Submitted to Bernoulli

Supplement to "An adaptive multiple-try Metropolis algorithm"

SIMON FONTAINE[1,2,*] and MYLÈNE BÉDARD[1]

[1] *Université de Montréal, Département de mathématiques et de statistique, 2920 chemin de la Tour, Montréal, QC, Canada, H3T 1J4.*

[2] *University of Michigan, Department of Statistics. West Hall, 1085 South University, Ann Arbor, MI, U.S.A., 48109. E-mail:* [*] simfont@umich.edu

Note. References to equations, figures and tables from the main text use the numbering from the main text. Equations, figures and tables in this supplementary material prepend an "S" to their respective numbering.

1. Additional details on aMTM variants

1.1. Sampling schemes

Table S1 contains an algorithmic description of different types of candidates used in the aMTM algorithm. The derivations for the EA candidates can be found in Fontaine (2019, Section 5.3.4.2); those for the RQMC candidates can be found in Fontaine (2019, Example 4.3). Independent and common random variable candidates have trivial formulations.

1.2. Update variations

Local updates A notable difference between (ASW)AM and RAM updates is the use of the running mean $\mu_n^{(k)}$. While the (ASW)AM update uses the difference between the new point and the current estimate for the mean $(x_{n+1} - \mu_n^{(k)})$ to update the covariance, the RAM update rather uses the proposed step $(y - x_n)$. The latter seems more appropriate to locally adjust proposal densities to the target distribution. Indeed, using a running mean will potentially produce marginal covariances that are all similar to one another; using the proposed step may prevent this uniformization. We thus propose to modify the (ASW)AM updates in (3.2) by making *local updates* in which $x_{n+1} - \mu_n^{(k)}$ is replaced by $y - x_n$. In that case, the running mean update (3.1)is no longer required.

Up to this point, the only marginal covariance updated in a given iteration is that of the selected candidate. We now propose two adaptation schemes imposing some conditions on the other marginal covariances.

Global proposal We propose to consider the first proposal density ($k = 1$) as a *global* one. Its marginal covariance is thus adapted at each iteration—using any of the three update rules—no matter if the candidate was generated from this proposal or not. Then, we expect that marginal covariance to approach the target's global covariance, while the covariance matrices of the other densities should explore more local properties of the target density. This approach is particularly well-suited to multimodal densities as the global density provides a way to jump between modes, while other densities propose jumps within specific modes. Computationally, adapting a second covariance at every iteration doubles the adaptation cost.

Scale adaptation In the ASWAM case, we propose to adapt the scale parameter $\lambda_n^{(k)}$ of densities that are not selected very often; indeed, these densities are rarely adapted and may therefore never recover from a bad initialization. Given a target floor selection rate $s_* \in [0, 1]$, we decrease the scale parameter whenever a proposal density's selection rate drops below s_*. Indeed, for importance weights (2.4) or weights proportional to the target (2.5), the fact of being selected too rarely is generally related to the scale being too large.

Type	**Candidates (Step (a))**	**Shadow points (Step (d))**
Independent	Sample $Y^{(j)} \sim q_\theta^{(j)}(\cdot\|x_n)$ independently for $j = 1, \ldots, K$.	Sample $Y^{(j)} \sim q_\theta^{(j)}(\cdot\|y)$ independently for $j \neq k$.
EA $\rho = \frac{-1}{K-1}$	Beforehand, compute the singular value decomposition (SVD) $\Psi_K = (X\Lambda^{1/2})(X\Lambda^{1/2})^\top$, where $\Psi_K = \rho I_K \otimes I_d + (1-\rho) I_{dK}$. For $k = 1, \ldots, K$: – Sample $Z^{(k)} \overset{\text{iid}}{\sim} \mathcal{N}_d(\mathbf{0}_d, I_d)$, – Compute $u^{(k)} = X\Lambda^{1/2} z^{(k)}$, – Compute $y^{(k)} = x_n + S^{(k)} u^{(k)}$.	Beforehand, compute the SVD $\Phi_{K-1} = (X'\Lambda'^{1/2})(X'\Lambda'^{1/2})^\top$, where $\Phi_{K-1} = \rho I_{K-1} \otimes I_d + I_{d(K-1)}$. For $j \neq k$: – Sample $Z^{(j)} \overset{\text{iid}}{\sim} \mathcal{N}_d(\mathbf{0}_d, I_d)$, – Compute $u_*^{(j)} = X'\Lambda'^{1/2} z^{(j)}$, – Compute $x_*^{(j)} = y + S^{(j)} \left(u_*^{(j)} - \rho u^{(k)}\right)$.
RQMC Koborov rule with $1 \leqslant a < K$	Sample $U \sim \text{Uniform}[0,1)^d$. For $k = 1, \ldots, K$: – Compute $u^{(k)} \equiv_1 \frac{k-1}{K}\left(1, a, \ldots, a^{d-1}\right) + u$, – Compute $z^{(k)} = F^{-1}(u^{(k)})$, – Compute $y^{(k)} = x_n + S^{(k)} z^{(k)}$.	Compute $u_* = F\left((S^{(k)})^{-1}(y - x_n)\right)$. For $j \neq k$: – Compute $u_*^{(j)} \equiv_1 \frac{j-1}{K}\left(1, a, \ldots, a^{d-1}\right) + u_*$, – Compute $z_*^{(j)} = F^{-1}(u_*^{(j)})$, – Compute $x_*^{(j)} = y + S^{(j)} z_*^{(j)}$.
Common RV	Sample $Z \sim \mathcal{N}_d(\mathbf{0}_d, I_d)$. For $k = 1, \ldots, K$, compute $y^{(k)} = x_n + S^{(k)} z$.	Compute $z_* = (S^{(k)})^{-1}(y - x_n)$. For $j \neq k$, compute $x_*^{(j)} = y + S^{(j)} z_*$.

Table S1. Summary of the different types of candidates used in the aMTM algorithm at the MTM sampling step; the rest of the MTM sampling remains unchanged. The sampling of candidates and shadow points is described algorithmically. Notation: x_n is the current state of the chain, F is the standard normal CDF, $y = y^{(k)}$ is the selected proposal, $\Sigma^{(k)} = S^{(k)} S^{(k)\top}$ is the square root decomposition and $\otimes$ denotes the usual Kronecker product.

2. Additional details on experiments

Region	x_1	x_2	x_4	x_5	iid weight
1 (orange)	–	–	$\geqslant 0$	$\geqslant 3$	3.5%
2 (blue)	–	–	< 0	$\geqslant 3$	3.5%
3 (green)	$\geqslant 0$	$\geqslant 3$	–	–	16.7%
4 (yellow)	< 0	$\geqslant 3$	–	–	16.6%
5 (black)	not in regions 1-4				59.7%

Table S2. Definition of regions used to compute the lower bound on the total variation distance between empirical distributions coming from MCMC samples and iid samples. Color codes refer to Figure S1.

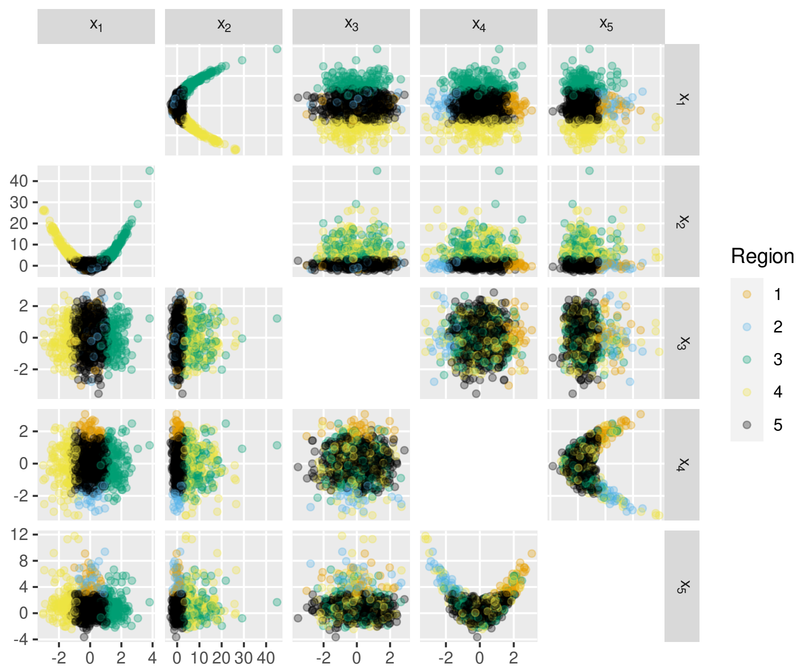

Figure S1. 1,000 iid samples from the banana target (5.1). Regions are defined in Table S2.

3. Proofs

3.1. Results on MTM transitions

Consider a MTM transition P with joint proposal density $q(\cdot|x)$. For $k = 1, \ldots, K$, the conditional density of $y^{(-k)}$ given $y^{(k)}$ is $q^{(-k)}(\cdot|y^{(k)}, x)$ and the marginal density of $y^{(k)}$ is $q^{(k)}(\cdot|x)$. The generalized MTM acceptance probability is given by

$$\alpha_{\mathrm{MTM}}\left(y, y^{(-k)}|x, x_*^{(-k)}\right) = \min\left\{1, \frac{\pi(y)q^{(k)}(x|y)\bar{w}^{(k)}(x, x_*^{(-k)}|y)}{\pi(x)q^{(k)}(y|x)\bar{w}^{(k)}(y, y^{(-k)}|x)}\right\} ,$$

where $x_*^{(-k)} \sim q^{(-k)}(\cdot|x, y)$ are the shadow points, and where

$$\bar{w}^{(k)}(y, y^{(-k)}|x) = \frac{w^{(k)}(y|x)}{\sum_{j=1}^{K} w^{(j)}(y^{(j)}|x)}$$

is the probability of choosing the k-th candidate.

