{P^{(1)} V(x)}{V(x)} < 1$ and that the ratio $\frac{P^{(1)} V(x)}{V(x)}$ is finite for all x. For the same function V if we can also prove that $\limsup_{\|x\| \to \infty} \frac{P^{(2)} V(x)}{V(x)} < 1$, and that the ratio $\frac{P^{(2)} V(x)}{V(x)}$ is finite for all x, then it follows that the mixture P^* is also

geometrically ergodic. Indeed,

$$\limsup_{\|x\|\to\infty} \frac{P^* V(x)}{V(x)} \le p \limsup_{\|x\|\to\infty} \frac{P^{(1)} V(x)}{V(x)} + (1-p) \limsup_{\|x\|\to\infty} \frac{P^{(2)} V(x)}{V(x)}$$

$$< p + (1-p) = 1,$$

and $v \in \mathcal{V}$ can be a limit point of small sets corresponding to P^*, since $P^*(x, A) \ge p P^{(1)}(x, A)$ for all x and all $A \in \mathcal{B}(\mathbb{R}^d)$, and all compact sets of $\mathbb{R}^d$ are small sets of $P^{(1)}$.

3.3 Distinctions between the roles of $P^{(2)}$ in P and P^*

3.3.1 *Geometric ergodicity of P requires geometric ergodicity of $P^{(2)}$.* The proof of geometric ergodicity of the essentially fully multiplicative mixture P will follow if we can show that $P^{(2)}$ is geometrically ergodic for π_2, where $\pi_2(\mathbb{N}_0) = 0$ by construction. Theorem 3.1 provides necessary and sufficient conditions for geometric ergodicity of $P^{(2)}$ under the super-exponential set-up, assuming $\pi_2(\mathbb{N}_0) = 0$, and that the proposal density $g^{(2)}$ gives zero probability to arbitrarily small compact neighborhoods of 0, -1 and $+1$. Since it is always possible to construct a proposal density $g^{(2)}$ with the requisite properties, the strategy of forming the mixture P is not restrictive, given the super-exponential set-up.

3.3.2 *Geometric ergodicity of P^* does not require geometric ergodicity of $P^{(2)}$ or the restriction $\pi(\mathbb{N}_0) = 0$.* Geometric ergodicity of the traditional mixture P^*, on the other hand, follows only if $\limsup_{\|x\|\to\infty} \frac{P^{(2)} V(x)}{V(x)} < 1$ and $\frac{P^{(2)} V(x)}{V(x)}$ is finite for all x; it does not require $\pi(\mathbb{N}_0) = 0$, π being the invariant target distribution for both $P^{(1)}$ and $P^{(2)}$ of P^*. That the former two conditions hold under the aforementioned assumptions, without the restriction $\pi(\mathbb{N}_0) = 0$, can be easily seen in the proof of Theorem 3.1. The implication is that, super-exponentiality of π and the aforementioned properties of the proposal distribution $g^{(2)}$ guarantee geometric ergodicity of P^*, even though $P^{(2)}$ does not individually converge to its invariant distribution π because the proposal density $g^{(2)}$ assigns zero probability to some arbitrarily small compact neighborhood of $\mathbf{0}$ (in the first place, the fact that P^* converges to π is clear because of irreducibility, aperiodicity and positive Harris recurrence of P^*). As before, assumption of such proposal density $g^{(2)}$ is not restrictive, and hence the strategy of forming the mixture P^* is not restrictive either, given the super-exponential set-up.

In the next section, we introduce our theorem characterizing geometric ergodicity of $P^{(2)}$.

3.4 Geometric ergodicity of $P^{(2)}$ for π_2

For multiplicative TMCMC, for a given move-type b, we define the acceptance region and the potential rejection region by $A^{(2)}(b, x) = \{\epsilon : \frac{\pi(T_b^{(2)}(\epsilon))}{\pi(x)} |J(b, \epsilon)| \ge 1\}$

and $R^{(2)}(b, x) = \{\epsilon : \frac{\pi(T_b^{(2)}(\epsilon))}{\pi(x)}|J(b, \epsilon)| < 1\}$, respectively. The overall acceptance region and the overall potential rejection region are $A^{(2)}(x) = \bigcup_{b_1,\dots,b_d} A^{(2)}(b, x)$ and $R^{(2)}(x) = \bigcap_{b_1,\dots,b_d} R^{(2)}(b, x)$, respectively. We also define $A^*(b, x) = \{\epsilon : \frac{\pi(T_b^{(2)}(\epsilon))}{\pi(x)} \geq 1\}$ and $R^*(b, x) = \{\epsilon : \frac{\pi(T_b^{(2)}(\epsilon))}{\pi(x)} < 1\}$, respectively.

Let $Q^{(2)}(x, B)$ denote the probability corresponding to the multiplicative TMCMC proposal of reaching the Borel set B from x in one step.

Then the following theorem characterizes geometric ergodicity of multiplicative TMCMC under the super-exponential set-up. For our convenience, we slightly abuse notation by referring to π_2 as π.

Theorem 3.1. *Suppose that π, the target density, is super-exponential and has contours that are nowhere piecewise parallel to $\{x : |x_1| = \dots = |x_d|\}$; also assume that there is an arbitrarily small compact neighborhood $\mathbb{N}_0$ such that $\pi(\mathbb{N}_0) = 0$. If there exist compact neighborhoods $\mathcal{N}_0$, $\mathcal{N}_{+1}$ and $\mathcal{N}_{-1}$ (all arbitrarily small) of 0, $+1$ and -1, respectively such that $g^{(2)}$ gives zero probability to $\mathcal{N}_0$, $\mathcal{N}_{+1}$ and $\mathcal{N}_{-1}$, then the multiplicative TMCMC chain satisfies geometric drift if and only if*

$$\liminf_{\|x\| \to \infty} Q^{(2)}(x, A^{(2)}(x)) > 0. \tag{39}$$

Proof. As before, let $C_{\pi(x)}$ be the contour of the density π corresponding to the value $\pi(x)$, and let the radial cone around $C_{\pi(x)}$ be $C_{\pi(x)}(\delta)$ given by (18). By (39), there exists $\gamma > 0$ such that

$$\limsup_{\|x\| \to \infty} Q^{(2)}(x, R^{(2)}(x)) \leq 1 - 5\gamma^{1/2}. \tag{40}$$

Once again, we take the belt of length δ such that the probability that a move from x falls within this δ belt is less than γ. This holds since the neighborhoods $\mathcal{N}_{+1}$ and $\mathcal{N}_{-1}$ of $+1$ and -1 receive zero probabilities under the proposal density $g^{(2)}$. The remaining arguments are similar as in the proof of Theorem 2.1. Hence, as before, there exists R_γ so that for any point y outside the δ bound around x, (22) holds.

As before, let $V(x) = \frac{c}{\sqrt{\pi(x)}}$, where $c > 0$ is chosen appropriately. Then it holds that

$$\frac{P^{(2)}V(x)}{V(x)} = \frac{1}{3^d} \sum_{b_1,\dots,b_d} \int_{A^{(2)}(x)} \left[\frac{\pi(x_1, \dots, x_d)}{\pi(T_b(x, \epsilon))} \right]^{1/2} g^{(2)}(\epsilon) \, d\epsilon$$

$$+ \frac{1}{3^d} \sum_{b_1,\dots,b_d} \int_{R^{(2)}(x)} \left[1 - \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \dots, x_d)} |J(b, \epsilon)| \right. \tag{41}$$

$$\left. + \left\{ \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \dots, x_d)} \right\}^{1/2} |J(b, \epsilon)| \right] g^{(2)}(\epsilon) \, d\epsilon$$

$$\leq \frac{1}{3^d} \sum_{b_1,\ldots,b_d} \left\{ \sum_{b_1,\ldots,b_d} \int_{A^{(2)}(b,x)} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(T_b(x,\epsilon))} \right]^{1/2} g^{(2)}(\epsilon)\,d\epsilon \right\}$$

$$+ \frac{1}{3^d} \sum_{b_1,\ldots,b_d} \int_{R^{(2)}(x)} \left[1 - \frac{\pi(T_b(x,\epsilon))}{\pi(x_1,\ldots,x_d)} |J(b,\epsilon)| \right. \tag{42}$$

$$\left. + \left\{ \frac{\pi(T_b(x,\epsilon))}{\pi(x_1,\ldots,x_d)} \right\}^{1/2} |J(b,\epsilon)| \right] g^{(2)}(\epsilon)\,d\epsilon.$$

We now break up the integrals on $A^{(2)}(b,x)$ as sums of the integrals on $A^{(2)}(b,x) \cap A^*(b,x)$ and $A^{(2)}(b,x) \cap R^*(b,x)$. Also, we break up the integrals on $R^{(2)}(x)$ as sums of integrals on $R^{(2)}(x) \cap A^*(b,x)$ and $R^{(2)}(x) \cap R^*(b,x)$. Since $R^{(2)}(x) = \bigcap_{b_1,\ldots,b_d} R^{(2)}(b,x)$, these involve the intersections $R^{(2)}(b,x) \cap A^*(b,x)$ and $R^{(2)}(b,x) \cap R^*(b,x)$, respectively.

Note that, since $|J(b,\epsilon)|$ is of the form $|\epsilon|^k$, for $k = -d,\ldots,-1,0,1,\ldots,d$, and $|\epsilon| \leq 1$ (almost surely), $A^{(2)}(b,x) \cap R^*(b,x)$ is either the null set $\emptyset$ (when $k = -d, -d+1, \ldots, -1, 0$), or of the form $A^{(2)}(b,x) \cap R^*(b,x) = \{\epsilon : |\epsilon|^k \leq \frac{\pi(T_b^{(2)}(x,\epsilon))}{\pi(x)} < 1\}$, for $k = 1, 2, \ldots, d$. Hence, for $\|x\| > R_\gamma$, by (22), $Q^{(2)}(x, A^{(2)}(b,x) \cap R^*(b,x) \cap [C_{\pi(x)}(\delta)]^c) \leq Q^{(2)}(|\epsilon|^k \leq \frac{\pi(T_b^{(2)}(x,\epsilon))}{\pi(x)} < \gamma) < \gamma/2$, and for δ sufficiently small, $Q^{(2)}(x, A^{(2)}(b,x) \cap R^*(b,x) \cap C_{\pi(x)}(\delta)) < \gamma/2$. Moreover, on $A^{(2)}(b,x) \cap R^*(b,x) \cap C_{\pi(x)}(\delta)$, $\frac{\pi(x_1,\ldots,x_d)}{\pi(T_b(x,\epsilon))}$ is bounded by a finite constant. By hypothesis, $\mathcal{N}_0$ has zero probability under $g^{(2)}$. This implies that the set

$$\mathcal{S} = \left\{ |\epsilon| \leq 1 : \exists b \text{ and set } \mathcal{S}_\epsilon \text{ such that for } x \in \mathcal{S}_\epsilon, \frac{\pi(x_1,\ldots,x_d)}{\pi(T_b(x,\epsilon))} > K, \forall K > 0 \right\}$$

has zero probability under $g^{(2)}$. Hence, $\frac{\pi(x_1,\ldots,x_d)}{\pi(T_b(x,\epsilon))}$ is almost surely bounded even on $A^{(2)}(b,x) \cap R^*(b,x) \cap C_{\pi(x)}^c(\delta)$.

