*The Annals of Statistics*
2021, Vol. 49, No. 4, 1958–1981
https://doi.org/10.1214/20-AOS2008


# PESKUN–TIERNEY ORDERING FOR MARKOVIAN MONTE CARLO: BEYOND THE REVERSIBLE SCENARIO

BY CHRISTOPHE ANDRIEU[1] AND SAMUEL LIVINGSTONE[2]

[1] *School of Mathematics, University of Bristol, c.andrieu@bristol.ac.uk*
[2] *Department of Statistical Science, University College London, samuel.livingstone@ucl.ac.uk*

Historically time-reversibility of the transitions or processes underpinning Markov chain Monte Carlo methods (MCMC) has played a key role in their development, while the self-adjointness of associated operators together with the use of classical functional analysis techniques on Hilbert spaces have led to powerful and practically successful tools to characterise and compare their performance. Similar results for algorithms relying on nonreversible Markov processes are scarce. We show that for a type of nonreversible Monte Carlo Markov chains and processes, of current or renewed interest in the physics and statistical literatures, it is possible to develop comparison results which closely mirror those available in the reversible scenario. We show that these results shed light on earlier literature, proving some conjectures and strengthening some earlier results.

**1. Introduction.** Markov chain Monte Carlo (MCMC) is concerned with the simulation of realisations of $\pi$-invariant and ergodic Markov chains, where $\pi$ is a probability distribution of interest defined on some appropriate measurable space $(\mathsf{X}, \mathscr{X})$. Such realisations can be used to produce samples of distributions arbitrarily close to $\pi$, or approximate expectations with respect to $\pi$. For a given probability distribution $\pi$ the choice of a Markov chain is not unique, and understanding the nature of the approximation associated to particular choices is therefore of importance and has generated a substantial body of literature, both in statistical science and physics among others and directly related to our work (Andrieu and Vihola (2016), Bacallado (2010), Bornn et al. (2017), Caracciolo, Pelissetto and Sokal (1990), Doucet et al. (2015), Hobert and Marchev (2008), Leisen and Mira (2008), Liu (1996), Maire, Douc and Olsson (2014), Mira (2001), Neal (2004), Peskun (1973), Rey-Bellet and Spiliopoulos (2016), Roberts and Rosenthal (2014), Rosenthal and Rosenthal (2015), Sakai and Hukushima (2016), Sherlock, Thiery and Lee (2017), Tierney (1998)). The present paper is a contribution to this literature and addresses a scenario currently barely covered by existing theory, despite recent interest motivated by applications.

Due to its wide applicability, the Metropolis–Hastings update (Hastings (1970), Metropolis et al. (1953)) is the cornerstone of the design of general purpose MCMC algorithms. The corresponding Markov transition satisfies the so-called detailed balance property, ensuring that $\pi$ is left invariant by this update, but also implies reversibility of the numerous algorithms of which it is a building block. An unintended benefit of reversibility is given at a theoretical level. Using the operator interpretation of Markov transitions, the properties of reversible Markov chains can be studied with well-established functional analysis techniques developed for self-adjoint operators. The celebrated result of Peskun and its extensions (Caracciolo, Pelissetto and Sokal (1990), Peskun (1973), Tierney (1998)) are an example (see Mira (2001) for a review), and allow for practical performance comparisons in numerous scenarios of interest (see Theorem 1 and its Corollary for a quick reference), providing in particular clear

Received July 2019; revised June 2020.
*MSC2020 subject classifications.* Primary 65C40, 65C05; secondary 62J10.
*Key words and phrases.* Markov chain Monte Carlo, Peskun ordering, piecewise deterministic Markov processes.

answers to questions concerned with the design of algorithms. While reversibility facilitates theoretical analysis and has historically enabled methodological developments, it is not necessarily a desirable property when performance is considered. Informally such processes have a tendency to "backtrack", slowing down exploration of the support of the target distribution $\pi$.

Recently, there has been renewed interest in the design of $\pi$-invariant Markov chains which are not reversible. In several specific scenarios it has been shown that departing from reversibility can both improve the speed of convergence of a Markov chain (Diaconis, Holmes and Neal (2000)), and reduce the asymptotic variance of resulting estimators (Sakai and Hukushima (2016)) (although counterexamples also exist (Roberts and Rosenthal (2015))). Certain nonreversible samplers have been known for some time (Gustafson (1998), Horowitz (1991)), but interest has been re-kindled more recently thanks to a suite of methods which are not instances of the Metropolis–Hastings class. All of these Markov transition probabilities share a common structure, illustrated here with a very simple example. Assume that $\mathsf{X} = \mathbb{Z}$, let $E := \mathsf{X} \times \{-1, 1\}$, embed the distribution of interest $\pi$ into $\mu(x, v) := \frac{1}{2}\pi(x)\mathbb{I}\{v \in \{-1, 1\}\}$ and consider the Markov transition

$$P(x, v; y, w) := \alpha(x, v)\mathbb{I}\{y = x + v, w = v\} + \mathbb{I}\{y = x, w = -v\}\bar{\alpha}(x, v), \tag{1}$$

where $\alpha(x, v) := \min\{1, \pi(x + v)/\pi(x)\}$, $\bar{\alpha}(x, v) = 1 - \alpha(x, v)$ and $\mathbb{I}S$ is the indicator function of set $S$. In words, starting at $(x, v)$, the first component of the Markov chain generated by $P$ will travel in the same direction $v$ in increments of size one until a rejection occurs and the direction is reversed. One can check that this does not satisfy detailed balance with respect to $\mu$ (or indeed $\pi$), but a similar looking property

$$\mu(x, v)P(x, v; y, w) = \mu(y, w)P(y, -w; x, -v),$$

for $(x, v), (y, w) \in E$. This is referred to as modified detailed balance in the literature (Fang, Sanz-Serna and Skeel (2014)) or skewed detailed balance (Hukushima and Sakai (2013)), and leads to what is known as Yaglom reversibility (Yaglom (1949)). It is instructive to write this identity in terms of the transition probability $Q(x, v; y, w) := \mathbb{I}\{y = x\}\mathbb{I}\{w = -v\}$, so that it now reads $\mu(x, v)P(x, v; y, w) = \mu(y, w)QPQ(y, w; x, v)$ where $QPQ$ is the composition of the three kernels. An interpretation of this identity is that the corresponding operator $QPQ$, not $P$, is the adjoint of $P$ as is the case in the self-adjoint scenario. This structure of the adjoint of $P$, together with the fact that $Q^2$ is the identity, play a central role in our analysis and covers a surprisingly large number of known scenarios and applications currently beyond the reach of earlier theory. Indeed our theory does not require the embedding $\mu$ of $\pi$ to be of the specific form above, and $Q$ is only required to be an isometric involution; see Section 2.1 for a precise definition in the present context. As we shall see, this structure allows us to develop a theory for performance comparison for this class of MCMC algorithms which parallels that existing for reversible algorithms; see Section 2.2. Applications are given in Section 3, and include the proof of conjectures concerned with the lifted Metropolis–Hastings method of Turitsyn, Chertkov and Vucelja (2011), Vucelja (2016) and improves and generalises the results of Sakai and Hukushima (2016), provide a direct and rigorous proof of Neal (2004) in a more general set-up and a connection to the results of Maire, Douc and Olsson (2014), which is generalised, permitting the characterisation of algorithms (e.g., Campos and Sanz-Serna (2015), Horowitz (1991)) currently not covered by existing theory.

We show that this structure is shared by nonreversible Markov process Monte Carlo (MPMC) methods, the continuous-time pendant of MCMC, such as the Bouncy Particle Sampler, the Zig-Zag process or event-chain processes which have recently attracted some attention (Bierkens, Fearnhead and Roberts (2019), Bierkens and Roberts (2017), Bou-Rabee and Sanz-Serna (2017), Bouchard-Côté, Vollmer and Doucet (2018), Ottobre (2016), Peters and de With (2012)). Characterisation of this property in the continuous-time setup is precisely formulated in Section 4.1 and a concrete example discussed in Section 5.2. In Section 4.2, we propose new tools which enable performance comparison for this class of processes and an application is presented in Section 5. All the proofs can be found in the Supplementary Material (Andrieu and Livingstone (2021)).

Throughout this paper, we will use the following standard notation. Let $(E, \mathscr{E})$ be a measurable space. For Markov kernels $T_1, T_2 \colon E \times \mathscr{E} \to [0,1]$, we let $T_1T_2(z, A) := \int T_1(z, \mathrm{d}z')T_2(z', A)$ for all $A \in \mathscr{E}$ and for any probability measure $\nu$ on $(E, \mathscr{E})$ and $f \in \mathbb{R}^E$ measurable, we let $\nu(f) := \int f \,\mathrm{d}\nu$, sometimes simplified to $\nu f$ when no ambiguity is possible and whenever this quantity exists. We denote by $T$ the associated operators acting on functions to the right as $Tf(z) := \int f(z')T(z, \mathrm{d}z')$ for $z \in E$, and on measures to the left as $\nu T(A) := \int_E \int_A \nu(\mathrm{d}z)T(z, \mathrm{d}z')$ for every $A \in \mathscr{E}$. Let $\mu$ be a probability distribution defined on some measurable space $(E, \mathscr{E})$. Whenever the following exist, for $f, g : E \to \mathbb{R}$, we define $\langle f, g\rangle_\mu := \int fg \,\mathrm{d}\mu$, $\|f\|_\mu := (\int f^2 \,\mathrm{d}\mu)^{1/2}$ and the Hilbert spaces $L^2(\mu) := \{f \in \mathbb{R}^E : \|f\|_\mu < \infty\}$, with $\mathbb{R}^E$ the set of functions $E \to \mathbb{R}$, and $L^2_0(\mu) := L^2(\mu) \cap \{f \in \mathbb{R}^E : \mu(f) = 0\}$. We let $|||T|||_\mu := \sup_{\|f\|_\mu=1} \|Tf\|_\mu$ and denote $T^*$ the adjoint of $T$, whenever it is well defined. For a set $S$, we let $S^c$ be its complement in the ambient space.

## 2. Discrete time scenario—general results.

2.1. *The notion of* $(\mu, Q)$*-self-adjointness.* Here, we formalise the notion of $(\mu, Q)$-self-adjointness, and discuss its consequences.

DEFINITION 1. We call a linear operator $Q \colon L^2(\mu) \to L^2(\mu)$ an *isometric involution* if:

(a) $\langle f, g\rangle_\mu = \langle Qf, Qg\rangle_\mu$ for all $f, g \in L^2(\mu)$,
(b) $Q^2 = \mathrm{Id}$, the identity operator.

REMARK 1. We note the simple properties for $f, g \in L^2(\mu)$:

- $Q$ is $\mu$-self-adjoint since $\langle f, Qg\rangle_\mu = \langle Qf, Q^2g\rangle_\mu = \langle Qf, g\rangle_\mu$,
- the operators $\Pi_+ := (\mathrm{Id} + Q)/2$ and $\Pi_- := (\mathrm{Id} - Q)/2$ are $\mu$-self-adjoint projectors and $f = \Pi_+ f + \Pi_- f$, $Q\Pi_+ f = \Pi_+ f$ and $Q\Pi_- f = -\Pi_- f$.
- for $\Pi$ an orthogonal projector, $Q = \pm(\mathrm{Id} - 2\Pi)$ is an isometric involution.

From now on, we will assume that $Q$ is a Markov operator, and again, we will use the same symbol for the associated Markov kernel $Q \colon E \times \mathscr{E} \to [0,1]$ and note that $\mu Q = \mu$. The following establishes that there exists an involution $\xi : E \to E$ such that for all $f \in \mathbb{R}^E$ and $z \in E$, $Qf(z) = f \circ \xi(z)$.

LEMMA 1. *Let* $T \colon E \times \mathscr{E} \to [0,1]$ *be a Markov transition such that for any* $z \in E$, $T^2(z, \{z\}) = 1$, *then there exists an involution* $\tau \colon E \to E$ *such that for* $z, z' \in E$, $T(z, \mathrm{d}z') = \delta_{\tau(z)}(\mathrm{d}z')$.

DEFINITION 2. We say a Markov operator $P \colon L^2(\mu) \to L^2(\mu)$ is $(\mu, Q)$-self-adjoint if there is an isometric involution $Q$ such that for all $f, g \in L^2(\mu)$ it holds that $\langle Pf, g\rangle_\mu = \langle f, QPQg\rangle_\mu$.

We will say that the corresponding kernel $P \colon E \times \mathscr{E} \to [0,1]$ is $(\mu, Q)$-reversible. When $Q = \mathrm{Id}$, we will simply say that $P$ is $\mu$-self-adjoint or $\mu$-reversible. The following is a simple but important characterisation of $(\mu, Q)$-self-adjoint operators.

PROPOSITION 1. *If the Markov operator $P$ is $(\mu, Q)$-self-adjoint (resp., $\mu$-self-adjoint) then $QP$ and $PQ$ are $\mu$-self-adjoint (resp., $(\mu, Q)$-self-adjoint). As a result a $(\mu, Q)$-self-adjoint Markov operator is always the composition of two $\mu$-self-adjoint Markov operators.*

DEFINITION 3. For $P$ a $(\mu, Q)$-self-adjoint operator, we call $QP$ and $PQ$ its left and right $Q$-symmetrisations.

2.2. *Ordering of asymptotic variances.* For an homogeneous Markov chain $\{Z_0, Z_1, \dots\}$ of transition kernel $P$ leaving $\mu$ invariant, started at equilibrium and any $\mu$–measurable $f : E \to \mathbb{R}$, we define the *asymptotic variance*

$$\mathrm{var}(f, P) := \lim_{n\to\infty} n \,\mathrm{var}\left(n^{-1}\sum_{i=0}^{n-1} f(Z_i)\right), \tag{2}$$

whenever the limit exists. This limit always exists, but may be infinite, when $P$ is $\mu$-reversible and $f \in L^2(\mu)$ (Tierney (1998)). Beyond this scenario, general criteria exist (Maigret (1978), Glynn and Meyn ((1996), Theorem 4.1)) and often require a bespoke analysis. A general question of interest is, given two Markov transitions $P_1$ and $P_2$ leaving $\mu$ invariant, can one find a simple criterion to establish that for some function $f$, $\mathrm{var}(f, P_1) \geq \mathrm{var}(f, P_2)$ or $\mathrm{var}(f, P_1) \leq \mathrm{var}(f, P_2)$, when these quantities are well defined. Provided the computational costs of implementing the two corresponding algorithms are comparable, this has the potential to provide practitioners with clear guidelines about which of the two kernels to use. When $P_1$ and $P_2$ are $\mu$-reversible, a criterion based on Dirichlet forms leads to a particularly simple solution. Beyond this scenario little is known in general.

