Non-reversible jump algorithms for Bayesian nested model selection — Supplementary material

Philippe Gagnon
and
Arnaud Doucet
Department of Statistics, University of Oxford

August 12, 2020

We present in Section 1 the proofs of Proposition 1, Theorem 1 and Corollary 1 of our paper. In Section 2, weak convergence results for the ideal samplers as the size of the state-space increases are presented. The details about the multiple change-point example of our paper are provided in Section 3

1 Proofs

Proof of Proposition 1. It suffices to prove that the probability to reach the state $k', \mathbf{y}_{k'} \in A, \nu'$ in one step is equal to the probability of this state under the target:

$$\sum_{k,\nu} \int \pi(k, \mathbf{x}_k) \times (1/2) \left(\int_A P((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu'))\, d\mathbf{y}_{k'} \right) d\mathbf{x}_k = \int_A \pi(k', \mathbf{y}_{k'}) \times (1/2)\, d\mathbf{y}_{k'}, \tag{1}$$

where P is the transition kernel. Note that we abuse notation here by denoting the measure $d\mathbf{y}_{k'}$ on the left-hand side (LHS) given that we in fact use the vector $\mathbf{u}_{k\mapsto k'}$ when switching models, which often do not have the same dimension as $\mathbf{y}_{k'}$.

We consider two distinct events: a model switch is proposed, that we denote S, or a parameter update is proposed (therefore denoted S^c). We know that the probabilities of these events are $1-\tau$ and τ, respectively, regardless of the current state of the Markov chain. We rewrite the LHS of (1) as

$$
\begin{aligned}
&\sum_{k,\nu} \int \pi(k, \mathbf{x}_k) \times (1/2) \left(\int_A P((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu'))\, d\mathbf{y}_{k'} \right) d\mathbf{x}_k \\
&\quad = \mathbb{P}(S) \times (1/2) \sum_{k,\nu} \int_A \int \pi(k, \mathbf{x}_k)\, P((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu') \mid S)\, d\mathbf{x}_k\, d\mathbf{y}_{k'} \\
&\quad\quad + \mathbb{P}(S^c) \times (1/2) \sum_{k,\nu} \int_A \int \pi(k, \mathbf{x}_k)\, P((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu') \mid S^c)\, d\mathbf{x}_k\, d\mathbf{y}_{k'}, \qquad (2)
\end{aligned}
$$

using Fubini's theorem. We analyse the two terms separately. We know that

$$
P((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu') \mid S^c) = \delta_{(k,\nu)}(k', \nu')\, P_{S^c}(\mathbf{x}_{k'}, \mathbf{y}_{k'}),
$$

where P_{S^c} is the transition kernel associated with the method used to update the parameters. Therefore, the second term on the right-hand side (RHS) of (2) is equal to

$$
\begin{aligned}
&\mathbb{P}(S^c) \times (1/2) \sum_{k,\nu} \int_A \int \pi(k, \mathbf{x}_k)\, P((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu') \mid S^c)\, d\mathbf{x}_k\, d\mathbf{y}_{k'} \\
&\quad = \mathbb{P}(S^c)\, \pi(k') \times (1/2) \int \pi(\mathbf{x}_{k'} \mid k') \left(\int_A P_{S^c}(\mathbf{x}_{k'}, \mathbf{y}_{k'})\, d\mathbf{y}_{k'} \right) d\mathbf{x}_{k'}.
\end{aligned}
$$

We also know that P_{S^c} leaves the conditional distribution $\pi(\,\cdot \mid k')$ invariant, implying that

$$
\begin{aligned}
&\mathbb{P}(S^c)\, \pi(k') \times (1/2) \int \pi(\mathbf{x}_{k'} \mid k') \left(\int_A P_{S^c}(\mathbf{x}_{k'}, \mathbf{y}_{k'})\, d\mathbf{y}_{k'} \right) d\mathbf{x}_{k'} \\
&\quad = \mathbb{P}(S^c)\, \pi(k') \times (1/2) \int_A \pi(\mathbf{y}_{k'} \mid k')\, d\mathbf{y}_{k'} = \mathbb{P}(S^c) \int_A \pi(k', \mathbf{y}_{k'}) \times (1/2)\, d\mathbf{y}_{k'}. \qquad (3)
\end{aligned}
$$

For the model switching case (the first term on the RHS of (2)), we use the fact that there is a connection between $P((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu') \mid S)$ and the kernel associated to a specific RJ. Consider that $g(k, k+1) = g(k, k-1)$ for all k and that all other proposal distributions are the same as the NRJ. In this case, $\alpha_{\mathrm{RJ}} = \alpha_{\mathrm{NRJ}}$. Given the reversibility of RJ, the probability to go from model k with parameters in B to model $k+1$ with parameters in A is

$$
\begin{aligned}
&\mathbb{P}(S) \int_B \pi(k, \mathbf{x}_k) \left(\int_A P_{\mathrm{RJ}}((k, \mathbf{x}_k), (k+1, \mathbf{y}_{k+1}) \mid S)\, d\mathbf{y}_{k+1} \right) d\mathbf{x}_k \\
&\quad = \mathbb{P}(S) \int_A \pi(k+1, \mathbf{y}_{k+1}) \left(\int_B P_{\mathrm{RJ}}((k+1, \mathbf{y}_{k+1}), (k, \mathbf{x}_k) \mid S)\, d\mathbf{x}_k \right) d\mathbf{y}_{k+1}, \qquad (4)
\end{aligned}
$$

where P_{RJ} is the transition kernel associated with the RJ. Note that

$$P_{\mathrm{RJ}}((k, \mathbf{x}_k), (k+1, \mathbf{y}_{k+1}) \mid S) = (1/2)\, P((k, \mathbf{x}_k, 1), (k+1, \mathbf{y}_{k+1}, 1) \mid S),$$

given that the difference between both kernels is that in RJ, once it is decided that a model switch is attempted, there is an additional probability of $1/2$ of trying model $k+1$. Analogously, $P_{\mathrm{RJ}}((k+1, \mathbf{y}_{k+1}), (k, \mathbf{x}_k) \mid S) = (1/2)\, P((k+1, \mathbf{y}_{k+1}, -1), (k, \mathbf{x}_k, -1) \mid S)$. Using that and taking B equals the whole parameter (and auxiliary) space in (4), we have

$$\begin{aligned} &\mathbb{P}(S) \int \pi(k, \mathbf{x}_k) \times (1/2) \left(\int_A P((k, \mathbf{x}_k, 1), (k+1, \mathbf{y}_{k+1}, 1) \mid S)\, d\mathbf{y}_{k+1} \right) d\mathbf{x}_k \\ &\quad = \mathbb{P}(S) \int_A \pi(k+1, \mathbf{y}_{k+1}) \times (1/2) \left(\int P((k+1, \mathbf{y}_{k+1}, -1), (k, \mathbf{x}_k, -1) \mid S)\, d\mathbf{x}_k \right) d\mathbf{y}_{k+1}. \end{aligned}$$

We thus analyse the probability to reach $k+1$ with parameters in A and direction $+1$. We know that the only other way of reaching this state (other than coming from k) is by being at $k+1$ with parameters in A and direction -1 and rejecting, which probability is

$$\mathbb{P}(S) \int_A \pi(k+1, \mathbf{y}_{k+1}) \times (1/2) \left(1 - \int P((k+1, \mathbf{y}_{k+1}, -1), (k, \mathbf{x}_k, -1) \mid S)\, d\mathbf{x}_k \right) d\mathbf{y}_{k+1}.$$

Therefore, the total probability to reach $k+1$ with parameters in A and direction $+1$ in one step (given that a model switch is proposed) is

$$\begin{aligned} &\mathbb{P}(S) \int \pi(k, \mathbf{x}_k) \times (1/2) \left(\int_A P((k, \mathbf{x}_k, 1), (k+1, \mathbf{y}_{k+1}, 1) \mid S)\, d\mathbf{y}_{k+1} \right) d\mathbf{x}_k \\ &+ \mathbb{P}(S) \int_A \pi(k+1, \mathbf{y}_{k+1}) \times (1/2) \left(1 - \int P((k+1, \mathbf{y}_{k+1}, -1), (k, \mathbf{x}_k, -1) \mid S)\, d\mathbf{x}_k \right) d\mathbf{y}_{k+1} \\ &\quad = \mathbb{P}(S) \int_A \pi(k+1, \mathbf{y}_{k+1}) \times (1/2)\, d\mathbf{y}_{k+1}. \end{aligned}$$

Combining this with (3) allows to conclude the proof. ■

Proof of Theorem 1. We show that Algorithm 2 converges towards its ideal version as $T \longrightarrow \infty$. As mentioned, for the ideal version, we consider the case where $q_{k \mapsto k'} := \pi(\,\cdot \mid k')$, the conditional distribution of the parameters of model k'. In this case, we set $\mathbf{y}_{k'} := \mathbf{u}_{k \mapsto k'}$ to be the proposal for the parameters of model k', and thus the function $T_{k \mapsto k'}$ to be the identity function.

To show the convergence, we use Theorem 1 in Karr (1975). We thus have to verify three assumptions, and this will allow to conclude that $\{(K, \mathbf{X}_K, \nu)_T(m) : m \in \mathbb{N}\} \Longrightarrow \{(K, \mathbf{X}_K, \nu)_{\text{ideal}}(m) : m \in \mathbb{N}\}$ as $T \longrightarrow \infty$. We focus on the movements involving model switches as the same parameter update schemes are used in both samplers. Here are the three assumptions.

1. The distributions that are used to initialise Algorithm 2 converge towards that used to initialise the ideal NRJ.

This is verified as we assume that the Markov chains produced by both Algorithm 2 and its ideal counterpart start at stationarity, i.e. $(K, \mathbf{X}_K, \nu)_T(0) \sim \pi \otimes \mathcal{U}\{-1, 1\}$ and $(K, \mathbf{X}_K, \nu)_{\text{ideal}}(0) \sim \pi \otimes \mathcal{U}\{-1, 1\}$.

2. For $h \in \bar{\mathcal{C}}^*$ (the space of bounded uniformly continuous functions), we have that

$$P_{\text{ideal}}(k, \mathbf{x}_k, \nu)h := \sum_{k',\nu'} \int h(k', \mathbf{y}_{k'}, \nu') P_{\text{ideal}}((k, \mathbf{x}_k, \nu), (k', \mathbf{y}_{k'}, \nu'))\, d\mathbf{y}_{k'}$$

is a bounded continuous function.

This kernel is such that

$$\begin{aligned} P_{\text{ideal}}(k, \mathbf{x}_k, \nu)h = \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right) \int h(k+\nu, \mathbf{u}_{k\mapsto k+\nu}, \nu)\, \pi(\mathbf{u}_{k\mapsto k+\nu} \mid k+\nu)\, d\mathbf{u}_{k\mapsto k+\nu} \\ + h(k, \mathbf{x}_k, -\nu)\left(1 - 1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right), \end{aligned}$$

which is bounded and continuous.