Hence, by the above arguments, for $\|x\| > R_\gamma$, and for sufficiently small $\xi > 0$,

$$\int_{A^{(2)}(b,x) \cap R^*(b,x)} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(T_b(x,\epsilon))} \right]^{1/2} g^{(2)}(\epsilon)\,d\epsilon$$

$$= \int_{A^{(2)}(b,x) \cap R^*(b,x) \cap C_{\pi(x)}(\delta)} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(T_b(x,\epsilon))} \right]^{1/2} g^{(2)}(\epsilon)\,d\epsilon$$

$$+ \int_{A^{(2)}(b,x) \cap R^*(b,x) \cap [C_{\pi(x)}(\delta)]^c} \left[\frac{\pi(x_1,\ldots,x_d)}{\pi(T_b(x,\epsilon))} \right]^{1/2} g^{(2)}(\epsilon)\,d\epsilon$$

$$< \xi/2.$$

By similar (in fact, somewhat simpler) arguments, it follows that for $\|x\| > R_\gamma$,

$$
\int_{A^{(2)}(b,x) \cap A^*(b,x)} \left[\frac{\pi(x_1, \ldots, x_d)}{\pi(T_b(x, \epsilon))} \right]^{1/2} g^{(2)}(\epsilon) \, d\epsilon
$$

$$
< \frac{\xi}{2} + \xi^{1/2} Q^{(2)}\left(x, A^{(2)}(b,x) \cap A^*(b,x) \cap \left[C_{\pi(x)}(\delta) \right]^c \right) \qquad (43)
$$

$$
< \frac{\xi}{2} + \xi^{1/2} Q^{(2)}\left(x, A^{(2)}(x)\right).
$$

The arguments required are somewhat simpler because on $A^{(2)}(b,x) \cap A^*(b,x)$, the ratio $\frac{\pi(x_1, \ldots, x_d)}{\pi(T_b(x,\epsilon))}$ is bounded above by 1. Hence, the first part of the expression for $P^{(2)}V(x)/V(x)$ given by (42) is less than $3^d(\xi + \xi^{1/2} Q^{(2)}(x, A^{(2)}(x)))$. Formally, for $\|x\| > R_\gamma$,

$$
\frac{1}{3^d} \sum_{b_1, \ldots, b_d} \left\{ \sum_{b_1, \ldots, b_d} \int_{A^{(2)}(b,x)} \left[\frac{\pi(x_1, \ldots, x_d)}{\pi(T_b(x, \epsilon))} \right]^{1/2} g^{(2)}(\epsilon) \, d\epsilon \right\}
$$

$$
< 3^d\left(\xi + \xi^{1/2} Q^{(2)}\left(x, A^{(2)}(x)\right)\right). \qquad (44)
$$

For sufficiently small $\xi > 0$ we can choose $\eta > 3^{2d}\xi$ so that

$$
3^d\left(\xi + \xi^{1/2} Q^{(2)}\left(x, A^{(2)}(x)\right)\right) < \eta + \eta^{1/2} Q^{(2)}\left(x, A^{(2)}(x)\right). \qquad (45)
$$

In the second part of the expression for $P^{(2)}V(x)/V(x)$, note that on $R^{(2)}(x)$,

$$
1 - \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} \left| J(b, \epsilon) \right| < 1,
$$

so that

$$
\int_{R^{(2)}(x)} \left[1 - \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} \left| J(b, \epsilon) \right| \right] g^{(2)}(\epsilon) \, d\epsilon < Q^{(2)}\left(x, R^{(2)}(x)\right), \qquad (46)
$$

and

$$
\int_{R^{(2)}(x)} \left\{ \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} \left| J(b, \epsilon) \right| g^{(2)}(\epsilon) \, d\epsilon \qquad (47)
$$

$$
= \int_{R^{(2)}(x) \cap A^*(b,x)} \left\{ \frac{\pi(T_b(x, \epsilon)) |J(b, \epsilon)|}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} \left| J(b, \epsilon) \right|^{1/2} g^{(2)}(\epsilon) \, d\epsilon \qquad (48)
$$

$$
+ \int_{R^{(2)}(x) \cap R^*(b,x)} \left\{ \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} \left| J(b, \epsilon) \right| g^{(2)}(\epsilon) \, d\epsilon. \qquad (49)
$$

Note that on $R^{(2)}(x) \cap A^*(b,x)$, $\frac{\pi(T_b(x,\epsilon))|J(b,\epsilon)|}{\pi(x_1,\ldots,x_d)} < 1$, and by our choice of the proposal density $g^{(2)}$, $\mathcal{N}_0$ has zero probability under $g^{(2)}$, so that the Jacobians $|J(b, \epsilon)|$ are bounded above by a finite constant, say K; we choose

$K > 1$. Hence, the first integral (48) in the break-up of the integral (47) is bounded above by $K Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x))$. Now, $Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x)) = Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x) \cap C_{\pi(x)}(\delta)) + Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x) \cap [C_{\pi(x)}(\delta)]^c)$, and we can achieve $Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x) \cap C_{\pi(x)}(\delta)) < \gamma^*/4$, for sufficiently small γ^*.

The sets of the form $R^{(2)}(b, x) \cap A^*(b, x)$ are again empty sets or of the form $\{\epsilon : |\epsilon|^k \leq \frac{\pi(T_b^{(2)}(x, \epsilon))}{\pi(x)} < 1\}; k = 1, 2, \ldots, d$. Hence, the sets $R^{(2)}(x) \cap A^*(b, x)$ are also either empty sets or intersections with sets of the form $\{\epsilon : |\epsilon|^k \leq \frac{\pi(T_b^{(2)}(x, \epsilon))}{\pi(x)} < 1\}$; $k = 1, 2, \ldots, d$. Hence, for $\|x\| > R_\gamma$, we can achieve $Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x) \cap [C_{\pi(x)}(\delta)]^c) < \gamma^*/4$. In other words, for $\|x\| > R_\gamma$,

$$
\begin{aligned}
\int_{R^{(2)}(x) \cap A^*(b,x)} &\left\{ \frac{\pi(T_b(x, \epsilon))|J(b, \epsilon)|}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} |J(b, \epsilon)|^{1/2} g^{(2)}(\epsilon) \, d\epsilon \\
&< K Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x) \cap C_{\pi(x)}(\delta)) \\
&\quad + Q^{(2)}(x, R^{(2)}(x) \cap A^*(b, x) \cap [C_{\pi(x)}(\delta)]^c) \\
&< K\gamma^*/2.
\end{aligned}
\tag{50}
$$

Now consider the second integral (49) in the break-up of the integral (47). We have

$$
\int_{R^{(2)}(x) \cap R^*(b,x)} \left\{ \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} |J(b, \epsilon)| g^{(2)}(\epsilon) \, d\epsilon
$$

$$
= \int_{R^{(2)}(x) \cap R^*(b,x) \cap C_{\pi(x)}} \left\{ \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} |J(b, \epsilon)| g^{(2)}(\epsilon) \, d\epsilon
\tag{51}
$$

$$
+ \int_{R^{(2)}(x) \cap R^*(b,x) \cap [C_{\pi(x)}]^c} \left\{ \frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} |J(b, \epsilon)|
$$

$$
\times g^{(2)}(\epsilon) \, d\epsilon.
\tag{52}
$$

Note that on $R^*(b, x)$, $\frac{\pi(T_b(x, \epsilon))}{\pi(x_1, \ldots, x_d)} < 1$. Hence, the first integral (51) in the above break-up is bounded above by $K Q^{(2)}(x, R^{(2)}(x) \cap R^*(b, x) \cap C_{\pi(x)}(\delta))$, which, in turn, is bounded above by $K\gamma^*/2$. For $\|x\| > R_\gamma$, the second integral (52) is bounded above by $K\gamma^{*1/2} Q^{(2)}(x, R^{(2)}(x) \cap R^*(b, x) \cap [C_{\pi(x)}(\delta)]^c)$, which, in turn, is bounded above by $K\gamma^{*1/2} Q^{(2)}(x, R^{(2)}(x))$. In other words,

$$
\begin{aligned}
\int_{R^{(2)}(x) \cap R^*(b,x)} &\left\{ \frac{\pi(T_b(x, \epsilon))|J(b, \epsilon)|}{\pi(x_1, \ldots, x_d)} \right\}^{1/2} |J(b, \epsilon)|^{1/2} g^{(2)}(\epsilon) \, d\epsilon \\
&< K\frac{\gamma^*}{2} + K\gamma^{*1/2} Q^{(2)}(x, R^{(2)}(x)).
\end{aligned}
\tag{53}
$$

Combining (50) and (53), we obtain that (47) is bounded above by $K\gamma^* + K\gamma^{*1/2}Q^{(2)}(x, R^{(2)}(x))$. With sufficiently small γ^* we have, for $\eta > K^2\gamma^*$,

$$K\gamma^* + K\gamma^{*1/2}Q^{(2)}(x, R^{(2)}(x)) < \eta + \eta^{1/2}Q^{(2)}(x, R^{(2)}(x)).$$

Combining this with (46) we get the following upper bound for the second term of (42):

$$\frac{1}{3^d} \sum_{b_1,\ldots,b_d} \int_{R^{(2)}(x)} \left[1 - \frac{\pi(T_b(x,\epsilon))}{\pi(x_1,\ldots,x_d)} |J(b,\epsilon)| \right.$$

$$\left. + \left\{ \frac{\pi(T_b(x,\epsilon))}{\pi(x_1,\ldots,x_d)} \right\}^{1/2} |J(b,\epsilon)| \right] g^{(2)}(\epsilon)\, d\epsilon \qquad (54)$$

$$< \eta + (1 + \eta^{1/2})Q^{(2)}(x, R^{(2)}(x)).$$

Combining (44) and (54), we obtain, for $\eta < \gamma$ (so that $\max\{3^{2d}\xi, K^2\gamma^*\} < \eta < \gamma$),

$$\limsup_{\|x\|\to\infty} \frac{P^{(2)}V(x)}{V(x)} \leq 2\eta + \eta^{1/2} + \limsup_{\|x\|\to\infty} Q^{(2)}(x, R^{(2)}(x))$$

$$< 2\eta + \eta^{1/2} + 1 - 5\eta^{1/2} \qquad \text{(by (40) and the fact that } \eta < \gamma)$$

$$= 1 - (2\eta)^{1/2} + 2\eta$$

$$< 1.$$

Hence, (12) holds. To see that condition (13) holds, in (41) observe that all the ratios in the integrands are bounded above by 1, while the terms $|J(b,\epsilon)|^{1/2}$ are almost surely bounded above by our choice of the proposal density $g^{(2)}$. Hence, $P^{(2)}V(x)/V(x)$ is finite for every x.