For $P$ and $f \in L^2(\mu)$, define the Dirichlet form

$$\mathcal{E}(f, P) := \langle f, (\mathrm{Id} - P)f\rangle_\mu = \frac{1}{2}\int \left[f(z') - f(z)\right]^2 \mu(\mathrm{d}z)P(z, \mathrm{d}z'), \tag{3}$$

and we let $\mathrm{Gap}_R(P) := \inf_{f\in L_0^2(\mu), \|f\|_\mu \neq 0} \mathcal{E}(f, P)/\|f\|_\mu^2$.

REMARK 2. Note that the expression (3) is a well-defined object whether or not $P$ is $\mu$-reversible, but is typically of less interest in the nonreversible scenario. In the present paper, however, nonreversible chains are compared through the Dirichet forms of their left and right $Q$-symmetrisations, which are in fact $\mu$-reversible.

THEOREM 1 (Caracciolo, Pelissetto and Sokal (1990), Tierney (1998)). *Let $\mu$ be a probability distribution on some measurable space $(E, \mathscr{E})$, and let $P_1$ and $P_2$ be two $\mu$-reversible Markov transitions. If for any $g \in L^2(\mu)$, $\mathcal{E}(g, P_1) \geq \mathcal{E}(g, P_2)$, then for any $f \in L^2(\mu)$*

$$\mathrm{var}(f, P_1) \leq \mathrm{var}(f, P_2) \quad \textit{and} \quad \mathrm{Gap}_R(P_1) \geq \mathrm{Gap}_R(P_2).$$

COROLLARY 1 (Peskun (1973)). *Whenever for any $z \in E$ and $A \in \mathscr{E}$ it holds that $P_1(z, A \cap \{z\}^c) \geq P_2(z, A \cap \{z\}^c)$ then for any $f \in L^2(\mu)$,*

$$\mathrm{var}(f, P_1) \leq \mathrm{var}(f, P_2) \quad \textit{and} \quad \mathrm{Gap}_R(P_1) \geq \mathrm{Gap}_R(P_2).$$

EXAMPLE 1. Consider two $\mu$–reversible Metropolis–Hastings kernels of the form $P_i(z, \mathrm{d}z') := T_i(z, \mathrm{d}z') + (1 - T_i(z, E))\delta_z(\mathrm{d}z')$ for $i \in \{1, 2\}$, where $T_i(z, \mathrm{d}z') := Q(z, \mathrm{d}z') \times \alpha_i(z, z')$, $Q$ is a Markov transition kernel, and each $\alpha_i : E \times E \to [0, 1]$ satisfies $\alpha_1(z, z') \geq \alpha_2(z, z')$ for all $(z, z') \in E \times E$. Then $2(\mathcal{E}(g, P_1) - \mathcal{E}(g, P_2)) = \int (g(z') - g(z))^2(\alpha_1(z, z') - \alpha_2(z, z'))\mu(\mathrm{d}z)Q(z, \mathrm{d}z')$, which is nonnegative for any $g \in L^2(\mu)$, meaning $\mathrm{var}(f, P_1) \leq \mathrm{var}(f, P_2)$ for any $f \in L^2(\mu)$ and $\mathrm{Gap}_R(P_1) \geq \mathrm{Gap}_R(P_2)$.

Our main result is that these conclusions extend in part to $(\mu, Q)$-reversible transitions. In order to ensure the existence of the quantities we consider, for any $\lambda \in [0, 1)$ we introduce the $\lambda$-asymptotic variance, defined for any $f \in L^2(\mu)$, with $\bar{f} := f - \mu(f)$, as

$$\mathrm{var}_\lambda(f, P) := \|\bar{f}\|_\mu^2 + 2\sum_{k\geq 1} \lambda^k \langle \bar{f}, P^k \bar{f}\rangle_\mu = 2\langle \bar{f}, [\mathrm{Id} - \lambda P]^{-1}\bar{f}\rangle_\mu - \|\bar{f}\|_\mu^2.$$

Whether $\lim_{\lambda\uparrow 1} \mathrm{var}_\lambda(f, P) = \mathrm{var}(f, P)$ when the latter exists is problem specific and not addressed here, but we note that this is always true in the reversible scenario (Tierney (1998)) and that a general sufficient condition is that $\sum_{k\geq 1} |\langle \bar{f}, P^k \bar{f}\rangle_\mu| < \infty$ (see Corollary 3 for a detailed discussion). Rather we focus on ordering $\mathrm{var}_\lambda(P_1, f)$ and $\mathrm{var}_\lambda(P_2, f)$ for $\lambda \in [0, 1)$ and leave the convergence to the asymptotic variances as a separate problem.

THEOREM 2. *Let $\mu$ be a probability distribution on some measurable space $(E, \mathscr{E})$, and let $P_1$ and $P_2$ be two $(\mu, Q)$-reversible Markov transitions. Assume that for any $g \in L^2(\mu)$, $\mathcal{E}(g, QP_1) \geq \mathcal{E}(g, QP_2)$, or for any $g \in L^2(\mu)$, $\mathcal{E}(g, P_1Q) \geq \mathcal{E}(g, P_2Q)$. Then for any $\lambda \in [0, 1)$ and $f \in L^2(\mu)$*:

(a) *satisfying $Qf = f$ it holds that* $\mathrm{var}_\lambda(f, P_1) \leq \mathrm{var}_\lambda(f, P_2)$,
(b) *satisfying $Qf = -f$ it holds that* $\mathrm{var}_\lambda(f, P_1) \geq \mathrm{var}_\lambda(f, P_2)$.

COROLLARY 2. *If $P_1$ and $P_2$ are such that for $\mu$-almost all $z \in E$ and every $A \in \mathscr{E}$, it holds that $P_1Q(z, A \cap \{z\}^c) \geq P_2Q(z, A \cap \{z\}^c)$, or $QP_1(z, A \cap \{z\}^c) \geq QP_2(z, A \cap \{z\}^c)$, then the conclusion of Theorem 2 holds.*

COROLLARY 3. *If for $i \in \{1, 2\}$ and $f \in L^2(\mu)$ the following limits exist and* $\lim_{\lambda\uparrow 1} \mathrm{var}_\lambda(f, P_i) = \mathrm{var}(f, P_i)$, *then* $\mathrm{var}(f, P_1) \leq \mathrm{var}(f, P_2)$. *This is satisfied if $\sum_{k\geq 1} |\langle \bar{f}, P^k \bar{f}\rangle_\mu| < \infty$, which can be established under fairly general conditions (Andrieu, Fort and Vihola* ((2015), *Theorem* 1, *Section* 4.1 *and* 4.2)): *essentially one requires the existence of $C \subset E$,*

(a) *$\epsilon > 0$ and a probability measure $\nu$ on $E$ such that for any $z \in C$ the inequality $P(z, \cdot) \geq \epsilon\nu(\cdot)$ is satisfied,*
(b) *a Lyapunov function $V: E \to [1, \infty)$, $\mu(V) < \infty$ and a concave function $\phi: [1, \infty) \to (0, \infty)$ such that the following holds*:

$$PV(z) \leq V(z) - \phi \circ V(z) + b\mathbb{I}\{z \in C\},$$

*and $\mu(|f|V) < \infty$.*

REMARK 3. We note that in contrast with the reversible scenario the result never provides us with information about the speed of convergence to equilibrium. The practical guideline resulting from the theorem is that after "burn-in" an algorithm should be tuned to maximise or minimise $\mathcal{E}(g, QP)$ or $\mathcal{E}(g, PQ)$ for all $g \in L^2(\mu)$.

REMARK 4. We expect the scenario where $f = Qf$ to be the only scenario relevant to statistical applications; $f = -Qf$ is provided for completeness only. From the proof it can be seen that ordering $\mathrm{var}_\lambda(f, P_1)$ and $\mathrm{var}_\lambda(f, P_2)$ for a specific $f \in L_0^2(\mu)$ such that $Qf = f$, only requires ordering Dirichlet forms for a particular subset of $L^2(\mu)$, namely the $\lambda$-solutions of the Poisson equation (see the proof of Theorem 2). Although these quantities are generally intractable, in some scenarios their structure may be exploited to order the Dirichlet forms involved. Such ideas have been extensively used in the reversible scenario, for example, in Andrieu and Vihola (2015, 2016) and we provide an example in the $(\mu, Q)$-self-adjoint setup in Section 3.1. Another consequence of this is that strict inequalities can be obtained when the Dirichlet forms are strictly ordered for nonconstant functions and these $\lambda$-solutions are nonconstant.

REMARK 5. More quantitative versions of this result, in the spirit of Caracciolo, Pelissetto and Sokal (1990) can also be replicated. If for $\alpha \in (0, 1]$ and all $g \in L^2(\mu)$, $\langle g, (\mathrm{Id} - QP_1)g\rangle_\mu \geq \alpha^{-1}\langle g, (\mathrm{Id} - QP_2)g\rangle_\mu$ then $\mathrm{var}_{\lambda\alpha(1-\lambda+\lambda\alpha)^{-1}}(f, P_1) \leq [1-\lambda+\lambda\alpha]\mathrm{var}_\lambda(f, P_2) - \lambda(1-\alpha)\|\bar{f}\|_\mu^2$. When the limits as $\lambda \uparrow 1$ exist this implies $\mathrm{var}(f, P_1) \leq \alpha \mathrm{var}(f, P_2) - (1 - \alpha)\|\bar{f}\|_\mu^2$; see the Supplementary Material.

**3. Discrete time scenario: Examples.** The notion of $(\mu, Q)$-reversibility, often described in terms of modified or skewed detailed balance, is known to hold for numerous processes of interest but its implications, beyond establishing that the corresponding Markov chain leaves $\mu$ invariant, are to the best of our knowledge unknown. In this section, we show that our framework contributes to filling this gap and revisit a wide range of simple, some foundational questions. In some scenarios, $(\mu, Q)$-reversibility is not immediately apparent for a specific problem and we present basic strategies to remedy this. More complex examples are possible, such as extension of Andrieu and Vihola (2015, 2016) or Andrieu (2016), for example, but beyond the scope of this paper.

3.1. *Links to* 2*-cycle based MCMC kernels.* Recently, Maire, Douc and Olsson (2014), have shown that results for ordering of asymptotic variances of reversible time-homogeneous Markov chains can be extended to certain inhomogeneous Markov chains arising naturally in the context of MCMC algorithms. Such chains are obtained by cycling between two reversible MCMC kernels (see Algorithm 1), and it is a natural question to ask whether improving either of the kernels in terms of individual Dirichlet forms improves performance of the inhomogeneous chain resulting from their combination. Theorem 3 below provides us with a simple and practical characterisation. We show that this result is in some sense dual to $(\mu, Q)$-reversibility and provide a generalisation which makes previously intractable analysis of some algorithms possible. For $\pi$ a probability distribution on some space $(\mathsf{X}, \mathscr{X})$, $P_1$ and $P_2$ two $\pi$-invariant Markov transitions and $f \in L^2(\pi)$, we extend the definition of $\lambda$-asymptotic variance, for $\lambda \in [0, 1)$ to the time inhomogeneous scenario

$$\begin{aligned}&\mathrm{var}_\lambda(f, \{P_1, P_2\}) + \|\bar{f}\|_\pi^2\\&\quad := [\mathrm{Id} - \lambda^2 P_1 P_2]^{-1}(\mathrm{Id} + \lambda P_1)\bar{f} + [\mathrm{Id} - \lambda^2 P_2 P_1]^{-1}(\mathrm{Id} + \lambda P_2)\bar{f},\end{aligned}$$

where $\bar{f} := f - \pi(f)$, which is well defined since for any $g \in L^2(\pi)$ $\|P_1 g\|_\pi \leq \|g\|_\pi$ and $\|P_2 g\|_\pi \leq \|g\|_\pi$. Under additional assumptions (see, e.g., Maire, Douc and Olsson (2014), Proposition 9) the following limits exist and satisfy:

$$\lim_{\lambda\uparrow 1} \mathrm{var}_\lambda(f, \{P_1, P_2\}) = \lim_{n\to\infty} n \ \mathrm{var}\left(n^{-1}\sum_{i=0}^{n-1} f(X_i)\right),$$

where here $\{X_0, X_1, \ldots\}$ is the time inhomogeneous Markov chain obtained by cycling through $P_1$ and $P_2$ and of initial distribution $\pi$, that is, for $A \in \mathscr{X}$, $\mathbb{P}(X_k \in A \mid X_0, \ldots, X_{k-1}) = P_{2-(k \mod 2)}(X_{k-1}, A)$ for $k \geq 1$ and $X_0 \sim \pi$. The following is a reformulation of Maire, Douc and Olsson ((2014), Theorem 4 and Lemma 25) combined with a generalisation of Maire, Douc and Olsson ((2014), Lemma 18).

**Algorithm 1** A 2-cycle based MCMC algorithm

- Initialisation $X_0$, $i = 0, n$. Require: $\pi$-reversible transition kernels $P_1$ and $P_2$.
- For $i = 1$ to $n$
  Draw $X_{i+1} \sim P_{2-((i+1) \mod 2)}(X_i, \cdot)$

THEOREM 3 (see Maire, Douc and Olsson ((2014), Theorem 4 and Lemma 25)). *Let $\pi$ be a probability distribution defined on $(\mathsf{X}, \mathscr{X})$. For $i, j \in \{1, 2\}$, let $P_{i,j} : \mathsf{X} \times \mathscr{X} \to [0, 1]$ be $\pi$-reversible Markov kernels such that for all $g \in L^2(\pi)$ and $i \in \{1, 2\}$ we have $\mathcal{E}(g, P_{1,i}) \geq \mathcal{E}(g, P_{2,i})$. Then, in the time inhomogeneous scenario, for any $f \in L^2(\pi)$ and $\lambda \in [0, 1)$*

$$\operatorname{var}_\lambda\big(f, \{P_{1,1}, P_{1,2}\}\big) \leq \operatorname{var}_\lambda\big(f, \{P_{2,1}, P_{2,2}\}\big).$$

*Further, for the time homogeneous scenario, if $f \in L^2(\pi)$ is such that $P_{i,1}f = f$ (or $P_{i,2}f = f$) for $i \in \{1, 2\}$, then*

$$\operatorname{var}_\lambda(f, P_{1,1}P_{1,2}) \leq \operatorname{var}_\lambda(f, P_{2,1}P_{2,2}).$$

COROLLARY 4. *Let $Q$ be an isometric involution and $P_1$ and $P_2$ be $(\pi, Q)$-reversible. Note that for $i \in \{1, 2\}$, $P_i = Q(QP_i)$ (resp. $P_i = (P_iQ)Q$) and that both $P_{i,1} := Q$ (resp., $P_{i,1} := P_iQ$) and $P_{i,2} := QP_i$ (resp., $P_{i,2} := Q$) are $\pi$-self-adjoint by Proposition 1. We can therefore apply Theorem 3 and the conclusion of Theorem 2 holds for $f \in L^2(\pi)$ such that $Qf = f$.*

Conversely one can show using a very simple argument that the first statement of Theorem 3 is a direct consequence of $(\mu, Q)$-reversibility of a particular time-homogeneous chain, where time is now part of the state, for a particular isometric involution. Apart from linking two seemingly unrelated ideas, an interest of the proof is that it highlights the difficulty with extending the results to $m$-cycles with $m \geq 3$.