We define some notation in order to simplify the MTM transition. The transition admits the following (pseudo-)density:

$$p(y|x) = a(y|x) + R(x)\delta_x(y) ,$$

where $a(y|x)$ is the density for transitioning from x to y using any of the K candidates and any shadow sample, and where $R(x) = 1 - \int_{\mathcal{X}} a(y|x)\,\mathrm{d}y$ is the integrated rejection probability. Since moving from x to y can be achieved through any of the K candidates, we therefore decompose

$$a(y|x) = \sum_{k=1}^{K} A^{(k)}(y|x)q^{(k)}(y|x) ,$$

where $A^{(k)}(y|x)$ is the density for accepting a move from x to y through the k-th candidate. We have

$$A^{(k)}(y|x) = \int_{\mathcal{X}^{K-1}}\int_{\mathcal{X}^{K-1}} q^{(-k)}(y^{(-k)}|y, x)\bar{w}(y; y^{(-k)}|x)\alpha_{\mathrm{MTM}}(y, y^{(-k)}|x, x_*^{(-k)}) \times q^{(-k)}(x_*^{(-k)}|y, x)\,\mathrm{d}y^{(-k)}\,\mathrm{d}x_*^{(-k)} .$$

Finally, we define the integrated probability of accepting the k-th candidate as the new point,

$$\overline{A}^{(k)}(x) = \int_{\mathcal{X}} A^{(k)}(y|x)q^{(k)}(y|x)\,\mathrm{d}y ,$$

so that we may write

$$R(x) = 1 - \sum_{k=1}^{K} \overline{A}^{(k)}(x) .$$

Naturally, when we want to make the dependence of a MTM transition on its set of parameters θ explicit, we simply index position each of the above definitions by θ.

Proposition 3.1. *Suppose that the marginal proposal densities satisfy*

$$q^{(k)}(y|x) > 0 \quad \Leftrightarrow \quad q^{(k)}(x|y) > 0 \ , \qquad k = 1, \dots, K \ .$$

Then, the MTM transition satisfies the detailed balance condition,

$$p(x|y)\pi(y) = p(y|x)\pi(x) \ , \qquad \forall\, x, y \in \mathcal{X} \ , \tag{S1}$$

for any weight function $w^{(k)}(y|x)$ that is positive whenever $x, y \in \mathcal{X}$.

Proof. If $x = y$, then (S1) trivially holds. Thus, we may assume that $y \neq x$, in which case $\delta_X(y) = 0$ and

$$\begin{aligned} p(y|x) &= a(y|x) \\ &= \sum_{k=1}^{K} \int_{\mathcal{X}^{K-1}} \int_{\mathcal{X}^{K-1}} q(y, y^{(-k)}|x)\bar{w}(y; y^{(-k)}|x)\alpha_{\mathrm{MTM}}(y, y^{(-k)}|x, x_*^{(-k)}) \\ &\qquad \times q^{(-k)}(x_*^{(-k)}|y, x)\, \mathrm{d}y^{(-k)}\, \mathrm{d}x_*^{(-k)} \ . \end{aligned}$$

Then, we decompose the joint proposal density as

$$q(y, y^{(-k)}|x) = q^{(-k)}(y^{(-k)}|y, x)q^{(k)}(y|x) \ .$$

We also rewrite the MTM acceptance probability in a more symmetric form,

$$\begin{aligned} \alpha_{\mathrm{MTM}}(y, y^{(-k)}|x, x_*^{(-k)}) &= \pi(y)q^{(k)}(x|y)\bar{w}^{(k)}(x; x_*^{(-k)}|y) \\ &\quad \times \min\left\{ \frac{1}{\pi(y)q^{(k)}(x|y)\bar{w}^{(k)}(x; x_*^{(-k)}|y)}, \frac{1}{\pi(x)q^{(k)}(y|x)\bar{w}^{(k)}(y; y^{(-k)}|x)} \right\} \ . \end{aligned}$$

We can now write

$$\begin{aligned} a(y|x)\pi(x) = \sum_{k=1}^{K} \int_{\mathcal{X}^{K-1}} \int_{\mathcal{X}^{K-1}} & \pi(x)q^{(-k)}(y^{(-k)}|y, x)q^{(k)}(y|x)\bar{w}(y; y^{(-k)}|x) \\ & \times \pi(y)q^{(k)}(x|y)\bar{w}^{(k)}(x; x_*^{(-k)}|y)q^{(-k)}(x_*^{(-k)}|y, x) \\ & \times \min\left\{ \frac{1}{\pi(y)q^{(k)}(x|y)\bar{w}^{(k)}(x; x_*^{(-k)}|y)}, \frac{1}{\pi(x)q^{(k)}(y|x)\bar{w}^{(k)}(y; y^{(-k)}|x)} \right\} \\ & \mathrm{d}y^{(-k)}\, \mathrm{d}x_*^{(-k)} \ . \end{aligned}$$

By direct inspection, we see that the expression is completely symmetric under the swap $(y, y^{(-k)}) \leftrightarrow (x, x_*^{(-k)})$. Hence,

$$a(y|x)\pi(x) = a(x|y)\pi(y) \ ,$$

and the detailed balance condition (S1) is satisfied. □

***Remark* 3.1.** Note that a similar result was obtained by Casarin, Craiu and Leisen (2013, in the appendix) for a slightly less general form of MTM acceptance probability.

Proposition 3.2. *Let π be a target density with connected support $\mathcal{X}$. Suppose that π and the weight function $w^{(k)}(\cdot|x)$ are bounded above on $\mathcal{X}$ and below on any compact subset of $\mathcal{X}$, for any fixed $x \in \mathcal{X}$. Further suppose that there exists δ, $\varepsilon > 0$ such that the marginal proposal densities are locally positive, that is,*

$$\|x - y\|_2 < \delta \quad \Rightarrow \quad q^{(k)}(y|x) > \varepsilon \ , \qquad k = 1\,, \ldots, K \ .$$

Then, the MTM transition is π-irreducible and aperiodic.

Proof. The proof of π-irreducilibity appeals to Meyn and Tweedie (2009, Proposition 4.2.1) which states that a transition P is ϕ-irreducible if and only if, for all $x \in \mathcal{X}$ and for all measurable B such that $\phi(B) > 0$, there exists $m \in \mathbb{N}$ with $P^m(B|x) > 0$. Thus, let us consider $x \in \mathcal{X}$ as well as a measurable set $B \subseteq \mathcal{X}$ with positive probability $\pi(B) > 0$. By connectedness of $\mathcal{X}$, we can find a path between x and any point in B. In particular, we can always find a path of length $m \in \mathbb{N}$ from x to some $x_m \in B$ such that each step is at most of size δ, i.e. $\|x_i - x_{i-1}\|_2 < \delta$ ($i = 1, \ldots, m$) and each x_i has positive density $\pi(x_i) > 0$. Around each x_i, we consider the ball of radius δ, denoted

$$B_\delta(x_i) = \{x \in \mathbb{R}^d \mid \|x_i - x\|_2 \leqslant \delta\} \ .$$

Since x_i is in the support of π, then $\pi\left(B_\delta(x_i)\right) > 0$ by the definition of a support. Now, we show that the transition from one ball to the next happens with positive probability. Consider $i \in \{0\,, \ldots, m-1\}$ and $x \in B_\delta(x_i)$. Then, the probability of landing in the next ball is bounded below by the probability of landing in the next ball through an accepted proposal, i.e.

$$P\left(B_\delta(x_{i+1})|x\right) \geqslant \sum_{k=1}^{K} \int_{B_\delta(x_{i+1})} A^{(k)}(y|x) q^{(k)}(y|x)\, \mathrm{d}y \ .$$

Now, $A^{(k)}$ is positive for any $y \in B_\delta(x_{i+1})$ since it is the expectation of a positive function ($\bar{w} > 0$ and $\alpha_{\mathrm{MTM}} > 0$ both follow from the assumptions). Then, since the marginal density $q^{(k)}$ is also positive on $B_\delta(x_{i+1})$ and since $B_\delta(x_{i+1})$ has positive probability, we find $P\left(B_\delta(x_{i+1})|x\right) > 0$. By induction, we can show that the i-step transition $P^i(B_\delta(x_i)|x)$ is positive for $i = 1\,, \ldots, m$. In particular, it holds for m so that $P^m(B_\delta(x_m)|x) > 0$ from which we find $P^m(B|x) > 0$ because $x_m \in B \cap \mathcal{X}$. By Meyn and Tweedie (2009, Proposition 4.2.1), P is π-irreducible.

To prove aperiodicity, we show that P is strongly aperiodic, meaning that there exists a $(\nu, 1)$-small measurable set B with $\nu(B) > 0$. A $(\nu, 1)$-small set is such that, for all $x \in B$ and for all measurable sets C,

$$P(C|x) \geqslant \nu(C) \ . \tag{S2}$$

We now consider $B = B_{\delta/2}(x)$ and construct a measure ν concentrated on B satisfying the minorization condition (S2). Thus, let us consider $x \in B$ and C measurable. Then, we can bound $P(C|x) \geqslant P(C \cap B|x)$; we can also bound the latter using accepted proposals, leading to

$$P(C|x) \geqslant \sum_{k=1}^{K} \int_{C\cap B} A^{(k)}(y|x) q^{(k)}(y|x)\,\mathrm{d}y \ .$$

We now define K partitions of the support, one for each candidate. Given the current state x, the candidates $y^{(-k)}$, and the shadow sample $x_*^{(-k)}$, this partition groups together all the proposals y that are automatically accepted, given that the k-th candidate was selected:

$$D^{(k)}(x) = \left\{ y \in \mathcal{X} \ \middle| \ \frac{\pi(y)\bar{w}^{(k)}(x; x_*^{(-k)}|y)}{\pi(x)\bar{w}^{(k)}(y; y^{(-k)}|x)} \leqslant 1 \right\} \ .$$

Note that, contrarily to what is suggested by the notation, $D^{(k)}(x)$ also is a function of $y^{(-k)}$ and $x_*^{(-k)}$. For $y \in D^{(k)}(x)$, we have

$$\alpha_{\mathrm{MTM}}(y, y^{(-k)}|x, x_*^{(-k)}) = \frac{\pi(y)\bar{w}^{(k)}(x; x_*^{(-k)}|y)}{\pi(x)\bar{w}^{(k)}(y; y^{(-k)}|x)} \ ,$$

while for $y \notin D^{(k)}$, we have $\alpha_{\mathrm{MTM}}(y, y^{(-k)}|x, x_*^{(-k)}) = 1$. We can now split the integral over $B \cap C$ into two parts over which the form of α_{MTM} is known.

