Efficient Markovian couplings: examples and counterexamples

Krzysztof Burdzy, Department of Mathematics,
University of Washington, Box 354350, Seattle,
WA 98195-4350, USA

Wilfrid S. Kendall, Department of Statistics,
University of Warwick, Coventry CV4 7AL, UK

November 4, 2005

Short title: Efficient Markovian couplings.

Abstract

In this paper we study the notion of an *efficient* coupling of Markov processes. Informally, an efficient coupling is one which couples at the maximum possible exponential rate, as given by the spectral gap. This notion is of interest not only for its own sake, but also of growing importance arising from the recent advent of methods of "perfect simulation": it helps to establish the "price of perfection" for such methods. In general one can always achieve efficient coupling if the coupling is allowed to "cheat" (if each component's behaviour is affected by *future* behaviour of the other component), but the situation is more interesting if the coupling is required to be *co-adapted*. We present an informal heuristic for the existence of an efficient coupling, and justify the heuristic by proving rigorous results and examples in the contexts of finite reversible Markov chains and of reflecting Brownian motion in planar domains.

Keywords: DIFFUSION, CHEN-OPTIMAL COUPLING, CO-ADAPTED COUPLING, COUPLING EXPONENT, EFFICIENT COUPLING, EFFICIENT COUPLING HEURISTIC, EXACT SIMULATION, MARKOV CHAIN, MIRROR COUPLING, MONOTONICITY, PERFECT SIMULATION, PRICE OF PERFECTION, REFLECTING BROWNIAN MOTION, SPECTRAL GAP, SYNCHRONOUS COUPLING.

AMS Subject Classification: 60J27, 60H30, 65U05

1 Introduction

We will call a diffusion-coupling "efficient" if it can be used to obtain a sharp estimate for the spectral gap of the operator which is the generator of the diffusion in question. The main results of this paper show that among well known couplings one can find both efficient and inefficient couplings. Moreover, we give examples of Markov processes for which there is no "efficient" Markovian coupling. We will present techniques which can be used to prove efficiency for many concrete examples of couplings.

Coupling techniques can be applied to obtain various estimates in probability and analysis, both in purely theoretical research and in situations directly related to applications ([32] provides a good introduction; see also [24, 25]). Their importance in applications has recently increased dramatically with the advent of coupling-based "perfect simulation" due to Propp and Wilson [40] and Fill [22]. It is now a matter of pressing importance better to understand the price which must be paid for using such coupling-based approaches – the "price of perfection". One measure of this price is the extent to which coupling occurs at a slower rate than the approach to equilibrium, and this therefore provides a strong motivation for the idea of efficiency and the explorations which we describe below. As pointed out to us by Terry Lyons, in high- or infinite-dimensional settings it is of more interest to consider the relationship of perfect simulation to log-Sobolev inequalities (see [18] for a useful expository article on log-Sobolev in the context of finite Markov chains), and we hope to consider this in later work.

The reader is advised that here we are considering only couplings of Markov chains or diffusions which are *co-adapted*, which is to say that either one of the random processes behaves as a Markov process when we take into account the past of *both* the random processes in question. This is an important point: it is possible (by rather soft arguments) to produce efficient couplings in which a process is allowed to "cheat" by looking ahead into the future of the other process. See for example [1, 2] (which contain much else of relevance to the general concerns which prompted our paper). The imposition of the co-adapted property turns efficiency into a non-trivial notion: it also corresponds to reasonable (though not entirely inevitable) assumptions about how one might implement actual couplings for example in a perfect simulation context.

The concept of *strong uniform times* [2, 1] is also motivated by the desire to get a handle on rates of convergence to stationarity, but uses randomized stopping times rather than coupling ideas, and delivers total variation bounds rather than the L^2-inspired arguments discussed below. Both the kind of coupling considered here and also strong uniform times have led to practical simulation algorithms (respectively "Coupling From The Past" (CFTP) [40] and a sophisticated rejection sampler [17, 22]). Note however that Matthews [35] uses spectral decomposition to obtain a near-optimal strong stationary time.

The idea of an efficient coupling is illustrated in this paper by two kinds of examples. First we consider continuous time Markov chains with finite state space. These results apply to many "attractive systems," similar to the Ehrenfest model discussed in Example 2.11 below. We restrict ourselves to Markov chains reversible with respect to counting measure (hence with symmetric transition probability functions): the ideas of this paper extend to irreversible chains and Rajesh Nandy is investigating this. The second family of examples is concerned with reflected Brownian motion in planar domains. This is related to work on applications of couplings to estimation of the spectral gap for diffusions on manifolds; recently [8, 10, 12, 48], though the basic idea dates back as far as [19]. An extensive bibliography of the notion of couplings as used in spectral gap theory is to be found in [11]: also see [43] for a useful introductory account of analytical approaches.

Wide applications of the coupling technique inevitably lead to the question, which couplings are "good" and which are not? Chen [9, 10] has contributed to this question, introducing a concept of "optimal" couplings. However the terminology is somewhat deceptive, as what is being optimized is a time-varying quantity rather auxiliary to any notion of rapid coupling. In general one expects there to be many different notions of good coupling, depending on whether one seeks high probability of early coupling, high probability of successful coupling, or low exponential moment $\mathbb{E}[\exp(\alpha\tau)]$ of coupling time τ. The notion of "efficient" couplings introduced below isolates those co-adapted couplings which can be used to give a sharp estimate for the "spectral gap." We will show that some Chen-optimal couplings are not efficient because there may be no efficient couplings for some Markov chains. It is natural to expect, although we do not prove it, that some efficient couplings are not Chen-optimal.

We note here that there are of course many other ways of estimating rates of convergence other than coupling: see [42, 41] for examples closely tied to the demands of Markov chain Monte Carlo.

We now present a brief and informal review of the concepts of coupling and spectral gap, and their relationship. Consider a positive-recurrent Markov process X, symmetric with respect to some reference measure m. For many processes the following eigenfunction expansion holds for the density $p_t(x, y)$ relative to m:

$$p(t, x, y) \quad = \quad c + g(x, y)e^{-\mu_2 t} + R(t, x, y)\,. \tag{1.1}$$

The first eigenvalue for the process generator is equal to 0 and the first eigenfunction is the constant function c while μ_2 stands for the second eigenvalue and g is a combination of corresponding eigenfunctions; $g(x, y) = \sum_\varphi \varphi(x)\varphi(y)$ where the φ are orthogonal eigenfunctions with eigenvalue μ_2. The remainder $R(t, x, y)$ converges to 0 faster than $e^{-\mu_2 t}$ as $t \to \infty$, uniformly in x and y for regular cases. Hence μ_2, the "spectral gap" between the first and second eigenvalues, determines the speed of convergence of the transition distribution (density $p(t, x, \cdot)$) to the stationary distribution as $t \to \infty$.

Notice that we may replace $g(x, y)$ by $\varphi(x)\varphi(y)$ when the second eigenvalue is multiplicity-free. Notice also that reversibility considerably simplifies the above analysis, since otherwise the multipliers of the $g(x, y)$ term may include a factor which is polynomial in t. Fortunately reversibility holds in many of the most important applications.

A "coupling" is a pair of (typically dependent) copies of the Markov process X, the first one, X^1, starting from x_1 and the second one, X^2, starting from x_2. "Good" couplings are characterized by small *coupling time* τ, the minimum time t for which X^1_t and X^2_t are equal. Applications of the coupling technique depend on the fact that we may, and we do, construct X^1 and X^2 in such a way that $X^1_t = X^2_t$ for all $t \geq \tau$.

The eigenfunction representation (1.1) now gives

$$\begin{aligned} p(t,x_1,y) - p(t,x_2,y) &= \\ (g(x_1,y)-g(x_2,y))e^{-\mu_2 t} &+ R(t,x_1,y)-R(t,x_2,y)\,, \end{aligned} \tag{1.2}$$

while the coupling yields

$$\begin{aligned} &|p(t,x_1,y)dy - p(t,x_2,y)dy| \\ &\quad= \left|\mathbb{P}(X^1_t \in dy \mid X^1_0 = x_1) - \mathbb{P}(X^2_t \in dy \mid X^2_0 = x_2)\right| \\ &\quad\leq \mathbb{P}(X^1_t \in dy, t<\tau \mid X^1_0 = x_1) + \mathbb{P}(X^2_t \in dy, t<\tau \mid X^2_0 = x_2)\,. \end{aligned} \tag{1.3}$$

Suppose that one can prove that $\mathbb{P}(\tau > t \mid X^1_t = x_1, X^2_t = x_2) \approx e^{-\mu^* t}$ for "generic" $x_1 \neq x_2$. Given a suitable sense for "generic", we can combine (1.2) and (1.3) to show that μ^* is a lower bound for μ_2 (and in most applications it is the lower bound which counts). We will call μ^* the *coupling exponent*.

The above argument has been used in various forms to estimate μ_2, as for example in [12, Theorem 1.7], or [48].

The following informal definition and heuristic capture the spirit of the results and examples of this paper.

> **Informal Definition of Efficiency**. We will call a coupling (X^1, X^2) an *efficient Markovian coupling* if (X^1, X^2) is a Markov process and $\mu^* = \mu_2$.
>
> **Informal Efficient Coupling Heuristic**. A coupling (X^1, X^2) is efficient if and only if, for all t, and given $\{t < \tau\}$, the conditional distributions of (X^1_t, X^2_t) and (X^2_t, X^1_t) are singular with respect to each other.

The above heuristic is *not* true in a rigorous sense, as will be shown in Section 2, but it works in sufficiently many circumstances to make it "almost true."

The efficient coupling heuristic is closely related to "monotonicity" in the sense of [31, §II.2]. The connection will be made more precise in Theorem 2.6 below. The importance of monotonicity or ordering for effective estimation of the rate of convergence of a Markov chain to its stationary distribution is clear, for example, in [34]. The results in Section 2 below are closely related to those in that paper except that our focus is different—the couplings are the main object of study in this paper, rather than just an effective technical tool. The literature on estimating the rate of convergence for Markov processes is enormous. The forthcoming book [1] or the articles [26, 36] may serve as starting points. The importance of monotonicity is explained in [22, §4].

Also note that [1, Chapter 14 §7.1] describes another way of measuring efficiency for a Markov coupling, in the sense of mean coupling time of a graph-based Markov chain of size n increasing fast with n.

We point out that our use of "monotonicity" is somewhat different from that in the sources quoted above. The difference is perhaps best explained by the example of obtuse and acute triangles, discussed in Section 3. The triangles in both families can be expressed as partially ordered sets in a similar way, but efficient couplings for reflected Brownian motion exist only in obtuse triangles, as far as we can tell.

On the practical side, being able to construct an efficient Markovian coupling does not guarantee having a good estimate for the rate of convergence of the process to the stationary distribution. Estimating the coupling exponent μ^* itself may be a hard task, especially when the state space and the transition probabilities do not have a simple structure. Note however that the ideas of perfect simulation [22, 40] finesse this problem away in suitably regular cases.

We will give several distinct formal definitions of efficiency, one for Markov processes with finite state space and continuous time, and two for reflected Brownian motion in planar domains. The goal of the paper is to introduce the idea of efficiency and some accompanying techniques, not to provide a rigid definition and a general theorem. We will adjust our definition to fit particular families of Markov processes and couplings.

The remainder of the paper consists of three sections. Section 2 is devoted to continuous time Markov chains with finite state space. Section 3 studies the mirror coupling of reflected Brownian motions in triangles. Section 4 presents a few informal examples involving mirror couplings for reflected Brownian motion in planar domains. This last section is specialized to a very narrow family of processes but it is at least partly justified by the fact that the technique developed for this case has been subsequently successfully applied in a different context; indeed the methods discussed below have already been used in [3] to prove a number of positive results on the "hot spots" conjecture of J. Rauch, and more recently to construct a counterexample to the conjecture [5, 6].

Acknowledgements: We are grateful to Antonio Galves for most helpful advice on the Ehrenfest model and other attractive systems; to Rajesh Nandy for helpful remarks about eigenvalues; to David Aldous, Jim Fill, and Persi Diaconis for useful suggestions about couplings and Markov chains; and to two referees for their careful reading and helpful comments. Visits by KB to Warwick to conduct research on this paper were enabled by NSF grant DMS-9322689; WSK's research was supported by EPSRC grants GR/K71677 and GR/L56831; further meetings were enabled by MSRI Berkeley as part of its 1997-1998 program on Stochastic Analysis.