3. For every $h \in \bar{\mathcal{C}}^*$, the Markov kernel associated with Algorithm 2 $P_T h$ converges towards $P_{\text{ideal}} h$ uniformly on each compact subset of the state-space as $T \longrightarrow \infty$.

We first show the pointwise convergence. Let us denote the conditional joint density of all the random variables involved in the proposal $\mathbf{y}_{k+\nu}^{(T-1)}$ given $(k, \mathbf{x}_k, \nu)$ by

$$q(\mathbf{u}_{k\mapsto k+\nu}^{(0)}, \mathbf{y}_{k+\nu}^{(1:T-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:T-1)}) := q_{k\mapsto k+\nu}(\mathbf{u}_{k\mapsto k+\nu}^{(0)}) \prod_{t=1}^{T-1} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)})),$$

where $K_{k\mapsto k+\nu}^{(t)}$ is a MH kernel reversible with respect to $\rho_{k\mapsto k+\nu}^{(t)}$. We have that

$$\begin{aligned} P_T h(k, \mathbf{x}_k, \nu) := \int h(k+\nu, \mathbf{y}_{k+\nu}^{(T-1)}, \nu)\, q(\mathbf{u}_{k\mapsto k+\nu}^{(0)}, \mathbf{y}_{k+\nu}^{(1:T-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:T-1)}) \\ \times \alpha_{\text{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{y}_{k+\nu}^{(T-1)}))\, d(\mathbf{u}_{k\mapsto k+\nu}^{(0)}, \mathbf{y}_{k+\nu}^{(1:T-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:T-1)}) \end{aligned}$$

$$+ h(k, \mathbf{x}_k, -\nu) \int q(\mathbf{u}^{(0)}_{k\mapsto k+\nu}, \mathbf{y}^{(1:T-1)}_{k+\nu}, \mathbf{u}^{(1:T-1)}_{k+\nu\mapsto k})$$
$$\times (1 - \alpha_{\mathrm{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{y}^{(T-1)}_{k+\nu})))\, d(\mathbf{u}^{(0)}_{k\mapsto k+\nu}, \mathbf{y}^{(1:T-1)}_{k+\nu}, \mathbf{u}^{(1:T-1)}_{k+\nu\mapsto k}).$$

Using the triangle inequality, we thus have that

$$\begin{aligned}
&|P_T h(k, \mathbf{x}_k, \nu) - P_{\mathrm{ideal}} h(k, \mathbf{x}_k, \nu)| \\
&\leq \Bigg| \int h(k+\nu, \mathbf{y}^{(T-1)}_{k+\nu}, \nu)\, q(\mathbf{u}^{(0)}_{k\mapsto k+\nu}, \mathbf{y}^{(1:T-1)}_{k+\nu}, \mathbf{u}^{(1:T-1)}_{k+\nu\mapsto k}) \\
&\qquad\qquad \times \alpha_{\mathrm{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{y}^{(T-1)}_{k+\nu}))\, d(\mathbf{u}^{(0)}_{k\mapsto k+\nu}, \mathbf{y}^{(1:T-1)}_{k+\nu}, \mathbf{u}^{(1:T-1)}_{k+\nu\mapsto k}) \\
&\quad - \int h(k+\nu, \mathbf{u}_{k\mapsto k+\nu}, \nu)\, \pi(\mathbf{u}_{k\mapsto k+\nu} \mid k+\nu) \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right) d\mathbf{u}_{k\mapsto k+\nu} \Bigg| \\
&+ \Bigg| h(k, \mathbf{x}_k, -\nu) \int q(\mathbf{u}^{(0)}_{k\mapsto k+\nu}, \mathbf{y}^{(1:T-1)}_{k+\nu}, \mathbf{u}^{(1:T-1)}_{k+\nu\mapsto k}) \\
&\qquad\qquad \times (1 - \alpha_{\mathrm{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{y}^{(T-1)}_{k+\nu})))\, d(\mathbf{u}^{(0)}_{k\mapsto k+\nu}, \mathbf{y}^{(1:T-1)}_{k+\nu}, \mathbf{u}^{(1:T-1)}_{k+\nu\mapsto k}) \\
&\quad - h(k, \mathbf{x}_k, -\nu) \left(1 - 1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right) \Bigg|.
\end{aligned} \tag{5}$$

We analyse the first absolute value on the RHS. We write the integrals as (conditional) expectations (given $(k, \mathbf{x}_k, \nu)$):

$$\begin{aligned}
&\Bigg| \mathbb{E}\left[h(k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}, \nu)\, \alpha_{\mathrm{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}))\right] \\
&\qquad\qquad - \mathbb{E}\left[h(k+\nu, \mathbf{U}_{k\mapsto k+\nu}, \nu) \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right)\right] \Bigg| \\
&\quad \leq \Bigg| \mathbb{E}\left[h(k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}, \nu)\, \alpha_{\mathrm{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}))\right] \\
&\qquad\qquad - \mathbb{E}\left[h(k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}, \nu) \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right)\right] \Bigg| \\
&\quad + \Bigg| \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right) \mathbb{E}\left[h(k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}, \nu)\right] - \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right) \mathbb{E}\left[h(k+\nu, \mathbf{U}_{k\mapsto k+\nu}, \nu)\right] \Bigg|,
\end{aligned}$$

using again the triangle inequality. We now show that both absolute values on the RHS converge towards 0. For the first one, we have

$$\begin{aligned}
&\Bigg| \mathbb{E}\left[h(k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}, \nu)\, \alpha_{\mathrm{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}))\right] \\
&\qquad\qquad - \mathbb{E}\left[h(k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu}, \nu) \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right)\right] \Bigg| \\
&\quad \leq M\, \mathbb{E}\left| \alpha_{\mathrm{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{Y}^{(T-1)}_{k+\nu})) - \left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right) \right| \longrightarrow 0,
\end{aligned}$$

using that there exists a positive constant M such that $|h| \leq M$ and that $r_{\text{NRJ2}}((k, \mathbf{x}_k), (k+\nu, \mathbf{Y}_{k+\nu}^{(T-1)})) \longrightarrow \pi(k+\nu)/\pi(k)$ in distribution (by assumption). The convergence of the expectation follows from the fact that if a random variable X_n converges towards a constant c in distribution, then $X_n - c$ converges towards 0 in probability and $\mathbb{E}|g(X_n) - g(c)| \longrightarrow 0$ for any bounded uniformly continuous function g ($\min(1, x)$ with $x \geq 0$ is a function having these characteristics). For the second absolute value, we have

$$\left(1 \wedge \frac{\pi(k+\nu)}{\pi(k)}\right) \left|\mathbb{E}\left[h(k+\nu, \mathbf{Y}_{k+\nu}^{(T-1)}, \nu)\right] - \mathbb{E}\left[h(k+\nu, \mathbf{U}_{k\mapsto k+\nu}, \nu)\right]\right|$$
$$\leq \left|\mathbb{E}\left[h(k+\nu, \mathbf{Y}_{k+\nu}^{(T-1)}, \nu)\right] - \mathbb{E}\left[h(k+\nu, \mathbf{U}_{k\mapsto k+\nu}, \nu)\right]\right| \longrightarrow 0,$$

if the (conditional) distribution of $\mathbf{Y}_{k+\nu}^{(T-1)}$ (given $(k, \mathbf{x}_k, \nu)$) converges towards $\pi(\,\cdot \mid k+\nu)$ given that h is a bounded continuous function.

Let us now prove this convergence in distribution. The conditional distribution of $\mathbf{Y}_{k+\nu}^{(T-1)}$ given $(k, \mathbf{x}_k, \nu)$ is written as

$$\mathbb{P}(\mathbf{Y}_{k+\nu}^{(T-1)} \in A \mid k, \mathbf{x}_k, \nu) := \int_{\mathbf{y}_{k+\nu}^{(T-1)} \in A} q_{k\mapsto k+\nu}(\mathbf{u}_{k\mapsto k+\nu}^{(0)})$$
$$\times \prod_{t=1}^{T-1} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))\, d(\mathbf{u}_{k\mapsto k+\nu}^{(0)}, \mathbf{y}_{k+\nu}^{(1:T-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:T-1)})$$
$$= \int_{\mathbf{y}_{k+\nu}^{(T-1)} \in A} q_{k\mapsto k+\nu}(\mathbf{u}_{k\mapsto k+\nu}^{(0)}) \prod_{t=1}^{t^*-1} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))$$
$$\times \prod_{t=t^*}^{T-1} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))\, d(\mathbf{u}_{k\mapsto k+\nu}^{(0)}, \mathbf{y}_{k+\nu}^{(1:T-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:T-1)}).$$

Under Assumption 3, one can show that t^* and T can be chosen such that $(T - t^*)/T$ is small and

$$\left|K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)})) - K_{k\mapsto k+\nu}^{(T)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))\right| < \frac{1}{T - t^*}\,\epsilon,$$

for all $t \geq t^*$ and any $\epsilon > 0$, where $K_{k\mapsto k+\nu}^{(T)}$ is the MH kernel for which $\rho_{k\mapsto k+\nu}^{(T)} := \pi(\,\cdot \mid k+\nu) \otimes q_{k+\nu\mapsto k}$ is used instead in the acceptance probability. One can thus show that

$$\left|\prod_{t=t^*}^{T-1} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)})) - \prod_{t=t^*}^{T-1} K_{k\mapsto k+\nu}^{(T)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))\right| < \epsilon,$$

and therefore,

$$\Bigg| q_{k\mapsto k+\nu}(\mathbf{u}_{k\mapsto k+\nu}^{(0)}) \prod_{t=1}^{t^*-1} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))$$

$$
\begin{aligned}
&\qquad\qquad \times \prod_{t=t^*}^{T-1} K^{(t)}_{k\mapsto k+\nu}((\mathbf{y}^{(t-1)}_{k+\nu}, \mathbf{u}^{(t-1)}_{k+\nu\mapsto k}), (\mathbf{y}^{(t)}_{k+\nu}, \mathbf{u}^{(t)}_{k+\nu\mapsto k})) \\
&- q_{k\mapsto k+\nu}(\mathbf{u}^{(0)}_{k\mapsto k+\nu}) \prod_{t=1}^{t^*-1} K^{(t)}_{k\mapsto k+\nu}((\mathbf{y}^{(t-1)}_{k+\nu}, \mathbf{u}^{(t-1)}_{k+\nu\mapsto k}), (\mathbf{y}^{(t)}_{k+\nu}, \mathbf{u}^{(t)}_{k+\nu\mapsto k})) \\
&\qquad\qquad \times \prod_{t=t^*}^{T-1} K^{(T)}_{k\mapsto k+\nu}((\mathbf{y}^{(t-1)}_{k+\nu}, \mathbf{u}^{(t-1)}_{k+\nu\mapsto k}), (\mathbf{y}^{(t)}_{k+\nu}, \mathbf{u}^{(t)}_{k+\nu\mapsto k}))\Bigg| \\
&< \epsilon.
\end{aligned}
$$