REMARK 6. Note that the instrumental Markov chains introduced in the proof are never ergodic, but can be ergodic marginally. It is possible to revisit this proof for $m$-cycles and $m \geq 3$, but the property $v_1 \oplus (-v_2) = v_1 \oplus v_2$ fails in this scenario, in general, and it is not possible to conclude.

Theorem 4 below extends Theorem 3 to 2-cycles of $(\mu, Q)$-reversible Markov kernels–applications of this result are given in Section 3.2.

THEOREM 4. *Let $\pi$ be a probability distribution defined on some probability space $(\mathsf{X}, \mathscr{X})$. For $i, j \in \{1, 2\}$, let $P_{i,j} : \mathsf{X} \times \mathscr{X} \to [0, 1]$ be $(\mu, Q)$-reversible Markov kernels for some isometric involution $Q$, and such that for all $i \in \{1, 2\}$ we have $\mathcal{E}(g, QP_{1,i}) \geq \mathcal{E}(g, QP_{2,i})$ for all $g \in L^2(\pi)$, or $\mathcal{E}(g, P_{1,i}Q) \geq \mathcal{E}(g, P_{2,i}Q)$ for all $g \in L^2(\pi)$. Then for any $f \in L^2(\pi)$ such that $Qf = f$ and $\lambda \in [0, 1)$*

$$\operatorname{var}_\lambda\big(f, \{P_{1,1}, P_{1,2}\}\big) \leq \operatorname{var}_\lambda\big(f, \{P_{2,1}, P_{2,2}\}\big).$$

*Further, if $f \in L^2(\pi)$ is such that $P_{i,1}f = f$ (or $P_{i,2}f = f$) for $i \in \{1, 2\}$, then*

$$\operatorname{var}_\lambda(f, P_{1,1}P_{1,2}) \leq \operatorname{var}_\lambda(f, P_{2,1}P_{2,2}).$$

3.2. *Construction of Markov kernels from time-reversible flows.* A generic way to construct $(\mu, Q)$-reversible Markov kernel consists of the following slight generalisation of Fang, Sanz-Serna and Skeel (2014), Horowitz (1991). For a probability distribution $m$ on $(E, \mathscr{E})$ and measurable mapping $\psi : E \to E$ we let for any $A \in \mathscr{E}$, $m^\psi(A) := m(\psi^{-1}(A))$. The presentation parallels that of Tierney ((1998), Section 2, second example) in order to avoid specificities concerned with densities and, for example, the presence of Jacobians.

PROPOSITION 2. *Let $\mu$ be a probability distribution on $(E, \mathscr{E})$,*

(a) *$\psi : E \to E$ be a bijection such that $\psi^{-1} = \xi \circ \psi \circ \xi$ for $\xi : E \to E$ corresponding to an isometric involution $Q$,*

**Algorithm 2** An MCMC algorithm constructed from time-reversible flows

- Initialisation $Z_0$, $i = 0$, $n$. Require: maps $\psi$, $\xi$, $r$ and $\phi$, as defined in Proposition 2.
- For $i = 1$ to $n$
  (a) Set $Z' = \xi \circ \psi(Z_i)$
  (b) Compute $r(Z_i)$ and the acceptance rate $\phi \circ r(Z_i)$
  (c) Draw $U \sim \mathcal{U}[0, 1]$
    (i) If $U \le \phi \circ r(Z_i)$ then set $Z_{i+1} = Z'$
    (ii) Otherwise set $Z_{i+1} = Z_i$

(b) $\phi \colon \mathbb{R}_+ \to [0, 1]$ *such that* $r\phi(r^{-1}) = \phi(r)$ *for* $r > 0$ *and* $\phi(0) = 0$,
(c) *define for* $z \in E$, *with* $\nu := \mu + \mu^{\xi\circ\psi}$,

$$r(z) := \begin{cases} \dfrac{\mathrm{d}\mu^{\xi\circ\psi}/\mathrm{d}\nu(z)}{\mathrm{d}\mu/\mathrm{d}\nu(z)} & \text{if } \mathrm{d}\mu^{\xi\circ\psi}/\mathrm{d}\nu(z) > 0 \text{ and } \mathrm{d}\mu/\mathrm{d}\nu(z) > 0, \\ 0 & \text{otherwise}. \end{cases}$$

*then the following kernel is* $(\mu, Q)$-*reversible*,

$$P(z, \mathrm{d}z') := \phi \circ r(z)\delta_{\psi(z)}(\mathrm{d}z') + \delta_{\xi(z)}(\mathrm{d}z')[1 - \phi \circ r(z)]. \tag{4}$$

The resulting scheme is described in Algorithm 2. See also Chapter 2 of Lelièvre, Rousset and Stoltz (2010) for a treatment of some related schemes.

REMARK 7. Note that $\phi(r)$, together with $r(z)$, define the probability of accepting a transition to the new state $\psi(z)$. Choices of $\phi$ include $\phi(r) = \min\{1, r\}$, which leads to the standard Metropolis–Hastings acceptance rule, or $\phi(r) = r/(1 + r)$ which corresponds to Barker's dynamic. It is well known that for any $\phi$ satisfying Proposition 2(b) one has $\phi(r) \le \min\{1, r\}$ and that for Barker's choice $\frac{1}{2}\min\{1, r\} \le \phi(r)$.

EXAMPLE 2. Assume $E = \mathsf{X} \times \mathsf{V}$ and for $(x, v) \in E$ and $f \in \mathbb{R}^E$ let $Qf(x, v) := f(x, -v)$. Then for any $t \in \mathbb{R}$, $\psi_t(x, v) = (x + tv, v)$ satisfies $\psi_t^{-1} = \xi \circ \psi_t \circ \xi$ and was considered in Gustafson (1998) to define the Guided Random Walk Metropolis. More general examples satisfying this condition include $\psi_t(x, v) = \psi^B_{t/2} \circ \psi^A_t \circ \psi^B_{t/2}(x, v)$ where $\psi^A_t(x, v) := (x + t\nabla_v H(x, v), v)$ and $\psi^B_t(x, v) := (x, v - t\nabla_x H(x, v))$ for a separable Hamiltonian $H \colon E \to \mathbb{R}$. This is the Störmer–Verlet scheme considered in Horowitz (1991) to define the Hybrid Monte Carlo algorithm in the situation where $H := -\log \mathrm{d}\mu/\mathrm{d}\lambda^{\mathrm{Leb}}$ is well defined and separable. More generally dynamical systems with the time reversal symmetry (e.g., Lamb and Roberts (1998) and also Faggionato, Gabrielli and Ribezzi Crivellari ((2009), Lemma 3.14)) provide ways of constructing such mappings (see also Campos and Sanz-Serna (2015), Ottobre (2016), Ottobre et al. (2016), Poncet (2017), Sohl-Dickstein, Mudigonda and DeWeese (2014) and Fang, Sanz-Serna and Skeel (2014)).

In order to be useful in practice a Markov transition of the type given in (4) must be combined with another transition in order to lead to an ergodic Markov chain (Gustafson (1998), Horowitz (1991)). We focus here on 2-cycles of $(\mu, Q)$-reversible Markov transitions.

THEOREM 5. *Let* $\mu$ *be a probability distribution on* $(E, \mathscr{E})$ *and let* $\psi$ *satisfy Proposition* 2(a) *for some isometric involution* $Q$. *Further for* $i \in \{1, 2\}$, *let* $P_{i,2}$ *be as in* (4) *for a mapping* $\psi_i = \psi$ *and some mapping* $\phi_i$ *satisfying Proposition* 2-(b) *and let* $P_{1,1} = P_{2,1}$ *be a*

*$(\mu, Q)$-reversible Markov transition. Assume that $\phi_1 \geq \phi_2$, then for any $f \in L^2(\mu)$ such that $Qf = f$ and $\lambda \in [0, 1)$ we have*

$$\mathrm{var}_\lambda\big(f, \{P_{1,1}, P_{1,2}\}\big) \leq \mathrm{var}_\lambda\big(f, \{P_{2,1}, P_{2,2}\}\big).$$

*In particular, $\phi(r) = \min\{1, r\}$ achieves the smallest $\lambda$-asymptotic variance.*

EXAMPLE 3 (Example 2 (ctd)). Assume here for presentational simplicity that $\mathsf{X} = \mathsf{V} = \mathbb{R}$ that $\mu$ has a density with respect to the Lebesgue measure and $\mu(x, v) = \pi(x)\varpi(v)$ where $\varpi$ is a $\mathcal{N}(0, \sigma^2)$ for some $\sigma^2 > 0$. In this setup a popular choice (Duane et al. (1987)) for $P_{1,1} = P_{2,1}$ is a momentum refreshment of the type, for some $\theta \in (0, \pi/2]$,

$$R_\theta\big((x, v); \mathrm{d}(y, w)\big) = \int \delta_{(x, v\cos\theta + v'\sin\theta)}\big(\mathrm{d}(y, w)\big)\varpi(\mathrm{d}v').$$

Lemma 2 below establishes that the corresponding operator is $(\mu, Q)$-self-adjoint. We can therefore apply Theorem 5 and deduce, for example, that the choice $\phi(r) = \min\{1, r\}$ for all $\theta \in (0, \pi/2]$ is optimum. Further since $R_\theta(x, v; \{x\} \times \mathsf{V}) = 1$ we note that the second statement of Theorem 3 holds, a result partially known for $\theta = \pi/2$ since in this case for $i \in \{1, 2\}$ $P_{i,1}P_{i,2}$ is $\mu$-reversible and Theorem 1 can be applied.

LEMMA 2. *For any $\theta \in (0, \pi/2]$, $R_\theta$ is $(\mu, Q)$-self-adjoint for $Q$ such that $Qf(x, v) = f(x, -v)$ for $f \in \mathbb{R}^E$.*

Another application of the results above is the extra chance HMC method presented in Campos and Sanz-Serna (2015), equivalent to the ideas of Sohl-Dickstein, Mudigonda and DeWeese (2014), which can be seen as an extension to Horowitz's scheme (Horowitz (1991)). Using the notation of Proposition 2, the main idea is to define a variation of (4) where transitions to $\xi \circ \psi(x, v), \xi \circ \psi \circ \psi(x, v), \dots$ are attempted in sequence until success.

EXAMPLE 4. Here, $\mathsf{X} = \mathbb{R}$ for simplicity and $\mu(\mathrm{d}(x, v)) = \pi(\mathrm{d}x)\varpi(\mathrm{d}v)$ where $\varpi(\mathrm{d}v)$ is a $\mathcal{N}(0, \sigma^2)$. With $Qf(x, v) = f(x, -v)$ for $f \in \mathbb{R}^E$ and $\psi$ as in Proposition 2(a), we let $\psi^0 = \mathrm{Id}$ and $\psi^k = \psi \circ \psi^{k-1}$ for $k \in \mathbb{N} \setminus \{0\}$. Define for $K \in \mathbb{N} \setminus \{0\}$,

$$P_K\big((x, v); \mathrm{d}(y, w)\big) := \sum_{k=1}^{K} \beta_k(x, v)\delta_{\psi^k(x,v)}\big(\mathrm{d}(y, w)\big) + \rho_K(x, v)\delta_{\xi(x,v)}\big(\mathrm{d}(y, w)\big),$$

where, with $\alpha_0(x, v) = 0$ and for $k = 1, \dots, K$ $\alpha_k(x, v) = \alpha_{k-1}(x, v) \vee \{1 \wedge r_k(x, v)\}$, with

$$r_k(x, v) := \begin{cases} \dfrac{\mathrm{d}\mu^{\xi\circ\psi^k}/\mathrm{d}\nu_k(z)}{\mathrm{d}\mu/\mathrm{d}\nu_k(z)} & \text{if } \mathrm{d}\mu^{\xi\circ\psi^k}/\mathrm{d}\nu_k(z) > 0 \text{ and } \mathrm{d}\mu/\mathrm{d}\nu_k(z) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and $\nu_k := \mu + \mu^{\xi\circ\psi^k}$, $\beta_k(x, v) = \alpha_k(x, v) - \alpha_{k-1}(x, v)$ and $\rho_K(x, v) := 1 - \sum_{k=1}^{K} \beta_k(x, v)$. It is shown in Campos and Sanz-Serna ((2015), Appendix A) that this update is $(\mu, Q)$-reversible, while it is pointed out that for $\omega \in (0, \pi/2]$, $R_\omega P_K$ is not. We can apply Theorem 4 to deduce that for any $f \in L^2(\mu)$ such that $Qf = f$ and any $\omega \in (0, \pi/2]$, the mapping $K \mapsto \mathrm{var}_\lambda(f, \{R_\omega, P_K\})$ is nonincreasing, since from Lemma 3 below, $K \mapsto \mathcal{E}(g, P_K Q)$ is nondecreasing. In fact, since $R_\omega(x, v; \{x\} \times \mathsf{V}) = 1$, for $f \in L^2(\pi)$ and $\breve{f}(x, v) := f(x)$ for $(x, v) \in E$, we also deduce that $K \mapsto \mathrm{var}_\lambda(\breve{f}, R_\omega P_K)$ is nonincreasing.

LEMMA 3. *For any $g \in L^2(\mu)$, $K \mapsto \mathcal{E}(g, P_K Q)$ is nondecreasing.*

REMARK 8. As pointed out by Campos and Sanz-Serna (2015), the rational behind the approach is that for $(x, v) \in E$, $k \mapsto H \circ \xi \circ \psi^k(x, v)$ typically fluctuates around $H(x, v)$. As a result, if there exist $(x_0, v_0) \in E$ and $k_0 \in \mathbb{N}_*$ such that $\min_{1\leq k\leq k_0} H \circ \xi \circ \psi^k(x_0, v_0) >$

$\max\{H(x_0, v_0), H \circ \xi \circ \psi^{k_0+1}(x_0, v_0)\}$ and, for example, $(x, v) \mapsto H \circ \xi \circ \psi^{k_0+1}(x, v)$ is continuous in a neighbourhood of $(x_0, v_0)$, then $\mu(\{\beta_{k_0+1}(X, V) > 0\}) > 0$ and $\mathcal{E}(g, P_{k_0+1}Q) - \mathcal{E}(g, P_{k_0}Q) > 0$ for $L^2(\mu) \ni g \neq g \circ \xi \circ \psi^{k_0+1}$ on the aforementioned neighbourhood, suggesting that the strict performance improvement observed numerically in Campos and Sanz-Serna (2015) for specific functions holds more generally. A more precise investigation of this point is far beyond the scope of the present work.