Now we prove that if multiplicative TMCMC is geometrically ergodic, then (39) is satisfied. As before, we prove that if $\limsup_{\|x\|\to\infty} Q^{(2)}(x, R^{(2)}(x)) = 1$, then $\limsup_{\|x\|\to\infty} P^{(2)}(x, \{x\}) = 1$. Again, we choose a compact set E such that $Q^{(2)}(x, E^c) \leq \eta$ and choose $\delta > 0$ small enough such that $\limsup_{\|x\|\to\infty} Q^{(2)}(x, C_{\pi(x)}(\delta) \cap E) \leq \eta$. Since $R^{(2)}(x) \cap A^*(b, x) \cap [C_{\pi(x)}(\delta)]^c$ is either null set or intersection with sets of the form $\{\epsilon : |\epsilon|^k \leq \frac{\pi(x)}{\pi(T_b(x,\epsilon))} < 1\}$, for $k = 1, 2, \ldots, d$, it follows from (22) that for any fixed b^*, if $\|x\| > R_\eta$,

$$Q^{(2)}(x, R^{(2)}(x) \cap A^*(b^*, x) \cap [C_{\pi(x)}(\delta)]^c)$$

$$\leq Q^{(2)}\left(x, \left\{ \epsilon : |\epsilon|^k \leq \frac{\pi(x)}{\pi(T_{b^*}(x,\epsilon))} < \eta \right\}\right) \leq \eta.$$

Hence,

$$\limsup_{\|x\|\to\infty} Q^{(2)}(x, R^{(2)}(x) \cap A^*(b^*, x) \cap [C_{\pi(x)}(\delta)]^c) \leq \eta.$$

Since $\limsup_{\|x\|\to\infty} Q^{(2)}(x, R^{(2)}(x)) = 1$, the above imply that

$$\limsup_{\|x\|\to\infty} Q^{(2)}\big(x, R^{(2)}(x) \cap R^*(b^*, x) \cap [C_{\pi(x)}(\delta)]^c\big) > 1 - 2\eta.$$

Moreover, since $|J(b, \epsilon)|$ are almost surely bounded by the choice of our proposal density, assume that there exists $0 < K < \infty$ such that $|J(b, \epsilon)| < K$ almost surely with respect to $g^{(2)}$.

These and the fact (22) imply that

$$\limsup_{\|x\|\to\infty} P^{(1)}(x, \{x\})$$

$$= \limsup_{\|x\|\to\infty} \frac{1}{3^d} \sum_{b_1,\ldots,b_d} \int_{R^{(2)}(x)} \left[1 - \frac{\pi(T_b(x, \epsilon))}{\pi(x)} |J(b, \epsilon)|\right] g^{(2)}(\epsilon) \, d\epsilon$$

$$\geq \limsup_{\|x\|\to\infty} \frac{1}{3^d} \sum_{b_1,\ldots,b_d} \int_{R^{(2)}(x)\cap R^*(b^*,x)\cap [C_{\pi(x)}(\delta)]^c} \left[1 - \frac{\pi(T_b(x, \epsilon))}{\pi(x)} |J(b, \epsilon)|\right]$$

$$\times g^{(2)}(\epsilon) \, d\epsilon \tag{55}$$

$$\geq (1 - \eta K) \limsup_{\|x\|\to\infty} Q\big(x, R^{(2)}(x) \cap R^*(b^*, x) \cap [C_{\pi(x)}(\delta)]^c\big)$$

$$\geq (1 - \eta K)(1 - 2\eta).$$

Since $\eta > 0$ is arbitrary, the proof is complete. $\qquad\square$

That it is easy to ensure geometric ergodicity of multiplicative TMCMC in super-exponential cases can be seen as follows. Select a move-type b^* such that $|J(b^*, \epsilon)| = |\epsilon|$. Then $A(b^*, x) = \{\epsilon : \frac{\pi(T_{b^*}(x,\epsilon))}{\pi(x)} |\epsilon| \geq 1\}$, and $A^*(b^*, x) = \{\epsilon : \frac{\pi(T_{b^*}(x,\epsilon))}{\pi(x)} \geq 1\}$. Then, since $|\epsilon| \leq 1$ almost surely,

$$A(b^*, x) \cap A^*(b^*, x) = \left\{\epsilon : \frac{\pi(x)}{\pi(T_{b^*}(x, \epsilon))} < |\epsilon| \leq 1\right\}. \tag{56}$$

If, for $\eta > 0$, $\|x\| > R_\eta$, then by (22),

$$\frac{\pi(x)}{\pi(T_{b^*}(x, \epsilon))} < \eta. \tag{57}$$

Equations (56) and (57) imply that for any given $\xi > 0$ it is possible to choose $\eta > 0$ such that for $\|x\| > R_\eta$, it holds that

$$Q^{(2)}\big(x, A^{(2)}(b^*, x) \cap A^*(b^*, x)\big) > 1 - \xi. \tag{58}$$

Hence, for $\|x\| > R_\eta$, we obtain using (58),

$$Q^{(2)}\big(x, A^{(2)}(x)\big) \geq Q^{(2)}\big(x, A^{(2)}(x) \cap A^*(b^*, x)\big)$$

$$\geq Q^{(2)}\big(x, A^{(2)}(b^*, x) \cap A^*(b^*, x)\big)$$

$$> 1 - \xi.$$

Hence, (39) holds, ensuring geometric ergodicity.

As we remarked earlier, we omit the proof of geometric ergodicity of additive–multiplicative TMCMC, since it is almost the same as that of multiplicative TM-CMC, provided above.

4 Illustration with simulation studies

There are several considerations in defining the accuracy or the efficiency of any MCMC-based approach. First, one important aspect is that the chain must have reasonably high acceptance rate. This has been an important consideration in our proposing TMCMC. It is to be noted that geometric ergodicity only tells us that convergence of our chain to the target density occurs at a geometric rate. However, if the value of ρ, the geometric rate in (9) is close to 1, then the algorithm in question, in spite of being geometrically ergodic, need not be efficient in practice. To test how efficient our TMCMC algorithms actually are in absolute terms and also relative to standard MCMC approaches, we need to define a measure of closeness of the nth order kernel $P^n(x, \cdot)$ with respect to the target density $\pi(\cdot)$, assuming that the latter can be empirically evaluated. The Kolmogorov–Smirnov (K–S) distance seems to be a suitable candidate in this regard, and the one that we adopt for our purpose. Corresponding to each MCMC algorithm, we consider N replicates of the chain starting from the same initial value, so that at each iteration t, we obtain a set of N many realizations of the chain. We then compute the empirical distribution of these N values and measure the K–S distance between the empirical distribution and the target distribution π. For the chain to be efficient, it must have K–S distance close to 0 after the chain has run for a large number of iterations (that is, when t is large). Moreover, the burn-in period is expected to be small for efficient MCMC algorithms.

4.1 First simulation experiment comparing RWMH and additive TMCMC

Table 1 presents the results of a simulation experiment comparing the performances of RWMH and additive TMCMC (Add.-TMCMC) chains for different dimensions, where, for our purpose we consider the target density π to be the multivariate normal distribution with mean vector **0** and covariance matrix **I**, the identity matrix. For RWMH, we consider two distinct scales for the normal random walk proposal for each of the co-ordinates—the optimal scale 2.4, and a sub-optimal scale 6. We consider the same scaling for additive TMCMC as well. Indeed, as shown in Dey and Bhattacharya (2013), for both additive TMCMC and RWMH, the optimal scaling parameter is very close to 2.4 but the optimal acceptance rate of additive TMCMC is around 0.439, which is significantly higher than 0.234, the optimal acceptance rate of the RWMH approach (Roberts and Tweedie (1996), Roberts, Gelman and Gilks (1997)). Moreover, the results of simulation

Table 1 *Performance evaluation of RWMH and additive TMCMC* (*Add.-TMCMC*) *chains for different dimensions*

Dim.	Scaling	Criteria			
		Acceptance rate (%)		Avg. K–S dist.	
		RWMH	Add.-TMCMC	RWMH	Add.-TMCMC
2	2.4	34.9	44.6	0.1651	0.1657
	6	18.66	29.15	0.1659	0.1655
5	2.4 (opt)	28.6	44.12	0.1659	0.1664
	6	2.77	20.20	0.1693	0.1674
10	2.4 (opt)	26.05	44.18	0.1652	0.1677
	6	1.19	20.34	0.1784	0.1688
100	2.4 (opt)	23.3	44.1	0.1594	0.1571
	6	0.32	20.6	0.1687	0.1645
200	2.4 (opt)	23.4	44.2	0.1596	0.1435
	6	0.38	20.7	0.1622	0.1484

experiments reported in Dey and Bhattacharya (2013) demonstrate superior performance of additive TMCMC over RWMH in terms of higher acceptance rates irrespective of dimensions and optimal or sub-optimal scale choices.

Referring to Table 1, since the K–S statistic is computed after burn in, the differences between additive TMCMC and RWMH in terms of the K–S distance do not appear to be pronounced in low dimensions, but for dimensions 100 and 200, the differences seem to be more pronounced, indicating somewhat better performance of TMCMC.

Figure 4 displays the K–S distances corresponding to RWMH and additive TMCMC when the target is a 30-dimensional normal distribution. It is clearly seen that additive TMCMC converges much faster than RWMH. In fact, the figures indicate that additive TMCMC takes around just 150 iterations to converge when the scale is optimal, and around 200 iterations when the scale is sub-optimal. On the other hand, in the case of optimal scaling, RWMH takes around 300 iterations to converge and for sub-optimal scaling it takes around 450 iterations. The mixing issue is quite pronounced in higher dimensions. Indeed, as seen in Figure 5, the K–S distances associated with additive TMCMC are almost uniformly smaller than those associated with RWMH, particularly when the scaling is sub-optimal. In fact, in the sub-optimal case it seems that additive TMCMC has converged within the first 2,000 iterations, whereas RWMH does not seem to show any sign of convergence even after 20,000 iterations (the K–S distances are significantly larger than those of additive TMCMC).