2 Couplings for symmetric Markov chains with finite state space

We devote this section to symmetric Markov processes with continuous time and a finite state space, where the reference measure m is counting measure. This simple case fully illustrates the main idea of our test for efficiency but avoids technical issues which arise when the state space is continuous.

Thus $X = \{X_t : t \geq 0\}$ is a continuous time symmetric Markov process with a finite state space D and transition probabilities $p(t, y, x) = p(t, x, y) = \mathbb{P}(X_{s+t} = y \mid X_s = x)$. The following eigenvalue expansion (1.1) holds for X [14, p. 183]:

$$p(t, x, y) \quad = \quad c + g(x, y)e^{-\mu_2 t} + R(t, x, y)\,. \tag{2.1}$$

Here $g(x, y)$ is a combination of eigenfunctions corresponding to the second eigenvalue μ_2, as in (1.1) and the remainder $R(t, x, y)$ converges to 0 faster than $e^{-\mu_2 t}$ when $t \to \infty$.

Suppose now that (X^1, X^2) is a Markovian coupling for the process X. That is to say, each of the three processes $\{(X^1_t, X^2_t) : t \geq 0\}$, $\{X^1_t : t \geq 0\}$ and $\{X^2_t : t \geq 0\}$ is Markov with respect to the filtration generated jointly by X^1 and X^2, and the processes X^1 and X^2 have the same transition probabilities as X. We call the time $\tau = \inf\{t \geq 0 : X^1_t = X^2_t\}$ the *coupling time* for X^1 and X^2. It is convenient to stipulate that $X^1_t = X^2_t$ for all $t \geq \tau$.

We also require that the coupling is invariant under the transposition of its components; this is to say that the transition probabilities for (X^1, X^2) are the same as for (X^2, X^1). In fact this entails no loss in generality, since standard stochastic control arguments (randomizing between an asymmetric coupling transition kernel and its transposition) show that coupling times are stochastically minorized by those obtained by transposition-invariant couplings.

Since the state space $D^2 = D \times D$ for the Markov process (X^1, X^2) is finite, we can apply the Perron-Frobenius theory for nonnegative matrices to the transition probability matrix of the coupling process (X^1, X^2). From this we deduce that there exists a $\mu' = \mu'(x_1, x_2)$, the *coupling exponent function*, such that for all $t \geq 0$ and for all $(x_1, x_2) \in D^2$

$$c_1(x_1, x_2)e^{-\mu' t} \quad \leq \quad \mathbb{P}(\tau > t \mid (X^1_0, X^2_0) = (x_1, x_2)) \quad \leq \quad c_2(x_1, x_2)e^{-\mu' t}\,.$$

(Of course for $x_1 = x_2 = x$ we take $c_1(x, x) = c_2(x, x) = 0$ and then $\mu'(x, x)$ is not well-defined.) We set

$$\mu^* \quad = \quad \min_{x_1, x_2 \in D} \mu'(x_1, x_2)\,, \tag{2.2}$$

the *coupling exponent*. The following simple inequality can be found for example in [12, Thm. 1.7], but we state and prove it here for the sake of completeness. Our main result about finite state space Markov processes uses a generalization of its proof.

Lemma 2.1 *If $g(x_1, \cdot)$ and $g(x_2, \cdot)$ are not identical then $\mu'(x_1, x_2) \leq \mu_2$. It follows that we always have $\mu^* \leq \mu_2$.*

Proof : Choose $x_1, x_2 \in D$ and $y \in D$ such that $g(x_1, y) \neq g(x_2, y)$. By equation (2.1),

$$\begin{aligned} p(t, x_1, y) - p(t, x_2, y) \quad &= \\ &[g(x_1, y) - g(x_2, y)]e^{-\mu_2 t} + R(t, x_1, y) + R(t, x_2, y)\,. \end{aligned} \tag{2.3}$$

Another representation for the same quantity comes from the coupling, namely,

$$\begin{aligned} |p(t, x_1, y) - p(t, x_2, y)| \;&=\; \left|\mathbb{P}(X_t^1 = y \mid X_0^1 = x_1) - \mathbb{P}(X_t^2 = y \mid X_0^2 = x_2)\right| \\ &= \;\left|\mathbb{P}(X_t^1 = y, t < \tau \mid X_0^1 = x_1) - \mathbb{P}(X_t^2 = y, t < \tau \mid X_0^2 = x_2)\right| \\ &\leq \;\mathbb{P}(t < \tau \mid X_0^1 = x_1, X_0^2 = x_2) \\ &\leq \;ce^{-\mu'(x_1,x_2)t}\,. \end{aligned}$$

Since this estimate and equation (2.3) both hold for arbitrarily large t, we can use the condition on our choices of x_1, x_2, y to show that $\mu'(x_1, x_2) \leq \mu_2$. □

This shows that the worst-case coupling exponential decay is never faster than the exponential decay rate of convergence to equilibrium (to wit, the second eigenvalue), and thus motivates our definition of an efficient coupling.

Definition 2.2 *Recall the coupling exponent $\mu^* = \min_{x_1,x_2\in D} \mu'(x_1, x_2)$ defined by equation (2.2). The coupling (X^1, X^2) will be called* efficient *if $\mu^* = \mu_2$.*

Before we state and prove some tests for efficiency, we discuss its definition. The definition is intended to encapsulate a desirable property of couplings in the context of spectral gap estimation. Suppose that one can prove that $\mu'(x_1, x_2) \geq a$ for some a and some pair of points $x_1, x_2 \in D$. Does it necessarily follow that $\mu_2 \geq a$? Example 2.3 below shows that the answer is "no." In order to prove this lower bound for the spectral gap it suffices, in view of Lemma 2.1, to show that $\mu'(x_1, x_2) \geq a$ for some $x_1, x_2 \in D$ with $g(x_1, \cdot)$ and $g(x_2, \cdot)$ not identical. A practical strategy might be to prove that $\mu'(x_1, x_2) \geq a$ for all distinct $x_1, x_2 \in D$. This will be illustrated in Example 2.11 below. We proceed with the aforementioned example showing that the bound $\mu'(x_1, x_2) \geq a$ for a single pair (x_1, x_2) does not necessarily imply the same bound for the spectral gap.

Example 2.3 *Let $D = \{0, 1, ..., 100\} \times \{0, 1, ..., 10\}$. Suppose the process X jumps only to its nearest neighbors in D, and let the jump rate be equal to 1 for every pair of neighbors. (So X is a reflected simple symmetric random walk on a rectangular portion of the planar square lattice $\mathbb{Z}^2$.) Consider a coupling (X^1, X^2) such that*

(a) the first components of X^1 and X^2 are collectively independent of the second components, so that we can describe the joint evolution of (X^1, X^2) by specifying how the first components behave and separately how the second components behave;

(b) the first components of X^1 and X^2 evolve independently till they first agree, after which they remain equal;

(c) the second components behave similarly (independent till they first agree, after which they stick together).

The components of X are independent so it is not hard to check that the second eigenvalue μ_2 for X is the same as for its first component, However, if $X_0^1 = (a,b)$ and $X_0^2 = (a,c)$ with $b \neq c$ then $\mu'((a,b),(a,c))$ is the same as the second eigenvalue for the second component of X and so it is strictly larger than μ_2.

Of course it can be checked directly here that, in the notation of Lemma 2.1, $g((a,b),\cdot) \equiv g((a,c),\cdot)$.

Definition 2.4 *We say that (y_1, y_2) is* accessible *from (x_1, x_2) if*

$$\mathbb{P}\left[(X_t^1, X_t^2) = (y_1, y_2) \mid (X_0^1, X_0^2) = (x_1, x_2)\right] \quad > \quad 0$$

for some (and, therefore, all) $t > 0$.

Accessibility is clearly a transitive property: if (y_1, y_2) is accessible from (x_1, x_2) and (z_1, z_2) is accessible from (y_1, y_2) then (z_1, z_2) is accessible from (x_1, x_2). Note that accessibility of states in D^2 is a property of the coupling and not just X: in fact it is the same as accessibility for the coupling Markov chain (X^1, X^2) restricted to $D^2 \setminus \Delta$, where $\Delta \subset D^2$ is the diagonal.

Definition 2.5 *We will say that the coupling (X^1, X^2) has the* transposition property *relative to x_1 if*

(A) for all x_2 with $x_1 \neq x_2$, and for all y_1 and y_2 with $y_1 \neq y_2$, the accessibility of (y_1, y_2) from (x_1, x_2) implies accessibility of both (x_1, x_2) and (x_2, x_1) from (y_1, y_2);

(B) for every $x_2 \neq x_1$, there is at least one pair (y_1, y_2), distinct from (x_1, x_2), which is accessible from (x_1, x_2) (so for (X_1, X_2) restricted to $D^2 \setminus \Delta$ there are no isolated states involving x_1).

We will say that D^2 is *irreducible* with respect to a given coupling if every state (y_1, y_2) with $y_1 \neq y_2$ is accessible from any other state (x_1, x_2) with $x_1 \neq x_2$.

Note that the transposition property relative to x_1 can also be reduced to considerations about state classification for the coupling chain (X^1, X^2), bearing in mind that we are only considering coupling chains which are symmetric under the permutation $(x_1, x_2) \leftrightarrow (x_2, x_1)$. It can be shown that the transposition property is *equivalent* to the requirement that, for any x_2 with $x_2 \neq x_1$, the communicating class of (x_1, x_2) under the chain (X^1, X^2) restricted to $D^2 \setminus \Delta$ is essential and is saturated under the symmetry $(x_1, x_2) \leftrightarrow (x_2, x_1)$.

The following fundamental result uses these properties to identify many chains which are efficient and many chains which are not.

Theorem 2.6

(i) *If the coupling* (X^1, X^2) *has the transposition property relative to a point* $x_1 \in D$ *then* $\mu'(x_1, x_2) < \mu_2$ *for some* $x_2 \in D^2$ *distinct from* x_1*, and so the coupling is not efficient.*

(ii) *Suppose that for some* $x_1, x_2 \in D$ *there exists a function* $f : D \to \mathbb{R}$ *with the property that* $f(X_t^1) - f(X_t^2)$ *almost surely remains strictly positive for* $t < \tau$*, given* $X_0^1 = x_1$ *and* $X_0^2 = x_2$*. Then* $\mu'(x_1, x_2) \geq \mu_2$*.*

The following statement follows immediately from Theorem 2.6.

Corollary 2.7

(i) *If* D^2 *is irreducible for the coupling* (X^1, X^2) *then this coupling is not efficient.*

(ii) *Suppose that for every pair of distinct points* $x_1, x_2 \in D$ *there exists a function* $f : D \to \mathbb{R}$ *with the property that* $f(X_t^1) - f(X_t^2)$ *almost surely remains strictly positive for* $t < \tau.$*, given* $X_0^1 = x_1$ *and* $X_0^2 = x_2$*. Then the coupling is efficient.*

Remark : An interesting example to which Corollary 2.7 applies is the *independence sampler* discussed in [46]. Here one can compute the eigenvalues (and indeed the transition matrix) explicitly [33, 45] and verify (at least for finite state space) that the Markov chain is efficient. As pointed out by Cai [7], this chain possesses a monotonicity structure. Cai uses this monotonicity to build a CFTP algorithm, but it also guarantees efficiency as above.

We will show in Examples 2.9 and 2.10 that neither of the conditions in parts (i) and (ii) of Theorem 2.6 is necessary. Theorem 2.6 should be compared with [40, §5], which implies that bounded monotone Markov chains are efficient.

Proof of Theorem 2.6:

(i) Suppose that the coupling (X^1, X^2) has the transposition property relative to some x_1. Fix any point $x_2 \in D$ distinct from x_1 and assume that $(X_0^1, X_0^2) = (x_1, x_2)$. Consider a coupling $(\tilde{X}^1, \tilde{X}^2)$ which is independent of (X^1, X^2), having the same transition probabilities as (X^1, X^2) but starting from (x_2, x_1) rather than from (x_1, x_2). Let $\tilde{\tau}$ denote the coupling time for $(\tilde{X}^1, \tilde{X}^2)$.