It is natural to try to assess the impact of $\theta \in (0, \pi/2]$ involved in the definition of $R_\theta$ on the performance of the type of algorithms presented in this section. In particular, a long-standing question is whether partial momentum refreshment is preferable to full refreshment, meaning replacing $R_\theta$ by $R_{\pi/2}$. Application of Theorem 4 requires establishing that $\langle g, Q(R_{\pi/2} - R_\theta)g\rangle_\mu$ does not change sign for all $g \in L^2(\mu)$. This, however, is not the case. For example, setting $g_1(x, v) := v$ then the quantity is positive but for $g_2(x, v) := v^2$ it is negative and we cannot conclude.

3.3. *Lifted MCMC algorithms.* Assume we are interested in sampling from $\pi$ defined on $(\mathsf{X}, \mathscr{X})$ and are given two sub-stochastic kernels $T_1$ and $T_{-1}$ such that for $x, y \in \mathsf{X}$ the following "skewed" detailed balance holds:

$$\pi(\mathrm{d}x)T_1(x, \mathrm{d}y) = \pi(\mathrm{d}y)T_{-1}(y, \mathrm{d}x). \tag{5}$$

A generic example, related to the Metropolis–Hastings algorithm, is as follows.

EXAMPLE 5. Let $\{q_1(x, \cdot), x \in \mathsf{X}\}$ and $\{q_{-1}(x, \cdot), x \in \mathsf{X}\}$ be two families of probability distributions on $(\mathsf{X}, \mathscr{X})$, then the kernel defined for $v \in \{-1, 1\}$ and $x, y \in \mathsf{X}$ as, with $\nu(\mathrm{d}(x, y)) := \gamma_v(\mathrm{d}(x, y)) + \gamma_{-v}(\mathrm{d}(y, x))$ and $\gamma_v(\mathrm{d}(x, y)) := \pi(\mathrm{d}x)q_v(x, \mathrm{d}y)$,

$$T_v(x, \mathrm{d}y) = 1 \wedge r_v(x, y)q_v(x, \mathrm{d}y), r_v(x, y) := \begin{cases} \dfrac{\mathrm{d}\gamma_{-v}/\mathrm{d}\nu(y, x)}{\mathrm{d}\gamma_v/\mathrm{d}\nu(x, y)} & \text{if } \dfrac{\mathrm{d}\gamma_v}{\mathrm{d}\nu}(x, y) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

satisfies (5) (see Algorithm 3).

A standard way of constructing a $\pi$-reversible Markov transition based on the above sub-kernels consists of the following mixture:

$$P(x, \mathrm{d}y) = \frac{1}{2}T_1(x, \mathrm{d}y) + \frac{1}{2}T_{-1}(x, \mathrm{d}y) + \delta_x(\mathrm{d}y)\left(1 - \frac{1}{2}T_1(x, \mathsf{X}) - \frac{1}{2}T_{-1}(x, \mathsf{X})\right). \tag{6}$$

The standard Metropolis–Hastings algorithm corresponds to the scenario where $T_1 = T_{-1}$. The aim of the lifting strategy is to stratify the choice between $T_1$ and $T_{-1}$ by embedding the sampling problem into that of sampling from $\mu(\mathrm{d}(x, v)) = \pi(\mathrm{d}x)\varpi(v) = \frac{1}{2}\pi(\mathrm{d}x)\mathbb{I}\{v \in \{-1, 1\}\}$ and using a Markov kernel defined on the corresponding extended space $E = \mathsf{X} \times \{-1, 1\}$ which promotes contiguous uses of $T_1$ or $T_{-1}$ along the iterations. As shown

**Algorithm 3** A lifted MCMC algorithm (taken from Example 5)

- Initialisation $Z_0 := (X_0, V_0)$, $i = 0, n$.
- For $i = 1$ to $n$

  Draw $X' \sim q_{V_i}(X_i, \cdot)$ and $U \sim \mathcal{U}[0, 1]$

  (a) If $U \leq \min\{1, r_{V_i}(X_i, X')\}$ then set $X_{i+1} = X'$ and $V_{i+1} = V_i$,
  (b) Else if $U > \min\{1, r_{V_i}(X_i, X')\} + \rho_{V_i,-V_i}(X_i)$ then set $X_{i+1} = X_i$ and $V_{i+1} = V_i$,
  (c) Otherwise set $X_{i+1} = X_i$ and $V_{i+1} = -V_i$,

in Turitsyn, Chertkov and Vucelja (2011), Vucelja (2016), one possible solution, imposing $P^{\text{lifted}}((x,v);(A\setminus\{x\})\times\{-v\})=0$ for any $A\in\mathscr{X}$, is

$$P^{\text{lifted}}\big((x,v);\mathrm{d}(y,w)\big)=\mathbb{I}\{w=-v\}\delta_x(\mathrm{d}y)\rho_{v,-v}(x) + \mathbb{I}\{w=v\}\big[T_v(x,\mathrm{d}y)+\delta_x(\mathrm{d}y)\big(1-T_v(x,\mathsf{X})-\rho_{v,-v}(x)\big)\big],$$

where $\rho_{1,-1}(x)$ and $\rho_{-1,1}(x)$ are free parameters, the "switching rates", required to satisfy for all $(x,v)\in E$ $0\le\rho_{v,-v}(x)\le 1-T_v(x,\mathsf{X})$ and

$$\rho_{v,-v}(x)-\rho_{-v,v}(x)=T_{-v}(x,\mathsf{X})-T_v(x,\mathsf{X}). \tag{7}$$

It is not difficult to check that under (5) and (7) $P^{\text{lifted}}$ is $(\mu,Q)$-self-adjoint, for $Q$ such that $Qf(x,v)=f(x,-v)$ for $f\in\mathbb{R}^E$. There are numerous known solutions to the condition above (Hukushima and Sakai (2013)), including $\tilde{\rho}_{v,-v}(x):=\max\{0,T_{-v}(x,\mathsf{X})-T_v(x,\mathsf{X})\}$. It is remarked as intuitive in Vucelja (2016) that among the possible solutions to (7) this choice should promote fastest exploration. We prove below that this is indeed true, in the sense that this choice minimises asymptotic variances, as a consequence of Theorem 2. We let $P^{\text{lifted},\rho}$ denote the transition probability which uses $\rho_{v,-v}$.

THEOREM 6. *For any switching rate $\rho_{v,-v}$ satisfying $0\le\rho_{v,-v}(x)\le 1-T_v(x,\mathsf{X})$ for all $(x,v)\in E$ and (7), any $f\in L^2(\mu)$ such that $Qf=f$ and $\lambda\in[0,1)$, we have*

$$\operatorname{var}_\lambda\big(f,P^{\text{lifted},\tilde{\rho}}\big)\le\operatorname{var}_\lambda\big(f,P^{\text{lifted},\rho}\big)\le\operatorname{var}_\lambda\big(f,P^{\text{lifted},1-T_v}\big).$$

REMARK 9. Readers familiar with the delayed rejection Metropolis–Hastings update may notice the similarity here since $P^{\text{lifted}}((v,x);\mathrm{d}(w,y))$ is

$$\mathbb{I}\{w=v\}T_v(x,\mathsf{X})\frac{T_v(x,\mathrm{d}y)}{T_v(x,\mathsf{X})}+\big[1-T_v(x,\mathsf{X})\big]\bigg[\mathbb{I}\{w=v\}\delta_x(\mathrm{d}y)\bigg(1-\frac{\rho_{v,-v}(x)}{1-T_v(x,\mathsf{X})}\bigg) + \mathbb{I}\{w=-v\}\delta_x(\mathrm{d}y)\frac{\rho_{v,-v}(x)}{1-T_v(x,\mathsf{X})}\bigg],$$

where we require the property

$$\big[1-T_v(x,\mathsf{X})\big]\bigg(1-\frac{\rho_{v,-v}(x)}{1-T_v(x,\mathsf{X})}\bigg)=\big[1-T_{-v}(x,\mathsf{X})\big]\bigg(1-\frac{\rho_{-v,v}(x)}{1-T_{-v}(x,\mathsf{X})}\bigg),$$

and notice that $1-\frac{\rho_{v,-v}(x)}{1-T_v(x,\mathsf{X})}=\min\{1,\frac{1-T_{-v}(x,\mathsf{X})}{1-T_v(x,\mathsf{X})}\}$.

The theorem above establishes that this latter form of acceptance probability for the second stage of the update is again optimum in this setup. The update however differs from the standard delayed rejection update in that here the accept/rejection probability is integrated, restricting implementability of the approach. We also note that our results can be used to established superiority of the standard delayed rejection strategy in the context of $(\mu,Q)$-reversible updates and that integration of the rejection probability in the scenario above is beneficial.

One can compare the performance of algorithms relying on $P^{\text{lifted}}$ and $P$. With a slight abuse of notation for any $\lambda\in[0,1)$ and $f\in L^2(\pi)$, we let $\operatorname{var}_\lambda(f,P^{\text{lifted}})=\operatorname{var}_\lambda(\breve{f},P^{\text{lifted}})$ where for $(x,v)\in E$ we let $\breve{f}(x,v):=f(x)$.

THEOREM 7. *For any $\lambda\in[0,1)$ and $f\in L^2(\pi)$, any switching rate $\rho_{v,-v}$ satisfying $\tilde{\rho}_{v,-v}(x,v)\le\rho_{v,-v}(x,v)\le 1-T_v(x,\mathsf{X})$ for all $(x,v)\in E$, $\operatorname{var}_\lambda(f,P^{\text{lifted},\rho})\le\operatorname{var}_\lambda(f,P)$, with $P$ given in (6).*

EXAMPLE 6. In the scenario where $\mathsf{X}=\mathbb{R}$ or $\mathsf{X}=\mathbb{Z}$ and $\pi$ has a density with respect to the Lebesgue or counting measure, Gustafson (1998) introduced the guided walk Metropolis, of kernel $P^{\mathrm{GRW}}((v, x); \mathrm{d}(w, y))$:

$$T_v^{\text{guided}}(x, \mathrm{d}y)\mathbb{I}\{w = v\} + \delta_x(\mathrm{d}y)\mathbb{I}\{w = -v\}\bigl[1 - T_v^{\text{guided}}(x, \mathsf{X})\bigr]$$

$$T_v^{\text{guided}}(x, \mathrm{d}y) := \int_{\mathsf{X}} \min\left\{1, \frac{\pi(x + |z|v)}{\pi(x)}\right\} q(\mathrm{d}z)\delta_{x+|z|v}(\mathrm{d}y),$$

for some symmetric distribution $q(\cdot)$ on $\mathsf{V} = \mathbb{R}$ or $\mathsf{V} = \{-1, 1\}$. It is straightforward to check that $T_v^{\text{guided}}$ satisfies (5), and hence we can construct a lifted version of Gustafson's algorithm. We also notice that $P$ corresponds in this case to the random walk Metropolis algorithm with proposal distribution $q(\cdot)$—we denote this algorithm $P^{\mathrm{RW}}$. Our two earlier results establish that for any switching rate $\rho_{v,-v}$, $f \in L^2(\pi)$ and $\lambda \in [0, 1)$,

$$\mathrm{var}_\lambda\bigl(f, P^{\text{lifted-GRW},\rho}\bigr) \le \mathrm{var}_\lambda\bigl(f, P^{\mathrm{GRW}}\bigr) \le \mathrm{var}_\lambda\bigl(f, P^{\mathrm{RW}}\bigr).$$

3.4. *Neal's scheme to avoid backtracking.* In Neal (2004), the author describes a generic way of modifying a reversible Markov chain defined on a finite state space $\mathsf{X}$ to reduce "backtracking" (a special case is also discussed in Diaconis, Holmes and Neal (2000)). More specifically, assume we are interested in sampling from some probability distribution $\pi$ defined on $\mathsf{X}$ and that we do so by using a $\pi$-reversible (first-order) Markov transition $T_2$ defined on $\mathsf{X}$. Informally the idea in Neal (2004) is to modify the first-order Markov chain of transition $T_2$ into a second-order Markov chain to ensure that given a realisation $X_0, X_1, \dots, X_{k-1}, X_k$ for some $k \ge 1$ the new chain samples $X_{k+1}$ conditional upon $X_k$ and $X_{k-1}$ and prevents the occurrence of the event $X_{k+1} = X_{k-1}$. A probabilistic argument is developed in Neal (2004) for $\mathsf{X}$ finite to establish that the resulting chain produces estimators with an asymptotic variance that cannot exceed that of estimators from the original chain. We show here that this holds more generally for countable spaces and is a direct consequence of $(\mu, Q)$-self-adjointness for a particular $Q$, the bivariate first-order representation of a second-order univariate Markov chain as used in Neal (2004) and the application of Theorem 2. For simplicity of exposition, we assume $0 < T_2(x_1, x_2) < 1$ for $x_1, x_2 \in \mathsf{X}$, but the extension is straightforward. First, define the extended probability distribution on $\mathsf{X} \times \mathsf{X}$:

$$\mu(x_1, x_2) := \pi(x_1)T_2(x_1, x_2) = \pi(x_2)T_2(x_2, x_1),$$

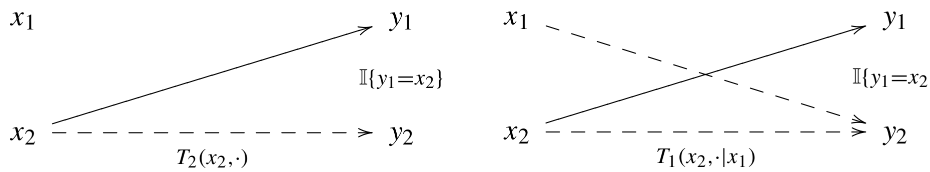
for $(x_1, x_2) \in \mathsf{X} \times \mathsf{X}$. Setting $Qf(x_1, x_2) := f(x_2, x_1)$, we notice that reversibility of $T_2$ implies that $Q$ is a $\mu$-isometric involution. Let for $(x_1, x_2), (y_1, y_2) \in \mathsf{X} \times \mathsf{X}$, $M_2((x_1, x_2); (y_1, y_2)) := \mathbb{I}\{y_1 = x_1\}T_2(x_1, y_2)$ and notice that $M_2$ is $\mu$-reversible. The Markov chain of transition $P_2 = QM_2$ is therefore $(\mu, Q)$-reversible from Proposition 1. Note that the first component of this process is simply a Markov chain of transition $T_2$. Following an idea of Liu (1996), it is suggested in Neal (2004) to use instead the transition $P_1 = QM_1$, where the $\mu$-reversible component $M_2$ is replaced with the $\mu$-reversible transition $M_1((x_1, x_2); (y_1, y_2)) := \mathbb{I}\{y_1 = x_1\}T_1(x_1, y_2 \mid x_2)$ where $T_1(x_1, y_2 \mid x_2)$ is the Metropolis–Hastings update:

$$T_1(x_1, y_2 \mid x_2) := U(x_1, y_2 \mid x_2) + \mathbb{I}\{y_2 = x_1\}\bigl(1 - U(x_1, \mathsf{X} \mid x_2)\bigr),$$

$$U(x_1, y_2 \mid x_2) := \frac{T_2(x_1, y_2)\mathbb{I}\{y_2 \ne x_2\}}{1 - T_2(x_1, x_2)} \min\left\{1, \frac{1 - T_2(x_1, x_2)}{1 - T_2(y_2, x_2)}\right\}.$$

The two resulting Markov transitions $P_1$ and $P_2$ are described graphically in Figure 1. The Markov transition $P_1 = QM_1$ is described algorithmically in Algorithm 4.