### 4.2 Performance comparison with "essentially fully" multiplicative TMCMC

In this case, we choose our neighborhood $\mathbb{N}_0$ in (37) to be $[-0.1, 0.1]^d$, where d is the dimension of the space. The method of estimation of the mixing probability

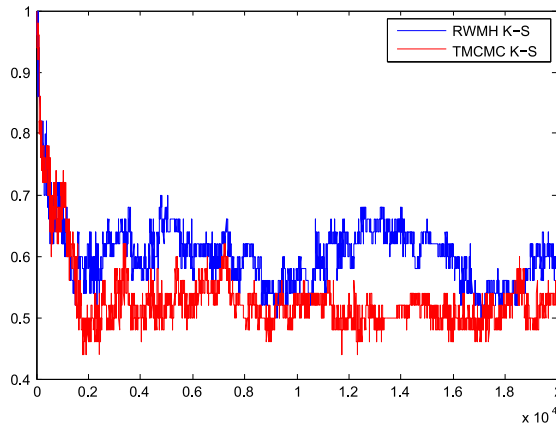(a) RWMH vs Add.-TMCMC (scale = 2.4)

(b) RWMH vs Add.-TMCMC (scale = 6)
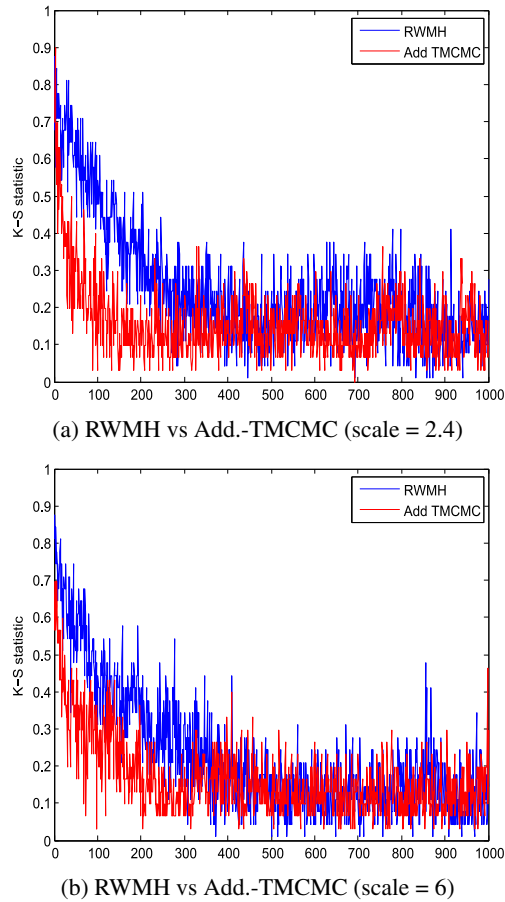
Figure 4 *Comparisons between K–S distances associated with additive TMCMC and RWMH for dimension = 30.*

$\pi(\mathbb{N}_0)$ is discussed in detail in Appendix C; however, in our simulation example, this probability is simply $\prod_{i=1}^{d}[2\Phi(0.1) - 1]$, Φ denoting the cumulative distribution function of $N(0, 1)$. This chain is basically as close as we can get to a fully multiplicative TMCMC chain on $\mathbb{R}^d$ ensuring that the geometric drift condition holds.

For our experiment, the scale of the additive TMCMC part of the mixture remains the same as before, that is, we consider the optimal scale 2.4, and the sub-optimal scale 6. We assume the proposal density $g^{(2)}$ is defined on a set of the form $[-l_2, -l_1] \cup [l_1, l_2]$ such that the interval $[l_1, l_2]$ is a proper subset of $[0, 1]$ minus small neighborhoods of 0 and 1. The distribution of the step ϵ is taken to be a mixture normal random variable such that $\epsilon \sim \frac{1}{2}N(\mu, \sigma^2)I_{[l_1, l_2]} + \frac{1}{2}N(-\mu, \sigma^2)I_{[-l_2, -l_1]}$ with mean $\mu \in [l_1, l_2]$ and variance σ^2. In our simulation experiment, we assumed $l_1 = 0.05$ and $l_2 = 0.95$ and optimal performance was
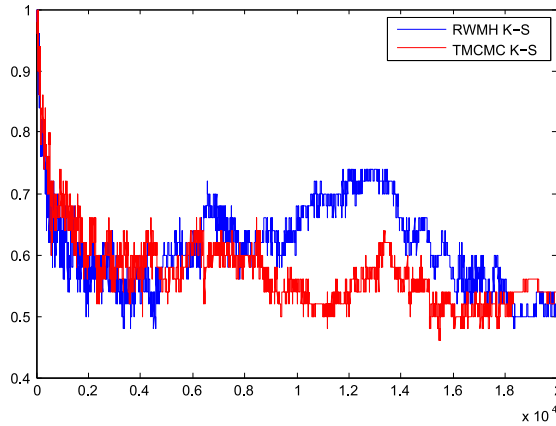
(a) RWMH vs Add.-TMCMC (scale = 2.4)

(b) RWMH vs Add.-TMCMC (scale = 6)

Figure 5 *Comparisons between K–S distances associated with additive TMCMC and RWMH for dimension* = 100.

observed when the mean μ is in the range 0.35 to 0.45, which is around halfway from both l_1 and l_2.

Table 2 provides a comparison of the performances between RWMH and essentially full multiplicative TMCMC with respect to acceptance rate and average K–S distance. Note that unlike additive TMCMC, we find here that the acceptance rate for essentially fully multiplicative TMCMC is poor compared to RWMH. Moreover, the K–S distances also suggest that RWMH is closer to the target distribution compared to essentially fully multiplicative TMCMC for most of the iterations considered. However, on inspection it is observed that the K–S distance initially drops faster for the latter compared to RWMH; see Figure 6. As shown by Dutta (2012), multiplicative TMCMC in one-dimensional situations are appropriate for certain heavy-tailed distributions. But in our current simulation study associated

Table 2 *Performance evaluation of RWMH and essentially fully multiplicative TMCMC (Mult.-TMCMC) chains for different dimensions*

		Criteria			
		Acceptance rate (%)		Avg. K–S dist.	
Dim.	Scaling	RWMH	Mult.-TMCMC	RWMH	Mult.-TMCMC
10	2.4 (opt)	26.05	16.86	0.1652	0.2097
	6	1.19	6.32	0.1784	0.2133
30	2.4 (opt)	23.5	15.74	0.1637	0.1828
	6	1.16	6.77	0.1711	0.1924
100	2.4 (opt)	23.4	15.46	0.1596	0.1812
	6	0.38	2.67	0.1622	0.1866

with high dimensions and a thin-tailed density, (essentially fully) multiplicative TMCMC did not seem to perform satisfactorily, although theoretically it is geometrically ergodic.

## 4.3 Performance comparison with the traditional mixture of additive and multiplicative TMCMC

Now we consider the traditional mixture chain of the form (38) with both additive and multiplicative moves. We assume that with probability $\frac{1}{2}$, we move by additive TMCMC and with probability $\frac{1}{2}$ by multiplicative TMCMC. The proposal mechanisms for additive and multiplicative TMCMC remain the same as in Section 4.2 associated with essentially fully multiplicative TMCMC.

Table 3 provides a comparison of the performances between RWMH and our traditional mixture TMCMC kernel with respect to acceptance rate and average K–S distance. Note that although the acceptance rate for the mixture kernel in our experiments is around 0.293 for $\mu = 0.35$ and $\sigma = 1$ which is quite low compared to additive TMCMC, it is of course still significantly higher than the optimal acceptance rate 0.234 for standard RWMH. To avoid any possible confusion it is important to emphasize that this acceptance rate for mixture kernel is not analytically derived as the optimal acceptance rate, rather it is the rate corresponding to the optimal value of μ, numerically obtained by varying over μ keeping σ fixed at 1 and computing the K–S distance and then choosing that μ for which the empirical average K–S distance was found to be the minimum. However, the average K–S distance for the mixture kernel is smaller compared to both RWMH and additive TMCMC, implying faster convergence. This improvement acts as a trade off for the low acceptance rate of the mixture kernel.

Figure 7 displays plots of K–S distances associated with RWMH and mixture TMCMC in the case of a 30-dimensional normal target distribution. The plot
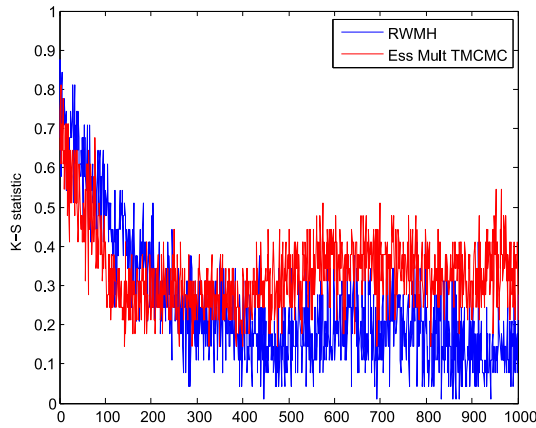
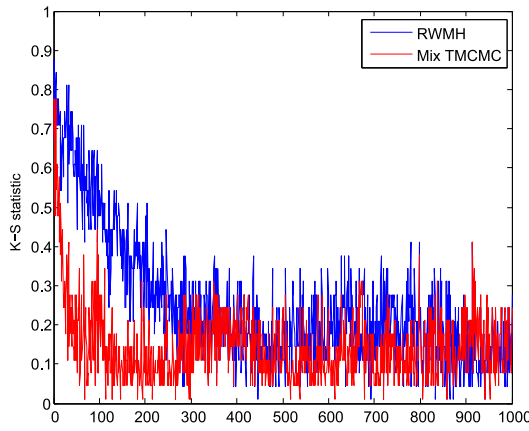(a) RWMH vs Mult.-TMCMC (RWMH scale = 2.4)
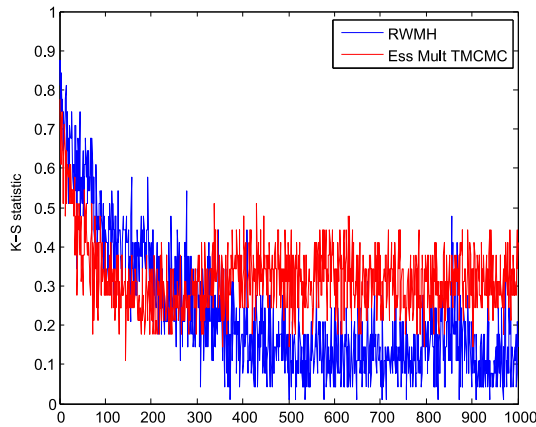
(b) RWMH vs Mult.-TMCMC (RWMH scale = 6)

Figure 6 *Comparisons between K–S distances associated with essentially Mult.-TMCMC and RWMH for dimension = 30.*

shows much faster convergence of mixture TMCMC compared to RWMH. From Figures 4 and 6, it is also clear that mixture TMCMC converges faster than even additive TMCMC and essentially fully multiplicative TMCMC. In fact, mixture TMCMC seems to converge in just about 100 iterations. This faster convergence may be attributed to the fact that the multiplicative steps allow the chain to take longer jumps and hence explore the space faster, while on the other hand the additive steps keep the acceptance rate high and enables the chain to move briskly. So, in other words, mixture TMCMC shares the positives of both the additive and the multiplicative chains and is found to outperform each of them individually.