For $y \in D^{(k)}$, the integrand takes the form

$$\begin{aligned} &\bar{w}^{(k)}(y; y^{(-k)}|x) \frac{\pi(y)\bar{w}^{(k)}(x; x_*^{(-k)}|y)}{\pi(x)\bar{w}^{(k)}(y; y^{(-k)}|x)} q^{(-k)}(y^{(-k)}|y, x) q^{(-k)}(x_*^{(-k)}|x, y) q^{(k)}(y|x) \\ &= \bar{w}^{(k)}(x; x_*^{(-k)}|y) \frac{\pi(y)}{\pi(x)} q^{(-k)}(y^{(-k)}|y, x) q^{(-k)}(x_*^{(-k)}|x, y) q^{(k)}(y|x) \ . \end{aligned}$$

Since we will integrate over $y^{(-k)}$ and $x_*^{(-k)}$, we try to bound all terms that are not the densities of these variables. In particular, we search a lower bound for

$$\bar{w}^{(k)}(x; x_*^{(-k)}|y) \frac{\pi(y)}{\pi(x)} q^{(k)}(y|x) \ .$$

When $x \in B$ and $y \in C \cap B \subseteq B$, we can bound each term by making use of the assumptions. Indeed, we have $\|y - x\|_2 \leqslant \delta$ so that $q^{(k)}(y|x) \geqslant \varepsilon$. Furthermore, from the conditions on the weight functions, there exists $0 < a < A < \infty$ such that $w^{(k)}(x|y) > a$ and $w^{(j)}(x_*^{(j)}|y) \leqslant A$ so that $\bar{w}^{(k)}(x; x_*^{(-k)}|y) \geqslant a/KA$ for all $\|y - x\|_2 \leqslant \delta$ and all $x_*^{(-k)}$. Hence, we find

$$\bar{w}^{(k)}(x; x_*^{(-k)}|y) \frac{\pi(y)}{\pi(x)} q^{(k)}(y|x) \geqslant \frac{a\varepsilon}{KA} \frac{\pi(y)}{\pi(x)} \geqslant \frac{a\varepsilon}{KA} \frac{\inf_{y\in B} \pi(y)}{\sup_{y\in B} \pi(y)} \ ,$$

which is positive because all quantities are positive (π is bounded below and above on $B = B_{\delta/2}(x)$ compact).

On $y \notin D^{(k)}$, the integrand takes the form

$$\bar{w}^{(k)}(y; y^{(-k)}|x) q^{(-k)}(y^{(-k)}|y,x) q^{(-k)}(x_*^{(-k)}|x,y) q^{(k)}(y|x) \ ,$$

which means we aim to bound

$$\bar{w}^{(k)}(y; y^{(-k)}|x) q^{(k)}(y|x) \ .$$

For the same reasons as before, we have $\bar{w}^{(k)}(x; x_*^{(-k)}|y) \geqslant a/KA$ and $q^{(k)}(y|x) \geqslant \varepsilon$. Then, we note that $\inf_B \pi / \sup_B \pi$ is always less than 1 so we find the same bound as in the case $y \in D^{(k)}$, i.e.

$$\bar{w}^{(k)}(y; y^{(-k)}|x) q^{(k)}(y|x) \geqslant \frac{a\varepsilon}{KA} \geqslant \frac{a\varepsilon}{KA} \frac{\inf_{y \in B} \pi(y)}{\sup_{y \in B} \pi(y)} \ .$$

We therefore find the following bound on $P(C|x)$:

$$\begin{aligned}
P(C|x) &\geqslant \sum_{k=1}^{K} \int_{C \cap B} A^{(k)}(y|x) q^{(k)}(y|x) \, \mathrm{d}y \\
&= \sum_{k=1}^{K} \left(\int_{C \cap B \cap D^{(k)}} + \int_{C \cap B \cap (D^{(k)})^c} \right) A^{(k)}(y|x) q^{(k)}(y|x) \, \mathrm{d}y \\
&\geqslant \sum_{k=1}^{K} \left(\int_{C \cap B \cap D^{(k)}} + \int_{C \cap B \cap (D^{(k)})^c} \right) \int_{\mathcal{X}^{K-1}} \int_{\mathcal{X}^{K-1}} \\
&\qquad q^{(-k)}(y^{(-k)}|y,x) q^{(-k)}(x_*^{(-k)}|x,y) \frac{a\varepsilon}{KA} \frac{\inf_B \pi}{\sup_B \pi} \, \mathrm{d}x_*^{(-k)} \, \mathrm{d}y^{(-k)} \, \mathrm{d}y \\
&= \sum_{k=1}^{K} \int_{C \cap B} \frac{a\varepsilon}{KA} \frac{\inf_B \pi}{\sup_B \pi} \int_{\mathcal{X}^{K-1}} \int_{\mathcal{X}^{K-1}} q^{(-k)}(y^{(-k)}|y,x) q^{(-k)}(x_*^{(-k)}|x,y) \, \mathrm{d}x_*^{(-k)} \, \mathrm{d}y^{(-k)} \, \mathrm{d}y \\
&= \sum_{k=1}^{K} \int_{C \cap B} \frac{a\varepsilon}{KA} \frac{\inf_B \pi}{\sup_B \pi} \, \mathrm{d}y \\
&= \frac{a\varepsilon}{A} \frac{\inf_B \pi}{\sup_B \pi} \lambda^{\mathrm{Leb}} (C \cap B) \ ,
\end{aligned}$$

where λ^{Leb} is the Lebesgue measure on $\mathbb{R}^d$. Since

$$\frac{a\varepsilon}{A} \frac{\inf_B \pi}{\sup_B \pi} =: c_0 > 0 \ ,$$

we have that

$$P(C|x) \geqslant \nu(C) \ ,$$

where $\nu(C) = c_0 \lambda^{\mathrm{Leb}} (C \cap B)$ is a non-trivial measure concentrated on B, as required. Finally, we note that $\nu(B) = c_0 \lambda^{\mathrm{Leb}}(B) > 0$ since $c_0 > 0$ and $\lambda^{\mathrm{Leb}}(B) > 0$, where B is a ball with positive radius $\delta/2 > 0$. $\square$

In the context of Markov transitions, a function $h : \mathbb{R}^d \to [0, \infty]$ is said to be *harmonic* for a transition P if $h = Ph$ everywhere, that is,

$$h(x) = \int_{\mathcal{X}} h(y)P(\,\mathrm{d}y|x) \;, \qquad x \in \mathbb{R}^d \;.$$

From Tierney (1994, Theorem 2), we know that a recurrent Markov transition P is Harris-recurrent if and only if every bounded harmonic function is a constant function. We use this result to show that recurrence and Harris-recurrence happen simultaneously for MTM transitions.

Proposition 3.3. *Let P be a MTM transition for a given target density π. If P is π-irreducible, then P is Harris-recurrent.*

Proof. By Proposition 3.1, the MTM transition satisfies the detailed balance condition. By Robert and Casella (2004, Theorem 6.46), the MTM transition admits π as its invariant distribution. By Tierney (1994, Theorem 1), the MTM transition is positive recurrent. From Nummelin (1984, Proposition 3.13), we know that a recurrent π-irreducible Markov transition P is such that every bounded harmonic function h is constant at least π-almost everywhere. Hence, we only require to extend that result to every $x \in \mathbb{R}^d$.

We define the set H as containing the points over which a function h is not constant, i.e.,

$$H = \{x \in \mathcal{X} \mid h(x) \neq \pi h\} \,.$$

By the above argument, we find $\pi(H) = 0$. Then, since the measure of H is null, the probability of transitioning from $x \in \mathcal{X}$ to H must also be 0:

$$a\,(H|x) = \int_H \sum_{k=1}^{K} A^{(k)}(y|x)q^{(k)}(y|x)\,\mathrm{d}y = 0 \;,$$

since $q^{(k)}$ is assumed to be a density and H has zero measure.

Now, since h is harmonic with respect to P, we can decompose

$$h(x) = \int_{\mathbb{R}^d} h(y)P(\,\mathrm{d}y|x) = \int_H h(y)P(\,\mathrm{d}y|x) + \int_{H^c} h(y)P(\,\mathrm{d}y|x) \;.$$

The former term satisfies

$$\begin{aligned}\int_H h(y)P(\,\mathrm{d}y|x) &= \int_H h(y) \sum_{k=1}^{K} \left[A^{(k)}(y|x)q^{(k)}(\,\mathrm{d}y|x)\right] + \int_H h(y)R(x)\delta_x(\,\mathrm{d}y) \\ &= 0 + h(x)R(x)\,\mathbb{I}\left(h(x) \neq \pi h\right) \;.\end{aligned}$$

For the latter term, we obtain

$$\begin{aligned}\int_{H^c} h(y)P(\,\mathrm{d}y|x) &= \int_{H^c} \pi h \sum_{k=1}^{K} \left[A^{(k)}(y|x)q^{(k)}(\,\mathrm{d}y|x)\right] + \int_{H^c} \pi h R(x)\delta_x(\,\mathrm{d}y) \\ &= \pi h(1 - R(x)) + \pi h R(x)\,\mathbb{I}\left(h(x) = \pi h\right) \;.\end{aligned}$$

Combining both expressions, we find that h must satisfy

$$h(x) = \pi h + R(x)(h(x) - \pi h)\,\mathbb{I}(h(x) \neq \pi h)\ .$$

Factorizing yield

$$0 = (h(x) - \pi h)(R(x)\,\mathbb{I}(h(x) \neq \pi h) - 1)\ . \tag{S3}$$

Thus, if $h(x) \neq \pi h$, we must have $R(x) = 1$, but this would contradict the π-irreducibility of the transition whenever $x \in \mathcal{X}$ because this means that the probability of staying at x is 1. Hence, the only points where we can have $h(x) \neq \pi h$ are $x \notin \mathcal{X}$. Then, by construction of MTM transitions, $0 < R(x) < 1$ so that $h(x) = \pi h$ must hold to satisfy (S3). This shows $h \equiv \pi h$. □

3.2. Results on adaptive MCMC

3.2.1. Additional background on adaptive algorithms

We say that a family of MCMC transitions $\{P_\theta\}_{\theta\in\Theta}$ satisfies the *uniform geometric ergodicity on compact sets* condition (Andrieu and Moulines, 2006, Assumption A1) if there exists a test function $V : \mathcal{X} \to [1, \infty)$ with $\sup_{\mathcal{X}} V < \infty$ such that, for any compact $\mathcal{K} \subseteq \Theta$, the following two conditions hold :

(i) *Minorisation.* There exists $C \in \mathcal{B}(\mathcal{X})$, $\delta > 0$ and a probability measure ν with $\nu(C) > 0$ such that

$$P_\theta(A|x) \geqslant \delta\nu(A)\ , \qquad \forall\, A \in \mathcal{B}(\mathcal{X}),\, \theta \in \mathcal{K},\, x \in C\ .$$

(ii) *Geometric drift.* There exists $\lambda \in [0, 1)$ and $b \in (0, \infty)$ such that

$$P_\theta V(x) \leqslant \begin{cases} \lambda V(x), & x \notin C, \\ b, & x \in C, \end{cases} \qquad \forall\, \theta \in \mathcal{K}\ ,$$

where $P_\theta V(x) = \int V(z) P_\theta(z|x)\,\mathrm{d}z$.