We have that the integral of the two functions in the absolute value converges towards 0 as well as a result of Scheffé's lemma (see Scheffé (1947)):

$$
\begin{aligned}
&\Bigg|\mathbb{P}(\mathbf{Y}^{(T-1)}_{k+\nu} \in A \mid k, \mathbf{x}_k, \nu) - \int_{\mathbf{y}^{(T-1)}_{k+\nu} \in A} q_{k\mapsto k+\nu}(\mathbf{u}^{(0)}_{k\mapsto k+\nu}) \prod_{t=1}^{t^*-1} K^{(t)}_{k\mapsto k+\nu}((\mathbf{y}^{(t-1)}_{k+\nu}, \mathbf{u}^{(t-1)}_{k+\nu\mapsto k}), (\mathbf{y}^{(t)}_{k+\nu}, \mathbf{u}^{(t)}_{k+\nu\mapsto k})) \\
&\qquad \times \prod_{t=t^*}^{T-1} K^{(T)}_{k\mapsto k+\nu}((\mathbf{y}^{(t-1)}_{k+\nu}, \mathbf{u}^{(t-1)}_{k+\nu\mapsto k}), (\mathbf{y}^{(t)}_{k+\nu}, \mathbf{u}^{(t)}_{k+\nu\mapsto k}))\, d(\mathbf{u}^{(0)}_{k\mapsto k+\nu}, \mathbf{y}^{(1:T-1)}_{k+\nu}, \mathbf{u}^{(1:T-1)}_{k+\nu\mapsto k})\Bigg| < \epsilon.
\end{aligned} \tag{6}
$$

We also have that

$$
\begin{aligned}
&\int_{\mathbf{y}^{(T-1)}_{k+\nu} \in A} \prod_{t=t^*}^{T-1} K^{(T)}_{k\mapsto k+\nu}((\mathbf{y}^{(t-1)}_{k+\nu}, \mathbf{u}^{(t-1)}_{k+\nu\mapsto k}), (\mathbf{y}^{(t)}_{k+\nu}, \mathbf{u}^{(t)}_{k+\nu\mapsto k}))\, d(\mathbf{y}^{(t^*:T-1)}_{k+\nu}, \mathbf{u}^{(t^*:T-1)}_{k+\nu\mapsto k}) \\
&\quad \leq \Bigg|\int_{\mathbf{y}^{(T-1)}_{k+\nu} \in A} \prod_{t=t^*}^{T-1} K^{(T)}_{k\mapsto k+\nu}((\mathbf{y}^{(t-1)}_{k+\nu}, \mathbf{u}^{(t-1)}_{k+\nu\mapsto k}), (\mathbf{y}^{(t)}_{k+\nu}, \mathbf{u}^{(t)}_{k+\nu\mapsto k}))\, d(\mathbf{y}^{(t^*:T-1)}_{k+\nu}, \mathbf{u}^{(t^*:T-1)}_{k+\nu\mapsto k}) \\
&\qquad - \mathbb{P}_{\rho^{(T)}_{k\mapsto k+\nu}}(\mathbf{Y}^{(T-1)}_{k+\nu} \in A)\Bigg| + \mathbb{P}_{\rho^{(T)}_{k\mapsto k+\nu}}(\mathbf{Y}^{(T-1)}_{k+\nu} \in A),
\end{aligned} \tag{7}
$$

where $\mathbb{P}_{\rho^{(T)}_{k\mapsto k+\nu}}$ is the probability measure using the density $\rho^{(T)}_{k\mapsto k+\nu}$. We choose t^* and T such that the absolute value above is smaller than ϵ which does not depend on $(\mathbf{y}^{(t^*-1)}_{k+\nu}, \mathbf{u}^{(t^*-1)}_{k+\nu\mapsto k})$. This is possible given that the time-homogeneous $\pi(\,\cdot \mid k+\nu) \otimes q_{k+\nu\mapsto k}$-reversible Markov chain associated with the proposal distribution $q^{k,\nu}_{\text{NRJ2}}$, $\{(\mathbf{Y}_{k+\nu}, \mathbf{U}_{k+\nu\mapsto k})(m) : m \in \mathbb{N}\}$, is uniformly ergodic (by assumption). This yields the convergence of the (conditional) distribution of $\mathbf{Y}^{(T-1)}_{k+\nu}$ (given $(k, \mathbf{x}_k, \nu)$) towards $\pi(\,\cdot \mid k+\nu)$.

It is proved that the second absolute value in (5) converges towards 0 using the same arguments, which allows to establish the pointwise convergence $P_T h(k, \mathbf{x}_k, \nu) \longrightarrow P_{\text{ideal}} h(k, \mathbf{x}_k, \nu)$. The uniform convergence on each compact subset of the state-space follows from the uniform ergodicity of the MH kernels. ■

We now highlight what modifications and which additional technical conditions are required if geometric ergodicity is instead assumed. The absolute value on the RHS in (7) is in this case bounded above by $M(\mathbf{y}_{k+\nu}^{(t^*-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*-1)})\, \rho^{T-1-t^*}$, where $M(\mathbf{y}_{k+\nu}^{(t^*-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*-1)})$ is finite for all $(\mathbf{y}_{k+\nu}^{(t^*-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*-1)})$ and $\rho < 1$. If the following integral is finite

$$\int q_{k\mapsto k+\nu}(\mathbf{u}_{k\mapsto k+\nu}^{(0)}) \prod_{t=1}^{t^*-1} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))$$
$$\times M(\mathbf{y}_{k+\nu}^{(t^*-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*-1)})\, d(\mathbf{u}_{k\mapsto k+\nu}^{(0)}, \mathbf{y}_{k+\nu}^{(1:t^*-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:t^*-1)}),$$

then we know that we have the same conclusion as above, i.e. we can choose t^* and T such that the absolute value on the RHS in (7) is smaller than ϵ. That integral shall be finite when the process associated with the kernels $K_{k\mapsto k+\nu}^{(t)}$ do not reach states $(\mathbf{y}_{k+\nu}^{(t^*-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*-1)})$ such that $M(\mathbf{y}_{k+\nu}^{(t^*-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*-1)})$ is extremely large (or at least if it does, it is with small enough probability).

This condition thus suffices to show the pointwise convergence $P_T h(k, \mathbf{x}_k, \nu) \longrightarrow P_{\text{ideal}} h(k, \mathbf{x}_k, \nu)$. To establish the uniform convergence under geometric ergodicity, we use the same strategy as that applied to show (6). We can choose t^* and T such that the first t^* steps (after having generated $\mathbf{u}_{k\mapsto k+\nu}^{(0)}$) with density $\prod_{t=1}^{t^*} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))$ are essentially MH steps with an invariant distribution given by $\rho_{k\mapsto k+\nu}^{(0)} := \pi(\,\cdot \mid k) \times q_{k\mapsto k+\nu} \times |J_{T_{k\mapsto k+\nu}}|^{-1}$. This implies that

$$\Bigg| \int \prod_{t=1}^{t^*} K_{k\mapsto k+\nu}^{(t)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))\, d(\mathbf{y}_{k+\nu}^{(1:t^*)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:t^*)})$$
$$- \int \prod_{t=1}^{t^*} K_{k\mapsto k+\nu}^{(0)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))\, d(\mathbf{y}_{k+\nu}^{(1:t^*)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:t^*)}) \Bigg| < \epsilon,$$

which in turns implies that

$$\Bigg| \int \prod_{t=1}^{t^*} K_{k\mapsto k+\nu}^{(0)}((\mathbf{y}_{k+\nu}^{(t-1)}, \mathbf{u}_{k+\nu\mapsto k}^{(t-1)}), (\mathbf{y}_{k+\nu}^{(t)}, \mathbf{u}_{k+\nu\mapsto k}^{(t)}))\, d(\mathbf{y}_{k+\nu}^{(1:t^*)}, \mathbf{u}_{k+\nu\mapsto k}^{(1:t^*)})$$
$$- \int \rho_{k\mapsto k+\nu}^{(0)}(\mathbf{y}_{k+\nu}^{(t^*)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*)})\, d(\mathbf{y}_{k+\nu}^{(t^*)}, \mathbf{u}_{k+\nu\mapsto k}^{(t^*)}) \Bigg| < M_1(\mathbf{y}_{k+\nu}^{(0)}, \mathbf{u}_{k+\nu\mapsto k}^{(0)})\, \rho_1^{t^*},$$

where $M(\mathbf{y}_{k+\nu}^{(0)}, \mathbf{u}_{k+\nu\mapsto k}^{(0)})$ is finite for all $(\mathbf{y}_{k+\nu}^{(0)}, \mathbf{u}_{k+\nu\mapsto k}^{(0)})$ and $\rho_1 < 1$. The uniform convergence $P_T h \longrightarrow P_{\text{ideal}} h$ on each compact subset of the state-space as $T \longrightarrow \infty$ follows if

$$\int q_{k\mapsto k+\nu}(\mathbf{u}_{k\mapsto k+\nu}^{(0)}) M_1(\mathbf{y}_{k+\nu}^{(0)}, \mathbf{u}_{k+\nu\mapsto k}^{(0)})\, d\mathbf{u}_{k\mapsto k+\nu}^{(0)}$$

is finite and continuous in $\mathbf{x}_k$ (recall that $\mathbf{x}_k$ and $\mathbf{u}^{(0)}_{k\mapsto k+\nu}$ are mapped to $(\mathbf{y}^{(0)}_{k+\nu}, \mathbf{u}^{(0)}_{k+\nu\mapsto k})$ using $T_{k\mapsto k+\nu}$).

Proof of Corollary 1. The proof is an application of Theorem 3.17 in Andrieu and Livingstone (2019) which will allow to establish that

$$\mathrm{var}_\lambda(f, P_{\mathrm{NRJ}}) \leq \mathrm{var}_\lambda(f, P_{\mathrm{RJ}}^{\mathrm{unif}}),$$

where $\mathrm{var}_\lambda(f, P_{\mathrm{NRJ}}) := \mathbb{E}[\{f(K(0), \mathbf{X}_K(0))\}^2] + 2\sum_{m>0}\lambda^m \mathbb{E}[f(K(0), \mathbf{X}_K(0)) f(K(m), \mathbf{X}_K(m))]$ of $\{(K(m), \mathbf{X}_K(m)) : m \in \mathbb{N}\}$ being a Markov chain of transition kernel P at equilibrium and $\lambda \in [0, 1)$. The limit of the RHS $\lim_{\lambda\longrightarrow 1} \mathrm{var}_\lambda(f, P_{\mathrm{RJ}}^{\mathrm{unif}})$ exists and is equal to $\mathrm{var}(f, P_{\mathrm{RJ}}^{\mathrm{unif}})$ because the Markov chain is reversible (see Andrieu and Livingstone (2019)). We will be able to conclude that the limit of the LHS exists as well using Lemma 1 that is presented after this proof.