Table 3 *Performance evaluation of RWMH and traditional Mixture TMCMC (Mix.-TMCMC) chains for different dimensions. For the multiplicative TMCMC part, we consider $\mu = 0.35$ and $\sigma = 1$*

Dim.	Scaling	Acceptance rate (%)		Avg. K–S dist.	
		RWMH	Mix.-TMCMC	RWMH	Mix.-TMCMC
10	2.4 (opt)	26.05	29.43	0.1652	0.1455
	6	1.19	11.26	0.1784	0.1576
30	2.4 (opt)	23.5	29.32	0.1637	0.1428
	6	1.16	16.33	0.1711	0.1529
100	2.4 (opt)	23.4	29.29	0.1596	0.1398
	6	0.38	10.67	0.1622	0.1412

(a) RWMH vs Mix.-TMCMC (scale = 2.4)
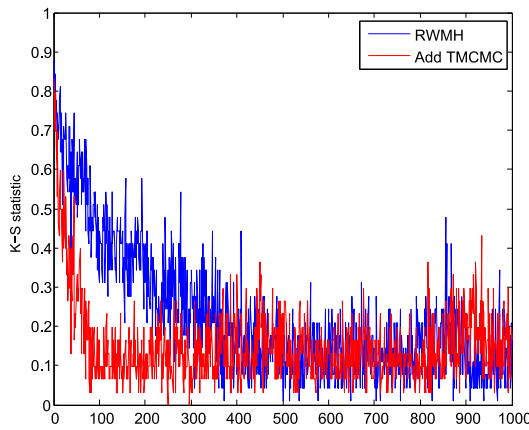
(b) RWMH vs Mix.-TMCMC (scale = 6)

Figure 7 *Comparisons between K–S distances associated with Mix.-TMCMC and RWMH for dimension = 30.*

5 Extensions of our geometric ergodicity results to target distributions that are not super-exponential

So far we have proved geometric ergodicity of additive and multiplicative TM-CMC when the target density π is super-exponential. It is natural to ask if our results go through when the super-exponential assumption does not hold.

5.1 Target density as mixture

Note that, if the target density π can be represented as a mixture of the form

$$\pi(x) = \int f_1(x|\theta) f_2(\theta) \, d\theta, \tag{59}$$

where $f_1(\cdot|\theta)$ is super-exponential for all θ and f_2 admits direct (exact) simulation, then the Markov transition kernel

$$P(x, A) = \int P(x, A|\theta) f_2(\theta) \, d\theta, \tag{60}$$

where $P(x, A|\theta)$ denotes either additive or multiplicative TMCMC-based Markov transition kernel conditional on θ, is geometrically ergodic for the target density π. The proof is essentially the same as the proof presented in Appendix B that the finite mixture Markov transition kernel (37) is geometrically ergodic for the mixture representation (35); only the summations need to be replaced with integrals. The kernel (60) will be implemented by first directly simulating $\theta \sim f_2$; then given θ, the transition mechanism $P(x, \cdot|\theta)$ has to be implemented.

Two popular examples of multivariate densities admitting mixture forms are multivariate t and multivariate Cauchy, both of which can be represented as univariate *Gamma*-distributed mixtures of multivariate normal distributions.

5.2 Change-of-variable idea

The general situation has been addressed by Johnson and Geyer (2012) using a change-of-variable idea. If π_β is the multivariate target density of interest, then one can first simulate a Markov chain having invariant density

$$\pi_\gamma(\gamma) = \pi_\beta\big(h(\gamma)\big)\big|\det \nabla h(\gamma)\big|, \tag{61}$$

where h is a diffeomorphism. If π_β is the density of the random vector β, then π_γ is the density of the random vector $\gamma = h^{-1}(\beta)$. Johnson and Geyer (2012) obtain conditions on h which make π_γ super-exponentially light. In more details, Johnson and Geyer (2012) define the following isotropic function $h : \mathbb{R}^d \mapsto \mathbb{R}^d$:

$$h(\gamma) = \begin{cases} f\big(\|\gamma\|\big)\dfrac{\gamma}{\|\gamma\|}, & \gamma \neq \mathbf{0}, \\ 0, & \gamma = \mathbf{0} \end{cases} \tag{62}$$

for some function $f : (0, \infty) \mapsto (0, \infty)$. Johnson and Geyer (2012) confine attention to isotropic diffeomorphisms, that is, to functions h where both h and h^{-1} are continuously differentiable, with the further property that $\det \nabla h$ and $\det \nabla h^{-1}$ are also continuously differentiable. In particular, they define $f : [0, \infty) \mapsto [0, \infty)$ as follows:

$$f(x) = \begin{cases} x, & x < R, \\ x + (x - R)^p, & x \geq R, \end{cases} \tag{63}$$

where $R \geq 0$ and $p > 2$.

Theorem 2 of Johnson and Geyer (2012) shows that if π_β is an exponentially light density (π_β is exponentially light if $\limsup_{\|x\| \to \infty} n(x)' \nabla \log \pi_\beta(x) < 0$) on $\mathbb{R}^d$, and h is defined by (62) and (63) then the transformed density π_γ given by (61) is super-exponentially light. Thus, this transformation transforms an exponential density to a super-exponential density. Theorem 3 of Johnson and Geyer (2012) provided conditions under which sub-exponential densities can be converted to exponential densities (π_β is sub-exponentially light if $\limsup_{\|x\| \to \infty} n(x)' \nabla \log \pi_\beta(x) = 0$). In particular, if π_β is a sub-exponentially light density on $\mathbb{R}^d$, there exist $\alpha > d$, $R < \infty$ such that

$$\left(\frac{\beta}{\|\beta\|} \right)' \nabla \log \pi_\beta(\beta) \leq -\frac{\alpha}{\|\beta\|}, \qquad \|\beta\| > R,$$

then h defined as (62) with $f : [0, \infty) \mapsto [0, \infty)$ given by

$$f(x) = \begin{cases} e^{bx} - \dfrac{e}{3}, & x > \dfrac{1}{b}, \\ x^3 \dfrac{b^3 e}{6} + x \dfrac{be}{2}, & x \leq \dfrac{1}{b}, \end{cases} \tag{64}$$

where $b > 0$, ensures that the transformed density π_γ of the form (61), is super-exponentially light.

In other words, starting from a sub-exponential target density, one can achieve a super-exponential density by first converting it to exponential using the transformation h (given by (62)) with f given by (64). Then one can convert the obtained exponential density to super-exponential using the transformation h and f given by (63). As an example, Johnson and Geyer (2012) show that the multivariate t distribution of the form

$$\pi_\beta(t) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{d/2} \det(\Sigma)} \left[1 + \frac{1}{\nu}(t - \mu)' \Sigma^{-1}(t - \mu) \right]^{-((\nu+d)/2)}, \tag{65}$$

is sub-exponential. This can be converted to super-exponential by applying the aforementioned transformations in succession.

Hence, we can run our geometric ergodic TMCMC algorithms for the super-exponentially light π_γ, and then transform the realizations $\{\gamma^{(k)}; k = 1, 2, \ldots\}$ to $\{h(\gamma^{(k)}); k = 1, 2, \ldots\}$. Then it easily follows (see Appendix A of Johnson and Geyer (2012)) that the transformed chain is also geometrically ergodic.

5.2.1 *Simulation studies comparing RWMH and additive TMCMC in the context of diffeomeophism based simulation from Cauchy and t-distributions.* We now compare diffeomorphism-based RWMH and additive TMCMC algorithms with respect to K–S distance, when the target distributions are d-dimensional Cauchy and t-distributions, the latter having ν degrees of freedom. We assume that the location vectors and scale matrices are $\boldsymbol{\mu} = \mathbf{0}_d$ and $\boldsymbol{\Sigma} = \text{diag}\{0.7\mathbf{1}'_d\} + 0.3\mathbf{1}_d\mathbf{1}'_d$, respectively, where $\mathbf{0}_d$ is a d-dimensional vector with all elements 0, and $\mathbf{1}_d$ is a d-dimensional vector with each component 1. We choose $d = 50$ for the illustrations. For both RWMH and additive TMCMC, we consider the scale of the proposal distribution to be 2.4.

Figure 8 compares the performances of diffeomorphism based RWMH and diffeomorphism based Add.-TMCMC with respect to the K–S distance when the target distributions are 50-variate Cauchy and 50-variate t respectively, with the aforementioned location vector and scale matrix. In both the cases, Add.-TMCMC quite significantly outperforms RWMH. Hence, the results are highly encouraging—additive TMCMC significantly outperforms RWMH when the high-dimensional target density is not super-exponential, and is highly dependent. Since a mixture of additive and multiplicative TMCMC is demonstrably more efficient than additive TMCMC, it is clear that the mixture will beat RWMH by a large margin. We have also carried out extensive simulation studies comparing RWMH and Add.-TMCMC when the target distributions are 50-dimensional i.i.d. Cauchy and 50-dimensional i.i.d. t with 10 degrees of freedom, that is, with $\boldsymbol{\mu} = \mathbf{0}_d$ and $\boldsymbol{\Sigma} = \mathbf{I_d}$, the latter standing for the identity matrix or order d, with $d = 50$. We do not present the results here due to lack of space, but Add.-TMCMC outperformed RWMH at least as significantly as in this reported dependent set-up.

As an aside, we also compare the gains of the diffeomorphism based approach over the usual, direct application of RWMH and TMCMC to the target densities. Figure 9 compares the performances of diffeomorphism based RWMH and direct RWMH when the targets are the above-defined 50-dimensional multivariate Cauchy and t (with 10 degrees of freedom). Likewise, Figure 10 compares the performances of diffeomorphism based Add.-TMCMC and direct Add.-TMCMC with the above 50-dimensional target densities. As is evident from the figures, the diffeomorphism based approaches quite significantly outperform the direct approaches.