The $(\mu, Q)$-reversible kernel $P_1$ is designed so that backtracking, the probability of returning to $x_1$ when sampling $y_2$ conditional upon $x_2$, of the chain is reduced, compared to $P_2$. Let $\{Z_k, k \ge 0\}$ denote a realisation of the homogeneous Markov chain of transition $P_i$ (for $i \in \{1, 2\}$) and arbitrary initial condition, one can check that its first component is a realisation $\{X_k, k \ge 0\}$ of the Markov chain of transition $T_i$, and in fact $Z_k = (X_k, X_{k+1})$ for $k \ge 0$.




(a) $P_2$: $x_1$ is not used (b) $P_1$: $T_1(x_2, \cdot \mid x_1)$ ensures $y_2 \neq x_1$

FIG. 1. *Neal's no backtracking strategy: both algorithms keep track of $x_2$ for the next iteration by copying it in $y_1$.*

With an abuse of notation, for any $\lambda \in [0, 1)$ and $f \in L^2(\pi)$ we let $\mathrm{var}_\lambda(f, T_1) := \mathrm{var}_\lambda(\check{f}, P_1)$ where for any $x_1, x_2 \in \mathsf{X}$, $\check{f}(x_1, x_2) := f(x_1)$.

THEOREM 8. *For any $g \in L^2(\mu)$ such that $Qg = g$ and $\lambda \in [0, 1)$, we have* $\mathrm{var}_\lambda(g, P_1) \leq \mathrm{var}_\lambda(g, P_2)$ *and as a consequence, for any $f \in L^2(\pi)$,*

$$\mathrm{var}_\lambda(f, T_1) \leq \mathrm{var}_\lambda(f, T_2).$$

**4. Continuous-time scenario—general results.** The continuous-time scenario follows in part ideas similar to those developed in the discrete time scenario, but requires the introduction of the generator of the semigroup associated with the continuous-time process, leading to additional technical complications. In Section 4.1, we develop a crucial result of practical interest, Theorem 9, which allows one to deduce that a (in general intractable) semigroup is $(\mu, Q)$-self-adjoint when its generator is $(\mu, Q)$-symmetric on a type of dense subset of its domain. In Section 4.2, we establish the continuous-time counterpart of Theorem 2, that is, show that ordering of tractable quantities involving the generators of two $(\mu, Q)$-reversible processes implies an order on their asymptotic variances (Theorem 10). We remark that while establishing order rigorously may appear complex and technical, checking the criterion suggesting order involves in general elementary calculations. To the best of our knowledge, no general result is available in the continuous-time reversible setup, that is when $Q = \mathrm{Id}$ in our setup, but note the works Leisen and Mira (2008), Roberts and Rosenthal (2014), focused on particular scenarios. Continuous-time processes with such structure are discussed in Chapter 4 of Lelièvre, Rousset and Stoltz (2010), and asymptotic variances of nonreversible processes are also treated in Chapter 4 of Komorowski, Landim and Olla (2012).

4.1. *Set-up and characterisation of $(\mu, Q)$-self-adjointness.* Let $\{Z_t, t \geq 0\}$ be a Markov process taking values in the space $\mathsf{D}(\mathbb{R}_+, E)$ of càdlàg functions endowed with the Skorokhod topology and corresponding probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote $\{P_t, t \geq 0\}$ the associated semigroup, assumed to have an invariant distribution $\mu$ defined on $(E, \mathscr{E})$ and let $(\mathcal{D}^2(L, \mu), L)$ be the generator associated with $\{P_t, t \geq 0\}$, that is, $L$ and $\mathcal{D}^2(L, \mu) \subset L^2(\mu)$ are such that, with Id the identity operator,

$$\mathcal{D}^2 := \Big\{ f \in L^2(\mu) \colon \exists l_f \in L^2(\mu), \lim_{t \downarrow 0} \big\| t^{-1}(P_t - \mathrm{Id}) f - l_f \big\|_\mu = 0 \Big\}, \tag{8}$$

**Algorithm 4** Neal's scheme to avoid backtracking

- Initialisation $Z_0 := (Z_{0,1}, Z_{0,2})$, $i = 0$ and $n$
- For $i = 1$ to $n$

  Set $Z_{i+1,1} = Z_{i,2}$, draw $X \sim T_1(Z_{i,2}, \cdot \mid Z_{i,1})$ and $U \sim \mathcal{U}[0, 1]$,

  (a) If $U \leq \min\{1, \frac{1 - T_2(Z_{i,2}, Z_{i,1})}{1 - T_2(X, Z_{i,1})}\}$ then set $Z_{i+1,2} = X$,
  (b) Otherwise set $Z_{i+1,2} = Z_{i,2}$.

and for any $f \in \mathcal{D}^2$ one lets $Lf := l_f$, which can easily be shown to be a linear operator and denote $\mathcal{D}^2(L,\mu) := \mathcal{D}^2$. From above, $\{P_t, t \geq 0\}$ is a strongly continuous contraction, $\mathcal{D}^2(L,\mu)$ is dense in $L^2(\mu)$ and $L$ is closed (Ethier and Kurtz ((2009), Corollary 1.6)). For any $t \in \mathbb{R}_+$, we let $P_t^*$ denote the $L^2(\mu)$-adjoint of $P_t$, and it is classical that $\{P_t^*, t \geq 0\}$ is a strongly continuous contraction of invariant distribution $\mu$ and generator $(\mathcal{D}^2(L^*,\mu), L^*)$, the adjoint of $L$ (Pedersen (2012)), that is, it holds that for $f \in \mathcal{D}^2(L,\mu)$ and $g \in \mathcal{D}^2(L^*,\mu)$, $\langle Lf, g\rangle_\mu = \langle f, L^* g\rangle_\mu$.

In order to avoid repetition, we group our basic assumptions on the triplet $(\mu, Q, \{P_t, t \geq 0\})$ used throughout this section.

(A1) (a) $\mu$ is a probability distribution defined on $(E, \mathscr{E})$,
(b) $\{P_t, t \geq 0\}$ is a strongly continuous Markov semigroup of invariant distribution $\mu$,
(c) $Q$ is a $\mu$-isometric involution.

DEFINITION 4. We will say that the semigroup $\{P_t, t \geq 0\}$ is $(\mu, Q)$-self-adjoint, if for all $f, g \in L^2(\mu)$ and $t \geq 0$ $\langle P_t f, g\rangle_\mu = \langle f, QP_tQg\rangle_\mu$.

We aim to characterise the adjoint of the generator of a $(\mu, Q)$-self-adjoint semigroup $\{P_t, t \geq 0\}$ and provide a practical simple condition to establish this property for a given semigroup. We preface our first results with a technical lemma. For two operators $(\mathcal{D}^2(A,\mu), A)$ and $(\mathcal{D}^2(B,\mu), B)$, $\mathcal{D}^2(AB,\mu) := \{f \in \mathcal{D}^2(B,\mu) \colon Bf \in \mathcal{D}^2(A,\mu)\}$.

LEMMA 4. *Let* $(\mu, Q, \{P_t, t \geq 0\})$ *satisfying* (A1) *and let* $\{T_t := QP_tQ, t \geq 0\}$. *Then*:

(a) $(\mu, Q, \{T_t, t \geq 0\})$ *satisfies* (A1),
(b) *the generator of* $\{T_t, t \geq 0\}$ *is* $(\mathcal{D}^2(QLQ,\mu), QLQ)$.

As a corollary, one can characterise the generator of a $(\mu, Q)$-self-adjoint semigroup.

PROPOSITION 3. *Let* $(\mu, Q, \{P_t, t \geq 0\})$ *satisfying* (A1) *be* $(\mu, Q)$-*self-adjoint*. *Then the generator of* $\{P_t^*, t \geq 0\}$ *is* $(\mathcal{D}^2(QLQ,\mu), L^* = QLQ)$.

The following allows one to check $(\mu, Q)$-self-adjointness of a semigroup from the restriction of its generator to a particular type of dense subspace. A subspace $\mathcal{A} \subset \mathcal{D}^2(L,\mu)$ is said to be a core for $L$ if the closure of the restriction $L_{|\mathcal{A}}$ of $L$ to $\mathcal{A}$ is $L$, where the closure is to be taken with respect to $\|(f,g)\|_\mu := \|f\|_\mu + \|g\|_\mu$ for $f, g \in L^2(\mu)$ on the graph $\mathcal{G}(L) = \{(f, Lf) \colon f \in \mathcal{D}^2(L,\mu)\}$.

THEOREM 9. *Let* $(\mu, \{P_t, t \geq 0\}, Q)$ *satisfying* (A1). *Assume that* $\mathcal{A}$ *is a core for* $(L, \mathcal{D}^2(L,\mu))$ *such that*:

(a) $f \in \mathcal{A}$ *implies* $Qf \in \mathcal{A}$,
(b) *for all* $f, g \in \mathcal{A}$ *we have* $\langle Lf, g\rangle_\mu = \langle f, QLQg\rangle_\mu$,

*then* $\{P_t, t \geq 0\}$ *is* $(\mu, Q)$-*self-adjoint*.

4.2. *Ordering of asymptotic variances*. For $f \in L^2(\mu)$ and $Z_0 \sim \mu$, we are interested in the limit of $\mathrm{var}(f, L) := \lim_{t\to\infty} \mathrm{var}(t^{-1/2}\int_0^t f(Z_s)\,\mathrm{d}s)$ when this quantity exists. In some circumstances (for instance, when a Foster–Lyapunov function can be found (Glynn and Meyn ((1996), Theorem 4.3)), the limit above exists and has the expression $\mathrm{var}(f, L) = 2\langle \bar{f}, R\bar{f}\rangle_\mu$, where, again, $\bar{f} := f - \mu(f)$ and for $g \in L_0^2(\mu)$, $Rg := \int_0^{+\infty} P_t g\,\mathrm{d}t$. For $\lambda > 0$

and $f \in L^2(\mu)$, we introduce $\mathrm{var}_\lambda(f, L) := 2\langle \bar{f}, R_\lambda \bar{f}\rangle_\mu$, where $R_\lambda$ is the bounded operator defined for $g \in L^2(\mu)$ as $R_\lambda g := \int_0^{+\infty} \exp(-\lambda t) P_t g \, \mathrm{d}t$, referred to as the resolvent from now on. It is classical that for any $f \in L^2(\mu)$, $(\lambda \mathrm{Id} - L) R_\lambda f = f$ and for $f \in \mathcal{D}(L, \mu)$, $R_\lambda(\lambda \mathrm{Id} - L) f = f$. As in the discrete time setup, we leave the issue of checking whether $\lim_{\lambda \downarrow 0} \mathrm{var}_\lambda(f, L) = \mathrm{var}(f, L)$ as separate. We note the following straightforward result.

LEMMA 5. *If* $(\mu, Q, \{P_t, t \geq 0\})$ *satisfies* (A1) *and is* $(\mu, Q)$*-self-adjoint, then for any* $\lambda > 0$ *the bounded operator* $R_\lambda$ *is also* $(\mu, Q)$*-self-adjoint.*

For two semigroups $\{P_{t,1}, t \geq 0\}$ and $\{P_{t,2}, t \geq 0\}$ leaving $\mu$ invariant and of generators $L_1$ and $L_2$ with domains $\mathcal{D}^2(L_1, \mu)$ and $\mathcal{D}^2(L_2, \mu)$, we are interested in ordering $\mathrm{var}_\lambda(f, L_1)$ and $\mathrm{var}_\lambda(f, L_2)$ for $\lambda > 0$. As in the discrete time set-up the comparison relies on the Dirichlet forms, defined as follows for a generator $L$ and $f \in \mathcal{D}^2(L, \mu)$, $\mathcal{E}(f, L) := \langle f, -Lf\rangle_\mu$. Our proof requires the introduction of interpolating processes, defined at the level of their generators. The unusual parametrisation of convex combinations used here ensures that $L(1) = L_1$ and $L(2) = L_2$.

(A2) $(\mu, Q, \{P_{t,1}, t \geq 0\})$ and $(\mu, Q, \{P_{t,2}, t \geq 0\})$ satisfy (A1) and are $(\mu, Q)$-self-adjoint. Their respective generators $(L_1, \mathcal{D}^2(L_1, \mu))$ and $(L_2, \mathcal{D}^2(L_2, \mu))$ are assumed:
  (a) to have a common core $\mathcal{A}$ dense in $L^2(\mu)$ such that $Q\mathcal{A} \subset \mathcal{A}$,
  (b) to be such that for any $\beta \in [1, 2]$ the operator $((2-\beta)L_1 + (\beta - 1)L_2, \mathcal{D}^2(L_1, \mu) \cap \mathcal{D}^2(L_2, \mu))$
    (i) has an extension defining a unique continuous contraction semigroup $\{P_t(\beta), t \geq 0\}$ on $L^2(\mu)$ of invariant distribution $\mu$ and of (closed) generator $(L(\beta), \mathcal{D}^2(L(\beta), \mu))$,
    (ii) and for any $f \in \mathcal{A}$ we have $P_t(\beta) f \in \mathcal{A}$ for any $t \geq 0$.

From Ethier and Kurtz ((2009), Proposition 3.3), the last assumption and density of $\mathcal{A}$ in $L^2(\mu)$ imply that $\mathcal{A}$ is a core for $L(\beta)$, $\beta \in [1, 2]$. Establishing that for $\beta \in [1, 2]$ the contraction semigroup $\{P_t(\beta), t \geq 0\}$ exists may require one to resort to the Hille–Yosida theory and/or perturbation theory results (Ethier and Kurtz (2009), Voigt (1977)), but turns out to be straightforward in some scenarios such as those treated in Section 5. For $\lambda > 0$ and $\beta \in [1, 2]$, we let $R_\lambda(\beta)$ be the corresponding resolvent operators. Differentiability of $\beta \to R_\lambda(\beta) f$ and the expression for the corresponding derivative are key to our result, as is the case in the discrete time scenario. The right derivatives of operators below are to be understood as limits in the Banach space $L^2(\mu)$ equipped with the norm $\|\cdot\|_\mu$. We only state the results for $f \in L_2(\mu)$ such that $Qf = f$ and note that the case $Qf = -f$ is straightforward.