In order to apply Theorem 3.17, we must verify that

$$\begin{aligned} P_{\mathrm{RJ}}^{\mathrm{unif}}((k, \mathbf{x}_k), (k', \mathbf{y}_{k'})) &= \frac{1}{2}T_{+1}((k, \mathbf{x}_k), (k', \mathbf{y}_{k'})) + \frac{1}{2}T_{-1}((k, \mathbf{x}_k), (k', \mathbf{y}_{k'})) \\ &\quad + \delta_{(k,\mathbf{x}_k)}(k', \mathbf{y}_{k'}) \left(1 - \frac{1}{2}T_{+1}((k, \mathbf{x}_k), (k', \mathbb{R}^{d_{k'}})) - \frac{1}{2}T_{-1}((k, \mathbf{x}_k), (k', \mathbb{R}^{d_{k'}}))\right), \end{aligned}$$

where T_{-1} and T_{+1} are two sub-stochastic kernels associated with accepted proposals when the current and next values for the direction variable are $\nu = -1$ and $\nu = +1$, respectively.

We set

$$T_\nu((k, \mathbf{x}_k), (k', \mathbf{y}_{k'})) := P_{S,\nu}((k, \mathbf{x}_k), (k + \nu, \mathbf{y}_{k+\nu}))\mathbb{P}(S) + \delta_k(k')\, P_{S^c,\nu}(\mathbf{x}_{k'}, \mathbf{y}_{k'})\mathbb{P}(S^c),$$

where S is used to denote that a model switch is proposed, $P_{S,\nu}$ is the conditional transition kernel given S and ν and $P_{S^c,\nu}$ is the conditional transition kernel given S^c and ν. Note that $P_{S^c,\nu}$ is in fact independent of ν (the parameters are updated in the same way whether $\nu = -1$ or $\nu = +1$), therefore we simplify the notation by denoting this kernel by $P_{S^c} := P_{S^c,\nu}$.

We thus have that

$$\frac{1}{2}T_{+1}((k, \mathbf{x}_k), (k', \mathbf{y}_{k'})) + \frac{1}{2}T_{-1}((k, \mathbf{x}_k), (k', \mathbf{y}_{k'}))$$

$$
\begin{aligned}
&= \frac{1}{2}(P_{S,+1}((k,\mathbf{x}_k),(k+1,\mathbf{y}_{k+1}))\mathbb{P}(S) + \delta_k(k')\, P_{S^c}(\mathbf{x}_{k'},\mathbf{y}_{k'})\mathbb{P}(S^c)) \\
&\quad + \frac{1}{2}(P_{S,-1}((k,\mathbf{x}_k),(k-1,\mathbf{y}_{k-1}))\mathbb{P}(S) + \delta_k(k')\, P_{S^c}(\mathbf{x}_{k'},\mathbf{y}_{k'})\mathbb{P}(S^c)) \\
&= \mathbb{P}(S)\left(\frac{1}{2}P_{S,+1}((k,\mathbf{x}_k),(k+1,\mathbf{y}_{k+1})) + \frac{1}{2}P_{S,-1}((k,\mathbf{x}_k),(k-1,\mathbf{y}_{k-1}))\right) \\
&\quad + \mathbb{P}(S^c)P_{S^c}(\mathbf{x}_{k'},\mathbf{y}_{k'}),
\end{aligned}
$$

which corresponds as explained in the proof of Proposition 1 to the sub-stochastic kernel associated with accepted proposals for standard RJ. This concludes the proof. ■

Lemma 1. *Assume that P_{NRJ} is uniformly ergodic. Then, for any real-valued bounded function f of $(k,\mathbf{x}_k)$,*

$$
\lim_{\lambda\to 1}\sum_{m>0}\lambda^m\mathbb{E}[f(K(0),\mathbf{X}_K(0))f(K(m),\mathbf{X}_K(m))] = \sum_{m>0}\mathbb{E}[f(K(0),\mathbf{X}_K(0))f(K(m),\mathbf{X}_K(m))],
$$

where $\{(K(m),\mathbf{X}_K(m)) : m\in\mathbb{N}\}$ is a Markov chain of transition kernel P_{NRJ} at equilibrium.

Proof. To simplify the notation, define $\langle f, P^m_{\mathrm{NRJ}}f\rangle := \mathbb{E}[f(K(0),\mathbf{X}_K(0))f(K(m),\mathbf{X}_K(m))]$. Define the sequence of functions $S_n : \lambda \mapsto \sum_{0<k\leq n}\lambda^k\langle f, P^k_{\mathrm{NRJ}}f\rangle$ defined for $\lambda\in[0,1)$ and its limit $S(\lambda) = \sum_{0<k}\lambda^k\langle f, P^k_{\mathrm{NRJ}}f\rangle$. We now show that the partial sum S_n converges uniformly to S on $[0,1)$, and given that for each $n\in\mathbb{N}$, the function $\lambda\to\lambda^n\langle f, P^n_{\mathrm{NRJ}}f\rangle$ admits a limit when $\lambda\to 1$, we have that S admits a limit when $\lambda\to 1$, given by

$$
\lim_{\lambda\to 1} S(\lambda) = S(1) = \sum_{k>0}\langle f, P^k_{\mathrm{NRJ}}f\rangle
$$

First, note that

$$
\begin{aligned}
\sup_{\lambda\in[0,1)}|S_n(\lambda)-S(\lambda)| = \sup_{\lambda\in[0,1)}\left|\sum_{k>n}\lambda^k\langle f, P^k_{\mathrm{NRJ}}f\rangle\right| &\leq \sup_{\lambda\in[0,1)}\sum_{k>n}\lambda^k\left|\langle f, P^k_{\mathrm{NRJ}}f\rangle\right| \\
&= \sum_{k>n}\left|\langle f, P^k_{\mathrm{NRJ}}f\rangle\right|.
\end{aligned}
$$

Thus, to prove that $\sup_{\lambda\in[0,1)}|S_n(\lambda)-S(\lambda)| \longrightarrow 0$, it is sufficient to prove that the series $\sum_{k>0}\left|\langle f, P^k_{\mathrm{NRJ}}f\rangle\right|$ converges.

Given that f is bounded we can consider without loss of generality that its expectation is 0 and that it takes values between -1 and $+1$ (we can re-normalise it). Because P_{NRJ}

is assumed to be uniformly ergodic, there exists constants $\rho \in (0,1)$ and $M \in (0,\infty)$ such that for any $m \in \mathbb{N}$,

$$\sup_{(k,\mathbf{x}_k,\nu)} \|\delta_{k,\mathbf{x}_k,\nu} P_{\text{NRJ}}^m - \pi \otimes \mathcal{U}\{-1,+1\}\|_{\text{tv}} \leq M\rho^m, \tag{8}$$

where for any signed measure μ, $\|\mu\|_{\text{tv}}$ denotes its total variation. Note that $\|\mu\|_{\text{tv}} = (1/2)\sup_{f:\mathcal{X}\to[-1,+1]} |\mu f|$ (see for instance Roberts and Rosenthal (2004), Proposition 3).

We have that

$$\begin{aligned} |\langle f, P_{\text{NRJ}}^k f\rangle| = |\mathbb{E} f(K,\mathbf{X}_k,\nu) P_{\text{NRJ}}^k f(K,\mathbf{X}_k,\nu)| &\leq \mathbb{E}|f(K,\mathbf{X}_k,\nu)||P_{\text{NRJ}}^k f(K,\mathbf{X}_k,\nu)| \\ &\leq \mathbb{E}|P_{\text{NRJ}}^k f(K,\mathbf{X}_k,\nu)| \\ &= \mathbb{E}|P_{\text{NRJ}}^k f(K,\mathbf{X}_k,\nu) - \pi f| \\ &\leq \mathbb{E}\sup_f \left|P_{\text{NRJ}}^k f(K,\mathbf{X}_k,\nu) - \pi f\right| \\ &\leq M\rho^k, \end{aligned}$$

which is clearly summable. As a consequence, S_n converges uniformly to S on $[0,1)$ which concludes the proof. ■

We now highlight what modifications and which additional technical conditions are required if geometric ergodicity is instead assumed. The constant M in (8) would depend on $(K,\mathbf{X}_k,\nu)$. Therefore, if $\mathbb{E}M(K,\mathbf{X}_k,\nu)$ is finite the result is also valid.

2 Weak convergence results for the ideal samplers

We analyse the asymptotic scenario in which the number of models grows to infinity. It will be noticed that the reversible and non-reversible Markov chains produced respectively by ideal RJ and NRJ have two distinct asymptotic behaviours which are consistent with what is observed for fixed numbers of models (see, e.g., Figure 1 in our paper), explaining their different state-space exploration speed.

We prove convergence towards continuous-time stochastic processes that take values on the real line. We thus need to consider functions of K to achieve that. Firstly, we consider that the model indicator K takes values in $\mathcal{K}^n := \{1,\ldots,\lfloor\sqrt{n}\log n\rfloor\}$, where $\lfloor\cdot\rfloor$ is the floor function. We added the superscript n to highlight the dependence on this variable. We select $\mathcal{K}^n$ in this way to obtain a random variable $S_K^n := (K^n - \psi(n))/\sqrt{n}$ that is (in the limit) continuous in addition to taking values on the real line, for a given function ψ (which can be thought of as the mean that can be for instance $\lfloor \sqrt{n} \log n \rfloor /2$). Imagine that the mode is around $\lfloor \sqrt{n} \log n \rfloor /2$ (so the mass is moving towards infinity), this transformation puts the mass around 0 and makes the different values of the centred variable ($-1, 0, 1$ and so on) close to each other (e.g. $|1-0|/\sqrt{n} \longrightarrow 0$). We squeeze the state-space as in the proof of existence of Brownian motion from random walks. We assume that $\pi^n(k) > 0$ for all $k \in \mathcal{K}^n$. For $t \geq 0$, we define the following rescaled stochastic process:

$$Z_{\mathrm{RJ}}^n(t) := \frac{K_{\mathrm{RJ}}^n(\lfloor nt \rfloor) - \psi(n)}{\sqrt{n}},$$

where $\{K_{\mathrm{RJ}}^n(m) : m \in \mathbb{N}\}$ is a Markov chain produced by the ideal RJ corresponding to the ideal NRJ described in Section 2.2 in our paper. We consider that this RJ updates parameters and switches models with probabilities τ and $1-\tau$, respectively, and that $g(k, k+1) = g(k, k-1) = 0.5(1-\tau)$, so it proposes to increase or decrease the model indicator with the same probability and $\alpha_{\mathrm{RJ}} = \alpha_{\mathrm{NRJ}}$. The continuous-time stochastic process $\{Z_{\mathrm{RJ}}^n(t) : t \geq 0\}$ is a sped up and modified version of $\{K_{\mathrm{RJ}}^n(m) : m \in \mathbb{N}\}$. The decreasing size of the jumps of $\{Z_{\mathrm{RJ}}^n(t) : t \geq 0\}$ as n increases (the size is $1/\sqrt{n}$), combined with its time acceleration, result in a continuous and non-trivial limiting process, as specified in Theorem 2. This time acceleration can be thought of as squeezing the time axis to make the iterations close to each other, again as in the proof of existence of Brownian motion from random walks.