We will estimate the chance that (X^1, X^2) and $(\tilde{X}^1, \tilde{X}^2)$ have not met before time s, given $\{\tau > s, \tilde{\tau} > s\}$. The invariance of the coupling under the transposition of its components, the transitivity of the accessibility property and the transposition property relative to x_1 can be used to show that the accessibility of any (y_1, y_2) from (x_1, x_2) implies accessibility of (y_2, y_1) from (x_1, x_2). Consider any integer $k > 0$. Since we have assumed that $(X_0^1, X_0^2) = (x_1, x_2)$ and $(\tilde{X}_0^1, \tilde{X}_0^2) = (x_2, x_1)$, at time t both processes can only take those pairs of values from which they can reach (x_1, x_2) (we are using the transposition property here). The state space is finite, so the processes can reach (x_1, x_2) within an arbitrarily small time, less than $1/4$ say, with some strictly positive probability not depending on their values at time $t = k$. We will need a stronger version of this statement. Condition the processes (X^1_t, X^2_t) and $(\tilde{X}^1_t, \tilde{X}^2_t)$ on their values at times $t = k$ and $t = k + 1$ (we consider only the values that can be taken with strictly positive probabilities). The transposition property can be used again to show that there are possible trajectories which take the processes to the intermediate state (x_1, x_2) for all $t \in [k+1/4, k+3/4]$. The finiteness of the state space can be now invoked to see that such an event has a probability bounded below. More precisely, there exists a probability $p > 0$ such that for all integers $k > 0$ and all $s \geq k + 1$,

$$\begin{aligned}
&\mathbb{P}\left[X^1_t = \tilde{X}^1_t = x_1, X^2_t = \tilde{X}^2_t = x_2 \text{ for } k + 1/4 \leq t \leq k + 3/4\right. \\
&\quad \left. \mid X^1_k, X^1_{k+1}, X^2_k, X^2_{k+1}, \tilde{X}^1_k, \tilde{X}^1_{k+1}, \tilde{X}^2_k, \tilde{X}^2_{k+1}, \tau > s, \tilde{\tau} > s\right] \quad > \quad p\,.
\end{aligned}$$

We are interested only in large s so we will assume that $s > 1$. Let j be the largest integer with $j \leq s$. Condition the processes (X^1, X^2) and $(\tilde{X}^1, \tilde{X}^2)$ on their values at times $0, 1, 2, \ldots, j$ and on the event $\{\tau > s, \tilde{\tau} > s\}$. It follows that

$$\begin{aligned}
&\mathbb{P}\left[(X^1_t, X^2_t) \neq (\tilde{X}^1_t, \tilde{X}^2_t) \text{ for } 0 \leq t < s \mid \tau > s, \tilde{\tau} > s\right] \\
&\leq \quad (1 - p)^j \leq e^{-\mu s}, \qquad (2.4)
\end{aligned}$$

for appropriate $\mu > 0$.

Let σ be the smallest t such that $(X^1_t, X^2_t) = (\tilde{X}^1_t, \tilde{X}^2_t)$. Thus inequality (2.4) gives

$$\mathbb{P}[\sigma > s \mid \tau > s, \tilde{\tau} > s, X^1_0 = x_1, X^2_0 = x_2] \quad \leq \quad e^{-\mu s}\,. \qquad (2.5)$$

We now make a simple observation about the relative behaviour of the processes (X^1, X^2) and $(\tilde{X}^1, \tilde{X}^2)$ after the meeting time σ. We use the strong Markov property applied to $(X^1, X^2, \tilde{X}^1, \tilde{X}^2)$ at time σ, to deduce the following:

$$\begin{aligned}
&\mathbb{P}\left[X^1_s = y, s \geq \sigma \mid X^1_\sigma, X^2_\sigma, \tilde{X}^1_\sigma, \tilde{X}^2_\sigma, \sigma, s < \tau, s < \tilde{\tau}, X^1_0 = x_1, \tilde{X}^1_0 = x_2\right] \\
&= \mathbb{P}\left[\tilde{X}^1_s = y, s \geq \sigma \mid X^1_\sigma, X^2_\sigma, \tilde{X}^1_\sigma, \tilde{X}^2_\sigma, \sigma, s < \tau, s < \tilde{\tau}, X^1_0 = x_1, \tilde{X}^1_0 = x_2\right]
\end{aligned}$$

(we suppress the conditioning on $X^2_0 = x_2$ and $\tilde{X}^2_0 = x_1$, since by definition we have $X^2_0 = \tilde{X}^1_0$ and $\tilde{X}^2_0 = X^1_0$). Hence we can use integration, and rewrite the conditioning in terms of (X^1_0, X^2_0), respectively $(\tilde{X}^1_0, \tilde{X}^2_0)$, to show that

$$\begin{aligned}
&\mathbb{P}\left[X^1_s = y, s \geq \sigma \mid s < \tau, s < \tilde{\tau}, X^1_0 = x_1, X^2_0 = x_2\right] \\
&\quad = \quad \mathbb{P}\left[\tilde{X}^1_s = y, s \geq \sigma \mid s < \tau, s < \tilde{\tau}, \tilde{X}^1_0 = x_2, \tilde{X}^2_0 = x_1\right]\,.
\end{aligned}$$

We now generalize the proof of Lemma 2.1. First we find points $x_2, y \in D$ such that $g(x_1, y) \neq g(x_2, y)$. Note that such points exist using the orthogonality of the eigenfunctions φ in $g(x, y) = \sum_\varphi \varphi(x)\varphi(y)$, and the fact that eigenfunctions corresponding to the second eigenvalue must change sign.

We recall (2.3), namely,

$$\begin{aligned} p(s, x_1, y) - p(s, x_2, y) \;&= \\ [g(x_1, y) - g(x_2, y)]e^{-\mu_2 s} &+ R(s, x_1, y) + R(s, x_2, y). \end{aligned} \tag{2.6}$$

The estimate based on coupling is more complicated in the present case. We use symmetry, inequality (2.5), and follow the method described in the proof of Lemma 2.1. We find

$$\begin{aligned}
&|p(s, x_1, y) - p(s, x_2, y)| \\
=\;& \left|\mathbb{P}(X^1_s = y \mid X^1_0 = x_1) - \mathbb{P}(X^2_s = y \mid X^2_0 = x_2)\right| \\
=\;& \big|\,\mathbb{P}(X^1_s = y, s < \tau \mid X^1_0 = x_1, X^2_0 = x_2) \\
&\quad - \mathbb{P}(X^2_s = y, s < \tau \mid X^1_0 = x_1, X^2_0 = x_2)\big| \\
=\;& \big|\,\mathbb{P}(X^1_s = y \mid s < \tau, X^1_0 = x_1, X^2_0 = x_2)\,\mathbb{P}(s < \tau \mid X^1_0 = x_1, X^2_0 = x_2) \\
&- \mathbb{P}(X^2_s = y \mid s < \tau, X^1_0 = x_1, X^2_0 = x_2)\,\mathbb{P}(s < \tau \mid X^1_0 = x_1, X^2_0 = x_2)\big| \\
\leq\;& \big|\,\mathbb{P}(X^1_s = y \mid s < \tau, X^1_0 = x_1, X^2_0 = x_2) \\
&\quad - \mathbb{P}(X^2_s = y \mid s < \tau, X^1_0 = x_1, X^2_0 = x_2)\big| \times c(x_1, x_2)e^{-\mu'(x_1,x_2)s} \\
=\;& \big|\,\mathbb{P}(X^1_s = y \mid s < \tau, s < \tilde\tau, X^1_0 = x_1, X^2_0 = x_2) \\
&\quad - \mathbb{P}(\tilde X^1_s = y \mid s < \tau, s < \tilde\tau, \tilde X^1_0 = x_2, \tilde X^1_0 = x_1)\big| \times ce^{-\mu' s}
\end{aligned}$$

This last step uses the fact that (X^1, X^2), $(\tilde X^1, \tilde X^2)$ are independent to justify the insertion of conditioning on *both* $s < \tau$ and $s < \tilde\tau$ for both probabilities. We now use the conditional probability identity noted above to cancel between the two conditional probabilities to yield:

$$\begin{aligned}
&\big|\,\mathbb{P}(X^1_s = y \mid s < \tau, s < \tilde\tau, X^1_0 = x_1, X^2_0 = x_2) \\
&\quad - \mathbb{P}(\tilde X^1_s = y \mid s < \tau, s < \tilde\tau, \tilde X^1_0 = x_2, \tilde X^2_0 = x_1)\big| \times ce^{-\mu' s} \\
=\;& \big|\,\mathbb{P}(X^1_s = y, s < \sigma \mid s < \tau, s < \tilde\tau, X^1_0 = x_1, X^2_0 = x_2) \\
&\quad - \mathbb{P}(\tilde X^1_s = y, s < \sigma \mid s < \tau, s < \tilde\tau, \tilde X^1_0 = x_2, \tilde X^2_0 = x_1)\big| \times ce^{-\mu' s} \\
&\leq \mathbb{P}(s < \sigma \mid s < \tau, s < \tilde\tau, X^1_0 = x_1, \tilde X^1_0 = x_2)ce^{-\mu' s} \\
&\leq e^{-\mu s}ce^{-\mu' s}\,.
\end{aligned}$$

The last estimate and (2.6) hold for arbitrarily large s, so $\mu_2 \geq \mu'(x_1, x_2) + \mu > \mu'(x_1, x_2)$ and we see that the coupling is not efficient.

(ii) Find a function $f : D \to \mathbb{R}$ corresponding to x_1, x_2, as in the statement of the theorem. Let $\rho = \inf\{f(x) - f(y) : f(x) > f(y)\}$. Note that $\rho > 0$

because D is finite. If $X_0^1 = x_1$ and $X_0^2 = x_2$ then $f(X_t^1) - f(X_t^2) \geq \rho$ for $t < \tau$.

Let n be the smallest index for an eigenvalue μ_n such that $\varphi_n(x_1) \neq \varphi_n(x_2)$. Such an index must exist because otherwise the eigenfunction expansions would be identical for $p(t, x_1, y)$ and $p(t, x_2, y)$ and, consequently, these functions would be identical; this is not the case since $x_1 \neq x_2$.

By replacing the function $f(x)$ with $f(x)\exp(\alpha f(x))$, for an appropriate α, if necessary, we may assume that $S = \sum_{y \in D} f(y)\varphi_n(y) \neq 0$. From an eigenfunction expansion similar to (2.1) but listing higher order terms we obtain,

$$\begin{aligned}
& \mathbb{E}\left[f(X_t^1) \mid X_0^1 = x_1\right] - \mathbb{E}\left[f(X_t^2) \mid X_0^2 = x_2\right] \\
= & \sum_{y \in D} f(y) p(t, x_1, y) - \sum_{y \in D} f(y) p(t, x_2, y) \\
= & \sum_{y \in D} f(y)\{[\varphi_n(x_1) - \varphi_n(x_2)]\varphi_n(y) e^{-\mu_n t} + \tilde{R}(t, x_1, y) + \tilde{R}(t, x_2, y)\} \\
= & S[\varphi_n(x_1) - \varphi_n(x_2)] e^{-\mu_n t} + \hat{R}(t, x_1, x_2, y)\,, \qquad (2.7)
\end{aligned}$$

where $\hat{R}(t, x_1, x_2, y)$ goes to 0 faster than $e^{-\mu_n t}$ as $t \to \infty$. Recalling the definition of μ',

$$\begin{aligned}
& \mathbb{E}\left[f(X_t^1) \mid X_0^1 = x_1\right] - \mathbb{E}\left[f(X_t^2) \mid X_0^2 = x_2\right] \\
& = \mathbb{E}\left[f(X_t^1)\mathrm{I}_{\{t<\tau\}} \mid X_0^1 = x_1\right] - \mathbb{E}\left[f(X_t^2)\mathrm{I}_{\{t<\tau\}} \mid X_0^2 = x_2\right] \\
& = \mathbb{E}\left[(f(X_t^1) - f(X_t^2))\mathrm{I}_{\{t<\tau\}} \mid X_0^1 = x_1, X_0^2 = x_2\right] \\
& \geq \rho\, \mathbb{E}\left[\mathrm{I}_{\{t<\tau\}} \mid X_0^1 = x_1, X_0^2 = x_2\right] \\
& \geq \rho c(x_1, x_2) e^{-\mu'(x_1, x_2) t}.
\end{aligned}$$

Comparing this with (2.7) for large t shows that $\mu'(x_1, x_2) \geq \mu_n \geq \mu_2$.

This concludes the proof of the theorem. □

Note that Corollary 2.7(ii) follows because $\mu^* = \min \mu'(x_1, x_2) \geq \mu_2$. Since we always have $\mu^* \leq \mu_2$, we see $\mu^* = \mu_2$ as required.