We say that a family of update functions $\{H_\theta\}_{\theta\in\Theta}$ is *V-Lipschitz* for some test function V (typically the same as in the geometric drift condition) if, for any compact $\mathcal{K} \subseteq \Theta$, we have

$$\sup_{\theta\in\mathcal{K}} \|H_\theta\|_V < \infty \qquad \text{and} \qquad \sup_{\theta\neq\theta'\in\mathcal{K}\times\mathcal{K}} \|\theta - \theta'\|_2^{-1} \|H_\theta - H_{\theta'}\|_V < \infty,$$

where $\|\mu\|_V$ defines the V-norm of the function f for some test function V, that is,

$$\|f\|_V = \sup_{x\in\mathcal{X}} \frac{\|f(x)\|_2}{V(x)}.$$

We say that a family of transitions is *V-Lipschitz on $\mathcal{K}$* if there exists $C < \infty$ such that, for all functions $f : \mathcal{X} \to \mathbb{R}$, with $\|f\|_V < \infty$, and all $r \in [0, 1]$,

$$\|P_\theta f - P_{\theta'} f\|_{V^r} \leqslant C \|f\|_{V^r} \|\theta - \theta'\|_2, \qquad \forall\, \theta, \theta' \in \mathcal{K}\,.$$

3.2.2. Diminishing adaptation

Define the V-norm of a (possibly signed) measure μ as

$$|||\mu|||_V = \sup_{g:|g|\leqslant V} |\mu(g)|,$$

where we use the triple bar notation to differentiate with the V-norm of a function defined earlier. Note that $|||\cdot|||_1$ is equivalent to $\|\cdot\|_{\mathrm{TV}}$:

$$\|\mu\|_{\mathrm{TV}} = \sup_{B\in\mathcal{B}(\mathcal{X})} |\mu(B)| = \frac{1}{2} \sup_{g:|g|\leqslant 1} |\mu(g)| = \frac{1}{2}|||\mu|||_1.$$

Proposition 3.4. *Suppose $\{P_\theta\}_{\theta\in\Theta}$ satisfies the uniform geometric ergodicity on compact sets, $\{H_\theta\}_{\theta\in\Theta}$ is V-Lipschitz and $\{P_\theta\}_{\theta\in\Theta}$ is V-Lipschitz on any compact subset of Θ for the same test function V. If $\sup_{\theta\in\mathcal{K}} \|H_\theta\|_V < \infty$ for any $\mathcal{K} \subseteq \Theta$ compact and $\{\theta_n\}_{n\geqslant 0}$ is bounded in probability, then the adaptive MCMC algorithm is such that*

$$\left|\left|\left| P_{\theta_{n+1}} - P_{\theta_n} \right|\right|\right|_V \xrightarrow{\mathbb{P}} 0, \qquad n \to \infty.$$

In particular, if $V \equiv 1$, then the algorithm satisfies the Diminishing Adaptation condition.

Proof. From the condition on H_θ, for any compact $\mathcal{K} \subseteq \Theta$, we have $\widetilde{H}(\mathcal{K}) := \sup_{\theta\in\mathcal{K}} \|H_\theta\|_V < \infty$. In particular, we have

$$\|H(\theta, x)\|_2 \leqslant \widetilde{H}(\mathcal{K})V(x) \ , \qquad \forall\, x \in \mathcal{X},\, \theta \in \mathcal{K} \ .$$

The uniform geometric ergodicity on compact sets ensures that $\{V(X_n)\}_{n\geqslant 0}$ is bounded in probability (Fontaine, 2019, Proposition 3.5). Hence, for all $\varepsilon > 0$, there exists $\widetilde{V} = \widetilde{V}(\varepsilon) < \infty$ such that $\mathbb{P}_{x_0,\theta_0}\left(V(X_n) \leqslant \widetilde{V}\right) \geqslant 1 - \frac{\varepsilon}{4}$ for all $n \geqslant 1$, where (x_0, θ_0) are the initial values of the joint chain $\{(X_n, \theta_n)\}_{n\geqslant 0}$. Then,

$$\mathbb{P}_{x_0,\theta_0}\left(\|H(\theta, X_n)\|_2 \leqslant \widetilde{H}(\mathcal{K})\widetilde{V} \mid \theta \in \mathcal{K}\right) \geqslant 1 - \frac{\varepsilon}{4} \ .$$

Then, for $\theta_{n+1} - \theta_n = \gamma_{n+1}H(\theta_n, X_n)$, we find

$$\mathbb{P}_{x_0,\theta_0}\left(\|\theta_{n+1} - \theta_n\|_2 \leqslant \gamma_{n+1}\widetilde{H}(\mathcal{K})\widetilde{V} \mid \theta_n \in \mathcal{K}\right) \geqslant 1 - \frac{\varepsilon}{4} \ , \qquad \forall\, \theta_n \in \mathcal{K} \ .$$

Since $\{\theta_n\}_{n\geqslant 0}$ is bounded in probability, there exists a compact $\mathcal{K} \subset \Theta$ such that

$$\mathbb{P}_{x_0,\theta_0}\left(\theta_n \in \mathcal{K}\right) \geqslant 1 - \frac{\varepsilon}{4} \ ,$$

from which we find

$$\mathbb{P}_{x_0,\theta_0}\left(\|\theta_{n+1} - \theta_n\|_2 \leqslant \gamma_{n+1}\widetilde{H}(\mathcal{K})\widetilde{V}\right) \geqslant \left(1 - \frac{\varepsilon}{4}\right)^2 \ .$$

The Lipschitz transition condition implies that there exists $C < \infty$ with

$$\left\|P_{\theta_{n+1}}f - P_{\theta_n}f\right\|_V \leqslant C\|f\|_V\|\theta_{n+1}-\theta_n\|_2, \qquad \forall\, x \in \mathcal{X},\, \theta_{n+1},\, \theta_n \in \mathcal{K}\,.$$

Thus, we can bound

$$\begin{aligned}
\left|\!\left|\!\left|P_{\theta_{n+1}}(\cdot \mid x) - P_{\theta_n}(\cdot \mid x)\right|\!\right|\!\right|_V &= \sup_{f:|f|\leqslant V}\left|(P_{\theta_{n+1}}f - P_{\theta_n}f)(x)\right| \\
&\leqslant \sup_{f:|f|\leqslant V}\sup_{y\in\mathcal{X}}\left|(P_{\theta_{n+1}}f - P_{\theta_n}f)(y)\right| \\
&= \sup_{g:|g|\leqslant 1}\sup_{y\in\mathcal{X}}\frac{1}{V(y)}\left|(P_{\theta_{n+1}}g - P_{\theta_n}g)(y)\right| \\
&= \sup_{g:|g|\leqslant 1}\left\|P_{\theta_{n+1}}g - P_{\theta_n}g\right\|_V \\
&\leqslant \sup_{g:|g|\leqslant 1} C\|g\|_V\|\theta_{n+1}-\theta_n\|_2 \\
&= C\|\theta_{n+1}-\theta_n\|_2.
\end{aligned}$$

We then find

$$\mathbb{P}_{x_0,\theta_0}\left(\left|\!\left|\!\left|P_{\theta_{n+1}} - P_{\theta_n}\right|\!\right|\!\right|_V \leqslant \gamma_{n+1}C\widetilde{H}(\mathcal{K})\widetilde{V} \mid \theta_{n+1}\in\mathcal{K}\right) \geqslant \left(1-\frac{\varepsilon}{4}\right)^2\ .$$

Using the boundedness in probability of $\theta_{n+1} \in \mathcal{K}$:

$$\mathbb{P}_{x_0,\theta_0}\left(\left|\!\left|\!\left|P_{\theta_{n+1}} - P_{\theta_n}\right|\!\right|\!\right|_V \leqslant \gamma_{n+1}C\widetilde{H}(\mathcal{K})\widetilde{V}\right) \geqslant \left(1-\frac{\varepsilon}{4}\right)^3\ .$$

Finally, $\gamma_n \to 0$ implies that, for any $\varepsilon' > 0$, there exists $M = M(\varepsilon') \in \mathbb{N}$ such that $\gamma_{n+1}C\widetilde{H}(\mathcal{K})\widetilde{V} \leqslant \varepsilon'$ whenever $n \geqslant M$. Hence, for all $n \geqslant M$, we have

$$\mathbb{P}_{x_0,\theta_0}\left(\left|\!\left|\!\left|P_{\theta_{n+1}} - P_{\theta_n}\right|\!\right|\!\right|_V \leqslant \varepsilon'\right) \geqslant \left(1-\frac{\varepsilon}{4}\right)^3 \geqslant 1-\varepsilon\ .$$

□

3.3. Results on the aMTM algorithm

3.3.1. Diminishing adaptation

Theorem 3.1. *Let π be a target density with compact support $\mathcal{X} \subseteq \mathbb{R}^d$. Consider a family of MTM transitions $\{P_\theta\}_{\theta\in\Theta}$ with compact parameter space Θ and satisfying the V-Lipschitz condition on Θ. An adaptive MTM algorithm on $\{P_\theta\}_{\theta\in\Theta}$ using the V-Lipschitz update family H_θ satisfying*

$$\sup_{\theta\in\mathcal{K}}\|H_\theta\|_V < \infty\ ,$$

with $V \equiv 1$ satisfies the diminishing adaptation condition.