Theorem 2 (Weak convergence of RJ)**.** *Assume that:*

(a) *the function ψ can be chosen such that S_K^n is asymptotically distributed as $f_S \in \mathcal{C}^1(\mathbb{R})$, a strictly positive probability density function (PDF), where $\mathcal{C}^1(\mathbb{R})$ denotes the space of real-valued functions on $\mathbb{R}$ with continuous first derivative;*

(b) *the function $(\log f_S(\cdot))'$ is Lipschitz continuous;*

(c) *ψ can be chosen such that*

$$\frac{1}{\pi^n(k)} \frac{\pi^n(k+1) - \pi^n(k)}{1/\sqrt{n}} - (\log f_S(S_k^n))' \tag{9}$$

is bounded for all n and converges towards 0 as $n \longrightarrow \infty$, for all k;

(d) $\lim_{n \longrightarrow \infty} \sqrt{n}\, \pi^n(1) = \lim_{n \longrightarrow \infty} \sqrt{n}\, \pi^n(\lfloor \sqrt{n} \log n \rfloor) = 0$.

If $K_{RJ}^n(0) \sim \pi^n$, then $\{Z_{RJ}^n(t) : t \geq 0\}$ converges weakly towards a Langevin diffusion as $n \longrightarrow \infty$, i.e.

$$\{Z_{RJ}^n(t) : t \geq 0\} \Longrightarrow \{Z_{RJ}(t) : t \geq 0\} \quad \textit{as} \quad n \longrightarrow \infty,$$

where the process $\{Z_{RJ}(t) : t \geq 0\}$ is such that $Z_{RJ}(0) \sim f_S$ and

$$dZ_{RJ}(t) = \frac{1-\tau}{2} (\log f_S(Z_{RJ}(t)))'\, dt + \sqrt{1-\tau}\, dB(t),$$

with $\{B(t) : t \geq 0\}$ being a Wiener process.

Proof. It is a straightforward adaptation of Theorem 1 in Gagnon et al. (2019). For sake of completeness, it is detailed in Section 2.1. ■

The notation "$\Longrightarrow$" represents here weak convergence of processes in the Skorokhod topology (see Section 3 of Ethier and Kurtz (1986) for more details about this type of convergence).

The two main assumptions are (a) and (c). The former requires to find a transformation of K^n such that the limit in distribution of the transformed random variable is a continuous random variable with density f_S. The latter requires that the "discrete version" of the derivative of $\log \pi^n$ share the same asymptotic behaviour as the derivative of $\log f_S$. Indeed, in Gagnon et al. (2019), it is explained that the left term in (9) can be seen as the discrete version of the derivative of $\log \pi^n$ because π^n is also the PMF of S_k^n (evaluated at a different point) and $S_{k+1}^n - S_k^n = 1/\sqrt{n}$. Assumption (b) is standard in the weak convergence literature; it ensures the existence of a unique strong solution to the stochastic differential equation given above. Assumption (d) is a regularity condition. In Gagnon et al. (2019), to illustrate how a PMF that satisfies the conditions looks like, the authors show one that is such that S_K^n converges in distribution towards a standard normal.

We now analyse the behaviour of the stochastic process produced by the ideal NRJ algorithm. We consider as before that $\mathcal{K}^n = \{1, \ldots, \lfloor \sqrt{n} \log n \rfloor\}$ and $\pi^n(k) > 0$ for all

$k \in \mathcal{K}^n$. For $t \geq 0$, we define the following rescaled stochastic process:

$$\mathbf{Z}^n_{\text{NRJ}}(t) := \left(\frac{K^n_{\text{NRJ}}(\lfloor \sqrt{n}t \rfloor) - \psi(n)}{\sqrt{n}}, \nu(\lfloor \sqrt{n}t \rfloor) \right), \tag{10}$$

where $\{(K^n_{\text{NRJ}}, \nu)(m) : m \in \mathbb{N}\}$ is a Markov chain produced by ideal NRJ described in Section 2.2 in our paper. Note that the distribution of ν does not change with n.

Theorem 3 (Weak convergence of NRJ)**.** *Assume that the same conditions (a)-(d) as in Theorem 2 are satisfied. Assume additionally that there exist two positive constants c and x_0 such that $|(\log f_S(x))'| \geq c$ for all $|x| \geq x_0$. If $(K^n_{NRJ}, \nu)(0) \sim \pi^n \otimes \mathcal{U}\{-1, 1\}$, then $\{\mathbf{Z}^n_{NRJ}(t) : t \geq 0\}$ converges weakly towards a piecewise deterministic Markov process (PDMP) as $n \longrightarrow \infty$, i.e.*

$$\{\mathbf{Z}^n_{NRJ}(t) : t \geq 0\} \Longrightarrow \{\mathbf{Z}_{NRJ}(t) : t \geq 0\} \quad \textit{as} \quad n \longrightarrow \infty,$$

where the process $\{\mathbf{Z}_{NRJ}(t) : t \geq 0\}$ is such that $\mathbf{Z}_{NRJ}(0) \sim f_S \otimes \mathcal{U}\{-1, 1\}$ with generator

$$Gh(x, y) := (1 - \tau) y h_x(x, y) + \max\{0, -y\,(\log f_S(x))'\}(1 - \tau)(h(x, -y) - h(x, y)),$$

where $h(\,\cdot\,, y) \in \mathcal{C}^1(\mathbb{R})$ and such that itself and $h_x(\,\cdot\,, y)$ vanish at infinity, for $y \in \{-1, 1\}$, h_x denoting the first derivative of h with respect to its first argument.

Proof. See Section 2.1. ■

The additional regularity condition on f_S in Theorem 3 essentially ensures that outside of a bounded set, this PDF decreases sufficiently quickly. Indeed, given that $(\log f_S(x))' = f_S'(x)/f_S(x)$ and f_S is strictly positive, it is required that the tail decay is bounded from below (relatively to f_S). This guarantees that the limiting PDMP has some important properties (e.g. non-explosiveness and $f_S \otimes \mathcal{U}\{-1, 1\}$ is an invariant distribution, see Bierkens and Roberts (2017)).

The PDMP in Theorem 3 corresponds to a zig-zag Markov process (Bierkens et al. (2019)), and in fact, a bouncy particle sampler (BPS, Bouchard-Côté et al. (2018)) given that they both coincide when the position variable is unidimensional. This position variable evolves with constant drift $1 - \tau$ either to the right or left of the real line depending on the direction variable, and changes direction with rate $\max\{0, -y\,(\log f_S(x))'\}(1-\tau)$ when the position is x and direction y. PDMP are known for being non-diffusive and having persistency-driven paths. We constructed NRJ to induce such a behaviour, but we do not know *a priori* when this will happen and how this will translate. An analysis was conducted in Section 4 in our paper to provide some answers. Theorem 3 and Theorem 2 indicate that in the (asymptotic) theoretical framework considered, the model indicator's paths produced by RJ and NRJ behave exactly as expected; the former show diffusive patterns and the latter not. This suggests that (at least under those conditions) NRJ outperform RJ. We even have a guarantee for the speed of convergence towards the target distribution for NRJ: Bierkens and Roberts (2017) prove that the PDMP in Theorem 3 is exponentially ergodic. We additionally know that the convergence is an order of magnitude slower for $\{K_{\text{RJ}}^n(m) : m \in \mathbb{N}\}$. Indeed, the different behaviour of $\{K_{\text{NRJ}}^n(m) : m \in \mathbb{N}\}$ compared with $\{K_{\text{RJ}}^n(m) : m \in \mathbb{N}\}$ requires to accelerate the time by a factor of only $\sqrt{n}$ in the definition of $\{\mathbf{Z}_{\text{NRJ}}^n(t) : t \geq 0\}$ comparatively to n in that of $\{Z_{\text{RJ}}^n(t) : t \geq 0\}$ to obtain non-trivial limiting stochastic processes. This highlights again that $\{K_{\text{NRJ}}^n(m) : m \in \mathbb{N}\}$ explores its state-space more quickly.

2.1 Proofs of Theorems 2 and 3

Proof of Theorem 2. In order to prove the result, we demonstrate the convergence of the finite-dimensional distributions of $\{Z_{\text{RJ}}^n(t) : t \geq 0\}$ to those of $\{Z_{\text{RJ}}(t) : t \geq 0\}$. To achieve this, we verify Condition (c) of Theorem 8.2 from chapter 4 of Ethier and Kurtz (1986). The weak convergence then follows from Corollary 8.6 of Chapter 4 of Ethier and Kurtz (1986). The remaining conditions of Theorem 8.2 and the conditions specified in Corollary 8.6 are either straightforward or easily derived from the proof given here.

The proof of the convergence of the finite-dimensional distributions relies on the convergence of (what we call) the "pseudo-generator", a quantity that we define as:

$$\varrho_{\text{RJ}}^n(t) := n\,\mathbb{E}[h(Z_{\text{RJ}}^n(t+1/n)) - h(Z_{\text{RJ}}^n(t)) \mid \mathcal{F}^{Z_{\text{RJ}}^n}(t)],$$

where $h \in \mathcal{C}_c^\infty(\mathbb{R})$, the space of infinitely differentiable functions on $\mathbb{R}$ with compact support. Theorem 2.1 from Chapter 8 of Ethier and Kurtz (1986) allows us to restrict our attention to this set of functions when studying the limiting behaviour of the pseudo-generator. In our situation, the pseudo-generator has a more precise expression:

$$\begin{aligned}\varrho_{\mathrm{RJ}}^n(t) = \frac{n(1-\tau)}{2}\left((h(S_{K+1}^n) - h(S_K^n))\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)}\right)\right) \\ + \frac{n(1-\tau)}{2}\left((h(S_{K-1}^n) - h(S_K^n))\left(1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)}\right)\right).\end{aligned} \tag{11}$$

Note that the Markov process $\{K_{\mathrm{RJ}}^n(m) : m \in \mathbb{N}\}$ is time-homogeneous, and because of this we replaced the random variable $Z_{\mathrm{RJ}}^n(t)$ by S_K^n and $Z_{\mathrm{RJ}}^n(t+1/n)$ by S_{K+1}^n or S_{K-1}^n given that we will work under expectations. Indeed, Condition (c) of Theorem 8.2 from chapter 4 of Ethier and Kurtz (1986) essentially reduces to the following convergence:

$$\mathbb{E}\left[|\varrho_{\mathrm{RJ}}^n(t) - Gh(Z_{\mathrm{RJ}}^n(t))|\right] \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty,$$

where G is the generator of the limiting diffusion with

$$Gh(Z_{\mathrm{RJ}}^n(t)) := \frac{1-\tau}{2}\,(\log f_S(Z_{\mathrm{RJ}}^n(t)))' h'(Z_{\mathrm{RJ}}^n(t)) + \frac{1-\tau}{2}\, h''(Z_{\mathrm{RJ}}^n(t)).$$

Note that there exists a positive constant M such that h and all its derivatives are bounded in absolute value by this constant. We choose M such that it is a Lipschitz constant for the function $(\log f_S(\cdot))'$.