THEOREM 10. *Assume* (A2) *and that for any* $\lambda > 0$, $\beta \in [1, 2]$ *and* $f \in \mathcal{A}$,

(a) $R_\lambda(\beta) f \in \mathcal{D}^2(L_1, \mu) \cap \mathcal{D}^2(L_2, \mu)$ *and there exists* $\{g_n(\beta) \in \mathcal{A}, n \in \mathbb{N}\}$ *such that* $\lim_{n\to\infty}(L_1 - L_2) g_n(\beta) = (L_1 - L_2) R_\lambda(\beta) f$,

(b) $[1, 2) \ni \beta \mapsto R_\lambda(\beta) f$ *is right differentiable with*

$$\partial_\beta R_\lambda(\beta) f = R_\lambda(\beta)(L_2 - L_1) R_\lambda(\beta) f, \tag{9}$$

*and* $\beta \mapsto \langle f, \partial_\beta R_\lambda(\beta) f\rangle_\mu$ *is continuous,*

(c) *either* $\mathcal{E}(g, QL_1 - QL_2) \geq 0$ *for any* $g \in \mathcal{A}$ *or* $\mathcal{E}(g, L_1 Q - L_2 Q) \geq 0$ *for any* $g \in \mathcal{A}$,

*then*

(a) *for any* $f \in \mathcal{A}$ *satisfying* $Qf = f$ *and* $\beta \in [1, 2)$,

$$\begin{aligned}&\partial_\beta \langle f, R_\lambda(\beta) f\rangle_\mu \\ &\quad = \mathcal{E}(Q R_\lambda(\beta) f, L_1 Q - L_2 Q) = \mathcal{E}(R_\lambda(\beta) f, QL_1 - QL_2) \geq 0,\end{aligned}$$

(b) *for any* $f \in L_2(\mu)$ *such that* $Qf = f$,

$$\operatorname{var}_\lambda(f, L_1) = 2\langle f, R_\lambda(1) f\rangle_\mu \le \operatorname{var}_\lambda(f, L_2) = 2\langle f, R_\lambda(2) f\rangle_\mu.$$

The following allows us to check the conditions of the theorem above.

LEMMA 6. *Assume* (A2) *and that for any* $\lambda > 0$, $\beta \in [1, 2]$ *and* $f \in \mathcal{A}$,

(a) $t \mapsto (L_2 - L_1)P_t(\beta)f$ *and* $t \mapsto (L_2 - L_1)QP_t(\beta)f$ *are continuous,*
(b) *there exists* $\delta(\beta) > 0$ *such that* $\int_0^\infty \exp(-\lambda t)\|(L_2 - L_1)P_t(\beta)f\|_\mu \,\mathrm{d}t < \infty$ *and* $\sup_{|\beta'-\beta|\le\delta(\beta)} \int_0^\infty \exp(-\lambda t)\|(L_2 - L_1)QP_t(\beta')f\|_\mu \,\mathrm{d}t < \infty$.

*Then for any* $\beta \in [1, 2)$ *and* $\lambda > 0$, *for any* $f \in \mathcal{A}$,

(a) $R_\lambda(\beta)f \in \mathcal{D}^2(L_1, \mu) \cap \mathcal{D}^2(L_2, \mu)$ *and there exists* $\{g_n(\beta) \in \mathcal{A}, n \in \mathbb{N}\}$ *such that* $\lim_{n\to\infty}(L_1 - L_2)g_n(\beta) = (L_1 - L_2)R_\lambda(\beta)f$,
(b) $[1, 2) \ni \beta \mapsto R_\lambda(\beta)f$ *is right differentiable of derivative given by* (9) *and* $\beta \mapsto \langle f, \partial_\beta R_\lambda(\beta)f\rangle_\mu$ *is continuous.*

**5. Continuous-time scenario—example.** In this section, we show how the results of the previous section can be applied to a particular class of processes designed to perform Monte Carlo simulation, which has recently received some attention (Section 5.1). In Section 5.2, we establish that most processes considered in the literature are indeed $(\mu, Q)$-self-adjoint; this includes, in particular, the Zig-Zag (ZZ) process. In Section 5.3, we show that with some smoothness conditions on the intensities involved in the definition of the ZZ process, then all the conditions required to apply our general results, namely Theorem 10 and Lemma 6, are satisfied. In Section 5.1, we apply our general theory and present some applications. In addition, we show how one can consider more general versions of ZZ relying on nonsmooth intensities using smooth approximation strategies which have the advantage of preserving the correct invariant distribution.

5.1. *PDMP-Monte Carlo.* We assume here that $E = \mathsf{X} \times \mathsf{V}$ and that the distribution $\mu$ of interest has density (also denoted $\mu$),

$$\mu(x, v) \propto \exp\bigl(-U(x)\bigr)\varpi(v) \tag{10}$$

with respect to some $\sigma$-finite measure denoted $\mathrm{d}(x, v)$, where $U : \mathsf{X} = \mathbb{R}^d \to \mathbb{R}$ is an energy function and $\varpi : \mathsf{V} \subset \mathbb{R}^d \to \mathbb{R}_+$ are such that $\mu$ induces a probability distribution. Piecewise deterministic Markov processes (PDMPs) (Davis (1993)) are continuous-time processes with various applications in engineering and science, but it has recently been shown (Bierkens, Fearnhead and Roberts (2019), Bierkens and Roberts (2017), Bou-Rabee and Sanz-Serna (2017), Bouchard-Côté, Vollmer and Doucet (2018), Faggionato, Gabrielli and Ribezzi Crivellari (2009), Peters and de With (2012)) that such processes can be used in order to sample from large classes of distributions defined as above. The particular cases derived for this purpose are known to be nonreversible, but we establish here that they are in fact $(\mu, Q)$-reversible for a specific isometric involution $Q$. This allows us to apply the theory developed in the previous section and to compare their performance in terms of some of their design parameters.

For $k \in \mathbb{Z}_+$, for $i \in [\![1, k]\!]$ define intensities $\lambda_i : E \to \mathbb{R}_+$, $\lambda := \sum_{i=1}^k \lambda_i$, for $(x, v) \in E$ and $t \ge 0$ $\Lambda_i(t, x, v) := \int_0^t \lambda_i(x + uv, v)\,\mathrm{d}u$, $\Lambda(t, x, v) := \sum_{i=1}^k \Lambda_i(t, x, v)$ and kernels $R_i : E \times \mathscr{E} \to [0, 1]$ such that for any $(x, v) \in E$, $R_i((x, v), \{x\} \times \mathsf{V}) = 1$. For any $x \in \mathsf{X}$ and $i \in [\![1, k]\!]$, we let $R_{x,i} : \mathsf{V} \times \mathscr{V} \to [0, 1]$ be such that $R_{x,i}(v, A) := R_i((x, v), \{x\} \times A)$ for $(v, A) \in \mathsf{V} \times \mathscr{V}$. For $\varsigma_1, \dots, \varsigma_k \in \mathbb{R}_+$, we let $\mathcal{P}(\varsigma_1, \dots, \varsigma_k)$ denote the probability distribution

**Algorithm 5** A piecewise deterministic Markov process to sample from $\mu$

- Initialisation $Z_0 = (X_0, v_0)$ , $T_0 = 0$ and $l = 0$.
- Repeat, $l \leftarrow l + 1$

  (a) Draw $T_l$ such that $\mathbb{P}(T_l \geq \tau \mid T_{l-1}) = \exp(-\Lambda(\tau - T_{l-1}, X_{T_{l-1}}, V_{T_{l-1}}))$,
  (b) $(X_t, V_t) = (X_{T_{l-1}} + (t - T_{l-1})V_{T_{l-1}}, V_{T_{l-1}})$ for $t \in [T_{l-1}, T_l)$,
  (c) $X_{T_l} = \lim_{t \uparrow T_l} X_t$ and with $M \sim \mathcal{P}(\lambda_1(Z_{T_l}), \dots, \lambda_d(Z_{T_l}))$ set $V_{T_l} \sim R_{X_{T_l}, M}(V_{T_{l-1}}, \cdot)$.

of the random variable $M$ such that $\mathbb{P}(M = m) \propto \varsigma_m$. The PDMPs of interest here can be described algorithmically as in Algorithm 5.

Davis (1993) (see also Durmus, Guillin and Monmarché (2018) for an alternative construction) shows that this defines a process, of corresponding semigroup $\{P_t, t \in \mathbb{R}_+\}$, as soon as the following standard two conditions on the intensity are satisfied (Davis ((1993), p. 62)):

(A3) For $i \in [\![1, k]\!]$,

  (a) $\lambda_i$ is measurable and $t \mapsto \lambda_i(x + tv, v)$ is integrable for all $(x, v) \in E$,
  (b) for any $t > 0$ and $(x, v) \in E$, $\mathbb{E}_{x,v}(\sum_{i=1}^{\infty} \mathbb{I}\{T_i \leq t\}) < \infty$.

Define for any $(x, v) \in E$ and $f \in \mathbb{R}^E$, whenever the limit exists,

$$Df(x, v) := \lim_{h \to 0} \frac{f(x + hv, v) - f(x, v)}{h},$$

then the extended generator of the process above, which solves the Martingale problem, is of the form

$$Lf := Df + \sum_{i=1}^{k} \lambda_i \cdot [R_i f - f], \tag{11}$$

for $f \in \mathcal{D}(L)$, a domain fully characterised by Davis ((1993), Theorem 26.14, p. 69 and Remark 26.16). Let $\mathbf{M}(E) \subset \mathbb{R}^E$ be the set of measurable functions and $\mathbf{B}(E) \subset \mathbf{M}(E)$ be the set of bounded measurable functions. It can be shown that $\{P_t, t \geq 0\}$ is a contraction semigroup on $\mathbf{B}(E)$ equipped with the $\|\cdot\|_\infty$ norm. Further with $\mathbf{B}_0(E) := \{f \in \mathbf{B}(E) \colon \lim_{t \downarrow 0} \|P_t f - f\|_\infty = 0\}$, one can show that $\{P_t, t \geq 0\}$ is a strongly continuous contraction semigroup on $\mathbf{B}_0(E)$ (Davis ((1993), pp. 28–29)) of strong generator $(\mathcal{D}_\infty(L_\infty), L_\infty)$, with $\mathcal{D}_\infty(L_\infty) \subset \mathcal{D}(L)$ and for any $f \in \mathcal{D}_\infty(L_\infty)$, $L_\infty f = Lf$. When $\mathsf{V} = \mathbb{R}^d$ (or such that $E$ is a Riemannian sub-manifold), we define $\mathbf{C}(E) := \mathbf{C}^0(E) := \{f \in \mathbb{R}^E \colon f \text{ is continuous}\}$ and $\mathbf{C}^1(E) := \{f \in \mathbb{R}^E \colon f \text{ is continuously differentiable}\}$, let $\mathbf{C}_c(E)$ and $\mathbf{C}_c^1(E)$ be their restrictions to compactly supported functions and $\mathbf{C}_0(E) \subset \mathbf{C}(E)$ the set containing functions vanishing at infinity. When $\mathsf{V}$ is finite, we let with $\mathbf{C}^0(\mathsf{X}) := \{f \in \mathbb{R}^{\mathsf{X}} \colon f \text{ is continuous}\}$ and for $i \in \mathbb{N}_+$ $\mathbf{C}^i(\mathsf{X}) := \{f \in \mathbb{R}^{\mathsf{X}} \colon f \text{ is } i \text{ times continuously differentiable}\}$, $\mathbf{C}^i(E) := \{f \in \mathbb{R}^E \colon \text{for any } v \in \mathsf{V}, x \mapsto f(x, v) \in \mathbf{C}^i(\mathsf{X})\}$, use the simplified notation $\mathbf{C}(E) := \mathbf{C}^0(E)$, and let $\mathbf{C}_c(E)$, $\mathbf{C}_c^1(E)$ be the corresponding restrictions to functions $x \mapsto f(x, v)$ of compact support for any $v \in \mathsf{V}$. We let $\mathbf{C}_0(E)$ be the set of $f \in \mathbf{C}(E)$ such that for any $\epsilon > 0$ there exists $M \in \mathbb{R}_+$ such that $|f(x, v)| \leq \epsilon$ for $(x, v) \in B^c(0, M) \times \mathsf{V}$ where $B(0, M) = \{x \in \mathsf{X} \colon \|x\| \leq M\}$ and $\|\cdot\|$ is the Euclidian norm.

5.2. *$(\mu, Q)$-Symmetry of some PDMP-Monte Carlo processes.* From now on, $Qf(x, v) = f(x, -v)$ for $f \in \mathbb{R}^E$ and $(x, v) \in E$. In the following, we establish simple conditions implying that $L$ is $(\mu, Q)$-symmetric on $\mathbf{C}_c^1(E)$, which cover most known scenarios. Hereafter, we will need the following assumption on the potential $U$:

(A4) $U \colon \mathsf{X} \to \mathbb{R}$ is $\mathbf{C}^2(\mathsf{X})$ and $\int [1 + \|\nabla U(x)\|] \exp(-U(x))\, \mathrm{d}x < \infty$.

The following was shown in Faggionato, Gabrielli and Ribezzi Crivellari ((2009), Proposition 3.2) for example.

LEMMA 7. *Assume* (A4). *Then for* $f, g \in \mathbf{C}_c^1(E)$,

$$\langle Df, g\rangle_\mu = \langle f, -Dg + DU \cdot g\rangle_\mu \quad \text{and} \quad -Df = QDQf.$$

The following establishes that a simple property on the family of operators $\{R_i, i \in [\![1,k]\!]\}$ ensures $(\mu, Q)$-symmetry of $L$, and hence invariance of $\mu$ if $\mathbf{C}_c^1(E)$ is a core.

THEOREM 11. *Let* $\mu$ *be a probability distribution defined on* $(E, \mathscr{E})$ *and consider the semigroup* $\{P_t, t \geq 0\}$ *with extended generator* $L$ *given in* (11). *Assume* (A4), $\lambda - Q\lambda = DU$ *and that for any* $i \in [\![1,k]\!]$ *the operator* $(\lambda_i \cdot R_i)$ *is* $(\mu, Q)$*-symmetric on* $\mathbf{C}_c^1(E)$. *Then* $L$ *is* $(\mu, Q)$*-symmetric on* $\mathbf{C}_c^1(E)$.