Proof. We verify the conditions of Proposition 3.4. When Θ and $\mathcal{X}$ are assumed to be compact, we can simply choose $V \equiv 1$ as the test function in the uniform geometric ergodicity condition. Since Θ is assumed to be compact, we directly have $\{\theta_n\}_{n\geqslant 0}$ bounded and therefore bounded in probability. The conditions on the family of updates are verified by hypothesis. □

3.3.2. Lipschitz transitions

Proposition 3.5. *Let $\{\varphi_\Sigma\}_{\Sigma\in\mathcal{S}}$ be a collection of d-dimensional Gaussian densities with mean $\mathbf{0}_d$ and covariance $\Sigma \in \mathcal{S} \subseteq \mathcal{C}_d^+$. If $\mathcal{S}$ is compact, then*

$$\int_{\mathbb{R}^d} |\varphi_\Sigma(z) - \varphi_{\Sigma'}(z)|\lambda(\,\mathrm{d}z) \leqslant \frac{d}{\lambda_{\min}} \|\Sigma - \Sigma'\|_F \ ,$$

where $\lambda_{\min} > 0$ is the smallest possible eigenvalue of a covariance $\Sigma \in \mathcal{S}$ and where $\|\cdot\|_F$ denotes the usual Frobenius norm.

Proof. Since $\mathcal{S}$ is compact, we can find $0 < \lambda_{\min} \leqslant \lambda_{\max} < \infty$ such that all eigenvalues of any $\Sigma \in \mathcal{S}$ are contained in $[\lambda_{\min}, \lambda_{\max}]$. Inspired by a step in the proof of Haario, Saksman and Tamminen (2001, Theorem 1), we consider the convex combination of $\Sigma, \Sigma' \in \mathcal{S}$, i.e.

$$\Sigma_t = (1-t)\Sigma + t\Sigma' = \Sigma + t(\Sigma' - \Sigma) \ .$$

While we do not require $\mathcal{S}$ to be convex, we know that $\mathcal{C}_+^d$ is indeed convex so that $\Sigma_t \in \mathcal{C}_+^d$ for any $t \in [0,1]$. In particular, φ_{Σ_t} is a well-defined d-dimensional Gaussian distribution for any $t \in [0,1]$. The purpose of this convex combination is the following identity, resulting from the fundamental theorem of calculus:

$$\int_0^1 \left(\frac{\partial}{\partial t}\varphi_{\Sigma_t}(z)\right) \mathrm{d}t = \varphi_{\Sigma_t}(z)\Big|_{t=0}^{t=1} = \varphi_\Sigma(z) - \varphi_{\Sigma'}(z) \ .$$

This identity holds as long as $\varphi_{\Sigma_t}(z)$ is differentiable w.r.t. t, but this will be verified implicitly in the following calculations. We then proceed to relate the previous identity to $\|\Sigma - \Sigma'\|_F$.

Logarithmic differentiation gives us

$$\begin{aligned}\frac{\partial}{\partial t}\varphi_{\Sigma_t}(z) &= \varphi_{\Sigma_t}(z)\frac{\partial}{\partial t}\log\varphi_{\Sigma_t}(z) \\ &= -\frac{1}{2}\varphi_{\Sigma_t}(z)\frac{\partial}{\partial t}\left[d\log(2\pi) + \log\det(\Sigma_t) + z^\top\Sigma_t^{-1}z\ .\right]\end{aligned}$$

Then, using matrix derivative identities (Petersen and Pedersen, 2008), we find

$$\frac{\partial}{\partial t}z^\top\Sigma_t^{-1}z = \operatorname{tr}\left(-\Sigma_t^{-1}zz^\top\Sigma_t^{-1}(\Sigma' - \Sigma)\right) \ ,$$

which yields

$$\frac{\partial}{\partial t}\varphi_{\Sigma_t}(z) = -\frac{1}{2}\varphi_{\Sigma_t}(z)\,\mathrm{tr}\left(\Sigma_t^{-1}(\Sigma' - \Sigma) - \Sigma_t^{-1}zz^\top\Sigma_t^{-1}(\Sigma' - \Sigma)\right) \ .$$

Therefore, we find

$$\frac{\partial}{\partial t}\log\varphi_{\Sigma_t}(z) = -\frac{1}{2}\,\mathrm{tr}\left(\Sigma_t^{-1}(\Sigma' - \Sigma) - \Sigma_t^{-1}zz^\top\Sigma_t^{-1}(\Sigma' - \Sigma)\right) \ ,$$

which can be bounded, using the triangle inequality, by

$$\left|\frac{\partial}{\partial t}\log\varphi_{\Sigma_t}(z)\right| \leqslant \left|\mathrm{tr}\left(\Sigma_t^{-1}(\Sigma' - \Sigma)\right)\right| + \left|\mathrm{tr}\left(\Sigma_t^{-1}zz^\top\Sigma_t^{-1}(\Sigma' - \Sigma)\right)\right| \ .$$

Now, we may use the general matrix norm inequality $|\mathrm{tr}(AB)| \leqslant \|A\|_F\|B\|_F$ to bound

$$\left|\mathrm{tr}\left(\Sigma_t^{-1}(\Sigma' - \Sigma)\right)\right| \leqslant \left\|\Sigma_t^{-1}\right\|_F\|\Sigma' - \Sigma\|_F \ ,$$

as well as

$$\begin{aligned}\left|\mathrm{tr}\left(\Sigma_t^{-1}zz^\top\Sigma_t^{-1}(\Sigma' - \Sigma)\right)\right| &\leqslant \left\|\Sigma_t^{-1}zz^\top\Sigma_t^{-1}\right\|_F\|\Sigma' - \Sigma\|_F \\ &\leqslant z^\top\Sigma_t^{-2}z\|\Sigma' - \Sigma\|_F \ .\end{aligned}$$

Hence,

$$\left|\frac{\partial}{\partial t}\log\varphi_{\Sigma_t}(z)\right| \leqslant \left(\left\|\Sigma_t^{-1}\right\|_F + z^\top\Sigma_t^{-2}z\right)\|\Sigma' - \Sigma\|_F \ .$$

Now, from the theory of Gaussian quadratic forms, we have

$$\int\left(z^\top\Sigma_t^{-2}z\right)\varphi_{\Sigma_t}(z)\lambda(\mathrm{d}z) = \mathrm{tr}\left(\Sigma_t^{-2}\Sigma_t\right) = \mathrm{tr}\left(\Sigma_t^{-1}\right) \ ,$$

which allows us to compute

$$\int\left(\left\|\Sigma_t^{-1}\right\|_F + z^\top\Sigma_t^{-2}z\right)\varphi_{\Sigma_t}(z)\lambda(\mathrm{d}z) = \left\|\Sigma_t^{-1}\right\|_F + \mathrm{tr}\left(\Sigma_t^{-1}\right) \ .$$

Finally, the bounded eigenvalues yield the following bounds,

$$\begin{aligned}\left\|\Sigma_t^{-1}\right\|_F^2 &= \sum_{i=1}^d\lambda_i^2(\Sigma_t^{-1}) = \sum_{i=1}^d\lambda_i^{-2}(\Sigma_t) \leqslant d\lambda_{\min}^{-2} \ , \\ \mathrm{tr}\left(\Sigma_t^{-1}\right) &= \sum_{i=1}^d\lambda_i(\Sigma_t^{-1}) \leqslant d\lambda_{\min}^{-1} \ ,\end{aligned}$$

which, in turn, give

$$\int \left(\left\|\Sigma_t^{-1}\right\|_F + z^\top \Sigma_t^{-2} z\right) \varphi_{\Sigma_t}(z)\lambda(\mathrm{d}z) \leqslant \sqrt{d}\lambda_{\min}^{-1} + d\lambda_{\min}^{-1} \leqslant 2d\lambda_{\min}^{-1} \ .$$

We conclude that

$$\begin{aligned}
\int |\varphi_\Sigma(z) - \varphi_{\Sigma'}(z)|\lambda(\mathrm{d}z) &= \int \left|\int_0^1 \frac{\partial}{\partial t}\varphi_{\Sigma_t}(z)\,\mathrm{d}t\right| \lambda(\mathrm{d}z) \\
&\leqslant \int \int_0^1 \frac{1}{2}\varphi_{\Sigma_t}(z)\left|\frac{\partial}{\partial t}\log\varphi_{\Sigma_t}(z)\right| \mathrm{d}t\lambda(\mathrm{d}z) \\
&= \frac{1}{2}\int_0^1 \int \left|\frac{\partial}{\partial t}\log\varphi_{\Sigma_t}(z)\right| \varphi_{\Sigma_t}(z)\lambda(\mathrm{d}z)\,\mathrm{d}t \\
&\leqslant \frac{1}{2}\int_0^1 2d\lambda_{\min}^{-1}\|\Sigma' - \Sigma\|_F\,\mathrm{d}t \\
&= \frac{d}{\lambda_{\min}}\|\Sigma' - \Sigma\|_F \ .
\end{aligned}$$

□

Proposition 3.6. *Consider a family of MTM transitions $\{P_\theta\}_{\theta\in\Theta}$ with Gaussian random walk marginal proposal densities whose covariances are contained in a compact subset of $\mathcal{C}_d^+$, the cone of symmetric positive-definite matrices. Suppose that the following Lipschitz condition holds: there exists $L < \infty$ such that, for all $x, y \in \mathcal{X}$*

$$\left|A_\theta^{(k)}(y|x) - A_{\theta'}^{(k)}(y|x)\right| \leqslant L\|\theta - \theta'\|_2 \ . \tag{S4}$$

Then, there exists $C < \infty$ such that, for all functions $f : \mathcal{X} \to \mathbb{R}$ with $\|f\|_1 < \infty$,

$$\|P_\theta f - P_{\theta'} f\|_1 \leqslant C\|f\|_1\|\theta - \theta'\|_2 \ .$$

In particular, $\{P_\theta\}_{\theta\in\Theta}$ is V-Lipschitz for $V \equiv 1$.