6 Concluding remarks

We presented a comprehensive comparative study of geometric ergodicity and convergence behavior of various versions of TMCMC: additive, "essentially full" multiplicative and mixture TMCMC. Additive TMCMC is the easiest to implement and as observed in the simulation study, has somewhat better convergence to the target distribution compared to RWMH. The essentially fully multiplicative
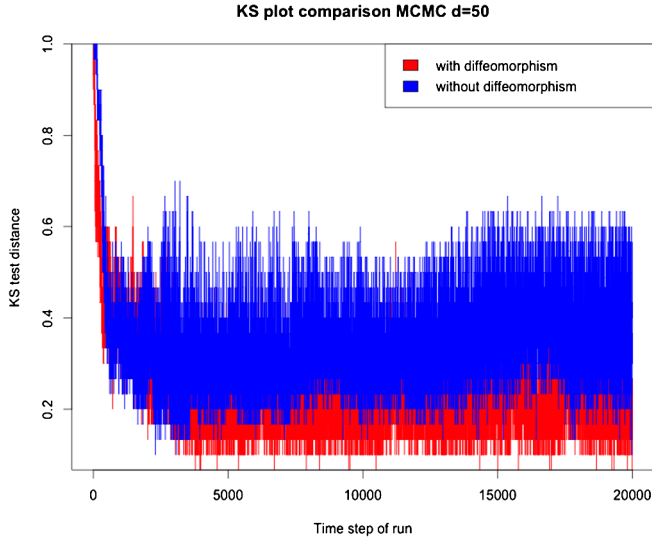
KS plot comparison
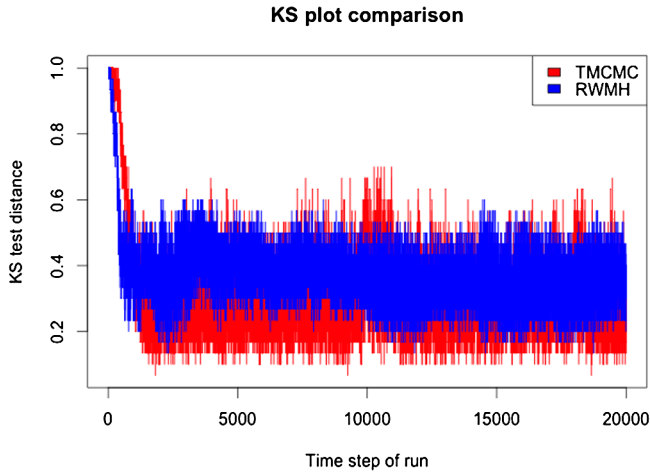
(a) Multivariate Cauchy Target: Diffeomorphism based RWMH vs diffeomorphism based Add.-TMCMC (scale = 2.4)
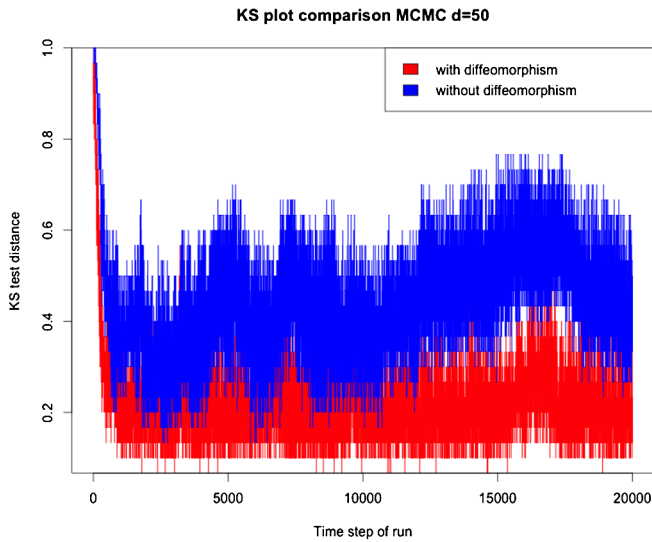
KS plot comparison
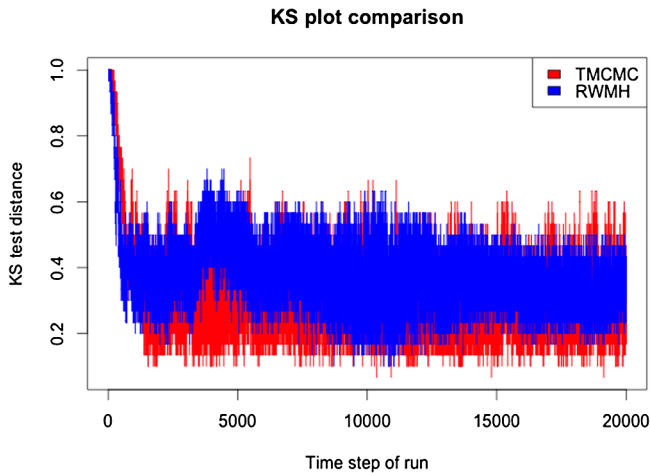
(b) Multivariate-*t* Target: Diffeomorphism based RWMH vs diffeomorphism based Add.-TMCMC (scale = 2.4)

Figure 8   50-*dimensional Cauchy and multivariate-t* (10 *degrees of freedom*) *targets*: *Comparisons between K–S distances associated with diffeomorphism based additive TMCMC and diffeomorphism based RWMH.*

TMCMC traverses the sample space more rapidly but we observed that it is relatively slow in convergence to the target density compared to the standard RWMH approach. The best convergence results are obtained for mixture TMCMC which combines the additive and the multiplicative moves in equal proportions.
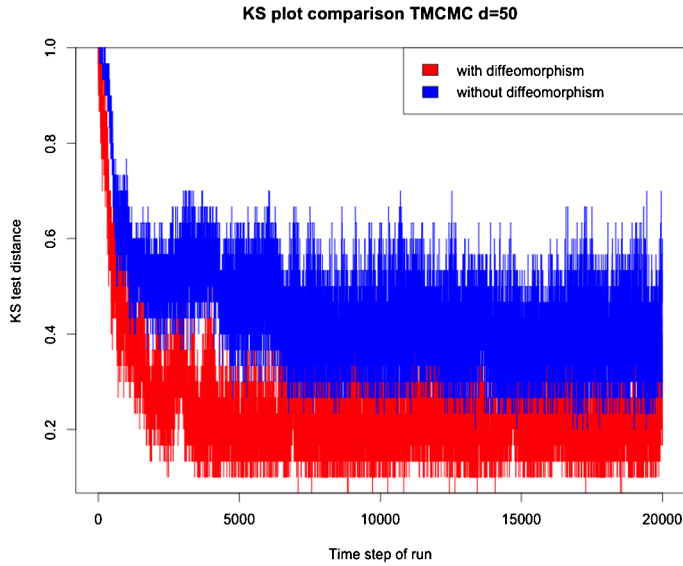
KS plot comparison MCMC d=50

(a) Multivariate Cauchy Target: RWMH with and without diffeomorphism (scale = 2.4)
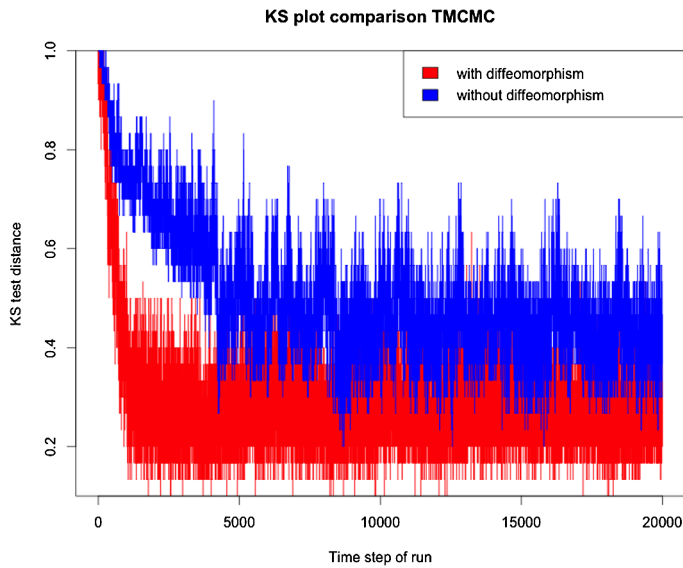
KS plot comparison MCMC d=50

(b) Multivariate-*t* Target: RWMH with and without diffeomorphism (scale = 2.4)

Figure 9 *50-dimensional Cauchy and multivariate-t* (10 *degrees of freedom*) *targets*: *Comparisons between K–S distances associated with RWMH implemented with and without diffeomorphism.*

Of considerable interest are situations when the high-dimensional target densities are not super-exponential but can be handled by the diffeomorphism based approach. The relevant simulation studies detailed in Section 5.2.1 demonstrate far superior convergence of additive TMCMC compared to RWMH. Since these simulation studies are conducted assuming high dependence structure of the tar-

KS plot comparison TMCMC d=50

(a) Multivariate Cauchy Target: Add.-TMCMC with and without diffeomorphism (scale = 2.4)

KS plot comparison TMCMC

(b) Multivariate-*t* Target: Add.-TMCMC with and without diffeomorphism (scale = 2.4)

Figure 10   50-*dimensional Cauchy and multivariate-t* (10 *degrees of freedom*) *targets*: *Comparisons between K–S distances associated with Add.-TMCMC implemented with and without diffeomorphism.*

get densities, the results are particularly encouraging and lead us to recommend TMCMC in general situations. Moreover, it is to be noted that in these simulation

studies we concern ourselves with only additive TMCMC. Since a mixture of additive and multiplicative TMCMC is seen to be more efficient in comparison with additive TMCMC, it is clear that such a mixture will outperform RWMH by even greater margins.

There are obviously some questions of further interest. We would definitely like to have quantitative rates of convergence for each of the three approaches to TMCMC. In this paper we considered the mixing proportion in mixture TMCMC to be 1/2 and we also observed in our simulation study that extremal mixing proportions (which correspond to additive and essentially fully multiplicative approaches) lead to slower convergence compared to uniform mixing. But it would be worth noting how this rate of convergence changes with the change in mixing proportion.

Optimal scaling of TMCMC methods is another area which is of considerable interest to us. The optimal scaling for additive TMCMC has been studied for a broad class of multivariate target densities (Dey and Bhattacharya (2013)), but the optimal scaling for mixture TMCMC and multiplicative or essentially fully multiplicative approaches are yet to be determined. The biggest challenge in dealing with this problem is that the generator functions for the associated time scaled diffusion process for these methods are hard to express in any simple analytic form.

One area we are currently focusing on is defining adaptive versions of the TMCMC approach (additive and multiplicative) and comparing the performances (convergence criterion and acceptance rate in particular) among various adaptive schemes and also with the typical non adaptive algorithms we considered here.

We are also trying to expand the scope of our approach beyond $\mathbb{R}^d$ by considering spheres and other Riemannian or Symplectic manifolds as the support of the target distributions and it would be interesting to investigate such properties like irreducibility, detailed balance and ergodicity properties of the TMCMC algorithms over such spaces.

Appendix A: Minorization condition for multiplicative TMCMC

For the one-dimensional case, minorization conditions of multiplicative TMCMC has been established by Dutta (2012). Here we generalize the results to arbitrary dimension. For simplicity we assume $p_i = q_i = 1/3$ for $i = 1, \ldots, d$. The following theorem establishes the minorization condition for multiplicative TMCMC.