The following notation will be used for the rest of the section. For distinct $d_1, d_2 \in D$, we will denote the jump rate from d_1 to d_2 by $q(d_1, d_2)$, i.e.,

$$q(d_1, d_2) \quad = \quad \lim_{s \to 0} \frac{1}{s}\,\mathbb{P}(X_{t+s} = d_2 \mid X_t = d_1).$$

Note that by symmetry we have $q(d_1, d_2) = q(d_2, d_1)$. Consider any coupling (X^1, X^2) for X. In a slight abuse of notation we will also use q for the transition rates for the coupling process (X^1, X^2): for distinct pairs (d_1, d_2) and (d_3, d_4)

we set

$$q((d_1,d_2),(d_3,d_4)) \quad = \quad \lim_{s\to 0}\frac{1}{s}\,\mathbb{P}\left[(X^1_{t+s},X^2_{t+s})=(d_3,d_4)\mid (X^1_t,X^2_t)=(d_1,d_2)\right].$$

Since the processes X^1 and X^2 are Markov and have the same transition probabilities as X, for all $d_1,d_2,d_3\in D$ with $d_1\neq d_3$ we must have

$$\sum_{d_4\in D} q((d_1,d_2),(d_3,d_4)) \quad = \quad q(d_1,d_3). \tag{2.8}$$

For the same reason, if $d_2\neq d_4$ then

$$\sum_{d_3\in D} q((d_1,d_2),(d_3,d_4)) \quad = \quad q(d_2,d_4)\,. \tag{2.9}$$

We will say that X^1 and X^2 make independent jumps from (d_1,d_2) if

$$q((d_1,d_2),(d_3,d_2)) \quad = \quad q(d_1,d_3)$$

and

$$q((d_1,d_2),(d_1,d_4)) \quad = \quad q(d_2,d_4)$$

for all $d_3\neq d_1$, $d_4\neq d_2$.

Consider a simple example with the state space $D=\{0,1,...,100\}^2$. Suppose that X is a continuous time Markov process on D such that its jumps form the simple random walk reflected on the "boundary" of D. Let X^1 and X^2 be run as independent copies of X until their coupling time τ. It is no surprise to note that (X^1,X^2) is not efficient, by Corollary 2.7 (i). However, this "independent" coupling is efficient when the state space is ordered. Moreover, a very weak condition ensures efficiency in this case, so that the family of efficient couplings is rather large: it is required only that the coupling maintains the ordering. We state the following result for skip-free chains:

Corollary 2.8 *Suppose the state space D is a finite subinterval of the integers $\mathbb{Z}$ and X can jump from x only to $x-1$ or $x+1$, for every x. (So X is a finite-state-space generalized birth-death process.) Assume that (X^1_t,X^2_t) almost surely never jumps to (X^2_{t-},X^1_{t-}). Then (X^1,X^2) is an efficient coupling. In particular, the coupling is efficient if X^1 and X^2 have independent jumps until the coupling time τ.*

Proof : The corollary follows from Corollary 2.7 (ii). It suffices to use either $f(x)\equiv x$ or $f(x)\equiv -x$. □

The next two examples show that *neither* of the conditions in parts (i) and (ii) of Theorem 2.6 is necessary.

Example 2.9 *We construct a Markov process, and a coupling with $\mu'(x_1, x_2) \geq \mu_2$ for a specific pair of points x_1 and x_2, although there is no function f satisfying condition (ii) of Theorem 2.6 for this pair of points. We also show that $\varphi_2(x_1) \neq \varphi_2(x_2)$, since otherwise this example would not be an improvement on Example 2.3 (where we have $\mu'(x_1, x_2) > \mu_2$ for some points but only for those with $\varphi_2(x_1) = \varphi_2(x_2)$).*

Fix some large n and let the state space of the process be

$$D = \{a_0^1, a_1^1, a_2^1, a_3^1, a_0^2, a_1^2, a_2^2, a_3^2, \dots, a_0^n, a_1^n, a_2^n, a_3^n\}.$$

Fig. 1 illustrates all possible jumps for the process. We take $q(a_0^j, a_0^{j+1}) = 1$ for $j = 1, 2, \dots, n-1$, and $q(d_1, d_2) = \tilde{q}$ for all other vertices d_1 and d_2 connected by a wedge in the graph. We will choose a value for $\tilde{q}$ later in the example.

[Figure 1 about here.]

Consider a coupling with the jump rates

$$q((a_m^j, a_m^k), (a_l^j, a_l^k)) = q(a_m^j, a_l^j)$$

for all j, k, l and m, with $j \neq k$. Suppose that

$$\begin{aligned}
q((a_0^j, a_0^k), (a_0^{j+1}, a_0^k)) &= 1, \qquad j \neq k,\ j = 1, \dots, n-1, \\
q((a_0^j, a_0^k), (a_0^{j-1}, a_0^k)) &= 1, \qquad j \neq k,\ j = 2, \dots, n, \\
q((a_0^j, a_0^k), (a_0^j, a_0^{k+1})) &= 1, \qquad j \neq k,\ k = 1, \dots, n-1, \\
q((a_0^j, a_0^k), (a_0^j, a_0^{k-1})) &= 1, \qquad j \neq k,\ k = 2, \dots, n.
\end{aligned}$$

We require that X^1 and X^2 make independent jumps from all other points (d_1, d_2).

If $(X_0^1, X_0^2) = (a_l^j, a_m^j)$ for some $l \neq m$ then $\mathbb{P}((X_s^1, X_s^2) = (a_m^j, a_l^j)) > 0$ for every $s > 0$, so there does not exist a function $f : D \to \mathbb{R}$ such that $f(X_t^1) - f(X_t^2)$ is strictly positive until the coupling time. However, we will argue that $\mu'(a_l^j, a_m^j) = \mu_2$ for some j and $l \neq m$, if $\tilde{q}$ is large enough.

If both X_t^1 and X_t^2 lie on the "spine" $\{a_0^1, a_0^2, \dots, a_0^n\}$ at some time s, then from this time on, they will make excursions into side alleys of the form $\{a_0^j, a_1^j, a_2^j, a_3^j\}$ at the same time and they will return from those excursions to the spine at the same time. It follows that if $(X_s^1, X_s^2) = (a_0^j, a_0^k)$ with $j < k$ then X_t^1 will lie to the left of X_t^2 for all $t \in [s, \tau)$. We let $f(a_m^j) = j$ and use Theorem 2.6 (ii) to see that $\mu'(a_0^j, a_0^k) \geq \mu_2$ for $j \neq k$.

We will estimate $\mu'(a_0^j, a_0^k)$. First suppose that the processes X^1 and X^2 start from distinct points on the spine. Then their evolution may be described as that of two independent copies of X along the spine except that they may make simultaneous excursions into the side alleys of the form $\{a_0^j, a_1^j, a_2^j, a_3^j\}$. Those side excursions can only delay the coupling time τ for X^1 and X^2 so τ is stochastically minorized by the coupling time for a pair of independent random walks on the spine, reflected at the endpoints of the spine. Hence, by Corollary 2.8,

$$\mathbb{P}\left[\tau > t \mid (X_0^1, X_0^2) = (a_0^j, a_0^k)\right] \quad \geq \quad c(j,k)e^{-\hat{\mu}t},$$

for all $j \neq k$, where $\hat{\mu}$ is the second eigenvalue for the process restricted to the spine. Note that $\hat{\mu}$ does not depend on $\tilde{q}$.

Next suppose that (X^1, X^2) starts from (a_l^j, a_m^j) for some j and $l \neq m$. If we choose sufficiently large $\tilde{q}$ then the processes X^1 and X^2 will rapidly and independently jump along the edges connecting the elements of the family $\{a_0^j, a_1^j, a_2^j, a_3^j\}$. It is clear that they will rapidly couple, before leaving this set. As a consequence, for sufficiently large $\tilde{q}$ one can choose c_1 such that for all j, l and m,

$$\mathbb{P}\left[\tau > t \mid (X_0^1, X_0^2) = (a_l^j, a_m^j)\right] \quad \leq \quad c_1 e^{-2\hat{\mu}t}.$$

It follows that for large t,

$$\begin{aligned}\mathbb{P}\left[\tau > t \mid (X_0^1, X_0^2) = (a_l^j, a_m^j)\right] \quad &\leq \quad c_1 e^{-2\hat{\mu}t} \quad \leq \quad \min_{r \neq k} c(r,k) e^{-\hat{\mu}t} \quad \leq \\ &\leq \quad \min_{r \neq k} \mathbb{P}\left[\tau > t \mid (X_0^1, X_0^2) = (a_0^r, a_0^k)\right] .\end{aligned}$$

Hence, $\min_j \min_{l \neq m} \mu'(a_l^j, a_m^j) \geq \max_{r \neq k} \mu'(a_0^r, a_0^k) \geq \mu_2$.

Finally we will show that $g(a_l^j, \cdot)$ and $g(a_m^j, \cdot)$ are not identical for some j and $l \neq m$. Suppose that the converse holds. Fix some j and note that if $g(a_l^j, \cdot) = g(a_m^j, \cdot)$ for all l and m then $g(a_1^j, \cdot)$ is an average in the first argument of the other $g(a_m^j, \cdot)$, and indeed g is harmonic in its first argument at a_1^j. As a function of its first argument, g is an eigenfunction corresponding to μ_2; therefore we must have $g(a_l^j, \cdot) = 0$ for all l. If this is true for all j then g is identically equal to 0, which is a contradiction.

We conclude that for some j and $l \neq m$ we have $\mu'(a_l^j, a_m^j) \geq \mu_2$ and $g(a_l^j, \cdot)$, $g(a_m^j, \cdot)$ not identical. However there is no function f which would satisfy condition (ii) of Theorem 2.6 for a_l^j and a_m^j.

Example 2.10 *We present a Markov process, a coupling and a pair of points* x_1, x_2 *with* $\mu'(x_1, x_2) < \mu_2$ *although neither* x_1 *nor* x_2 *satisfies the transposition property (i) of Theorem 2.6 (i). Let* $D = \{-100, -99, ..., 100\} \times \{0, 1, ..., 50\}$ *be a bounded portion of* $\mathbb{Z}^2$. *Let the process* X *be able to jump only to its nearest neighbors in* D, *and let the jump rate be equal to 1 for every pair of neighbors. We consider a coupling* (X^1, X^2) *with the jump rates*

$$q(((-j,k),(j,k)),\, ((-l,m),(l,m)) \quad = \quad 1$$

for $j \geq 1$, $|j - l| + |k - m| = 1$. *The processes* X^1 *and* X^2 *have independent jumps from all points which are not symmetric with respect to the vertical axis* K.

Suppose that X^1 and X^2 start from distinct points x_1 and x_2 which are not symmetric with respect to K. Then these processes are independent copies of X until the time T when X^1 hits X^2 or the symmetric image of X^2 with respect to K. The second eigenvalue μ_2 for X is the same as the second eigenvalue for its first component. The following few assertions are quite clear but we will not go into a detailed proof of them as it would take too much space. We have $\mathbb{P}(T > t) \geq ce^{-\mu t}$ for some c and $\mu < \mu_2$. This is justified by comparing T to U, where U is the time when X hits K, and by the fact that $\mathbb{P}(U > t) \leq ce^{-\mu_2 t}$. We obtain $\mu'(x_1, x_2) \leq \mu < \mu_2$.

Let x_3 be the point symmetric to x_1 with respect to K. Note that (x_1, x_3) is accessible from (x_1, x_2) for this particular coupling but neither (x_1, x_2) nor (x_2, x_1) is accessible from (x_1, x_3). Hence neither x_1 nor x_2 has the transposition property.

Our next example is a continuous version of [21, Exercise 5.9].

Example 2.11 *(Ehrenfest model or random walk on hypercube) Consider two urns with n marked balls distributed among them. At every arrival time for a Poisson process, a ball is randomly chosen from among all balls in both urns and moved to the other urn.*

A formal description of the model is the following. The state space D for our process is the set of all binary sequences $(i_1, i_2, \dots, i_n)$ of length n, so each i_j is equal to 0 or to 1. Let $U_k, k \geq 1$, be independent exponential (mean 1) random variables and set $T_k = U_1 + \cdots + U_k$. Consider random variables N_k which are independent of each other and of the T_k's, and which are uniformly distributed over the fixed range $\{1, 2, \dots, n\}$. Finally, let $\{J_k\}$ be a sequence of random variables, independent of each other, of the T_k's, and of the N_k's, and such that $\mathbb{P}(J_k = 0) = \mathbb{P}(J_k = 1) = 1/2$. We define the process X on D by prescribing its initial value $X_0 = (i_1, i_2, \dots, i_n)$ and by specifying its jumps; X jumps at times T_k (and only at these times), the jump at time T_k taking the process from $X_{T_k-} = (j_1, j_2, \dots, j_{N_k}, \dots, j_n)$ to $X_{T_k} = (j_1, j_2, \dots, J_k, \dots, j_n)$. If the process jumped instead to $(j_1, j_2, \dots, 1 - j_{N_k}, \dots, j_n)$ at time T_k then we would have obtained a model directly corresponding to the informal "urn" representation. However, the two processes X, corresponding to two kinds of jumps, can be transformed into each other by speeding up or slowing down the clock for X.