Proof. By definition, we have

$$\|P_\theta f - P_{\theta'} f\|_1 = \sup_{x\in\mathcal{X}} |P_\theta f(x) - P_{\theta'} f(x)| \ .$$

For $\|f\|_1 < \infty$, we have

$$\frac{|f(x)|}{\|f\|_1} \leqslant 1 \ , \qquad \forall\, x \in \mathcal{X} \ . \tag{S5}$$

We first consider the following development of $P_\theta f(x) - P_{\theta'} f(x)$:

$$\begin{aligned}
P_\theta f(x) - P_{\theta'} f(x) &= \int_{\mathcal{X}} f(y) P_\theta(y|x)\lambda(\,\mathrm{d}y) - \int_{\mathcal{X}} f(y) P_{\theta'}(y|x)\lambda(\,\mathrm{d}y) \\
&= \int_{\mathcal{X}} f(y)\left[P_\theta(y|x) - P_{\theta'}(y|x)\right]\lambda(\,\mathrm{d}y) \\
&= \int_{\mathcal{X}} f(y)\left[R_\theta(x)\delta_x(y) + p_\theta(y|x) - R_{\theta'}(x)\delta_x(y) - p_{\theta'}(y|x)\right]\lambda(\,\mathrm{d}y) \\
&= \int_{\mathcal{X}} f(y)\left[\left(R_\theta(x) - R_{\theta'}(x)\right)\delta_x(y) + \left(p_\theta(y|x) - p_{\theta'}(y|x)\right)\right]\lambda(\,\mathrm{d}y)\ .
\end{aligned}$$

Then, using properties of integrals and the inequality (S5), we find

$$\begin{aligned}
\frac{|P_\theta f(x) - P_{\theta'} f(x)|}{\|f\|_1} &\leqslant \int_{\mathcal{X}} \frac{f(y)}{\|f\|_1}\left[|R_\theta(x) - R_{\theta'}(x)|\delta_x(y) + |p_\theta(y|x) - p_{\theta'}(y|x)|\right]\lambda(\,\mathrm{d}y) \\
&\leqslant \int_{\mathcal{X}} \left[|R_\theta(x) - R_{\theta'}(x)|\delta_x(y) + |p_\theta(y|x) - p_{\theta'}(y|x)|\right]\lambda(\,\mathrm{d}y) \\
&= |R_\theta(x) - R_{\theta'}(x)| + \int_{\mathcal{X}} |p_\theta(y|x) - p_{\theta'}(y|x)|\lambda(\,\mathrm{d}y)\ .
\end{aligned}$$

Now, we note that

$$\begin{aligned}
|R_\theta(x) - R_{\theta'}(x)| &= \left|1 - \int_{\mathcal{X}} p_\theta(y|x)\lambda(\,\mathrm{d}y) - 1 + \int_{\mathcal{X}} p_{\theta'}(y|x)\lambda(\,\mathrm{d}y)\right| \\
&= \left|\int_{\mathcal{X}} \left[p_{\theta'}(y|x) - p_\theta(y|x)\right]\lambda(\,\mathrm{d}y)\right| \\
&\leqslant \int_{\mathcal{X}} |p_\theta(y|x) - p_{\theta'}(y|x)|\lambda(\,\mathrm{d}y)\ ,
\end{aligned}$$

which allows us to bound

$$|P_\theta f(x) - P_{\theta'} f(x)| \leqslant 2\|f\|_1 \int_{\mathcal{X}} |p_\theta(y|x) - p_{\theta'}(y|x)|\lambda(\,\mathrm{d}y)\ . \tag{S6}$$

Rearranging terms, we can write

$$\begin{aligned}
p_\theta(y|x) - p_{\theta'}(y|x) &= \sum_{k=1}^{K} A_\theta^{(k)}(y|x) q_\theta^{(k)}(y|x) - \sum_{k=1}^{K} A_{\theta'}^{(k)}(y|x) q_{\theta'}^{(k)}(y|x) \\
&= \sum_{k=1}^{K} \left[A_\theta^{(k)} q_\theta^{(k)} - A_{\theta'}^{(k)} q_{\theta'}^{(k)}\right](y|x) \\
&= \sum_{k=1}^{K} \left[A_\theta^{(k)} q_\theta^{(k)} - A_\theta^{(k)} q_{\theta'}^{(k)} + A_\theta^{(k)} q_{\theta'}^{(k)} - A_{\theta'}^{(k)} q_{\theta'}^{(k)}\right](y|x) \\
&= \sum_{k=1}^{K} \left[A_\theta^{(k)}\left(q_\theta^{(k)} - q_{\theta'}^{(k)}\right) + \left(A_\theta^{(k)} - A_{\theta'}^{(k)}\right) q_{\theta'}^{(k)}\right](y|x)\ .
\end{aligned} \tag{S7}$$

In that last expression, it is possible to directly bound the first term. Indeed, $A_\theta^{(k)} \leqslant 1$ and $q_\theta^{(k)} - q_{\theta'}^{(k)}$ may be bounded by Proposition 3.5:

$$\begin{aligned}
\int_{\mathcal{X}} \left|\sum_{k=1}^K A_\theta^{(k)} \left(q_\theta^{(k)} - q_{\theta'}^{(k)}\right)(y|x)\right| \lambda(\mathrm{d}y) &\leqslant \sum_{k=1}^K \int_{\mathcal{X}} 1 \cdot \left|q_\theta^{(k)}(y|x) - q_{\theta'}^{(k)}(y|x)\right| \lambda(\mathrm{d}y) \\
&\leqslant \sum_{k=1}^K \int_{\mathcal{X}} |\varphi_{\Sigma^{(k)}}(z) - \varphi_{\Sigma'^{(k)}}(z)| \lambda(\mathrm{d}z) \\
&\leqslant \frac{d}{\lambda_{\min}} \sum_{k=1}^K \left\|\Sigma^{(k)} - \Sigma'^{(k)}\right\|_F \\
&\leqslant \frac{d}{\lambda_{\min}} \sum_{k=1}^K \|\theta - \theta'\|_2 , \\
&= \frac{Kd}{\lambda_{\min}} \|\theta - \theta'\|_2 , \qquad \text{(S8)}
\end{aligned}$$

where $\lambda_{\min} > 0$ is the smallest eigenvalue over covariances in $\mathcal{K}$ compact. The second term of (S7) can be bounded using the Lipschitz condition on $A_\theta^{(k)}$:

$$\int_{\mathcal{X}} \left|A_\theta^{(k)} - A_{\theta'}^{(k)}\right| q_{\theta'}^{(k)}(y|x)\lambda(\mathrm{d}y) \leqslant \int_{\mathcal{X}} L\|\theta - \theta'\|_2 q_{\theta'}^{(k)}(y|x)\lambda(\mathrm{d}y) = L\|\theta - \theta'\|_2 . \qquad \text{(S9)}$$

Combining (S8) and (S9), we can finally bound the integral in (S6). Indeed, we find

$$\begin{aligned}
\int_{\mathcal{X}} |p_\theta(y|x) - p_{\theta'}(y|x)| \lambda(\mathrm{d}y) &\leqslant \sum_{k=1}^K \int_{\mathcal{X}} A_\theta^{(k)} \left|q_\theta^{(k)} - q_{\theta'}^{(k)}\right|(y|x)\lambda(\mathrm{d}y) \\
&\quad + \sum_{k=1}^K \int_{\mathcal{X}} \left|A_\theta^{(k)} - A_{\theta'}^{(k)}\right| q_{\theta'}^{(k)}(y|x)\lambda(\mathrm{d}y) \\
&\leqslant \frac{Kd}{\lambda_{\min}} \|\theta - \theta'\|_2 + KL\|\theta - \theta'\|_2 \\
&\leqslant K\left(\frac{d}{\lambda_{\min}} + L\right) \|\theta - \theta'\|_2 ,
\end{aligned}$$

which concludes the proof. □

The Lipschitz condition on the acceptance probability (S4) highly depends on the specific instance of the aMTM algorithm implemented. In particular, the expression $A_\theta^{(k)}$ involves the weight function $w_\theta^{(k)}$, the acceptance probability $\alpha_\theta^{(k)}$ and the conditional densities $q_\theta^{(k)}$. Hence, the choices of weights and covariance structure among candidates influence how we can verify (S4) so we must resort to a case-by-case approach. Fontaine (2019, Section 5.5.2) contains all the details so we only report the general ideas here.

First, when the weight function is independent of θ (e.g. proportional to the target density) and candidates are chosen to be independent, then we do not require any additional assumption.

When the weight function depends on θ, then we can extract some general sufficient conditions. We require that the weight function and acceptance probability be Lipschitz in θ uniformly over their arguments; for any $\theta \neq \theta' \in \Theta^2$,

$$\sup_{y,y^{(-k)},x,\theta\neq\theta'} \frac{\left|\bar{w}_\theta(y,y^{(-k)}|x) - \bar{w}_{\theta'}(y,y^{(-k)}|x)\right|}{\|\theta-\theta'\|_2} < \infty \ ,$$

$$\sup_{y,y^{(-k)},x,x_*^{(k)},\theta\neq\theta'} \frac{\left|\alpha_\theta(y,y^{(-k)}|x,x_*^{(k)}) - \alpha_{\theta'}(y,y^{(-k)}|x,x_*^{(k)})\right|}{\|\theta-\theta'\|_2} < \infty \ .$$

In the independent case, such conditions are easily verified by choosing functions that have continuous gradients and by supposing Θ compact and convex. In the extremely antithetic case, we can use similar arguments, but the details are more tedious since the conditional densities $q^{(k)}$ lie in some strict subspace of $\mathcal{X}^{K-1}$. When candidates are deterministic (e.g. RQMC or common random variable), these conditions become simpler as the conditional densities become degenerate: we can then drop the dependence on $y^{(-k)}$ and on $x_*^{(-k)}$.

3.3.3. Bounded updates

The set of parameters is given by

$$\theta = \left(\theta^{(1)}, \ldots, \theta^{(K)}\right)$$

where, in general, each component consists of a moving average, a covariance and a scale factor:

$$\theta^{(k)} = \left(\mu^{(k)}, \Sigma^{(k)}, l^{(k)}\right), \qquad k = 1, \ldots, K,$$

where $l^{(K)} = \log \lambda^{(K)}$. We denote by $\|\cdot\|_2$ the $\mathcal{L}_2$-norm; for elements of θ that are matrices, the contribution to $\|\theta\|$ will thus be the Frobenius norm $\|\cdot\|_F$ which corresponds to the $\mathcal{L}_2$-norm of the vectorized matrix. At iteration n, the available information to be used by the adaptation function is given by

$$\Xi_n = \left(k_n, y^{(1:K)}, x_*^{(1:K)}\right) \ .$$

Hence, we can describe the update function as

$$H_\theta\left(\Xi_n\right) = \begin{pmatrix} H_\theta^{(1)}\left(\Xi_n\right) \\ \vdots \\ H_\theta^{(K)}\left(\Xi_n\right) \end{pmatrix} \ ,$$

where $H_\theta^{(k)}$ corresponds to the update of $\theta^{(k)}$ as introduced in (4.1). Then

$$H_\theta^{(k)}(\Xi_n) = \begin{pmatrix} H_{\mu,\Sigma}^{(k)}(\Xi_n) \\ H_{l,\alpha}^{(k)}(\Xi_n) \end{pmatrix} ,$$

where $H_{\mu,\Sigma}^{(k)}$ corresponds to the (joint) update of $\mu^{(k)}$ and $\Sigma^{(k)}$, and where $H_{l,\alpha}^{(k)}$ corresponds to the update of $l^{(k)}$ using the acceptance probability.