REMARK 10. With an abuse of notation, for $f \in \mathbb{R}^{\mathsf{V}}$ let $Qf := Q\breve{f}$ where for $(x, v) \in E$ $\breve{f}(x, v) = f(v)$. For $f \in \mathbb{R}^E$ and for $x \in \mathsf{X}$, denote $f_x(\cdot) = f(x, \cdot)\colon \mathsf{V} \to \mathbb{R}$. Let $i \in [\![1,k]\!]$. If for any $x \in \mathsf{X}$, $\lambda_{x,i} \cdot R_{x,i}$ is $(\varpi(\cdot), Q)$-symmetric on $\mathbf{B}_c(\mathsf{V})$, then for $f, g \in \mathbf{C}_c^1(E)$,

$$\begin{aligned}\langle(\lambda_i \cdot R_i)f, g\rangle_\mu &= \int \langle(\lambda_{i,x} \cdot R_{i,x})f_x, g_x\rangle_\varpi \pi(\mathrm{d}x)\\ &= \int \langle f_x, Q(\lambda_{i,x} \cdot R_{i,x})Qg_x\rangle_\varpi \pi(\mathrm{d}x) = \langle f, Q\lambda_i \cdot QR_iQg\rangle_\mu,\end{aligned}$$

that is, $(\lambda_i \cdot R_i)$ is $(\mu, Q)$-symmetric on $\mathbf{C}_c^1(E)$.

The most popular PDMP-MC processes satisfy the properties of Theorem 11 and are covered by the following examples. For notational simplicity, we may drop the index $i$ below.

EXAMPLE 7. Let $x \mapsto n(x)$ be a unit vector field and assume that for any $(x, v) \in E$ we have $v - 2\langle n(x), v\rangle n(x) \in \mathsf{V}$. Consider the operator such that for any $f \in \mathbb{R}^E$ and $(x, v) \in E$, $Rf(x, v) := f(x, v - 2\langle n(x), v\rangle n(x))$ and assume that the property $R\lambda = Q\lambda$ holds. Note that $R^2 = \mathrm{Id}$ and that for any $f \in \mathbb{R}^E$ and $(x, v) \in E$, $RQf(x, v) = f(x, -v + 2\langle n(x), v\rangle n(x))$, and hence $QRQf = Rf$. Therefore, for any $f, g \in \mathbf{C}_c^1(E)$,

$$\langle(\lambda \cdot R)f, g\rangle_\mu = \langle Rf, \lambda \cdot g\rangle_\mu = \langle f, R\lambda \cdot Rg\rangle_\mu = \langle f, Q\lambda \cdot QRQg\rangle_\mu.$$

Now let $\{n_i\colon \mathsf{X} \to \mathbb{R}^d, i \in [\![1,k]\!]\}$ be unitary vector fields and $\{a_i\colon \mathsf{X} \to \mathbb{R}, i \in [\![1,k]\!]\}$ such that $\nabla U = \sum_{i=1}^k a_i n_i$. Assume that for $i \in [\![1,k]\!]$ the intensities are of the form $\lambda_i(x, v) = \varphi(a_i(x)\langle n_i(x), v\rangle)$ for $\varphi\colon \mathbb{R} \to \mathbb{R}_+$ such that $\varphi(s) - \varphi(-s) = s$ and $R_i f(x, v) := f(x, v - 2\langle n_i(x), v\rangle n_i(x))$ for $f \in \mathbb{R}^E$ and $(x, v) \in E$. Possible choices of $\varphi$ are discussed later on and include $\varphi(s) = \max\{0, s\}$. Then for $i \in [\![1,k]\!]$, $R_i\lambda_i = Q\lambda_i$, $(\lambda_i \cdot R_i)$ is $(\mu, Q)$-symmetric on $\mathbf{C}_c^1(E)$ and $\lambda - Q\lambda = DU$. Therefore, Theorem 11 holds. This covers the Zig-Zag and Bouncy Particle Sampler processes, for example, Andrieu et al. (2018).

EXAMPLE 8. The choice $Rf(x, v) = \int f(x, w)\varpi(\mathrm{d}w)$ for $(x, v) \in E$ and $f \in L^2(\mu)$, the "refreshment" operator, is such that for any $f, g \in L^2(\mu)$, $\langle Rf, g\rangle_\mu = \langle f, Rg\rangle_\mu$, $RQf = Rf$ and $QRf = Rf$ since for any $x \in \mathsf{X}$, $v \mapsto Rf(x, v)$ is constant. If for any $x \in \mathsf{X}$ the mapping $v \mapsto \bar{\lambda}(x, v)$ is constant (implying $\bar{\lambda} - Q\bar{\lambda} = 0$), we deduce that for any $f, g \in \mathbf{C}_c^1(E)$,

$$\langle(\bar{\lambda} \cdot R)f, g\rangle_\mu = \langle Rf, \bar{\lambda} \cdot g\rangle_\mu = \langle f, \bar{\lambda} \cdot Rg\rangle_\mu = \langle f, Q\bar{\lambda} \cdot QRQg\rangle_\mu,$$

that is, $(\bar{\lambda} \cdot R)$ is $(\mu, Q)$-symmetric on $\mathbf{C}_c^1(E)$. In fact, from the proof of Lemma 2 we note that $R$ can be taken to be Horowitz's refreshment operator.

EXAMPLE 9. The choice $R_x f_x(v) \propto \int f_x(w) Q\lambda(x,w)\varpi(\mathrm{d}w)$, with $R_x\mathbf{1}(v) = 1$ when possible, for any $(x,v) \in E$ and $f \in \mathbf{C}_c^1(E)$ has been suggested in Fearnhead et al. (2018). It is such that for $f, g \in \mathbf{C}_c^1(E)$ and $x \in \mathsf{X}$,

$$\int f_x(v)g_x(w)\lambda(x,v)Q\lambda(x,w)\varpi(\mathrm{d}w)\varpi(\mathrm{d}v)$$
$$= \int Qf_x(v)g_x(w)Q\lambda(x,v)Q\lambda(x,w)\varpi(\mathrm{d}w)\varpi(\mathrm{d}v),$$

and we conclude that $(\lambda_x \cdot R_x)$ is $(\varpi, Q)$-self-adjoint.

REMARK 11. We note that Theorem 11 holds more generally when the operator $D$ is replaced with the generator $D_F$ of a dynamic with time-reversal symmetry (Lamb and Roberts (1998)) for which $\langle D_F f, g\rangle_\mu = \langle f, QD_FQg + D_FU \cdot g\rangle_\mu$, which is the case for the Liouville operator for an arbitrary potential $H(x,v)$, and the condition on the total intensity rate adjusted accordingly. We do not pursue this here for brevity.

5.3. *Zig-Zag*: *Generator and semigroup properties*. Zig-Zag (ZZ) is a particular continuous-time Markov process designed to sample from $\mu$ and described in Algorithm 5. The name was coined in Bierkens and Roberts (2017) and further extended in Bierkens, Fearnhead and Roberts (2019), and can be interpreted as being a particular case of the process studied in Faggionato, Gabrielli and Ribezzi Crivellari (2009). In this scenario, $k = d+1$, $\mathsf{V} := \{-1, 1\}^d$, $\varpi$ is the uniform distribution and, with $\{\mathbf{e}_i \in \mathbb{R}^d, i \in [\![1,d]\!]\}$ the canonical basis of $\mathbb{R}^d$, for $i \in [\![1,d]\!]$ and $(x,v) \in E$ we let $R_i f(x,v) := f(x, v - 2v_i\mathbf{e}_i)$ where $v_i := \langle v, \mathbf{e}_i\rangle$. Note that this corresponds to $n_i(x) = \mathbf{e}_i$ in Example 7. For $i = d+1$, we let $\lambda_{d+1}(x,v) = \bar{\lambda}$ for $\bar{\lambda} \in \mathbb{R}_+$ and $R_{d+1}$ is as in Example 8. We require the following assumptions on the intensities:

(A5) For any $i \in [\![1,d]\!]$ and $(x,v) \in E$, we have:

- (a) $\lambda_i \in \mathbf{C}^1(E)$ and $\lambda_i > 0$,
- (b) $\lambda_i(x,v) - Q\lambda_i(x,v) = \partial_i U(x)v_i$,
- (c) $R_i\lambda_i(x,v) = Q\lambda_i(x,v)$.

The following establishes the existence of such intensities.

PROPOSITION 4. *Assume* (A4). *Let* $\phi : \mathbb{R} \to [0,1]$ *be such that* $r\phi(r^{-1}) = \phi(r)$ *for* $r \geq 0$ *and define for any* $(x,v) \in E$ *and* $i \in [\![1,d]\!]$, $\lambda_i^\phi(x,v) := -\log(\phi(\exp(\partial_i U(x)v_i))) \geq 0$. *If further* $\phi < 1$ *and* $\phi \in \mathbf{C}^1(\mathbb{R})$, *then* $\{\lambda_i, i \in [\![1,d]\!]\}$ *satisfies* (A5).

COROLLARY 5. *The choice* $\phi(r) = r/(1+r)$ *satisfies the assumptions of Proposition* 4, *but this is not the case for the canonical choice* $\phi(r) = \min\{1, r\}$.

We now establish properties required of $\{P_t, t \geq 0\}$ and its generator in order to check (A2) and apply Theorem 10 and Lemma 6.

PROPOSITION 5. *Let* $L$ *be the extended generator of the ZZ process and assume* (A4)–(A5). *Then* $L$ *is* $(\mu, Q)$-*symmetric on* $\mathbf{C}_c^1(E)$.

For $\mathbf{A}, \mathbf{B} \subset \mathbf{M}(E)$ and $t > 0$, we let $P_t\mathbf{A} \subset \mathbf{B}$ mean that for any $f \in \mathbf{A}$ such that $P_t f$ exists, then $P_t f \in \mathbf{B}$. A semigroup $\{P_t, t \geq 0\}$ is said to be Feller if $\mathbf{C}_0(\mathsf{E}) \subset \mathbf{B}_0(\mathsf{E})$ and $\{P_t, t \geq 0\}$ is a strongly continuous contraction semigroup on $\mathbf{C}_0(\mathsf{E})$ equipped with $\|\cdot\|_\infty$, that is, for any $s, t > 0$, $P_{s+t} = P_sP_t$, $\|P_t f\|_\infty \leq \|f\|_\infty$, $P_t\mathbf{C}_0(\mathsf{E}) \subset \mathbf{C}_0(\mathsf{E})$ and for any $f \in \mathbf{C}_0(\mathsf{E})$, $\lim_{t\downarrow 0}\|P_t f - f\|_\infty = 0$.

THEOREM 12. *Consider a ZZ process of intensities* $\{\lambda_i, i \in [\![1, d+1]\!]\}$ *satisfying* (A5). *Then*:

(a) $\{P_t, t \geq 0\}$ *is Feller*,
(b) *for any* $t > 0$, $P_t \mathbf{C}_c^1(E) \subset \mathbf{C}_c^1(E)$,
(c) $\mathbf{C}_c^1(E)$ *is a core for the strong generator of the semigroup* $\{P_t, t \geq 0\}$,
(d) $\mu$ *is invariant for* $\{P_t, t \geq 0\}$,
(e) $\{P_t, t \geq 0\}$ *can be extended to a strongly continuous semigroup on* $L^2(\mu)$ *equipped with* $\|\cdot\|_\mu$,
(f) $\mathbf{C}_c^1(E)$ *is a core for the strong generator of the extended semigroup on* $L^2(\mu)$.

5.4. *Zig-Zag—main results and some examples.* The main result of this section is the ordering of Theorem 13, which we illustrate with two examples.

THEOREM 13. *Assume* (A4) *and consider two ZZ processes of intensities* $\{\lambda_{1,i}, i \in [\![1, d+1]\!]\}$ *and* $\{\lambda_{2,i}, i \in [\![1, d+1]\!]\}$ *satisfying* (A5) *such that for* $i \in [\![1, d+1]\!]$, $\|\lambda_{1,i} - \lambda_{2,i}\|_\mu < \infty$. *Then if for all* $g \in \mathbf{C}_c^1(E)$,

$$\begin{aligned}\langle g, -(L_1 - L_2)Qg\rangle_\mu = &\sum_{i=1}^{d+1}\langle g, (\lambda_{1,i} - \lambda_{2,i}) \cdot [\mathrm{Id} - R_i Q]g\rangle_\mu \\ &- \langle g, (\lambda_1 - \lambda_2) \cdot [\mathrm{Id} - Q]g\rangle_\mu \geq 0,\end{aligned} \tag{12}$$

*then* $\mathrm{var}_\lambda(L_1, f) \leq \mathrm{var}_\lambda(L_2, f)$ *for* $\lambda > 0$ *and* $f \in L^2(\mu)$ *such that* $Qf = f$.

REMARK 12. The assumption on the intensity is satisfied as soon as for some $c, C > 0$, for any $i \in [\![1, d+1]\!]$ and $(x, v) \in E$, $\lambda_{1,i}(x, v) \leq c + C\|\nabla U(x)\|$ and $\int \|\nabla U\|^2 \,\mathrm{d}\pi < \infty$. As we shall see, this can be checked for various examples.

EXAMPLE 10. When $d = 1$, $R_1 = Q$ and, therefore, $R_1 Q = \mathrm{Id}$. If we further assume that $\lambda_{1,2} = \lambda_{2,2} = \bar{\lambda}$ then for all $g \in \mathbf{C}_c^1(E)$ $\langle g, -(L_1 - L_2)Qg\rangle_\mu = -\langle g, (\lambda_{1,1} - \lambda_{2,2}) \cdot [\mathrm{Id} - Q]g\rangle_\mu \geq 0$, whenever $\lambda_{2,1} \geq \lambda_{1,1}$, a result similar to that of Bierkens and Duncan (2017).

The situation where the total event rate is constant, that is $\lambda_1 = \lambda_2$ in the expression above, but distributed differently between updates of the velocity leads to the following.

EXAMPLE 11. Let $d = 2$ and for $g \in \mathbf{C}_c^1(E)$ let $Lg = Dg + \sum_{i=1}^{2} \lambda_i \cdot [R_i - \mathrm{Id}]g$ and consider the ZZ processes of generators, for $\mathbf{C}^1(\mathsf{X}) \ni \bar{\gamma} : \mathsf{X} \to \mathbb{R}_+$, $L_1 g = Lg + \bar{\gamma} \cdot [\mathit{\Pi} - \mathrm{Id}]g$ and $L_2 g = Lg + \bar{\gamma}/2 \cdot \sum_{i=1}^{2}[R_i - \mathrm{Id}]g$. Then for $g \in \mathbf{C}_c^1(E)$, $\langle g, -(L_1 - L_2)Qg\rangle_\mu = \langle g, \bar{\gamma}/2 \cdot \sum_{i=1}^{2}[\mathrm{Id} - R_i]g\rangle_\mu - \langle g, \bar{\gamma} \cdot [\mathrm{Id} - \mathit{\Pi}]g\rangle_\mu \geq 0$, where the equality follows from $R_1 Q = R_2$, $R_2 Q = R_1$, Lemma 9 and O'Donnell ((2014), p. 52). We therefore conclude that in this setup partial refreshment of the velocity is superior to full refreshment in terms of asymptotic variance.