We construct a coupling (X^1, X^2) by using just one family of random variables $\{T_k, N_k, J_k\}_{k \geq 1}$ for both processes X^1 and X^2. Specifically, the transition probabilities for (X^1, X^2) are specified by the requirement that the process (X^1, X^2) jumps at times T_k (and only at these times) from

$$(X^1_{T_k-}, X^2_{T_k-}) = ((j^1_1, j^1_2, \dots, j^1_{N_k}, \dots, j^1_n), (j^2_1, j^2_2, \dots, j^2_{N_k}, \dots, j^2_n))$$

to

$$(X^1_{T_k}, X^2_{T_k}) = ((j^1_1, j^1_2, \dots, J_k, \dots, j^1_n), (j^2_1, j^2_2, \dots, J_k, \dots, j^2_n)).$$

It is immediate that (X^1, X^2) and also X^1 and X^2 are all Markov processes; moreover the latter two have the same transition probabilities as X.

Suppose that $X_0^1 = (j_1^1, j_2^1, \ldots, j_n^1)$. Let $f(i_1, i_2, \ldots, i_n)$ be the number of k such that $i_k = j_k^1$. It is elementary to check that if $X_0^1 \neq X_0^2$ then $f(X_t^1) - f(X_t^2)$ stays strictly positive for all $t < \tau$. By Corollary 2.7 (ii), our coupling is efficient.

Recall μ^* defined before Lemma 2.1 in equation (2.2). In order to estimate μ^* we consider the worst case scenario, i.e., that initially no components of X_0^1 and X_0^2 are equal. For a fixed k, the waiting time for the k-th components of X^1 and X^2 to meet is exponential with mean n. Once these components meet, they will be equal to each other forever, although they will not be constant. It is an easy consequence of the theory of Poisson point processes with independent marks that the waiting times for different components are independent. The probability that a specified pair of components have not merged by time t is equal to $e^{-t/n}$ so the probability that there exists at least one such pair is equal to $1 - (1 - e^{-t/n})^n$ which is between $(1/2)ne^{-t/n}$ and $ne^{-t/n}$ for large t. We conclude that $\mu^* = 1/n$, and since our coupling is efficient, we see that the spectral gap is also equal to $1/n$.

Consider now the asymptotics when $n \to \infty$. In this example, the mean time to coupling is not of order $1/\mu^* = n$. The expected waiting time between the k-th coupling of a pair of components of X^1 and X^2 and the $(k+1)$-st coupling is $n/(n-k)$, so that the expected time until all components are coupled is of order $\sum_{k=0}^{n-1} n/(n-k) \approx n \log n$, which is larger than $1/\mu^*$ by a factor of $\log n$.

Example 2.12 *We construct a symmetric Markov process X_t for which there are no efficient Markovian couplings. The state space consists of* 11 *points,*

$$D = \{a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, \ldots, c_5\}.$$

The edges in Fig. 2 show all possible jumps for the process; in other words, points $d_1, d_2 \in D$ are connected by an edge in Fig. 2 if and only if $q(d_1, d_2) > 0$. The jump rates are $q(a_1, a_2) = 1$, $q(a_2, a_3) = 2$, and $q(d_1, d_2) = 4$ for all other edges in Fig. 2.

[Figure 2 about here.]

Consider any coupling (X^1, X^2) for this process. Suppose that $d_1, d_2 \in D \setminus \{a_2\}$ are such that $q(d_1, d_2) > 0$, and, therefore, $q(d_1, d_2) = 4$. Since $q(a_2, a_1) = 1$ and $q(a_2, a_3) = 2$, the identities (2.8) and (2.9) imply that $q((d_1, a_2), (d_2, a_1)) \leq 1$ and $q((d_1, a_2), (d_2, a_3)) \leq 2$. By another application of equations (2.8)-(2.9) we deduce from these inequalities that $q((d_1, a_2), (d_2, a_2)) > 0$, since $q(a_2, a_1) = 1$, $q(a_2, a_3) = 2$, and the sum of all $q((d_1, a_2), (d_2, a_i))$ equals 4.

We will describe an evolution of the process (X^1, X^2) before the coupling time which may happen with positive probability, no matter what coupling is used.

First we consider the case when one of the processes X^1 or X^2 starts from a_2. Without loss of generality suppose that $X_0^1 = a_2$ and $X_0^2 = d_0 \neq a_2$. It follows from what we have just proved that $q((d_0, a_2), (d_2, a_2)) > 0$, where d_2 is any "neighbor" of d_0 in the "rectangle" $D \setminus \{a_2\}$. Hence, there is a positive probability that X^2 will move through a sequence of points in $D \setminus \{a_2\}$ and reach b_2 at time t_1, before the time when X^1 leaves a_2.

We note that for any coupling (X^1, X^2) and any position (d_1, d_2) of this process at time s, if X can jump with a positive probability from d_1 to d_3, then there is a positive probability that the first jump of X^1 after time s will take it to d_3, and, moreover, the jump of X^1 will occur before or at the same time when X^2 makes its first jump after time s. Thus, there is a positive probability that X^1 will jump to a_1 at some time $t_2 > t_1$, but X^2 will not jump within interval $(t_1, t_2]$ at all or it will have only one jump, at time t_2. Hence, there is positive probability that for some t_2 we have $X^1_{t_2} = a_1$ and $X^2_{t_2} = b_k$ for some k.

The same argument shows that, with a positive probability, X^1 will jump after time t_2 from a_1 to c_1, c_2 and c_3, at times t_3, t_4 and t_5, while X^2_t will have at most 3 jumps for $t \in (t_2, t_5]$, and, moreover, all of the jumps of X^2 will occur at times t_3, t_4 and/or t_5. It may happen that X^2 hits a_1 or a_3 at some time $t_6 \in (t_2, t_5]$. If this is the case then $X^2_{t_6} \in \{a_1, a_3\}$ while $X^1_{t_6} \in \{c_1, c_2, c_3\}$.

If X^2 does not hit a_1 or a_3 before or at the time t_5 then $X^2_{t_5} \in \{b_1, b_2, b_3\}$ and $X^1_{t_5} = c_3$. In this case, a possible evolution of the process after time t_5 is that X^2 will make one or two jumps that will take this process to either a_1 or a_3, whichever is closer to $X^2_{t_5}$. Let t_7 be the time when X^2 hits a_1 or a_3. With positive probability, X^1 will make at most two jumps during the time interval $(t_5, t_7]$ and so we will have $X^1_{t_7} \in \{c_1, c_2, c_3, c_4, c_5\}$. We see that with positive probability, there is a finite time t_8, equal either to t_6 or t_7, such that $X^2_{t_8} \in \{a_1, a_3\}$ and $X^1_{t_8} \in \{c_1, c_2, c_3, c_4, c_5\}$.

With positive probability, the process X^2 can jump to a_2 at time $t_9 > t_8$, while X^1 will not jump during (t_8, t_9). We will have $X^2_{t_9} = a_2$ and $X^1_{t_9} \in D \setminus \{a_2\}$. Then X^1_t may reach d_0 (the initial position of X^2) at time $t_{10} > t_9$, before X^2 leaves a_2, by the argument presented earlier in the proof. We have shown that with a positive probability, the coupling (X^1, X^2) may go from (a_2, d_0) to (d_0, a_2), before the coupling time.

Next we consider an arbitrary initial position (d_1, d_2) for (X^1, X^2) with $d_1 \neq d_2$. We assume that $d_1, d_2 \in D \setminus \{a_2\}$ because the other case has been discussed in the first part of the proof. With positive probability, the process X^1 may keep jumping in the clockwise direction around $D \setminus \{a_2\}$, while X^2 will jump only at the same times when X^1 jumps. The same remark applies to jumps of X^1 in the counter-clockwise direction; it also applies when we reverse the roles of X^1 and X^2. An elementary argument now shows that for any starting position, one or both processes may hit $\{a_1, a_3\}$ before the coupling time. First suppose that only one of them hits the set $\{a_1, a_3\}$. Then this process may jump to a_2 before or at the same time when the other process jumps. At this instant, one of the processes will be at a_2 while the other will be in $D \setminus \{a_2\}$. The other possibility is that both processes hit $\{a_1, a_3\}$ at the same instant, before the coupling time. Then, since $q(a_3, a_2) > q(a_1, a_2)$, and using equations (2.8)-(2.9), there is a positive chance that the process at the point a_3 will jump to a_2 before the other process jumps to a_2. Hence, just as in the first part of the proof, we will have one of the processes at a_2 and the other at a point of $D \setminus \{a_2\}$. The process which happens to be in $D \setminus \{a_2\}$ may go to any other point of $D \setminus \{a_2\}$ strictly before the other process leaves a_2.

We have proved that for every (d_1, d_2), and every $d_3 \in D \setminus \{a_2\}$, either (a_2, d_3) is accessible from (d_1, d_2) or (d_3, a_2) is accessible from the same point. By the first part of the proof, (a_2, d_3) is accessible from (d_3, a_2) and vice versa, so by transitivity, both (a_2, d_3) and (d_3, a_2) are accessible from (d_1, d_2). In particular, if (d_1, d_2) is accessible from (a_2, d_3) then both (a_2, d_3) and (d_3, a_2) are accessible from (d_1, d_2). In other words, every coupling (X_t^1, X_t^2) has the transposition property relative to a_2. We conclude that no coupling is efficient for this Markov chain, by Theorem 2.6 (i). Since for every (d_1, d_2) with $d_1 \neq d_2$, (a_2, d_3) is accessible from (d_1, d_2), it easily follows that $\mu'(d_1, d_2) < \mu_2$ for every pair of distinct points $d_1, d_2 \in D$.

Remark : We list three open problems inspired by Example 2.12.

1. Give necessary and sufficient conditions in terms of $q(x, y)$ for the existence of an efficient coupling for a continuous-time symmetric Markov process with finite state space.

2. A quantitative version of Problem 1 is the following. Let $\overline{\mu} = \sup \mu^*$, where the supremum is taken over all couplings for a given Markov process. Can we have $\overline{\mu} < \mu_2$? If so, how can one calculate $\mu_2 - \overline{\mu}$ starting from $q(x, y)$?

3. If the state space D has only 3 elements then there exists an efficient coupling (we omit an easy proof). Does an efficient coupling exist for every Markov process whose state space is a loop: namely, the state space is $\{0, 1, \dots, n\}$ for some n and $q(j, k) > 0$ if and only if $|j - k| = 1$ or n? We conjecture that the answer is no if $n \geq 3$ (Mountford and Cranston [37] have now produced a counterexample to the question, as well as discussing many other interesting related questions, so our conjecture is correct.).

As with many coupling problems, it may help to think of this problem in terms of a game. Suppose that player $\mathcal{K}$ begins at state x_1 and player $\mathcal{W}$ begins at a different state x_2. $\mathcal{W}$ can make various moves according to possibilities admitted by the Q-matrix of the Markov process in question, and weighted by the off-diagonal terms of the Q-matrix. $\mathcal{K}$ must predeclare his moves in terms of either what $\mathcal{W}$ might do or electing to move independently, dividing weights between various possibilities so as (a) to add up to the weights prescribed by the off-diagonal terms of the Q-matrix, and (b) such that a move predeclared in terms of a $\mathcal{W}$-move must have weight no greater than that of the $\mathcal{W}$-move. (Thus a viable strategy for $\mathcal{K}$ corresponds to an admissible coupling.) Player $\mathcal{W}$ wins the game if he can choose a sequence of $\mathcal{W}$- and $\mathcal{K}$-moves compatible with $\mathcal{K}$'s predeclared options, and such that at some stage the positions of $\mathcal{W}$ and $\mathcal{K}$ interchange from what they were at a previous occasion. By Theorem 2.6(i), there is no efficient coupling exactly when for some initial point x_1 for $\mathcal{K}$ it is the case that player $\mathcal{W}$ can win the game whatever viable strategy is chosen by $\mathcal{K}$.