Theorem A.1. *Let the target density π be bounded and positive on compact sets. Then there exists a nonzero measure ν, a positive integer m, $\delta > 0$, and a small set E^* such that*

$$\{P^{(2)}\}^m(x, \mathbb{A}) \geq \delta \nu(\mathbb{A}), \qquad \forall x \in E^* \text{ and for all Borel sets } \mathbb{A}. \qquad (66)$$

Proof. Observe that, from $x = (x_1, \ldots, x_d)$ it is possible to move to any Borel set $\mathbb{A}$ in at least d steps using those multiplicative TMCMC move types $b =$

$(b_1, \ldots, b_d)$ which update only one co-ordinate at a time. Hence, for our purpose it is sufficient to confine attention to these moves.

Let E^* denote a compact subset of $\mathbb{R}^d$. Also, let $\mathbb{C}$ be a compact set containing E^*. Let $\mathbb{A}^* = \mathbb{A} \cap \mathbb{C}$. For the simplicity of presentation we present the proof of minorization for $d = 2$.

Let $\mathbb{A}_1 = \{(\epsilon_1, \epsilon_2) : (x_1\epsilon_1, x_2\epsilon_2) \in \mathbb{A}^*\}$, $\mathbb{A}_2 = \{(\epsilon_1, \epsilon_2) : (x_1/\epsilon_1, x_2/\epsilon_2) \in \mathbb{A}^*\}$, $\mathbb{A}_3 = \{(\epsilon_1, \epsilon_2) : (x_1\epsilon_1, x_2/\epsilon_2) \in \mathbb{A}^*\}$, and $\mathbb{A}_4 = \{(\epsilon_1, \epsilon_2) : (x_1/\epsilon_1, x_2\epsilon_2) \in \mathbb{A}^*\}$.

For $x \in E^*$, we have

$$
\begin{aligned}
&\{P^{(2)}\}^2(x, \mathbb{A}) \\
&\geq \{P^{(2)}\}^2(x, \mathbb{A}^*) \\
&\geq \frac{1}{3^4} \int_{\mathbb{A}_1} \min\left\{1, \frac{\pi(x_1\epsilon_1, x_2)|\epsilon_1|}{\pi(x_1, x_2)}\right\} \\
&\quad \times \min\left\{1, \frac{\pi(x_1\epsilon_1, x_2\epsilon_2)|\epsilon_2|}{\pi(x_1\epsilon_1, x_2)}\right\} g^{(2)}(\epsilon_1) g^{(2)}(\epsilon_2)\, d\epsilon_1\, d\epsilon_2 \\
&\quad + \frac{1}{3^4} \int_{\mathbb{A}_2} \min\left\{1, \frac{\pi(x_1/\epsilon_1, x_2)|\epsilon|^{-1}}{\pi(x_1, x_2)}\right\} \\
&\quad \times \min\left\{1, \frac{\pi(x_1/\epsilon_1, x_2/\epsilon_2)|\epsilon_2|^{-1}}{\pi(x_1/\epsilon_1, x_2)}\right\} g^{(2)}(\epsilon_1) g^{(2)}(\epsilon_2)\, d\epsilon_1\, d\epsilon_2 \\
&\quad + \frac{1}{3^4} \int_{\mathbb{A}_3} \min\left\{1, \frac{\pi(x_1\epsilon_1, x_2)|\epsilon_1|}{\pi(x_1, x_2)}\right\} \\
&\quad \times \min\left\{1, \frac{\pi(x_1\epsilon_1, x_2/\epsilon_2)|\epsilon_2|^{-1}}{\pi(x_1\epsilon_1, x_2)}\right\} g^{(2)}(\epsilon_1) g^{(2)}(\epsilon_2)\, d\epsilon_1\, d\epsilon_2 \\
&\quad + \frac{1}{3^4} \int_{\mathbb{A}_4} \min\left\{1, \frac{\pi(x_1/\epsilon_1, x_2)|\epsilon_1|^{-1}}{\pi(x_1, x_2)}\right\} \\
&\quad \times \min\left\{1, \frac{\pi(x_1/\epsilon_1, x_2\epsilon_2)|\epsilon_2|}{\pi(x_1/\epsilon_1, x_2)}\right\} g^{(2)}(\epsilon_1) g^{(2)}(\epsilon_2)\, d\epsilon_1\, d\epsilon_2.
\end{aligned} \tag{67}
$$

Let $r = \inf_{y \in \mathbb{C}} \pi(y)$ and $R = \sup_{y \in \mathbb{C}} \pi(y)$. Also note that each integral on $\mathbb{A}_i$; $i = 1, 2, 3, 4$, can be split into $\mathbb{A}_i = \{\mathbb{A}_i \cap \mathbb{S}_\eta\} \cup \{\mathbb{A}_i \cap \mathbb{S}_\eta^c\}$, where $\mathbb{S}_\eta = \{(\epsilon_1, \epsilon_2) : \eta < |\epsilon_1| \leq 1, \eta < |\epsilon_2| \leq 1\}$, for some $\eta > 0$, and $\mathbb{S}_\eta^c$ denotes the complement of $\mathbb{S}_\eta$. Let G denote the probability measure corresponding to the distribution $\epsilon_1, \epsilon_2 \overset{\text{i.i.d.}}{\sim} g^{(2)}$.

On $\mathbb{A}_i \cap \mathbb{S}_\eta^c$, for $i = 1, 3, 4$, the corresponding integrands have infimum zero; hence zero is the lower bound of the respective integrals on $\mathbb{A}_i \cap \mathbb{S}_\eta^c$, for $i = 1, 3, 4$. On $\mathbb{A}_2 \cap \mathbb{S}_\eta^c$, the integrand of the second integral has infimum equal to 1; hence, the corresponding integral is bounded below by $\frac{1}{3^4} G(\mathbb{A}_2 \cap \mathbb{S}_\eta^c)$. Note that $G(\mathbb{A}_2 \cap \mathbb{S}_\eta^c)$ can be made arbitrarily small by choosing η to be as small as desired.

On $\mathbb{A}_i \cap \mathbb{S}_\eta$, each of the integrals are bounded below by $\frac{\eta^2}{3^4}(\frac{r}{R})^2 G(\mathbb{A}_i \cap \mathbb{S}_\eta)$. Hence,

$$
\begin{aligned}
\{P^{(2)}\}^2(x, \mathbb{A}) &\geq \{P^{(2)}\}^2(x, \mathbb{A}^*) \\
&\geq \frac{\eta^2}{3^4}\left(\frac{r}{R}\right)^2 \sum_{i=1}^{4} G(\mathbb{A}_i \cap \mathbb{S}_\eta) \\
&\geq \frac{\eta^2}{3^4}\left(\frac{r}{R}\right)^2 G\left(\left\{\bigcup_{i=1}^{4} \mathbb{A}_i\right\} \cap \mathbb{S}_\eta\right) \\
&= \frac{\eta^2}{3^4}\left(\frac{r}{R}\right)^2 G(\mathbb{A}^* \cap \mathbb{S}_\eta). \\
&= \frac{\eta^2}{3^4}\left(\frac{r}{R}\right)^2 G(\mathbb{S}_\eta) \times \frac{G(\mathbb{A}^* \cap \mathbb{S}_\eta)}{G(\mathbb{S}_\eta)} \\
&= \delta \nu(\mathbb{A}^*),
\end{aligned}
\tag{68}
$$

with

$$
\delta = \frac{\eta^2}{3^4}\left(\frac{r}{R}\right)^2 G(\mathbb{S}_\eta) \quad \text{and} \quad \nu(\mathbb{A}^*) = \frac{G(\mathbb{A}^* \cap \mathbb{S}_\eta)}{G(\mathbb{S}_\eta)}.
$$

Hence, minorization holds for multiplicative TMCMC, and E^* is the small set. The same ideas of the proof go through for any finite dimension d. $\qquad\square$

We next show that vectors in the set

$$
\mathcal{V} = \{(v_1, \ldots, v_d) \in \mathbb{R}^d : v_i = 0 \text{ for at least one } i \in \{1, \ldots, d\}\},
$$

cannot be limit points of small sets. For our purpose, we need a lemma which can be seen as a generalization of Lemma 1 of Dutta (2012) to arbitrary dimensions and for vectors in $\mathcal{V}$.

Lemma A.1. *Fix $v = (v_1, \ldots, v_d) \in \mathcal{V}$. For $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, d\}$, where $k \leq d$, let $v_{i_j} = 0$, for $j = 1, \ldots, k$. Let $\{x_n\}$ be a sequence of positive (negative) numbers decreasing (increasing) to zero. Consider the sequence $v_n = (v_{1,n}, \ldots, v_{d,n})'$, where $v_{j,n} = x_n$ for $j = i_1, \ldots, i_k$, and $v_{j,n} = v_j$ for $j \in \{1, \ldots, d\}\backslash\{i_1, \ldots, i_k\}$. If $v_i = 0$ for $i = 1, \ldots, d$, then $v_n = (x_n, \ldots, x_n)'$ may also be considered. Then,*

$$
P^{(2)}(v_n, \mathbb{A}) \to 0,
\tag{69}
$$

for all Borel sets $\mathbb{A}$ such that $\mathbb{A} \cap \{(v_1, \ldots, v_d) \in \mathbb{R}^d : v_{i_j} = 0; j = 1, \ldots, k\} = \emptyset$.

Proof. Without loss of generality, we present the proof for $d = 2$. Let us fix $v = (v_1, v_2)$, where $v_1 = 0$ and $v_2 \in \mathbb{R}$. Let $v_n = (x_n, v_2)$. Note that for moving from

x_n to $z \in \mathbb{R}$, where $|x_n| \leq |z|$ for all n, we must simulate $\epsilon = x_n/z$ and take the backward move $z = x_n/\epsilon$. The move $z = x_n\epsilon$, with $\epsilon = z/x_n$ cannot be valid in this case, since $x_n \to 0$ implies that for large n, $\epsilon \notin [-1, 1]$.

Since the acceptance probability is bounded above by 1, we have, for $y < 0$,

$$P^{(2)}(v_n, (-\infty, y] \times (-\infty, \infty)) \leq \frac{1}{3^2} \int_{x_n/y}^0 g(\epsilon) \, d\epsilon$$
$$\to 0. \tag{70}$$

If $y > 0$, then

$$P^{(2)}(v_n, [y, \infty) \times (-\infty, \infty)) \leq \frac{1}{3^2} \int_0^{x_n/y} g(\epsilon) \, d\epsilon$$
$$\to 0. \tag{71}$$

Hence, (69) holds when $d = 2$. The proof clearly goes through for any dimension d.