The key here is to use Taylor expansions in (11) to obtain derivatives of h as in G. By noting that $S_{K+1}^n = S_K^n + 1/\sqrt{n}$ and $S_{K-1}^n = S_K^n - 1/\sqrt{n}$, and using Taylor expansions of h around S_K^n, we obtain

$$\begin{aligned}h(S_K^n + 1/\sqrt{n}) - h(S_K^n) &= \frac{1}{\sqrt{n}}\, h'(S_K^n) + \frac{1}{2n}\, h''(S_K^n) + \frac{1}{6n^{3/2}} h'''(W), \\ h(S_K^n - 1/\sqrt{n}) - h(S_K^n) &= -\frac{1}{\sqrt{n}}\, h'(S_K^n) + \frac{1}{2n}\, h''(S_K^n) - \frac{1}{6n^{3/2}} h'''(T),\end{aligned}$$

where W and T belong to $(S_K^n, S_K^n + 1/\sqrt{n})$ and $(S_K^n - 1/\sqrt{n}, S_K^n)$, respectively. We also note that the first term on the RHS of (11) equals 0 when $K^n = \lfloor\sqrt{n}\log n\rfloor$ because $\pi^n(\lfloor\sqrt{n}\log n\rfloor + 1) = 0$. For the analogous reason, the second term on the RHS of (11) equals 0 when $K^n = 1$. Therefore,

$$\varrho_{\mathrm{RJ}}^n(t) - Gh(Z_{\mathrm{RJ}}^n(t)) = \mathbb{1}(2 \leq K^n \leq \lfloor\sqrt{n}\log n\rfloor - 1)\,\frac{1-\tau}{2}\, h'(S_K^n)$$

$$
\begin{aligned}
&\quad \times \left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)}\right) - (\log f_S(S_K^n))'\right) \\
&+ \mathbb{1}(K^n = 1)\,\frac{1-\tau}{2}\,h'(S_K^n)\left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)}\right) - (\log f_S(S_K^n))'\right) \\
&- \mathbb{1}(K^n = \lfloor \sqrt{n}\log n \rfloor)\,\frac{1-\tau}{2}\,h'(S_K^n)\left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)}\right) - (\log f_S(S_K^n))'\right) \\
&+ \mathbb{1}(2 \leq K^n \leq \lfloor \sqrt{n}\log n \rfloor - 1)\,\frac{1-\tau}{4}\,h''(S_K^n) \\
&\quad \times \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} + 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} - 2\right) \\
&+ \mathbb{1}(K^n = 1)\,\frac{1-\tau}{4} h''(S_K^n)\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 2\right) \\
&+ \mathbb{1}(K^n = \lfloor \sqrt{n}\log n \rfloor)\,\frac{1-\tau}{4} h''(S_K^n)\left(1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} - 2\right) \\
&+ \frac{1-\tau}{12\sqrt{n}}\,h'''(W)\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)}\right)\mathbb{1}(1 \leq K^n \leq \lfloor \sqrt{n}\log n \rfloor - 1) \\
&- \frac{1-\tau}{12\sqrt{n}}\,h'''(T)\left(1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)}\right)\mathbb{1}(2 \leq K^n \leq \lfloor \sqrt{n}\log n \rfloor).
\end{aligned} \tag{12}
$$

We now prove that expectation of the absolute value of each term on the RHS in (12) converges towards 0 as $n \longrightarrow \infty$. We start with the last terms and make our way up. It is clear that the expectation of the absolute value of each of the last two terms converges towards 0 as $n \longrightarrow \infty$ given that $|h'''| \leq M$ and $0 \leq 1 \wedge x \leq 1$ for positive x. We now analyse the fourth one (starting from the bottom). As $n \longrightarrow \infty$,

$$
\mathbb{E}\left[\left|\mathbb{1}(K^n = 1)\,\frac{1-\tau}{2} h''(S_K^n)\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 2\right)\right|\right] \leq \frac{(1-\tau)M}{2}\mathbb{P}(K^n = 1) \longrightarrow 0,
$$

using $|h''| \leq M$ and

$$
0 \leq \left|1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 2\right| \leq 2.
$$

Recall that $\mathbb{P}(K^n = 1) \longrightarrow 0$ by assumption. The proof for the third term (starting from the bottom) is similar.

Applying Lemmas 2 to 4 (that follow), each of the remaining terms is seen to converge towards 0 in L^1 as $n \longrightarrow \infty$, which concludes the proof. ∎

Lemma 2. *As $n \longrightarrow \infty$, we have*

$$
\mathbb{E}\left[\left|\mathbb{1}(2 \leq K^n \leq \lfloor \sqrt{n}\log n \rfloor - 1)\,\frac{1-\tau}{4}\,h''(S_K^n)\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} + 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} - 2\right)\right|\right] \longrightarrow 0.
$$

Proof. We have

$$
\begin{aligned}
&\mathbb{E}\left[\left|\mathbb{1}(2 \leq K^n \leq \lfloor \sqrt{n}\log n \rfloor - 1)\, \frac{1-\tau}{4}\, h''(S_K^n) \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} + 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} - 2\right)\right|\right] \\
&\leq \frac{(1-\tau)M}{4}\, \mathbb{E}\left[\left|\mathbb{1}(2 \leq K^n \leq \lfloor \sqrt{n}\log n \rfloor - 1) \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} + 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} - 2\right)\right|\right],
\end{aligned}
$$

because $|h''| \leq M$. We show that

$$
\frac{\pi^n(k+1)}{\pi^n(k)} \longrightarrow 1 \quad \text{for all } k \in \{1, \ldots, \lfloor \sqrt{n}\log n \rfloor - 1\},
$$

which allows to conclude using the triangle inequality, the continuity of the function $1 \wedge x$, and the Lebesgue's dominated convergence theorem. We have

$$
\begin{aligned}
\left|\frac{\pi^n(k+1)}{\pi^n(k)} - 1\right| &= \left|\frac{1}{\pi^n(k)}\, \frac{\pi^n(k+1) - \pi^n(k)}{1/\sqrt{n}} - (\log f_S(S_k^n))' + (\log f_S(S_k^n))'\right| \frac{1}{\sqrt{n}} \\
&\leq \left|\frac{1}{\pi^n(k)}\, \frac{\pi^n(k+1) - \pi^n(k)}{1/\sqrt{n}} - (\log f_S(S_k^n))'\right| \frac{1}{\sqrt{n}} \\
&\quad + |(\log f_S(S_k^n))'|\, \frac{1}{\sqrt{n}},
\end{aligned}
$$

using again the triangle inequality. By assumption, we have that

$$
\left|\frac{1}{\pi^n(k)}\, \frac{\pi^n(k+1) - \pi^n(k)}{1/\sqrt{n}} - (\log f_S(S_k^n))'\right| \frac{1}{\sqrt{n}} \longrightarrow 0.
$$

We also have that

$$
\begin{aligned}
|(\log f(S_k^n))'|\, \frac{1}{\sqrt{n}} &= |(\log f_S(S_k^n))' - (\log f_S(0))' + (\log f_S(0))'|\, \frac{1}{\sqrt{n}} \\
&\leq |(\log f_S(S_k^n))' - (\log f_S(0))'|\, \frac{1}{\sqrt{n}} + \frac{|(\log f_S(0))'|}{\sqrt{n}} \\
&\leq M \left|\frac{k - \psi(n)}{\sqrt{n}}\right| \frac{1}{\sqrt{n}} + \frac{|(\log f_S(0))'|}{\sqrt{n}},
\end{aligned}
$$

using first the triangle inequality, and next the fact that $(\log f_S(\cdot))'$ is Lipschitz continuous. We have that $|(\log f_S(0))'|/\sqrt{n} \longrightarrow 0$ because $f_S \in \mathcal{C}^1(\mathbb{R})$. Also,

$$
\left|\frac{k - \psi(n)}{\sqrt{n}}\right| \frac{1}{\sqrt{n}} \leq 2\, \frac{\lfloor \sqrt{n}\log n \rfloor}{n} \longrightarrow 0,
$$

using the triangle inequality and the fact that $k, \psi(n) \leq \lfloor \sqrt{n}\log n \rfloor$. ■

Lemma 3. *As $n \longrightarrow$, we have*

$$\mathbb{E}\left[\left|\mathbb{1}(K^n = 1)\,\frac{1-\tau}{2}\,h'(S^n_K)\left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)}\right) - (\log f_S(S^n_K))'\right)\right|\right] \longrightarrow 0,$$

and

$$\mathbb{E}\left[\left|\mathbb{1}(K^n = \lfloor\sqrt{n}\log n\rfloor)\,\frac{1-\tau}{2}\,h'(S^n_K)\left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)}\right) - (\log f_S(S^n_K))'\right)\right|\right] \longrightarrow 0.$$

Proof. We have that

$$\begin{aligned}
&\mathbb{E}\left[\left|\mathbb{1}(K^n = 1)\,\frac{1-\tau}{2}\,h'(S^n_K)\left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)}\right) - (\log f_S(S^n_K))'\right)\right|\right] \\
&\quad \leq \frac{(1-\tau)M}{2}\,\mathbb{E}\left[\mathbb{1}(K^n = 1)\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)}\right)\right] + \frac{(1-\tau)M}{2}\,\mathbb{E}\left[\mathbb{1}(K^n = 1)\,|(\log f_S(S^n_K))'|\right],
\end{aligned}$$

using that $|h'| \leq M$ and the triangle inequality. The first term on the RHS converges towards 0 by assumption because $0 \leq 1 \wedge x \leq 1$ for positive x. Using the same mathematical arguments as in the proof of Lemma 2, we have that

$$|(\log f_S(S^n_K))'| \leq 2M\,\frac{\lfloor\sqrt{n}\log n\rfloor}{\sqrt{n}} + |(\log f_S(0))'|\,.$$