3 Reflected Brownian motion in a triangle

We will illustrate the concept of efficiency for Markovian couplings for continuous processes by a detailed analysis of two couplings for reflected Brownian motion in a triangle. We have chosen this example as the role of "partial ordering" is clear in this case. Moreover, our methods developed for this example have laid a foundation for some results on the "hot spots" conjecture of J. Rauch [3, 6, 5]. We remark in passing that not much can be said about the exact values of eigenvalues for reflected Brownian motion in a triangle: see [38, 39] for the equilateral case.

We will discuss "synchronous" and "mirror" couplings. The synchronous couplings have been studied, for example, by Cranston and Le Jan [15, 16]. The mirror couplings seem to be a more effective tool than the synchronous couplings for estimating the spectral gap [48]. We will start with synchronous couplings. Our results are not as complete in this case as in the case of mirror couplings and for this reason this part of the presentation is less technical.

First we will give a construction of the synchronous coupling. Note that the notation is changed from the last section in the following respect: previously X^1 and X^2 denoted copies of X, while in this section they will stand for the components of the two-dimensional process X. The coupling process will consist of X and an identically distributed (but not independent!) copy Y.

Let $\tilde{X} = (\tilde{X}^1, \tilde{X}^2)$ be a 2-dimensional Brownian motion with $\tilde{X}_0 = (x^1, x^2)$ where $x^2 > 0$. Let $L^X_t = -(0 \wedge \min_{s \le t} \tilde{X}^2_s)$. Then the Skorokhod Lemma [28, Lemma 3.6.14] implies that the process X, defined by $X_t = (\tilde{X}^1_t, \tilde{X}^2_t + L^X_t)$, is a reflected Brownian motion in the upper half-plane. Let $\tilde{Y}_t = (\tilde{Y}^1_t, \tilde{Y}^2_t) = (\tilde{X}^1_t + (y^1 - x^1), \tilde{X}^2_t + (y^2 - x^2))$, where $y_2 > 0$. Then $\tilde{Y}$ is a Brownian motion starting from (y^1, y^2). Arguing as before, $L^Y_t = -(0 \wedge \min_{s \le t} \tilde{Y}^2_s)$ can be used to define a reflected Brownian motion Y by means of the formula $Y_t = (\tilde{Y}^1_t, \tilde{Y}^2_t + L^Y_t)$. Let K_t be the straight line passing through the two planar points X_t and Y_t and let $\angle K_t \in (-\pi/2, \pi/2]$ be the angle between K_t and the horizontal axis. (If $X_t = Y_t$ then let K_t be the horizontal line passing through X_t.) If $\angle K_0 = \pi/2$ then K_t will stay perpendicular to the horizontal axis until X_t and Y_t meet at some time u and then $X_t = Y_t$ for $t > u$. If $\angle K_0 \neq \pi/2$ then $\angle K_t$ will converge in a monotone way to 0 as $t \to \infty$ and, moreover, $\angle K_t$ will be constantly equal to 0 after both X_t and Y_t hit the horizontal axis. The pair (X, Y) of reflected Brownian motions will be called a *synchronous coupling*.

The above construction generalizes in a straightforward way to polygonal domains D: given a pair of starting points $x, y \in \overline{D}$ we can define a pair of reflected Brownian motions (X, Y) in D with $(X_0, Y_0) = (x, y)$, and such that $X - Y$ remains constant on every interval during which both processes stay in the interior of the domain. It should be noted that neither of X and Y can hit any vertices of ∂D; this follows either by the results of Varadhan and Williams [47] or indeed by viewing X, Y in polar coordinates centered at each of the finitely many polygonal vertices. Finally it is not hard to prove (for example using Brownian excursion theory based on excursions from the side visited immediately prior to coupling) that with positive probability there will be u such that $X_u = Y_u$ if and only if K_0 is perpendicular to one of the sides of the polygon ∂D or if ∂D contains a perpendicular pair of line segments. If such a u exists then $X_t = Y_t$ for all $t \geq u$.

Chen explains how the spectral gap can be estimated using couplings [10, Theorem 6.2]. We will discuss two concrete implementations of Chen's theorem in the case of reflecting Brownian motion in a triangle.

Let $p(t, x, y)$ be the transition densities for reflecting Brownian motion in D. We have

$$p(t, x, y) = \sum_{n=1}^{\infty} \varphi_n(x)\varphi_n(y)e^{-\mu_n t}, \tag{3.1}$$

where μ_n is the n-th eigenvalue for the Laplacian in D with Neumann boundary conditions and φ_n is the corresponding eigenfunction. Recall that $\mu_1 = 0$ and φ_1 is a constant function. It is a classical result ("Mercer's Theorem") that

$$p(t, x, y) = c_1 + \varphi_2(x)\varphi_2(y)e^{-\mu_2 t} + R(t, x, y), \tag{3.2}$$

where $R(t, x, y)$ converges to 0 faster than $e^{-\mu_2 t}$ as $t \to \infty$ uniformly in x and y (see [3, Proposition 2.1] for a recent proof). To be more precise, the function φ_2 in (3.2) is an eigenfunction corresponding to μ_2 but in the case of eigenvalue multiplicity it may be a linear combination $g(x, y) = \sum_\varphi \varphi(x)\varphi(y)$ of several eigenfunctions in (3.1) corresponding to μ_2, analogous to g in equation (1.1). However for the sake of simplicity, and because the generic case will not exhibit such multiplicity, we consider only the case $g(x, y) = \varphi_2(x)\varphi_2(y)$ in the following. Suppose that one can prove that for some $\mu \geq 0$ and $x, y \in D$,

$$\mathbb{E}(|X_t - Y_t| \mid X_0 = x, Y_0 = y) \leq c(x, y)e^{-\mu t}, \qquad t \geq 0. \tag{3.3}$$

(We choose to consider $\mathbb{E}[|X_t - Y_t| \mid X_0 = x, Y_0 = y]$ rather than $\mathbb{P}[\tau > t \mid X_0 = x, Y_0 = y]$ because typically $\mathbb{P}[\tau > t \mid X_0 = x, Y_0 = y] = 0$ for synchronous couplings: see [16].) One would expect that μ is then a lower bound for μ_2. We examine this assertion in the next example and lemma.

Example 3.1 *Consider the long rectangle $D = [0, 1] \times [0, 100]$ and let $x = (1/4, 1)$, $y = (3/4, 1)$. For these x and y, and the synchronous coupling, (3.3) will hold with $\mu = \pi$. This follows from the fact that the line K_t will always stay parallel to the horizontal axis and so we are effectively dealing with a 1-dimensional Neumann problem on the interval $[0, 1]$ for which π is the second eigenvalue. The second eigenvalue for the Laplacian in D is the same as for the interval $[0, 100]$, i.e., $\mu_2 = \pi/100$. Hence, (3.3) may hold for some $\mu > \mu_2$ and* some $x, y \in D$.

Lemma 3.2 *If $\varphi_2(x) \neq \varphi_2(y)$ and (3.3) holds, then $\mu \leq \mu_2$.*

Proof : Consider $x, y \in D$ such that $\varphi_2(x) > \varphi_2(y)$. Suppose that $X_0 = x$ and $Y_0 = y$. The function φ_2 is not identically equal to 0 so,

$$\int_D \exp(c_1 z_1 + c_2 z_2)\varphi_2(z_1, z_2) dz_1 dz_2 > 0,$$

for some $c_1, c_2 \in \mathbb{R} \setminus \{0\}$. Since D is bounded there exists $c_3 > 0$ such that c_3^{-1} is a Lipschitz constant for $\exp(c_1 x^1 + c_2 x^2)$, and so

$$\mathbb{E}[|X_t - Y_t|] \geq c_3 \Big[\mathbb{E}[\exp(c_1 X_t^1 + c_2 X_t^2)] - \mathbb{E}[\exp(c_1 Y_t^1 + c_2 Y_t^2)] \Big].$$

Then, by equation (3.2),

$$\begin{aligned}
& \mathbb{E}(|X_t - Y_t| \mid X_0 = x, Y_0 = y) \\
\geq \;& c_3 \Big[\mathbb{E} \exp(c_1 X_t^1 + c_2 X_t^2) - \mathbb{E} \exp(c_1 Y_t^1 + c_2 Y_t^2) \Big] \\
= \;& c_3 \int_D \exp(c_1 z_1 + c_2 z_2) p(t, x, z) dz - c_3 \int_D \exp(c_1 z_1 + c_2 z_2) p(t, y, z) dz \\
= \;& c_3 \int_D [\varphi_2(x) - \varphi_2(y)] \exp(c_1 z_1 + c_2 z_2)\varphi_2(z) e^{-\mu_2 t} dz + R(t, x, y) \\
= \;& c_4(x, y) e^{-\mu_2 t} + R(t, x, y),
\end{aligned}$$

where $c_4(x, y) > 0$. (Note that here $R(t, x, y)$ is actually the integrated sum of two terms of the form of $R(t, x, y)$ in equation (3.2).) Since this estimate and inequality (3.3) hold for arbitrarily large t, we see that $\mu \leq \mu_2$. □

In view of Example 3.1 and Lemma 3.2, we propose the following definition.

Definition 3.3 *We will call a synchronous coupling (X, Y) of reflected Brownian motions in D* efficient *if for some x and y with $\varphi_2(x) \neq \varphi_2(y)$, the estimate (3.3) holds with $\mu = \mu_2$.*

Note that the above definition of efficiency is different from that given in Section 2 for Markov processes with finite state space. We require in Definition 3.3 that (3.3) holds with $\mu = \mu_2$ only for some x and y with $\varphi_2(x) \neq \varphi_2(y)$, *not* for all. This is as opposed to the obvious extension of Definition 2.2, which would require $\mu = \mu_2$ for all x, y with $\varphi_2(x) \neq \varphi_2(y)$. The change of the definition is dictated by technical considerations: the stronger condition seems to be very hard to verify with the exception of some trivial examples.

Theorem 3.4 *If a triangle D has an obtuse angle (i.e., strictly greater than $\pi/2$) then the synchronous coupling for the reflected Brownian motion in D is efficient.*

Proof : This follows the idea of Corollary 2.7 (ii). Suppose that D is an obtuse triangle. We will suppose without loss of generality that the longest side of ∂D lies on the horizontal axis. By convention, the angle $\angle L$ between a straight line L and the horizontal axis will lie in $(-\pi/2, \pi/2]$. Let the angles formed by the sides I_1, I_2 and I_3 of the triangle with the horizontal axis be 0, $\alpha \in (-\pi/2, 0)$ and $\beta \in (0, \pi/2)$.

A remark following [23, Corollary (6.31)] may be applied to the operator $\Delta + \mu_2$. Since this operator has analytic coefficients, the eigenfunctions must be analytic. In particular, an eigenfunction cannot be constant on a non-empty open set unless it is identically equal to 0. Since φ_2 is not constant, we can find $x, y \in D$ (with $\varphi_2(x) \neq \varphi_2(y)$) such that the angle $\angle K$ between the line K passing through these points and the horizontal axis belongs to (α, β). Moreover, we choose the points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ so that $x_1 < y_1$.

Let (X, Y) be a synchronous coupling of reflecting Brownian motions in D which starts from (x, y). Let K_t be the line passing through X_t and Y_t. It follows easily from the construction of the synchronous coupling at the beginning of the section that $\angle K_t$ will monotonically move towards α, as long as one of the processes is reflecting on the side I_2. Likewise, $\angle K_t$ can monotonically approach 0 or β, depending on the side where the reflection is taking place. We conclude that $\angle K_t$ will stay within $[\alpha, \beta]$ for all t. This part of the proof uses the obtuse property of the triangle in a crucial way.