Bounding $H_{l,\alpha}^{(k)}$ is easily achieved. In the ASWAM case, we have

$$H_{l,\alpha}^{(k)}(\Xi_n) = \mathbb{I}(\{k_n = k\}) \left[\alpha_\theta\left(y; y^{(-k)}|x_n; x_*^{(-k)}\right) - \alpha_*\right] ,$$

which can be bounded by

$$\left|H_{l,\alpha}^{(k)}(\Xi_n)\right| \leqslant \mathbb{I}(\{k_n = k\}) \left|\alpha_\theta\left(y; y^{(-k)}|x_n; x_*^{(-k)}\right) - \alpha_*\right| \leqslant 1 .$$

For AM or RAM updates, $H_{l,\alpha}^{(k)} = 0$.

Lemma 3.1. *Let $H_{\mu,\Sigma}^{(k)}$ denote the AM or ASWAM update of $(\mu^{(k)}, \Sigma^{(k)})$. Then, if the sample space $\mathcal{X}$ and parameter space Θ are both compact,*

$$\sup_{\theta\in\mathcal{K}} \left\|H_{\mu,\Sigma}^{(k)}\right\|_1 < \infty .$$

Proof. The AM and ASWAM updates are such that

$$H_{\mu,\Sigma}^{(k)}(\Xi_n) = \mathbb{I}(\{k_n = k\}) \begin{pmatrix} x_{n+1} - \mu^{(k)} \\ (x_{n+1} - \mu^{(k)})(x_{n+1} - \mu^{(k)})^\top - \Sigma^{(k)} \end{pmatrix} .$$

Thus, $H_{\mu,\Sigma}^{(k)}$ only depends on θ, k_n, and x_{n+1}.

By definition, we have

$$\left\|H_{\mu,\Sigma}^{(k)}\right\|_1 = \sup_{(x_{n+1},k_n)\in\mathcal{X}\times\{1,\ldots,K\}} \left\|H_{\mu,\Sigma}^{(k)}(\Xi_n)\right\|_2 .$$

Obviously, the supremum over $k_n \in \{1, \ldots, K\}$ is attained for $k_n = k$ because of the term $\mathbb{I}(\{k_n = k\})$. Hence, we find the following bound

$$\begin{aligned}\left\|H_{\mu,\Sigma}^{(k)}\right\|_2 &\leqslant \left\|x_{n+1} - \mu^{(k)}\right\|_2 + \left\|(x_{n+1} - \mu^{(k)})(x_{n+1} - \mu^{(k)})^\top - \Sigma^{(k)}\right\|_F \\ &\leqslant \|x_{n+1}\|_2 + \left\|\mu^{(k)}\right\|_2 + \left\|x_{n+1}x_{n+1}^\top\right\|_F \\ &\quad + 2\left\|\mu^{(k)}x_{n+1}^\top\right\|_F + \left\|\mu^{(k)}\mu^{(k)\top}\right\|_2 + \left\|\Sigma^{(k)}\right\|_F .\end{aligned}$$

Assuming $\mathcal{X}$ and Θ compact, then x_{n+1}, $\mu^{(k)}$, and $\Sigma^{(k)}$ are all bounded so that $\left\|H_{\mu,\Sigma}^{(k)}\right\|_2$ is uniformly bounded for $(x, k) \in \mathcal{X} \times \{1, \ldots, K\}$, as well as for $\theta \in \mathcal{K}$. □

For the RAM update, we rewrite it as a Robbins-Monro recursion:

$$\begin{aligned}\Sigma_{n+1}^{(k)} &= S_n^{(k)}\left(I_d + \gamma_{n+1}\,\mathbb{I}\left(\{k_n = k\}\right)\left[\alpha_\theta(y;y^{(-k)}|x_n;x_*^{(-k)}) - \alpha_*\right]\frac{u_{n+1}u_{n+1}^\top}{\|u_{n+1}\|_2^2}\right)S_n^{(k)\top}\\ &= S_n^{(k)}S_n^{(k)\top} + \gamma_{n+1}\,\mathbb{I}\left(\{k_n = k\}\right)S_n^{(k)}\left(\left[\alpha_\theta(y;y^{(-k)}|x_n;x_*^{(-k)}) - \alpha_*\right]\frac{u_{n+1}u_{n+1}^\top}{\|u_{n+1}\|_2^2}\right)S_n^{(k)\top}\\ &=: \Sigma_n^{(k)} + \gamma_{n+1}H_{\Sigma_n}^{(k)}\left(\Xi_n\right)\ ,\end{aligned}$$

where

$$H_{\Sigma_n}^{(k)}\left(\Xi_n\right) = \mathbb{I}\left(\{k_n = k\}\right)S_n^{(k)}\left(\left[\alpha_\theta(y;y^{(-k)}|x_n;x_*^{(-k)}) - \alpha_*\right]\frac{u_{n+1}u_{n+1}^\top}{\|u_{n+1}\|_2^2}\right)S_n^{(k)\top}\ ,$$

with $u_{n+1} = y - x_n$ and $\Sigma^{(k)} = S^{(k)}S^{(k)\top}$.

Lemma 3.2. *Let $H_\Sigma^{(k)}$ denote the RAM update function of $\Sigma^{(k)}$. Then, if the sample space $\mathcal{X}$ and parameter space Θ are both compact,*

$$\sup_{\theta\in\mathcal{K}}\left\|H_\Sigma^{(k)}\right\|_1 < \infty\ .$$

Proof. The norm of $H_\Sigma^{(k)}$ can be bounded as follows:

$$\begin{aligned}\left\|H_\Sigma^{(k)}\left(\Xi_n\right)\right\|_2 &\leqslant \left\|S^{(k)}\left(\left[\alpha_\theta(y;y^{(-k)}|x_n;x_*^{(-k)}) - \alpha_*\right]\frac{u_{n+1}u_{n+1}^\top}{\|u_{n+1}\|_2^2}\right)S^{(k)\top}\right\|_2\\ &\leqslant \left\|S^{(k)}\right\|_2\left\|\left[\alpha_\theta(y;y^{(-k)}|x_n;x_*^{(-k)}) - \alpha_*\right]\frac{u_{n+1}u_{n+1}^\top}{\|u_{n+1}\|_2^2}\right\|_2\left\|S^{(k)\top}\right\|_2\\ &\leqslant \left\|S^{(k)}\right\|_2\frac{\|u_{n+1}u_{n+1}^\top\|_2}{\|u_{n+1}\|_2^2}\left\|S^{(k)}\right\|_2\\ &\leqslant \left\|S^{(k)}\right\|_2\frac{\|u_{n+1}\|_2^2}{\|u_{n+1}\|_2^2}\left\|S^{(k)}\right\|_2 = \left\|S^{(k)}\right\|_2^2\ .\end{aligned}$$

For Θ compact, we find $\|S^{(k)}\|_2$ to be uniformly bounded, which implies that $\left\|H_\Sigma^{(k)}\right\|_1$ is uniformly bounded for $\theta \in \Theta$ as well as for all Ξ_n. That is, $\sup_{\theta\in\mathcal{K}}\left\|H_\Sigma^{(k)}\right\|_1 < \infty$. □

3.3.4. Continuity of the convergence metric

Recall the metric used to compare the iterated transition to the target density:

$$\Delta_n(x,\theta) = \|P_\theta^n(\cdot|x) - \pi(\cdot)\|_{\mathrm{TV}}\ .$$

Lemma 3.3. *Let $F : \mathcal{W} \to \mathbb{R}$ be a function defined by*

$$F(w) = \int_{\mathcal{T}} f(w,t)\lambda(\,\mathrm{d}t)\ ,$$

where $f : \mathcal{W} \times \mathcal{T} \to \mathbb{R}$ is continuous w.r.t. (w,t) and where λ denotes the Lebesgue measure on $\mathbb{R}$. Suppose there exists a function $g : \mathcal{T} \to \mathbb{R}$ such that $|f(w,t)| \leqslant |g(t)|$ for all $(w,t) \in \mathcal{W} \times \mathcal{T}$ with $\lambda(|g|) < \infty$. Then, F is a continuous function of w on the whole of $\mathcal{W}$.

Proof. The function F is continuous on the whole of $\mathcal{W}$ if and only if $\lim_{n\to\infty} F(w_n) = F(w)$ for any sequence $w_n \to w$. Then, let $\{w_n\}_{n\in\mathbb{N}} \subseteq \mathcal{W}$ be an arbitrary sequence with $w_n \to w \in \mathcal{W}$ and define, for all $n \in \mathbb{N}$, $f_n : \mathcal{T} \to \mathbb{R}$ by $f_n(t) = f(w_n, t)$. By the continuity of f w.r.t. w, we know that $f_n(t) \to f(w,t)$ point-wise. By hypothesis, we have

$$|f_n(t)| = |f(w_n,t)| \leqslant |g(t)|\ , \qquad n \in \mathbb{N}\ .$$

Now, write

$$\lim_{n\to\infty} F(w_n) = \lim_{n\to\infty} \int_{\mathcal{T}} f(w_n,t)\lambda(\,\mathrm{d}t) = \lim_{n\to\infty} \int_{\mathcal{T}} f_n(t)\lambda(\,\mathrm{d}t) = \lim_{n\to\infty} \lambda(f_n(\cdot))\ .$$

By the Monotone Convergence Theorem, we find

$$\lim_{n\to\infty} F(w_n) = \lambda\left(\lim_{n\to\infty} f_n(\cdot)\right) = \lambda\left(f(w,\cdot)\right) = F(w)\ ,$$

which concludes the proof. □

Lemma 3.4. *Let P_θ be a MTM transition using a given set parameters $\theta \in \Theta$. Then, the acceptance probability through candidate k from the current state x to some other state y, $A_\theta^{(k)}(y|x)$, is a continuous function of (x,y,θ) assuming that each of $q_\theta^{(-k)}$, $\bar{w}^{(k)}$ and α_{MTM} are continuous functions of their arguments and parameters, and that the conditional densities $q_\theta^{(-k)}$ are uniformly bounded above by some integrable function $q^+ : \mathcal{X}^{K-1} \to \mathbb{R}_{\geqslant 0}$.*