Therefore, using the triangle inequality

$$\mathbb{E}\left[\mathbb{1}(K^n = 1)\,|(\log f_S(S^n_K))'|\right] \leq \mathbb{P}(K^n = 1)\left(2M\,\frac{\lfloor\sqrt{n}\log n\rfloor}{\sqrt{n}} + |(\log f_S(0))'|\right) \longrightarrow 0,$$

by assumption (and because $f_S \in \mathcal{C}^1(\mathbb{R})$). The proof that

$$\mathbb{E}\left[\left|\mathbb{1}(K^n = \lfloor\sqrt{n}\log n\rfloor)\,\frac{1-\tau}{2}\,h'(S^n_K)\left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)}\right) - (\log f_S(S^n_K))'\right)\right|\right] \longrightarrow 0$$

is similar. ∎

Lemma 4. *As $n \longrightarrow \infty$, we have*

$$\begin{aligned}
&\mathbb{E}\Bigg[\Bigg|\mathbb{1}(2 \leq K^n \leq \lfloor\sqrt{n}\log n\rfloor - 1)\,\frac{1-\tau}{2}\,h'(S^n_K) \\
&\qquad \times\left(\sqrt{n}\left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)}\right) - (\log f_S(S^n_K))'\right)\Bigg|\Bigg] \longrightarrow 0.
\end{aligned}$$

Proof. First, we have that

$$\mathbb{E}\Bigg[\Bigg|\mathbb{1}(2 \leq K^n \leq \lfloor\sqrt{n}\log n\rfloor - 1)\,\frac{1-\tau}{2}\,h'(S^n_K)$$

$$\times \left(\sqrt{n} \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} \right) - (\log f_S(S_K^n))' \right) \Big| \Big]$$

$$\leq \frac{(1-\tau)M}{2} \mathbb{E} \Big[\Big| \mathbb{1}(2 \leq K^n \leq \lfloor \sqrt{n} \log n \rfloor - 1)$$

$$\times \left(\sqrt{n} \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} \right) - (\log f_S(S_K^n))' \right) \Big| \Big],$$

because $|h'| \leq M$. We now consider four cases for K^n:

1. $\pi^n(K^n+1)/\pi^n(K^n) < 1$ and $\pi^n(K^n-1)/\pi^n(K^n) \geq 1$,
2. $\pi^n(K^n+1)/\pi^n(K^n) \geq 1$ and $\pi^n(K^n-1)/\pi^n(K^n) < 1$,
3. $\pi^n(K^n+1)/\pi^n(K^n) \geq 1$ and $\pi^n(K^n-1)/\pi^n(K^n) \geq 1$,
4. $\pi^n(K^n+1)/\pi^n(K^n) < 1$ and $\pi^n(K^n-1)/\pi^n(K^n) < 1$.

In Case 1, we have that

$$\begin{aligned}
&\sqrt{n} \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} \right) - (\log f_S(S_K^n))' \\
&\quad = \sqrt{n} \left(\frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \right) - (\log f_S(S_K^n))' \\
&\quad = \frac{1}{\pi^n(K^n)} \frac{\pi^n(K^n+1) - \pi^n(K^n)}{1/\sqrt{n}} - (\log f_S(S_K^n))' \longrightarrow 0,
\end{aligned}$$

by assumption. We can prove that it converges towards 0 in Case 2 in the same way. Case 3 corresponds to a local minimum. In this case,

$$\sqrt{n} \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \wedge \frac{\pi^n(K^n-1))}{\pi^n(K^n)} \right) = 0,$$

for all n, and $(\log f_S(S_K^n))' = f'_{Z_{\mathrm{RJ}}}(S_K^n)/f_S(S_K^n) \longrightarrow 0$. Case 4 corresponds to a local (or global) maximum. Again, $(\log f_S(S_K^n))' \longrightarrow 0$. Additionally,

$$\begin{aligned}
&\sqrt{n} \left(1 \wedge \frac{\pi^n(K^n+1)}{\pi^n(K^n)} - 1 \wedge \frac{\pi^n(K^n-1)}{\pi^n(K^n)} \right) - (\log f_S(S_K^n))' \\
&\quad = \sqrt{n} \left(\frac{\pi^n(K^n+1) - \pi^n(K^n)}{\pi^n(K^n)} - \frac{\pi^n(K^n-1) - \pi^n(K^n)}{\pi^n(K^n)} \right) - (\log f_S(S_K^n))' \\
&\quad = \frac{1}{\pi^n(K^n)} \frac{\pi^n(K^n+1) - \pi^n(K^n)}{1/\sqrt{n}} - (\log f_S(S_K^n))' \\
&\qquad - \frac{1}{\pi^n(K^n)} \frac{\pi^n(K^n-1) - \pi^n(K^n)}{1/\sqrt{n}},
\end{aligned}$$

but both terms converge towards 0. Consequently, Lebesgue's dominated convergence theorem allows to conclude the proof. ■

Proof of Theorem 3. Analogously to the proof of Theorem 2, we demonstrate the convergence of the finite-dimensional distributions of $\{\mathbf{Z}^n_{\text{NRJ}}(t) : t \geq 0\}$ to those of $\{\mathbf{Z}_{\text{NRJ}}(t) : t \geq 0\}$. The same strategy as in that proof is employed: we verify Condition (c) of Theorem 8.2 from chapter 4 of Ethier and Kurtz (1986). The weak convergence then follows from Corollary 8.6 of Chapter 4 of Ethier and Kurtz (1986). The remaining conditions of Theorem 8.2 and the conditions specified in Corollary 8.6 are either straightforward or easily derived from the proof given here.

Beforehand, we note that the additional assumption on f_S (about the lower bound on $|(\log f_S(\cdot))'|$ outside of a bounded set) implies that Assumption 3 in Section 5 of Bierkens and Roberts (2017) is satisfied. In that paper, it is proved that it implies that the PDMP defined in Theorem 3 is a non-explosive strong Markov process. The authors also demonstrate that the Markov transition semigroup to which the generator corresponds is Feller.

For this proof, the time acceleration factor is different, and accordingly, the pseudo-generator is defined as:

$$
\begin{aligned}
\varrho^n_{\text{NRJ}}(t) &:= \sqrt{n}\,\mathbb{E}[h(\mathbf{Z}^n_{\text{NRJ}}(t + 1/\sqrt{n})) - h(\mathbf{Z}^n_{\text{NRJ}}(t)) \mid \mathcal{F}^{\mathbf{Z}^n_{\text{NRJ}}}(t)] \\
&= \sqrt{n}(1-\tau)(h(S^n_{K+\nu}, \nu) - h(S^n_K, \nu)) \left(1 \wedge \frac{\pi^n(K^n + \nu)}{\pi^n(K^n)}\right) \\
&\quad + \sqrt{n}(1-\tau)(h(S^n_K, -\nu) - h(S^n_K, \nu)) \left(1 - 1 \wedge \frac{\pi^n(K^n + \nu)}{\pi^n(K^n)}\right).
\end{aligned}
$$

As in the proof of Theorem 2, we replaced $\mathbf{Z}^n_{\text{NRJ}}(t)$ by (S^n_K, ν) and $\mathbf{Z}^n_{\text{NRJ}}(t + 1/\sqrt{n})$ by $(S^n_{K+\nu}, \nu)$ or $(S^n_K, -\nu)$ given that the Markov process $\{(K^n_{\text{NRJ}}, \nu)(m) : m \in \mathbb{N}\}$ is time-homogeneous and we will work under expectations. Recall that Condition (c) of Theorem 8.2 from chapter 4 of Ethier and Kurtz (1986) is essentially

$$
\mathbb{E}\left[|\varrho^n_{\text{NRJ}}(t) - Gh(\mathbf{Z}^n_{\text{NRJ}}(t))|\right] \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty,
$$

where G is in this case the generator expressed in Theorem 3. We have that

$$
\begin{aligned}
&\mathbb{E}\left[|\varrho^n_{\text{NRJ}}(t) - Gh(\mathbf{Z}^n_{\text{NRJ}}(t))|\right] \\
&\quad \leq \mathbb{E}\left[\left|\sqrt{n}(1-\tau)(h(S^n_{K+\nu}, \nu) - h(S^n_K, \nu)) \left(1 \wedge \frac{\pi^n(K^n + \nu)}{\pi^n(K^n)}\right) - (1-\tau)\nu h_x(S^n_K, \nu)\right|\right] \\
&\qquad + \mathbb{E}\left[\left|\sqrt{n}(1-\tau)(h(S^n_K, -\nu) - h(S^n_K, \nu)) \left(1 - 1 \wedge \frac{\pi^n(K^n + \nu)}{\pi^n(K^n)}\right)\right.\right.
\end{aligned}
$$

$$- \max\{0, -\nu\,(\log f_S(S_K^n))'\}(1-\tau)(h(S_K^n, -\nu) - h(S_K^n, \nu))|]\,, \tag{13}$$

using the triangle inequality. We analyse the two terms separately. We start with the first one. By the mean value theorem and using that $S_{K+\nu}^n - S_K^n = \nu/\sqrt{n}$, we have that

$$h(S_{K+\nu}^n, \nu) - h(S_K^n, \nu) = \frac{\nu}{\sqrt{n}}\, h_x(T, \nu),$$

where T is in $(S_K^n, S_{K+\nu}^n)$ or $(S_{K+\nu}^n, S_K^n)$. We therefore also know that $T \longrightarrow S_K^n$ with probability 1. In the proof of Lemma 2, it is shown that

$$1 \wedge \frac{\pi^n(K^n+\nu)}{\pi^n(K^n)} \longrightarrow 1 \quad \text{with probability 1,}$$

and consequently,

$$\mathbb{E}\left[\left|\sqrt{n}(1-\tau)(h(S_{K+\nu}^n, \nu) - h(S_K^n, \nu))\left(1 \wedge \frac{\pi^n(K^n+\nu)}{\pi^n(K^n)}\right) - (1-\tau)\nu h_x(S_K^n, \nu)\right|\right] \longrightarrow 0,$$

using Lebesgue's dominated convergence theorem (given that the quantity in the expectation is bounded, because $h_x(\cdot, \nu)$ is bounded for $\nu \in \{-1, 1\}$). For the second term in (13), we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\sqrt{n}(1-\tau)(h(S_K^n, -\nu) - h(S_K^n, \nu))\left(1 - 1 \wedge \frac{\pi^n(K^n+\nu)}{\pi^n(K^n)}\right)\right.\right. \\
&\qquad - \max\{0, -\nu\,(\log f_S(S_K^n))'\}(1-\tau)(h(S_K^n, -\nu) - h(S_K^n, \nu))|] \\
&\qquad \leq (1-\tau)2M\,\mathbb{E}\left[\left|\sqrt{n}\left(1 - 1 \wedge \frac{\pi^n(K^n+\nu)}{\pi^n(K^n)}\right) - \max\{0, -\nu\,(\log f_S(S_K^n))'\}\right|\right], \qquad (14)
\end{aligned}$$

because there exists a positive constant M such that $|h(\cdot, \nu)| \leq M$ for $\nu \in \{-1, 1\}$ (recall that $h(\cdot, \nu)$ is continuous and vanishes at infinity for any value of ν).