Since $\angle K_t \in [\alpha, \beta] \subset (-\pi/2, \pi/2)$, we have $|X_t - Y_t| \leq k(Y_t^1 - X_t^1)$ for some $k < \infty$. By equation (3.2),

$$
\begin{aligned}
& \mathbb{E}(|X_t - Y_t| \mid X_0 = x, Y_0 = y) \\
\leq \quad & k\,\mathbb{E}(Y_t^1 \mid X_0 = x, Y_0 = y) - k\,\mathbb{E}(X_t^1 \mid X_0 = x, Y_0 = y) \\
= \quad & k\left(\int_D p(t, y, z) z_1 dz - \int_D p(t, x, z) z_1 dz\right) \\
= \quad & k\left(\int_D [\varphi_2(x) - \varphi_2(y)] z_1 \varphi_2(z) e^{-\mu_2 t} dz\right) + R(t, x, y) \\
\leq \quad & c_1(x, y) e^{-\mu_2 t}\,.
\end{aligned}
$$

It follows that inequality (3.3) holds with $\mu = \mu_2$, and so the coupling is efficient. □

Remark : We will prove in Theorem 3.7 below that the mirror coupling is not efficient for reflected Brownian motion in a triangle if all its angles are acute (smaller than $\pi/2$) and distinct. One may ask if the same is true for the synchronous coupling. We presently do not know the answer although we guess that synchronous couplings are also inefficient for triangles with acute angles. We will outline below an argument which shows that the synchronous coupling has a property similar to the "transposition property" discussed in Section 2, which is the basis for the proofs of inefficiency in Theorems 2.6 (i) and 3.7 (ii). Then we will show why this property alone is not sufficient to complete the proof of inefficiency.

Assume that all angles of the triangle D are less than $\pi/2$. Let us consider two possible scenarios for motions of the line K_t passing through X_t and Y_t. Suppose that $|X_0 - Y_0|$ is small and both starting points are close to the center of D. In the first scenario the particles X_t and Y_t move around each other by the angle π during a time interval (t_1, t_4). This happens thus: one of the processes first reflects on the side I_1 so that K_t becomes parallel to this side at time t_1, then one reflects on I_2 until K_t becomes parallel to I_2 at time t_2, then one reflects on I_3 until K_t becomes parallel to I_3 at time t_3, and finally one reflects on I_1, until K_t again becomes parallel to I_1 at time t_4. From elementary geometry and the acuteness of all angles of D it follows that $X^1_{t_1} - Y^1_{t_1}$ and $X^1_{t_4} - Y^1_{t_4}$ have opposite signs. Let α_j be the angle of D opposite to I_j. Then trigonometry and the synchronized property of the coupling combine to show that

$$|X_{t_2} - Y_{t_2}| \quad = \quad \cos\alpha_3 \; |X^1_{t_1} - Y^1_{t_1}|.$$

By repeating this remark we see that

$$|X_{t_4} - Y_{t_4}| \quad = \quad \cos\alpha_2 \cos\alpha_1 \cos\alpha_3 \; |X_{t_1} - Y_{t_1}|.$$

In the second scenario the particles do not revolve around each other. The processes instead reflect on I_1 until K_t is parallel to I_1 at time t_5, then they reflect on I_2, and finally on I_1, so that K_t is again parallel to I_1 at time t_6. Then $X^1_{t_5} - Y^1_{t_5}$ and $X^1_{t_6} - Y^1_{t_6}$ will have the same sign and

$$|X_{t_6} - Y_{t_6}| = \cos\alpha_3 \cos\alpha_3 \; |X_{t_5} - Y_{t_5}|.$$

Now suppose that the angles α_j are such that for some integers n and m (n odd) we have

$$(\cos\alpha_2 \cos\alpha_1 \cos\alpha_3)^n \quad = \quad (\cos\alpha_3)^{2m}.$$

The family of triplets $(\alpha_1, \alpha_2, \alpha_3)$ with this property is dense in the set of all possible acute angles with $\alpha_1+\alpha_2+\alpha_3 = \pi$. To show this amounts to considering angles for which the quantity

$$\frac{1}{2}\left(1 + \frac{\cos(\alpha_1 - \alpha_2)}{\cos\alpha_3}\right)^n / (\cos\alpha_3)^{2m-n}$$

is equal to 1: the required density statement follows by noting that x^n/y^{2n-m} is dense in $(0,\infty)$ for $x, y \in (0,1)$ with x/y irrational.

If the process (X, Y) repeats n times the motions described in the first scenario, then for some t_7,

$$|X_{t_7} - Y_{t_7}| = (\cos\alpha_2 \cos\alpha_1 \cos\alpha_3)^n \; |X_{t_1} - Y_{t_1}|,$$

and $X^1_{t_7} - Y^1_{t_7}$ and $X^1_{t_1} - Y^1_{t_1}$ will have opposite sign (since n is odd). If the second scenario is repeated m times then for some t_8,

$$|X_{t_8} - Y_{t_8}| = (\cos\alpha_3)^{2m} \; |X_{t_5} - Y_{t_5}| = (\cos\alpha_2 \cos\alpha_1 \cos\alpha_3)^n \; |X_{t_5} - Y_{t_5}|,$$

and $X^1_{t_8} - Y^1_{t_8}$ and $X^1_{t_5} - Y^1_{t_5}$ will have the same sign. Hence, under the two scenarios the moving particles may become parallel to I_1 and at the same distance from each other, but with their "order" reversed. Since the densities for the processes (X_t, Y_t) and (Y_t, X_t) are both continuous, the corresponding distributions must be mutually absolutely continuous on a part of the state space. This is a version of the "transposition property" used in our proofs of inefficiency.

Next we indicate why it is difficult to derive inefficiency of the synchronous coupling directly from the "transposition property." Suppose that the points x and y belong to D, the line passing through them is parallel to I_1, and $X_0 = x, Y_0 = y$. Consider the event A_u that neither process X_t or Y_t touches I_2 or I_3 before time u. Note that given A_u, we have $|X_u - Y_u| = |X_0 - Y_0|$. Hence,

$$\mathbb{E}(|X_u - Y_u| \mid X_0 = x, Y_0 = y) \geq |X_0 - Y_0|\, \mathbb{P}(A_u \mid X_0 = x, Y_0 = y).$$

If the points x and y are very close to each other, the event A_u is "almost identical" to the event A^1_u that X_t does not hit I_2 or I_3 before time u. For large u, the probability of A^1_u is well approximated by $c(x,y)\exp(-u\mu^*)$, where μ^* is the first eigenvalue for the Laplacian in D with Neumann boundary conditions on I_1 and Dirichlet conditions on $I_2 \cup I_3$. Hence, we have a heuristic estimate $\mathbb{P}(A_u) \approx \mathbb{P}(A^1_u) \approx c\exp(-u\mu^*)$. It is conceivable that $|X_0 - Y_0|\,\mathbb{P}(A_u)$ is the main contribution to $\mathbb{E}(|X_u - Y_u|)$, since the length of the vector $X_t - Y_t$ is shortened exponentially fast over long intervals of time owing to reflection of the processes X_t and Y_t on ∂D. If this is the case then we might have $\mathbb{E}(|X_u - Y_u|) \approx c\exp(-u\mu^*)$, for large u. So now the question is whether μ^* might be equal to the second Neumann eigenvalue μ_2. This is rather doubtful but at present we do not know how to prove that the two eigenvalues are different. It should be noted that for an arbitrary convex planar domain and arbitrary division of the boundary into the "Neumann" and "Dirichlet" parts, the first mixed eigenvalue can be smaller than, equal to or larger than the second Neumann eigenvalue. Hence, there is no general principle that would show that $\mu^* \neq \mu_2$.

For the remaining part of the section, we switch our attention to the "mirror" coupling for reflected Brownian motion in planar domains. The mirror coupling seems to be the most natural coupling for diffusions in $\mathbb{R}^d$ and for reflected Brownian motion in particular. One feels that the mirror coupling is optimal from the point of view of efficiency but we do not have any rigorous results to this effect. See [10, Theorem 5.3] for results on other versions of optimality for couplings of diffusions.

Mirror couplings for reflected processes have been constructed in [29]. We will present a new construction of mirror couplings which is particularly well suited for the study of those of its properties which are important in this paper. We will start with the discussion of the mirror coupling in very simple domains and then (in Section 4) we will progress towards more complicated domains.

First we discuss the mirror coupling for free Brownian motions in $\mathbb{R}^2$. Suppose that $x, y \in \mathbb{R}^2$, $x \neq y$, and that x and y are symmetric with respect to a line M. Let X be a Brownian motion starting from x and let τ be the first time t with $X_t \in M$. Then we let Y be the mirror image of X with respect to

M for $t \leq \tau$, and we let $Y_t = X_t$ for $t > \tau$. The process Y is a Brownian motion starting from y and (X, Y) is called the *mirror coupling* for (non-reflecting) Brownian motion.

We start the discussion of the mirror coupling for reflected Brownian motions with the simplest case, that of a half-plane D. Suppose $x, y \in D$ and let M be the line of symmetry for x and y. The case when M is parallel to ∂D can be easily handled using Skorokhod's lemma [28, Lemma 3.6.14], so we focus on the case when M intersects ∂D. By performing rotation and translation, if necessary, we may suppose that D is the upper halfplane. Let h be the point of intersection between the boundary ∂D and the line of symmetry M. We write $x = (r^x, \theta^x)$ and $y = (r^y, \theta^y)$ in polar coordinates based on h. The points x and y are at the same distance from h so $r^x = r^y$. Suppose without loss of generality that $\theta^x < \theta^y$. We first generate a 2-dimensional Bessel process R_t starting from r^x. Then we generate two coupled one-dimensional processes on the "half-circle" as follows. Let W be a 1-dimensional Brownian motion starting from 0. We construct $\tilde{\Theta}^x$ as the reflected Brownian motion in $[0, \pi]$ started at θ^x, solving the Skorokhod equation

$$\tilde{\Theta}^x_t \quad = \quad \theta^x + W_t + L^o(\tilde{\Theta}^x)_t - L^\pi(\tilde{\Theta}^x)_t$$

where $L^o(\tilde{\Theta}^x)$, $L^\pi(\tilde{\Theta}^x)$ are the local time "pushes" for $\tilde{\Theta}^x$ at 0 and π (the minimal increasing processes required to keep $\tilde{\Theta}^x$ nonnegative and no greater than π). We construct $\hat{\Theta}^y$ similarly but using a mirror-reflected driving Brownian motion:

$$\hat{\Theta}^y_t \quad = \quad \theta^y - W_t + L^o(\hat{\Theta}^y)_t - L^\pi(\hat{\Theta}^y)_t \,.$$

These reflecting Brownian motions have to be time-changed in order to serve as the angular parts of reflected Brownian motion in D: fortunately we can use the same time-change in each case, namely

$$\sigma(t) \quad = \quad \int_0^t R_s^{-2}\, ds \,.$$

We set

$$X_t \quad = \quad \left(R_t, \tilde{\Theta}^x_{\sigma(t)}\right), \qquad Y_t \quad = \quad \left(R_t, \hat{\Theta}^y_{\sigma(t)}\right) .$$

This is a generalization of the skew-product representation of planar Brownian motion, as described in [27]. Both X and Y can be viewed as obtained from *free* Brownian motions using reflection in the boundary of D. Indeed the processes X_t and Y_t behave like free Brownian motions coupled by the mirror coupling as long as they are both strictly inside D. The processes will stay together after the first time they meet. The pair (X_t, Y_t) will be called the mirror coupling for reflected Brownian motions in a half-plane.

The line of symmetry for X_t and Y_t will be denoted M_t if $X_t \neq Y_t$. For definiteness, we let M_t be the horizontal line passing through X_t if $X_t = Y_t$.

The most important property of the above coupling is that by construction the distances of X_t and Y_t from h remain equal to each other as time t varies. This property manifests itself in more general domains in the following way.

First of all, suppose that D is an arbitrary halfplane, and x and y belong to D. Let M_t be the line of symmetry for X_t and Y_t, constructed as above as a mirror coupling of reflected Brownian motion begun at x, y respectively. Suppose that M_0 intersects ∂D. Then for every t, the distance from X_t to $M_t \cap \partial D$ is the same as for Y_t. Note that M_t may move, but only in a continuous way, while the point $M_t \cap \partial D$ will never move. We will call M_t the *mirror* and the intersection point h of M_t and ∂D will be called the *hinge*. The absolute value of the angle between the mirror and the normal vector to ∂D at h can only decrease; thus if M_t is parallel to ∂D then it will stay parallel to ∂D until the coupling time. In this case, M_t can move only away from ∂D and only in a continuous fashion.