If $v = (0, 0)$, we can consider $v_n = (x_n, 0)'$ or $v_n = (x_n, x_n)'$. Then, in addition to (70) and (71), which clearly hold, the following also hold true: if $y < 0$

$$P^{(2)}(v_n, (-\infty, \infty) \times (-\infty, y]) \to 0,$$

and

$$P^{(2)}(v_n, (-\infty, \infty) \times [y, \infty)) \to 0,$$

if $y > 0$. These imply that for dimension $d = 2$,

$$P^{(2)}(v_n, \cdot) \to I_{\{0\}}(\cdot). \tag{72}$$

The above result (72) clearly holds for any dimension d for $v = (0, 0, \ldots, 0)'$ and $v_n = x_n \mathbf{1}$, where $\mathbf{1} = (1, 1, \ldots, 1)'$ is the d-component vector of ones. □

Now, if $v \in \mathcal{V}$ is a limit point of E^*, then there exists a sequence v_n as in Lemma A.1, converging to v. This, and Lemma A.1 imply that for any fixed integer $m > 1$, and for any Borel set $\mathbb{A}$,

$$\{P^{(2)}\}^m(v_n, \mathbb{A}) = \int_{\mathbb{R}^d} \{P^{(2)}\}^{m-1}(z, \mathbb{A}) P^{(2)}(v_n, dz)$$
$$\to 0, \tag{73}$$

if $\mathbb{A} \cap \{(v_1, \ldots, v_d) \in \mathbb{R}^d : v_{i_j} = 0; \, j = 1, \ldots, k\} = \emptyset$.

In particular, if $\mathbf{0}$ is a limit point of E, then for any fixed integer $m > 1$, and for any Borel set $\mathbb{A}$,

$$\{P^{(2)}\}^m(x_n \mathbf{1}, \mathbb{A}) \to I_{\{0\}}(\mathbb{A}). \tag{74}$$

Both (73) and (74) contradict the minorization inequality (66).

Now consider the case of additive–multiplicative TMCMC. Let the coordinates with indices $\{j_1, j_2, \ldots, j_\ell\} \subset \{1, 2, \ldots, d\}$ be given the multiplicative transformation and let the remaining co-ordinates be given the additive transformation. Here, let $\mathcal{V}(j_1, \ldots, j_\ell) = \{(v_1, \ldots, v_d)' \in \mathbb{R}^d : v_j = 0 \text{ for at least one } j \in \{j_1, j_2, \ldots, j_\ell\}\}$. Then vectors $v \in \mathcal{V}(j_1, \ldots, j_\ell)$ cannot be limit points of small sets associated with additive–multiplicative TMCMC. In particular, $\mathbf{0}$ cannot be a limit point. The proof is the same as in the case of multiplicative TMCMC, and hence omitted.

Appendix B: Proof of geometric ergodicity of the Markov transition kernel $P = \pi(\mathbb{N}_0) P^{(1)} + \pi(\mathbb{N}_0^c) P^{(2)}$

Let us first introduce an auxiliary random variable Z, with

$$\Pr(Z = 1) = \pi(\mathbb{N}_0) \quad \text{and} \quad \Pr(Z = 2) = 1 - \Pr(Z = 1). \tag{75}$$

Note that for $i = 1, 2$,

$$P(x, A | Z = i) = P^{(i)}(x, A) \quad \text{and} \quad \pi(A | Z = i) = \pi_i. \tag{76}$$

Also note that, since $P^{(i)}$ is geometrically ergodic when the target density is π_i, we must have

$$\left\| \{P^{(i)}\}^n(x, \cdot) - \pi_i(\cdot) \right\| \le M_i(x) \rho_i^n, \tag{77}$$

for $i = 1, 2$, for some $M_1(x), M_2(x) < \infty$ and $0 < \rho_1, \rho_2 < 1$.

Now,

$$\left\| P^n(x, \cdot) - \pi(\cdot) \right\|_{\text{TV}}$$

$$= \sup_{A \in \mathcal{B}(\mathbb{R}^d)} \left| P^n(x, A) - \pi(A) \right|$$

$$= \sup_{A \in \mathcal{B}(\mathbb{R}^d)} \left| P^n(x, A | Z = 1) \Pr(Z = 1) + P^n(x, A | Z = 2) \Pr(Z = 2) \right.$$

$$\left. - \left(\pi(A | Z = 1) \Pr(Z = 1) + \pi(A | Z = 2) \Pr(Z = 2) \right) \right|$$

$$= \sup_{A \in \mathcal{B}(\mathbb{R}^d)} \left| \{P^{(1)}\}^n(x, A) \Pr(Z = 1) + \{P^{(2)}\}^n(x, A) \Pr(Z = 2) \right.$$

$$\left. - \left(\pi_1(A) \Pr(Z = 1) + \pi_2(A) \Pr(Z = 2) \right) \right| \qquad \text{by (76)}$$

$$\le \Pr(Z = 1) \left\| \{P^{(1)}\}^n(x, \cdot) - \pi_1(\cdot) \right\| + \Pr(Z = 2) \left\| \{P^{(2)}\}^n(x, \cdot) - \pi_2(\cdot) \right\|$$

$$\le \Pr(Z = 1) M_1(x) \rho_1^n + \Pr(Z = 2) M_2(x) \rho_2^n \qquad \text{by (77)}$$

$$\le M(x) \rho^n,$$

where $M(x) \ge \max\{M_1(x), M_2(x)\}$, and $\rho \ge \max\{\rho_1, \rho_2\}$. Hence, P is geometrically ergodic when the target density is π.

Note that the proof employed in Section 3.2 for showing geometric ergodicity of the alternative mixture Markov transition kernel P^*, is also valid for showing geometric ergodicity of P, but the current proof (with slight modification; replacing the summations with integrations) is appropriate for proving geometric ergodicity of continuous mixture kernels of the form (60) for continuous mixture target densities of the form (59) since a single function V need not be appropriate for (uncountably) infinite number of mixture components.

Appendix C: Discussion on estimation of the mixing probability $\pi(\mathbb{N}_0)$

In order to implement the Markov transition kernel P, for each $k = 1, 2, \ldots$, we are required to draw $u \sim U(0, 1)$; if $u < \pi(\mathbb{N}_0)$, we select $x_1^{(k)}$, else we select $x_2^{(k)}$. Note that $\pi(\mathbb{N}_0)$ is not known, and needs to be estimated numerically. Direct estimation using TMCMC samples from π will generally not be reliable, since the region $\mathbb{N}_0$, being arbitrarily small, can be easily missed by any MCMC method. However, this may be reliably estimated using importance sampling as follows.

Let $\pi(x) = c\ell(x)$, where $c = 1/\int \ell(y)\, dy$ is the unknown normalizing constant. Also, let $h(x) = |\mathbb{N}_0|^{-1} I_{\mathbb{N}_0}(x)$ be the uniform distribution on $\mathbb{N}_0$, where $|\mathbb{N}_0|$ denotes the Lebesgue measure of the set $\mathbb{N}_0$. We may use h as the importance sampling density in the region $\mathbb{N}_0$. For the region $\mathbb{N}_0^c$ we may consider some thick-tailed importance sampling density $g(x)$, for example, a d-variate t-density, but adjusting the support to be $\mathbb{N}_0^c$. Then

$$\pi(\mathbb{N}_0) = \frac{\int_{\mathbb{N}_0} \ell(x)\, dx}{\int \ell(x)\, dx} = \frac{\int_{\mathbb{N}_0} \frac{\ell(x)}{h(x)} h(x)\, dx}{\int_{\mathbb{N}_0} \frac{\ell(x)}{h(x)} h(x)\, dx + \int_{\mathbb{N}_0^c} \frac{\ell(x)}{g(x)} g(x)\, dx}$$

$$\approx \frac{\frac{1}{N_1} \sum_{j=1}^{N_1} \frac{\ell(x^{(j)})}{h(x^{(j)})}}{\frac{1}{N_1} \sum_{j=1}^{N_1} \frac{\ell(x^{(j)})}{h(x^{(j)})} + \frac{1}{N_2} \sum_{k=1}^{N_2} \frac{\ell(y^{(k)})}{g(y^{(k)})}} = \hat{\pi}(\mathbb{N}_0) \qquad \text{(say)},$$

where $\{x^{(j)}; j = 1, \ldots, N_1\}$ are i.i.d. realizations drawn from the uniform distribution h and $\{y^{(k)}; k = 1, \ldots, N_2\}$ are i.i.d. or TMCMC realizations from g, depending on the complexity of the form of g. The parameters of g may be chosen by variational methods; see http://www.gatsby.ucl.ac.uk/vbayes/ for a vast repository of papers, softwares and links on variational methods.

Observe that even though we are proposing to estimate $\pi(\mathbb{N}_0)$ by $\hat{\pi}(\mathbb{N}_0)$, implementation of the mixture kernel P with $\hat{\pi}(\mathbb{N}_0)$ as the mixing probability is expected to be exactly the same as the mixture kernel P with the true mixing probability $\pi(\mathbb{N}_0)$. This is because even if $\hat{\pi}(\mathbb{N}_0)$ is only a reasonably accurate estimate of $\pi(\mathbb{N}_0)$, it is expected that for any $u \sim U(0, 1)$, $u < \pi(\mathbb{N}_0)$ if and only if $u < \hat{\pi}(\mathbb{N}_0)$. For instance, if $\hat{\pi}(\mathbb{N}_0) = \pi(\mathbb{N}_0) + \eta$, for some $\eta > 0$, then

$\Pr(\pi(\mathbb{N}_0) < u < \hat{\pi}(\mathbb{N}_0)) = \eta$. Even if η is not extremely small, the above probability is still reasonably small, for reasonably small values of η. In other words, a very high degree of accuracy of the estimate $\hat{\pi}(\mathbb{N}_0)$ is not that important in this case.

Acknowledgments

References

Dey, K. K. and Bhattacharya, S. (2016). A brief tutorial on Transformation based Markov Chain Monte Carlo and optimal scaling of the additive transformation. *Brazilian Journal of Probability and Statistics*. To appear. Available at arXiv:1307.1446.

Dutta, S. (2012). Multiplicative random walk Metropolis–Hastings on the real line. *Sankhya B* **74**, 315–342. MR3046902

Dutta, S. and Bhattacharya, S. (2014). Markov Chain Monte Carlo based on deterministic transformations. *Statistical Methodology* **16**, 100–116. Also available at arXiv:1106.5850. Supplement available at arXiv:1306.6684. MR3110892

Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and their Applications* **85**, 341–361. MR1731030

Jarner, S. F. and Roberts, G. O. (2002). Polynomial convergence rates of Markov chains. *Annals of Applied Probability* **12**, 224–247. MR1890063

Jarner, S. F. and Roberts, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scandinavian Journal of Statistics* **34**, 781–815. MR2396939

Johnson, L. T. and Geyer, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *The Annals of Statistics* **40**, 3050–3076. MR3097969

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science* **16**, 312–334. MR1888447

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics* **24**, 101–121. MR1389882

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. London: Springer. MR1287609

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

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**, 110–120. MR1428751

Department of Statistics
University of Chicago
Chicago, Illinois 60637
USA
E-mail: kdey@galton.uchicago.edu

Indian Statistical Institute
203 B. T. Road
Kolkata
India
E-mail: bhsourabh@gmail.com