We now consider four cases for K^n and ν:

1. $\nu = +1$ and $\pi^n(K^n+\nu)/\pi^n(K^n) \geq 1$ (we are going to the right on the real line and in this direction the PMF increases),
2. $\nu = +1$ and $\pi^n(K^n+\nu)/\pi^n(K^n) < 1$ (we are going to the right on the real line and in this direction the PMF decreases),
3. $\nu = -1$ and $\pi^n(K^n+\nu)/\pi^n(K^n) \geq 1$ (we are going to the left on the real line and in this direction the PMF increases),

4. $\nu = -1$ and $\pi^n(K^n + \nu)/\pi^n(K^n) < 1$ (we are going to the left on the real line and in this direction the PMF decreases).

In Case 1,

$$\left(1 - 1 \wedge \frac{\pi^n(K^n + \nu)}{\pi^n(K^n)}\right) = 0,$$

for all n and $-\nu\,(\log f_S(S_K^n))' = -(\log f_S(S_K^n))'$ is negative in the limit because $f_S'(S_K^n)$ is positive in the limit. Therefore, $\max\{0, -\nu\,(\log f_S(S_K^n))'\} \longrightarrow 0$. Using Lebesgue's dominated convergence theorem, we thus know that the expectation at the RHS in (14) converges towards 0 when restricted to Case 1. We can prove that it converges towards 0 in Case 3 in the same way. In Case 2,

$$\sqrt{n}\left(1 - 1 \wedge \frac{\pi^n(K^n + \nu)}{\pi^n(K^n)}\right) = -\frac{1}{\pi^n(K^n)}\,\frac{\pi^n(K^n + \nu) - \pi^n(K^n)}{1/\sqrt{n}}.$$

By assumption, we know that this behaves asymptotically like $(\log f_S(S_K^n))'$. We also know that $-\nu\,(\log f_S(S_K^n))' = -(\log f_S(S_K^n))'$ is positive in the limit because $f_S'(S_K^n)$ is negative in the limit. Therefore, $-\max\{0, -\nu\,(\log f_S(S_K^n))'\}$ behaves like $(\log f_S(S_K^n))'$ in the limit. Using Lebesgue's dominated convergence theorem, we thus know that the expectation at the RHS in (14) converges towards 0 when restricted to Case 2 (recall the assumed boundedness of the limiting quantity in the expectation). We can prove that it converges towards 0 in Case 4 in the same way. ∎

3 Details about the multiple change-point example

It is assumed that the Poisson process has been observed on the time interval $[0, L]$, where $L > 0$ is known. The starting point for each step is denoted by $s_{j,k}$, $j = 0, \ldots, k$, to which we add the endpoint of the last step $s_{k+1,k}$, where these are subject to the constraint $0 =: s_{0,k} < s_{1,k} < \ldots < s_{k+1,k} := L$. The height of the j-th step is denoted by $h_{j,k}$, $j = 1, \ldots, k+1$. The log-likelihood of model k is

$$\log \mathcal{L}(\mathbf{x}_k \mid k, \mathbf{t}) := \sum_{i=1}^{n} \log(\lambda_k(t_i \mid \mathbf{x}_k)) - \int_0^L \lambda_k(t \mid \mathbf{x}_k)\,dt,$$

where $\lambda_k(t \mid \mathbf{x}_k) := \sum_{j=0}^{k} h_{j+1,k} \mathbb{1}_{[s_{j,k}, s_{j+1,k})}(t)$ for $t \in [0, L]$ and $\mathbf{x}_k := (s_{1,k}, \ldots, s_{k,k}, h_{1,k}, \ldots, h_{k+1,k})^T$, $\mathbb{1}$ being the indicator function.

We use the same prior structure as Green (1995). The prior on K is a Poisson distribution with parameter $\lambda > 0$, but conditioned on $K \leq \mathrm{K}_{\max}$. Given $K = k$, the starting points $s_{1,k}, \ldots, s_{k,k}$ are *a priori* distributed as the even-numbered order statistics from $2k+1$ points uniformly distributed on $[0, L]$, and the heights are independently and identically distributed as $\Gamma(\alpha, \beta)$, where $\alpha > 0$ and $\beta > 0$ are the shape and rate parameters, respectively. In Green (1995), the hyperparameters are set to $\lambda := 3, \mathrm{K}_{\max} := 30, \alpha := 1$, and $\beta := 200$.

As done in Section 5.1 of our paper, one may take advantage of the information at its disposal about the problem and model to design the sampler. Green (1995) follows this approach. We design the RJ and the corresponding NRJ as this author. For parameter updates, we randomly choose to modify either one of the heights $h_{j,k}$ or one of the starting points $s_{j,k}$. We modify a starting point $s_{j,k}$ by proposing a new value uniformly between $s_{j-1,k}$ and $s_{j+1,k}$. We modify a height $h_{j,k}$ by proposing a new value $h'_{j,k}$ that is such that $\log(h'_{j,k}/h_{j,k}) \sim \mathcal{U}[-1/2, 1/2]$. For model switches, we randomly choose to either add or withdraw a step. When we add a step, we first generate its starting point $s^* \sim \mathcal{U}[0, L]$. Deterministically, given s^*, we know which step will be splitted in two, in the sense that the proposal for the starting points is: $(s_{0,k}, \ldots, s_{j^*,k}, s^*, s_{j^*+1,k}, \ldots, s_{k+1,k})$, where $s_{0,k} < \ldots < s_{j^*,k} < s^* < s_{j^*+1,k} < \ldots < s_{k+1,k}$ (the step $(s_{j^*,k}, s_{j^*+1,k}]$ is splitted in two). We perturb as follows the height of this step $h_{j^*+1,k}$ to obtain proposals for the two heights $h'_{j^*+1,k+1}$ and $h'_{j^*+2,k+1}$ in the proposed model: generate $u_p \sim \mathcal{U}[0, 1]$ which is such that $h'_{j^*+2,k+1}/h'_{j^*+1,k+1} = (1 - u_p)/u_p$, and set the height proposals such that

$$(h'_{j^*+1,k+1})^{\frac{s^* - s_{j^*,k}}{s_{j^*+1,k} - s_{j^*,k}}} (h'_{j^*+2,k+1})^{\frac{s_{j^*+1,k} - s^*}{s_{j^*+1,k} - s_{j^*,k}}} = h_{j^*+1,k}.$$

The height proposals are $(h_{1,k}, \ldots, h_{j^*,k}, h'_{j^*+1,k+1}, h'_{j^*+2,k+1}, h_{j^*+2,k}, \ldots, h_{k+1,k})$. When we withdraw a step, we proceed with the reverse move, which is deterministic after having generated $j^* \sim \mathcal{U}\{0, \ldots, k-1\}$ (starting from model k). See Green (1995) for the acceptance probabilities and more details.

For implementing Algorithm 3 and the corresponding RJ, we proceed as in Karagiannis and Andrieu (2013) for generating the paths. More precisely, when switching from model k to model $k+1$, we use the same strategy as above to set the starting point of the path to $(s_{0,k}, \ldots, s_{j^*,k}, s^*, s_{j^*+1,k}, \ldots, s_{k+1,k})$ and $(h_{1,k}, \ldots, h_{j^*,k}, h'_{j^*+1,k+1}, h'_{j^*+2,k+1}, h_{j^*+2,k}, \ldots, h_{k+1,k})$, which are parameters in model $k+1$. We next update the parameters in model $k+1$ using blockwise MCMC sweeps. In a random order, we modify one of the heights $h_{j,k+1}$ and one of the starting points $s_{j,k+1}$ as when updating the parameters in Algorithm 1 (as explained above), and we update j^*. Note that when updating $h_{j,k+1}$ and $s_{j,k+1}$, we update the corresponding parameters in model k as they are linked through deterministic functions. When updating j^* given the rest, the parameters in model k may be updated as we may change which step is splitted in two. The intermediate distributions are

$$\rho^{(t)}_{k \mapsto k+1}(\mathbf{x}^{(t)}_k, \mathbf{u}^{(t)}_{k \mapsto k'}) \propto \left[\pi(k, \mathbf{x}^{(t)}_k)\,\frac{1}{L}\,\frac{h^{(t)}_{j^*+1,k}}{(h^{(t)}_{j^*+2,k+1} + h^{(t)}_{j^*+1,k+1})^2}\right]^{1-t/T} \left[\pi(k+1, \mathbf{y}^{(t)}_{k+1})\,\frac{1}{k+1}\right]^{t/T}.$$

See Karagiannis and Andrieu (2013) for more details.

We finish this section by explaining how we established the number of iterations for the vanilla samplers. In Algorithm 2 when we switch models (see Step 2.(a)), we first generate the starting point of the path, which is done as in Algorithm 1, and next we generate the path using $T-1$ MCMC steps. In each MCMC step, we try to modify one of the heights, one of the starting points and we sample j^*. Each MCMC step is thus essentially equivalent to 3 parameter updates, which is in turn essentially equivalent to 3 model switching attempts. To one model switching attempt in Algorithm 1 we thus essentially need to add $3(T-1)$ model switching attempts to obtain an equivalent cost.

On average, in one Algorithm 2 run, there are $I(1-\tau)$ model switching attempts. Thus they correspond in Algorithm 1 to

$$I(1-\tau) + I(1-\tau)3(T-1) \leq I(1-\tau)3T$$

model switching attempts. To identify the equivalence between Algorithm 3 and Algorithm 1, we need to multiply the number above by 1.5, as explained in Section 3.2 of our paper. Therefore, if in one run of Algorithm 3 there are on average $I(1-\tau)$ model switching attempts, then they correspond to $I(1-\tau)4.5T$ model switching attempts in Algorithm 1. Algorithm 1 must thus be run for $I\tau + I(1-\tau)4.5T$ iterations and τ in this algorithm must be set to

$$\tau := \frac{I\tau}{I\tau + I(1-\tau)4.5T}.$$

References

Andrieu, C. and S. Livingstone (2019). Peskun-Tierney ordering for Markov chain and process Monte Carlo: beyond the reversible scenario. *arXiv:1906.06197*.

Bierkens, J., P. Fearnhead, G. Roberts, et al. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist. 47*(3), 1288–1320.

Bierkens, J. and G. Roberts (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab. 27*(2), 846–882.

Bouchard-Côté, A., S. J. Vollmer, and A. Doucet (2018). The Bouncy Particle Sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.*, 1–13.

Ethier, S. N. and T. G. Kurtz (1986). *Markov Processes: Characterization and Convergence.* Wiley.

Gagnon, P., M. Bédard, and A. Desgagné (2019). Weak convergence and optimal tuning of the reversible jump algorithm. *Math. Comput. Simulation 161*, 32–51.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4), 711–732.

Karagiannis, G. and C. Andrieu (2013). Annealed importance sampling reversible jump MCMC algorithms. *J. Comp. Graph. Stat. 22*(3), 623–648.

Karr, A. F. (1975). Weak convergence of a sequence of Markov chains. *Z. Wahrsch. Verw. Gebiete 33*(1), 41–48.

Roberts, G. O. and J. S. Rosenthal (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv. 1*, 20–71.

Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.*, 434–438.