Proof. This result is a direct consequence of Lemma 3.3. The complete argument may be found in Fontaine (2019, Lemma 5.11). □

Lemma 3.5. *Let P_θ be a MTM transition using a given set parameters $\theta \in \Theta$. Then, the integrated acceptance probability through candidate k from the current state x, $\overline{A}_\theta^{(k)}(x)$, is a continuous function of (x,θ) assuming that $A_\theta^{(k)}(y|x)$ is a continuous function of (x,y,θ) and assuming that each $q_\theta^{(k)}(y|x)$ is a density, with respect to the Lebesgue measure on $\mathbb{R}^d$, such that there exists an integrable function $q^+ : \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ with $q_\theta^{(k)}(y|x) \leqslant q^+(y)$ uniformly for (x,θ,k). Furthermore, the rejection probability $R_\theta(x)$ is also a continuous function of (x,θ).*

Proof. This result is a direct consequence of Lemma 3.3. The complete argument may be found in Fontaine (2019, Lemma 5.12). □

For a Markov transition taking the form of a MH density, i.e.

$$P(\,\mathrm{d}y|x) = p(y|x)\lambda(\,\mathrm{d}y) + R(x)\delta_x(\,\mathrm{d}y)\ , \tag{S10}$$

we can write the iterated transition using the following recursion,

$$P^n(\,\mathrm{d}y|x) = p^n(y|x)\lambda(\,\mathrm{d}y) + R^n(x)\delta_x(\,\mathrm{d}y),$$

where

$$p^n(y|x) = \int_{\mathcal{X}} p^{n-1}(y|z)p(z|x)\lambda(\,\mathrm{d}z)\ ,$$

with the convention $p^0(y|x) = \delta_x(y)$.

Corollary 3.1. *Under the setup and conditions of Lemma 3.5, the iterated MTM transition, $p^n_\theta(y|x)$, is a continuous function of (x, y, θ) for all $n \in \mathbb{N}$.*

Proof. We proceed by induction over $n \geqslant 1$ to show that $p^n_\theta(y|x)$ is continuous with respect to (x, y, θ) and is uniformly bounded by $K^n\bar{q}^{n-1}q^+(y)$, where $\bar{q} = \sup_{\mathcal{X}} q^+ < \infty$.

For $n = 1$, we have

$$p^1_\theta(y|x) = \int_{\mathcal{X}} \delta_z(y)p_\theta(z|x)\lambda(\,\mathrm{d}z) = p_\theta(y|x) = \sum_{k=1}^{K} A^{(k)}_\theta(y|x)q^{(k)}_\theta(y|x)\ ,$$

which is a sum of products of continuous functions and is therefore continuous. The uniform bound is direct:

$$\left|p^1_\theta(y|x)\right| \leqslant \sum_{k=1}^{K} \left|A^{(k)}_\theta(y|x)q^{(k)}_\theta(y|x)\right| \leqslant \sum_{k=1}^{K} 1\cdot q^+(y) = K\cdot q^+(y) = K^1\bar{q}^{1-1}\cdot q^+(y)\ .$$

For $n > 1$, we suppose that $p^{n-1}_\theta(y|x)$ is continuous and uniformly bounded by $K^{n-1}\bar{q}^{n-2}q^+(y)$. We use Lemma 3.3; to this end, we let

$$F(w) = p^n_\theta(y|x) = \int_{\mathcal{X}} p^{n-1}_\theta(y|z)p_\theta(z|x)\lambda(\,\mathrm{d}z)\ ,$$

the variables $(x, y, \theta) = w \in \mathcal{W}$ with $\mathcal{W} = \mathcal{X}^2 \times \Theta$, the integrated variable $z = t \in \mathcal{T}$ with $\mathcal{T} = \mathcal{X}$, and the integrand

$$f(w, t) = p^{n-1}_\theta(y|z)p_\theta(z|x)\ .$$

Since $p_\theta^{n-1}(y|z)$ is continuous w.r.t. (y,z,θ) by induction hypothesis and since $p_\theta(z|x)$ is continuous w.r.t. (x,z,θ) by assumption (see the case $n=1$), we find that f is a continuous function of all its arguments. Also, the induction hypothesis implies the following uniform bound

$$\left|p_\theta^{n-1}(y|z)\right| \leqslant K^{n-1}\bar{q}^{n-2}q^+(y) \leqslant K^{n-1}\bar{q}^{n-2}\sup_{\mathcal{X}} q^+ = K^{n-1}\bar{q}^{n-1} \ .$$

Hence, we find

$$|f(w,t)| \leqslant \left|p_\theta^{n-1}(y|z)\right||p_\theta(z|x)| \leqslant K^n\bar{q}^{n-1}q^+(z) =: g(t) \ .$$

Since q^+ is integrable and $K, \bar{q} < \infty$, we find that g is integrable for each fixed n. Finally, Lemma 3.3 implies that F is continuous w.r.t. w, that is, $p_\theta^n(y|x)$ is continuous w.r.t. $(x,y,\theta) \in \mathcal{X}^2 \times \Theta$ for each fixed n. □

Theorem 3.2. *Let $\{P_\theta\}_{\theta\in\Theta}$ be a family of MTM transitions indexed by its set of parameters θ, and suppose that the target density π and each P_θ satisfy the conditions of Proposition 3.2. Further suppose that Θ is compact and that the resulting chain $\{X_n\}_{n\in\mathbb{N}}$ is bounded in probability. Then, if the conditions of Lemma 3.5 hold, the adaptive chain satisfies the bounded convergence condition.*

Proof. We use a result by Craiu et al. (2015, Proposition 23) restated as Theorem 4.1 in the main text.

All conditions of the result are verified except the continuity of $(x,\theta) \mapsto \Delta_n(x,\theta)$. Indeed, the MTM transitions all admit π as their invariant distribution because of the detailed balance condition (Proposition 3.1). They are also ergodic with respect to π by Proposition 3.2 and then Harris-ergodic with respect to π by Proposition 3.3.

To verify the continuity of Δ_n, we proceed in a similar fashion to Roberts and Rosenthal (2007, Corollary 11) in the MH case. We develop Δ_n using the decomposition of the iterated transition (S10):

$$\begin{aligned}
\Delta_n(x,\theta) &= \|P_\theta^n(\cdot|x) - \pi(\cdot)\|_{\mathrm{TV}} \\
&= \sup_{B\in\mathcal{B}(\mathcal{X})} |P_\theta^n(B|x) - \pi(B)| \\
&= \sup_{B\in\mathcal{B}(\mathcal{X})} \left|\int_B P_\theta^n(\,\mathrm{d}y|x) - \int_B \pi(\,\mathrm{d}y)\right| \\
&= \sup_{B\in\mathcal{B}(\mathcal{X})} \left|R_\theta^n(x)\delta_x(B) + \int_B p_\theta^n(\,\mathrm{d}y|x) - \int_B \pi(\,\mathrm{d}y)\right| \\
&= R_\theta^n(x) + \frac{1}{2}\int_{\mathcal{X}} |p_\theta^n(y|x) - \pi(y)|\lambda(\,\mathrm{d}y) \ .
\end{aligned}$$

By inspection of the last expression, we can show that Δ_n is indeed a continuous function of its arguments.

By Lemma 3.5, we know that $R_\theta^n(x)$ is a continuous function of (x,θ).

We then use Lemma 3.3 to show that the integral is indeed continuous w.r.t. (x, θ). By Corollary 3.1, we know that $p_\theta^n(y|x)$ is continuous w.r.t. (x, y, θ). Since π is assumed to be a density w.r.t. the Lebesgue measure, we have that $|p_\theta^n(y|x) - \pi(y)|$ is continuous w.r.t. (x, y, θ). Thus, we only need a (x, θ)-uniform and integrable bound on $|p_\theta^n(y|x) - \pi(y)|$. By the triangle inequality, we have

$$|p_\theta^n(y|x) - \pi(y)| \leqslant p_\theta^n(y|x) + \pi(y) \ .$$

By Corollary 3.1, we find a uniform and integrable bound on the first term; the target density is independent of (x, θ) and integrable:

$$|p_\theta^n(y|x) - \pi(y)| \leqslant K^n \bar{q}^{n-1} q^+(y) + \pi(y) \in L_1(\lambda) \ .$$

Hence, all conditions of Lemma 3.3 are verified, which implies that $\int_{\mathcal{X}} |p_\theta^n(y|x) - \pi(y)| \lambda(\mathrm{d}y)$ is continuous w.r.t. (x, θ). We conclude that Δ_n is continuous w.r.t. (x, θ) since it corresponds to a linear combination of continuous functions. $\square$

References

Andrieu, C. and Moulines, É. (2006). On the Ergodicity Properties of Some Adaptive MCMC Algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070

Casarin, R., Craiu, R. and Leisen, F. (2013). Interacting Multiple Try Algorithms with Different Proposal Distributions. *Stat. Comput.* **23** 185–200. MR3016937

Craiu, R. V., Gray, L., Ł atuszyński, K., Madras, N., Roberts, G. O. and Rosenthal, J. S. (2015). Stability of Adversarial Markov Chains, with an Application to Adaptive MCMC Algorithms. *Ann. Appl. Probab.* **25** 3592–3623. MR3404645

Fontaine, S. (2019). MCMC adaptatifs à essais multiples, Master's thesis, Department of Mathematics and Statistics, University of Montreal.

Haario, H., Saksman, E. and Tamminen, J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli* **7** 223–242. MR1828504

Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*, Second ed. Cambridge University Press, Cambridge With a prologue by Peter W. Glynn. MR2509253

Nummelin, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. *Cambridge Tracts in Mathematics* **83**. Cambridge University Press, Cambridge. MR776608

Petersen, K. B. and Pedersen, M. S. (2008). The Matrix Cookbook. *Technical University of Denmark* **7** 510.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, second ed. *Springer Texts in Statistics*. Springer-Verlag, New York. MR2080278

Roberts, G. O. and Rosenthal, J. S. (2007). Coupling and Ergodicity of Adaptive Markov Chain Monte Carlo Algorithms. *J. Appl. Probab.* **44** 458–475. MR2340211

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Ann. Statist.* **22** 1701–1762. With discussion and a rejoinder by the author. MR1329166