We note that checking (12) involves the difference of two nonnegative terms (from Lemma 9) and may be challenging to establish for this class of processes. For example, we have not been able to extend the result of Example 10 to the situation where $d \geq 2$, yet. We have not explored comparisons involving other updates $R_i$, which would require establishing Theorem 12 for this setup, and rather focus on the following issue. Intensities of interest may not satisfy (A5) and we may not be able to apply Theorem 12. This is the case for the so-called canonical choice $\lambda(x, v) = (\partial U(x)v)_+$, which may however be of interest as suggested by the following. In the following discussion, we assume $d = 1$ for presentational simplicity, but the approach is valid for $d \geq 1$.

PROPOSITION 6. *Let* $\lambda: E \to \mathbb{R}_+$ *be an intensity satisfying* $\lambda(x,v) - Q\lambda(x,v) = \partial U(x)v$, *then* $\lambda(x,v) \geq (\partial U(x)v)_+$.

A natural question is whether we can establish that the choice $\lambda^0(x,v) := (\partial U(x)v)_+$ is optimum in terms of asymptotic variance. Our argument relies on the existence of regularising intensities satisfying the following properties:

(A6) The intensities $\{\lambda^\epsilon, \epsilon \geq 0\}$ satisfy for any $(x,v) \in E$ and $\epsilon > 0$,

- (a) $\lambda^\epsilon \in \mathbf{C}^1(E)$ and $\lambda^\epsilon > 0$,
- (b) $\epsilon \mapsto \lambda^\epsilon(x,v)$ is nonincreasing,
- (c) $\lambda^\epsilon(x,v) - Q\lambda^\epsilon(x,v) = \partial U(x)v$,
- (d) $\lim_{\epsilon \downarrow 0} \sup_{(x,v)\in E} |\lambda^\epsilon(x,v) - \lambda^0(x,v)| = 0$.

Intensities satisfying these properties exist.

PROPOSITION 7. *Assume* (A4) *and for any* $\epsilon > 0$ *define the intensities such that for* $(x,v) \in E$, $\lambda^\epsilon(x,v) := -\log(\phi_\epsilon(\exp(\partial U(x)v)))$, *where for* $r > 0$ $\phi_\epsilon(r) := r[1 - \Phi(\epsilon/2 + \log(r)/\epsilon)] + [1 - \Phi(\epsilon/2 - \log(r)/\epsilon)] > 0$, *with* $\Phi(\cdot)$ *the cumulative distribution function of the* $\mathcal{N}(0,1)$. *Then* $\{\lambda^\epsilon, \epsilon > 0\}$ *satisfies* (*A*6).

THEOREM 14. *Let* $d = 1$, *assume* (A4) *and* $\int \|\nabla U\|^2 \,\mathrm{d}\pi < \infty$, *and consider two ZZ processes of common invariant distribution and of intensities* $\lambda_1$ *and* $\lambda_2$ *where* $\lambda_1(x,v) := (\partial U(x) \cdot v)_+$ *and* $\lambda_2(x,v) := \lambda_1(x,v) + \gamma(x,v)$ *with* $0 \leq \gamma \leq c + C\|\nabla U\|$ *for* $c, C > 0$, $\gamma \in \mathbf{C}^1(E)$ *and such that* $\gamma - Q\gamma = 0$. *Then for any* $f \in L^2(\mu)$ *such that* $Qf = f$ *and* $\lambda \in [0,1)$ $\mathrm{var}_\lambda(f, L_1) \leq \mathrm{var}_\lambda(f, L_2)$.

One can consider more general forms for $\lambda_2$ in the theorem above. For example, the result will hold when $\gamma$ can be uniformly approximated by a sequence $\{\gamma^\epsilon \in \mathbf{C}^1(E), \epsilon > 0\}$ such that $\gamma^\epsilon \geq 0$ for $\epsilon_0 > \epsilon > 0$ for some $\epsilon_0 > 0$. Another possibility is to consider generalisations of the ideas of Proposition 7: for example, with $\tilde{\phi}(r) = r/(1+r)$ instead of $\phi(r) = \min\{1, r\}$ as a starting point in Proposition 8 one can analogously define a family of acceptance ratios which is automatically such that $\tilde{\phi}_\epsilon(r) \leq \phi_\epsilon(r)$ for $\epsilon > 0$ and $r \geq 0$, define the corresponding intensities, and then proceed as above to compare the processes with intensities derived from $\tilde{\phi}(\cdot)$ and $\phi(\cdot)$.

**Acknowledgments.** The authors would like to thank Florian Maire for pointing out an error in Remark 5 and suggesting the correct formulation of the result. The authors are grateful to Anthony Lee for sharing his LaTeX code to handle Supplementary Material efficiently.

**Funding.** The authors acknowledge support from EPSRC "Intractable Likelihood: New Challenges from Modern Applications (ILike)", (EP/K014463/1). CA acknowledges support from EPSRC "Computational Statistical Inference for Engineering and Security (CoSInES)", (EP/R034710/1).

## SUPPLEMENTARY MATERIAL

**Supplement to: "Peskun–Tierney order: Beyond the reversible scenario"** (DOI: 10.1214/20-AOS2008SUPP; .pdf). This supplement contains all the proofs and a discussion.

## REFERENCES

ANDRIEU, C. (2016). On random- and systematic-scan samplers. *Biometrika* **103** 719–726. MR3551794 https://doi.org/10.1093/biomet/asw019

ANDRIEU, C., FORT, G. and VIHOLA, M. (2015). Quantitative convergence rates for subgeometric Markov chains. *J. Appl. Probab.* **52** 391–404. MR3372082 https://doi.org/10.1239/jap/1437658605

ANDRIEU, C. and LIVINGSTONE, S. (2021). Supplement to "Peskun–Tierney ordering for Markovian Monte Carlo: Beyond the reversible scenario." https://doi.org/10.1214/20-AOS2008SUPP

ANDRIEU, C. and VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25** 1030–1077. MR3313762 https://doi.org/10.1214/14-AAP1022

ANDRIEU, C. and VIHOLA, M. (2016). Establishing some order amongst exact approximations of MCMCs. *Ann. Appl. Probab.* **26** 2661–2696. MR3563190 https://doi.org/10.1214/15-AAP1158

ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2018). Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* to appear. arXiv:1808.08592.

BACALLADO, S. (2010). A version of Peskun's theorem for higher-order Markov chains. Master's thesis, Stanford.

BIERKENS, J. and DUNCAN, A. (2017). Limit theorems for the zig-zag process. *Adv. in Appl. Probab.* **49** 791–825. MR3694318 https://doi.org/10.1017/apr.2017.22

BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

BIERKENS, J. and ROBERTS, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.* **27** 846–882. MR3655855 https://doi.org/10.1214/16-AAP1217

BORNN, L., PILLAI, N. S., SMITH, A. and WOODARD, D. (2017). The use of a single pseudo-sample in approximate Bayesian computation. *Stat. Comput.* **27** 583–590. MR3613586 https://doi.org/10.1007/s11222-016-9640-7

BOU-RABEE, N. and SANZ-SERNA, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. MR3693523 https://doi.org/10.1214/16-AAP1255

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

CAMPOS, C. M. and SANZ-SERNA, J. M. (2015). Extra chance generalized hybrid Monte Carlo. *J. Comput. Phys.* **281** 365–374. MR3281978 https://doi.org/10.1016/j.jcp.2014.09.037

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

DAVIS, M. H. A. (1993). *Markov Models and Optimization*. *Monographs on Statistics and Applied Probability* **49**. CRC Press, London. MR1283589 https://doi.org/10.1007/978-1-4899-4483-2

DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978 https://doi.org/10.1214/aoap/1019487508

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

DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

DURMUS, A., GUILLIN, A. and MONMARCHÉ, P. (2018). Piecewise deterministic Markov processes and their invariant measure. arXiv e-prints arXiv:1807.05421.

ETHIER, S. N. and KURTZ, T. G. (2009). *Markov Processes*: *Characterization and Convergence* **282**. John Wiley & Sons. MR0838085 https://doi.org/10.1002/9780470316658

FAGGIONATO, A., GABRIELLI, D. and RIBEZZI CRIVELLARI, M. (2009). Non-equilibrium thermodynamics of piecewise deterministic Markov processes. *J. Stat. Phys.* **137** 259–304. MR2559431 https://doi.org/10.1007/s10955-009-9850-x

FANG, Y., SANZ-SERNA, J.-M. and SKEEL, R. D. (2014). Compressible generalized hybrid Monte Carlo. *J. Chem. Phys.* **140** 174108.

FEARNHEAD, P., BIERKENS, J., POLLOCK, M. and ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33** 386–412. MR3843382 https://doi.org/10.1214/18-STS648

GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

GUSTAFSON, P. (1998). A guided walk Metropolis algorithm. *Stat. Comput.* **8** 357–364.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97

HOBERT, J. P. and MARCHEV, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *Ann. Statist.* **36** 532–554. MR2396806 https://doi.org/10.1214/009053607000000569

HOROWITZ, A. M. (1991). A generalized guided Monte Carlo algorithm. *Phys. Lett. B* **268** 247–252.

HUKUSHIMA, K. and SAKAI, Y. (2013). An irreversible Markov-chain Monte Carlo method with skew detailed balance conditions. *J. Phys., Conf. Ser.* **473** 012012.

KOMOROWSKI, T., LANDIM, C. and OLLA, S. (2012). *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **345**. Springer, Heidelberg. MR2952852 https://doi.org/10.1007/978-3-642-29880-6

LAMB, J. S. and ROBERTS, J. A. (1998). Time-reversal symmetry in dynamical systems: A survey. *Phys. D, Nonlinear Phenom.* **112** 1–39. MR1605826 https://doi.org/10.1016/S0167-2789(97)00199-1

LEISEN, F. and MIRA, A. (2008). An extension of Peskun and Tierney orderings to continuous time Markov chains. *Statist. Sinica* **18** 1641–1651. MR2469328

LELIÈVRE, T., ROUSSET, M. and STOLTZ, G. (2010). *Free Energy Computations: A Mathematical Perspective*. Imperial College Press, London. MR2681239 https://doi.org/10.1142/9781848162488

LIU, J. S. (1996). Peskun's theorem and a modified discrete-state Gibbs sampler. *Biometrika* **83** 681–682. MR1423883 https://doi.org/10.1093/biomet/83.3.681

MAIGRET, N. (1978). Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive. *Ann. Inst. Henri Poincaré B, Calc. Probab. Stat.* **14** 425–440 (1979). MR0523221

MAIRE, F., DOUC, R. and OLSSON, J. (2014). Comparison of asymptotic variances of inhomogeneous Markov chains with application to Markov chain Monte Carlo methods. *Ann. Statist.* **42** 1483–1510. MR3262458 https://doi.org/10.1214/14-AOS1209

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

MIRA, A. (2001). Ordering and improving the performance of Monte Carlo Markov chains. *Statist. Sci.* **16** 340–350. MR1888449 https://doi.org/10.1214/ss/1015346319

NEAL, R. M. (2004). Improving asymptotic variance of MCMC estimators: Non-reversible chains are better. arXiv preprint math/0407281.

O'DONNELL, R. (2014). *Analysis of Boolean Functions*. Cambridge Univ. Press, New York. MR3443800 https://doi.org/10.1017/CBO9781139814782

OTTOBRE, M. (2016). Markov chain Monte Carlo and irreversibility. *Rep. Math. Phys.* **77** 267–292. MR3517569 https://doi.org/10.1016/S0034-4877(16)30031-3

OTTOBRE, M., PILLAI, N. S., PINSKI, F. J. and STUART, A. M. (2016). A function space HMC algorithm with second order Langevin diffusion limit. *Bernoulli* **22** 60–106. MR3449777 https://doi.org/10.3150/14-BEJ621

PEDERSEN, G. K. (2012). *Analysis Now. Graduate Texts in Mathematics* **118**. Springer, New York. MR0971256 https://doi.org/10.1007/978-1-4612-1007-8

PESKUN, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** 607–612. MR0362823 https://doi.org/10.1093/biomet/60.3.607

PETERS, E. A. J. F. and DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* (3) **85** 026703.

PONCET, R. (2017). Generalized and hybrid MCMC overdamped Langevin algorithms. arXiv Preprint arXiv:1701.05833.

REY-BELLET, L. and SPILIOPOULOS, K. (2016). Improving the convergence of reversible samplers. *J. Stat. Phys.* **164** 472–494. MR3519206 https://doi.org/10.1007/s10955-016-1565-1

ROBERTS, G. O. and ROSENTHAL, J. S. (2014). Minimising MCMC variance via diffusion limits, with an application to simulated tempering. *Ann. Appl. Probab.* **24** 131–149. MR3161644 https://doi.org/10.1214/12-AAP918

ROBERTS, G. O. and ROSENTHAL, J. S. (2015). Surprising convergence properties of some simple Gibbs samplers under various scans. *International Journal of Statistics and Probability* **5** 51.

ROSENTHAL, J. S. and ROSENTHAL, P. (2015). Spectral bounds for certain two-factor non-reversible MCMC algorithms. *Electron. Commun. Probab.* **20** no. 91, 10. MR3434208 https://doi.org/10.1214/ECP.v20-4528

SAKAI, Y. and HUKUSHIMA, K. (2016). Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions. *Phys. Rev. E* (3) **93** 043318. https://doi.org/10.1103/PhysRevE.93.043318

SHERLOCK, C., THIERY, A. H. and LEE, A. (2017). Pseudo-marginal Metropolis–Hastings sampling using averages of unbiased estimators. *Biometrika* **104** 727–734. MR3694593 https://doi.org/10.1093/biomet/asx031

SOHL-DICKSTEIN, J., MUDIGONDA, M. and DEWEESE, M. R. (2014). Hamiltonian Monte Carlo without detailed balance. arXiv Preprint arXiv:1409.5191.

TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401 https://doi.org/10.1214/aoap/1027961031

TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D, Nonlinear Phenom.* **240** 410–414.

VOIGT, J. (1977). On the perturbation theory for strongly continuous semigroups. *Math. Ann.* **229** 163–171. MR0445335 https://doi.org/10.1007/BF01351602

VUCELJA, M. (2016). Lifting: A nonreversible Markov chain Monte Carlo algorithm. *Am. J. Phys.* **84** 958–968.

YAGLOM, A. M. (1949). On the statistical reversibility of Brownian motion. *Mat. Sb.* (*N.S.*) **24(66)** 457–492. MR0030718