The mirror coupling of reflected Brownian motions in a convex polygonal domain D can be described as follows. Suppose that X_t and Y_t start from x and y inside the domain D. As soon as one of the particles hits a side I of ∂D, the processes will evolve according to the coupling described in the previous paragraph. To be more precise, let K be the straight line containing I where I is the side of ∂D most recently hit by one of the particles. Since the process which hits I does not "feel" the shape of D except for the direction of I, it follows that X_t and Y_t will remain at the same distance from the hinge $\{h_t\} = M_t \cap K$, as long as the particles do not hit a side different from I. The mirror M_t can move but the hinge h_t will remain constant as long as I remains the side of ∂D where the reflection takes place. The hinge h_t will jump when the reflection location moves from I to another side of ∂D. Since D is convex, h_t will be always on ∂D or outside D.

Remark : We remark in passing a point of methodological interest: this representation was first discovered by accident as we explored the system of mirror-coupled reflecting Brownian motions using computer algebra, specifically the implementation *Itovsn3* of stochastic calculus in the computer algebra package *REDUCE*. For details (as implemented in the *Mathematica* version of *Itovsn3*) see the *Mathematica* notebook `reflect` in [30]. Of course with hindsight the properties mentioned above now appear obvious ....

Recall that $p(t, x, y)$ denotes the transition densities for reflecting Brownian motion in the triangle D and recall from (3.2) the following one-term eigenfunction expansion for $p(t, x, y)$,

$$p(t, x, y) = c_1 + \varphi_2(x)\varphi_2(y)e^{-\mu_2 t} + R(t, x, y), \tag{3.4}$$

where $R(t, x, y)$ converges to 0 faster than $e^{-\mu_2 t}$ as $t \to \infty$. The coupling time is denoted by τ. Suppose that one can prove that for some $\mu \geq 0$ and $x, y \in D$,

$$\mathbb{P}(\tau > t \mid X_0 = x, Y_0 = y) \leq c(x, y)e^{-\mu t}, \qquad t \geq 0. \tag{3.5}$$

It is reasonable to expect that μ is then a lower bound for μ_2. However, Example 3.1 applies to the mirror coupling as well and we see that (3.5) may hold for some $\mu > \mu_2$ and some $x, y \in D$. The following lemma is entirely analogous to Lemma 3.2.

Lemma 3.5 *If $\varphi_2(x) \neq \varphi_2(y)$ and (3.5) holds, then $\mu \leq \mu_2$.*

Proof : Consider $x, y \in D$ such that $\varphi_2(x) \neq \varphi_2(y)$. Let $A = \{v \in D : \varphi_2(v) > 0\}$ and note that A must have positive measure. By (3.4),

$$\begin{aligned}\int_A p(t,x,z)dz - \int_A p(t,y,z)dz \quad &= \textstyle\int_A [\varphi_2(x) - \varphi_2(y)]\varphi_2(z)e^{-\mu_2 t}dz + R(t,x,y) \\ &= c_1(x,y)e^{-\mu_2 t} + R(t,x,y)\,,\end{aligned}$$

where $c_1(x,y) \neq 0$. On the other hand,

$$\begin{aligned}&\left|\int_A p(t,x,z)dz - \int_A p(t,y,z)dz\right| \\ &= \quad |\,\mathbb{P}(X_t \in A \mid X_0 = x) - \mathbb{P}(Y_t \in A \mid Y_0 = y)| \\ &= \quad |\,\mathbb{P}(X_t \in A, t < \tau \mid X_0 = x) - \mathbb{P}(Y_t \in A, t < \tau \mid Y_0 = y)| \\ &\leq \quad \mathbb{P}(t < \tau \mid X_0 = x, Y_0 = y) \quad \leq \quad c_2(x,y)e^{-\mu t}\,.\end{aligned}$$

Since this inequality and (3.6) hold for arbitrarily large t, we see that $\mu \leq \mu_2$. □

Definition 3.6 *A mirror coupling (X, Y) of reflected Brownian motions in D is said to be* efficient *if the estimate (3.5) holds with $\mu = \mu_2$ for some x and y with $\varphi_2(x) \neq \varphi_2(y)$.*

Our main theorem for mirror coupling in triangles identifies the cases of inefficiency and efficiency for this coupling in simple geometric terms.

Theorem 3.7

(i) If a triangle D has an obtuse angle (which is to say, strictly greater than $\pi/2$) then the mirror coupling for the reflected Brownian motion in D is efficient.

(ii) If all angles of the triangle D are distinct from each other and acute (which is to say, strictly less than $\pi/2$) then the mirror coupling for the reflected Brownian motion in D is not efficient.

Remark :

(i) Note that in Theorem 3.7 (ii) we assume that all angles of D are distinct. This technical assumption is probably unnecessary, but would be tedious to lift.

(ii) Example 2.12 and Theorem 3.7 (ii) naturally lead to the following open question: Are there no efficient Markovian couplings for reflected Brownian motion in generic acute triangles?

Proof of Theorem 3.7 (i): Suppose that an obtuse triangle D is oriented so that its longest side lies on the horizontal axis, its leftmost vertex is at the origin and the triangle is contained in the first quadrant (see Fig. 3). The angle formed by any straight line K with the horizontal axis will be denoted $\angle K$. Let the angles formed by the sides I_2 and I_1 of the triangle D with the horizontal axis be $\alpha \in (-\pi/2, 0)$ and $\beta \in (0, \pi/2)$ (see Fig. 3).

[Figure 3 about here.]

Fix any two points $x = (x_1, x_2) \in D$ and $y = (y_1, y_2) \in D$, such that $x_1 < y_1$ and $\angle M \in \mathcal{A} = (\pi/2 + \alpha, \pi/2 + \beta)$, where M is the line of symmetry for x and y. Consider a mirror coupling (X_t, Y_t) with $X_0 = x$, $Y_0 = y$, and recall that M_t denotes the mirror, i.e., the line of symmetry for X_t and Y_t. Since $M_0 = M$ we have $\angle M_0 \in \mathcal{A}$. We will argue that $\angle M_t$ will not leave the interval $\mathcal{A}$ until the coupling time. Let K_j denote the straight line containing the side I_j of the triangle. Suppose that for some s, the angle $\angle M_s$ is within this interval and the hinge h_t belongs to K_3 for $t \in (s, u)$. Then $|\pi/2 - \angle M_t|$ will be decreasing on the interval (s, u), and so $\angle M_t \in \mathcal{A}$ for all $t \in (s, u)$. Next consider the case when $\angle M_s \in \mathcal{A}$ and $h_t \in K_1$ for $t \in (s, u)$. Then $|(\pi/2 + \beta) - M_t|$ is decreasing for $t \in (s, u)$ and so $\angle M_t$ must stay in $\mathcal{A}$ for $t \in (s, u)$. The final case, when the hinge belongs to K_2, may be treated in the same way. We have shown that $\angle M_t$ does not leave $\mathcal{A}$ before the coupling time.

By acuteness of the triangle D, the interval $\mathcal{A}$ lies strictly inside $(0, \pi)$, so there exists $k > 0$ such that $Y_t^1 - X_t^1 \geq k|Y_t - X_t|$ for $t < \tau$ (and so for all t). We will now analyze the distance between X_t and Y_t. Up to the coupling time, the process $\rho = |Y - X|$ is a one-dimensional Brownian motion with twice the standard variance as long as both X and Y are strictly inside D. When one of the processes X or Y is reflecting on ∂D, then ρ gets a "push" determined by the local time spent by X or Y on ∂D and by the direction of M relative to the reflecting side of D. Since D is convex, the direction of the push for ρ always points towards 0. This shows that for any ρ, the hitting distribution of 0 for the process ρ starting from ρ_0 is stochastically majorized by the hitting distribution of 0 for the one-dimensional Brownian motion with twice the standard variance and starting from ρ_0. Hence, we may fix arbitrarily small $p_0 > 0$ and find $\hat{\rho} > 0$ such that if $\rho_t \leq \hat{\rho}$ (but the positions of X_t and Y_t are otherwise arbitrary) then $\mathbb{P}(\rho_{t+1} > 0) < p_0$. Choose p_0 such that $(2p_0)^j < e^{-2\mu_2 j}$, and find a corresponding $\hat{\rho} > 0$ with $\hat{\rho} < |x - y|$.

Recall from the proof of Theorem 3.4 that an eigenfunction must be analytic. In particular, an eigenfunction cannot vanish on a non-empty open set unless it is identically equal to 0. Fix any $x \in D$ and find $y \in D$ such that $\varphi_2(x) \neq \varphi_2(y)$ and $\angle M \in (\pi/2 + \alpha, \pi/2 + \beta)$, where M is the line of symmetry for x and y. Such a point y must exist because otherwise φ_2 would be constant, and, therefore, it would vanish, on a non-empty open set inside D, which is impossible.

Consider t such that

$$\mathbb{P}(\rho_t > \hat{\rho} \mid t < \tau, X_0 = x, Y_0 = y) \quad \geq \quad p_0. \tag{3.6}$$

Recall that $Y^1_t - X^1_t \geq k|Y_t - X_t| = k\rho_t$ for $t < \tau$. This and (3.6) show that

$$\mathbb{P}(Y^1_t - X^1_t > k\hat{\rho} \mid t < \tau, X_t = x, Y_t = y) \quad \geq \quad p_0. \tag{3.7}$$

Since $Y^1_t \geq X^1_t$ for all t, (3.7) implies that

$$\mathbb{E}(Y^1_t \mid t < \tau, X_0 = x, Y_0 = y) \quad \geq \quad \mathbb{E}(X^1_t \mid t < \tau, X_0 = x, Y_0 = y) + p_0 k\hat{\rho},$$

and therefore we have

$$\begin{aligned} &\mathbb{E}(Y^1_t \mid X_0 = x, Y_0 = y) - \mathbb{E}(X^1_t \mid X_0 = x, Y_0 = y) \\ &\quad\geq \quad p_0 k\hat{\rho}\, \mathbb{P}(t < \tau \mid X_0 = x, Y_0 = y)\,. \end{aligned} \tag{3.8}$$

By (3.4),

$$\begin{aligned} &\mathbb{E}\left[Y^1_t \mid X_0 = x, Y_0 = y\right] - \mathbb{E}\left[X^1_t \mid X_0 = x, Y_0 = y\right] \\ &= \int_D p(t, y, z) z_1 dz - \int_D p(t, x, z) z_1 dz \\ &= \int_D [\varphi_2(x) - \varphi_2(y)] z_1 \varphi_2(z) e^{-\mu_2 t} dz + R(t, x, y)\,. \end{aligned}$$

It follows from this and inequality (3.8) that

$$\mathbb{P}(t < \tau \mid X_0 = x, Y_0 = y) \quad \leq \quad c(x, y) e^{-\mu_2 t}. \tag{3.9}$$

Next we consider a t for which inequality (3.6) fails. Let s be the supremum of times less than t for which (3.6) holds. Let j_0 and j_1 be the smallest and largest integers in (s, t). If there are no such j_0, j_1 then

$$\begin{aligned} \mathbb{P}[t \leq \tau \mid X_0 = x, Y_0 = y] \quad &\leq \quad \mathbb{P}[s \leq \tau \mid X_0 = 0, Y_0 = y] \\ &\leq \quad c(x, y) e^{-\mu_2 s} \quad \leq \quad c'(x, y) e^{-\mu_2 t} \end{aligned}$$

On the other hand if ℓ is an integer in $[j_0, j_1 - 1]$, then by the definition of s, (3.6) fails for ℓ. If there is no coupling by time ℓ and $\rho_\ell \leq \hat{\rho}$ then the probability of no coupling by time $\ell + 1$ is less than p_0. This and the failure of inequality (3.6) at time ℓ imply that

$$\begin{aligned} \mathbb{P}(\ell + 1 < \tau \mid X_0 = x, Y_0 = y) \quad &\leq \\ (p_0 + \mathbb{P}[\rho_\ell > \hat{\rho} \mid \ell < \tau, X_0 = x, Y_0 = y]) \quad &\times \quad \mathbb{P}[\ell < \tau \mid X_0 = x, Y_0 = 0] \\ &\leq \quad 2p_0\, \mathbb{P}(\ell < \tau \mid X_0 = x, Y_0 = y). \end{aligned}$$

Thus, applying (3.9) to s,

$$\begin{aligned} \mathbb{P}(t < \tau \mid X_0 = x, Y_0 = y) \quad &\leq \quad \mathbb{P}(j_1 < \tau \mid X_0 = x, Y_0 = y) \\ &\leq \quad (2p_0)^{j_1 - j_0}\, \mathbb{P}(j_0 < \tau \mid X_0 = x, Y_0 = y) \\ &\leq \quad (2p_0)^{t - s - 2}\, \mathbb{P}(s < \tau \mid X_0 = x, Y_0 = y) \\ &\leq \quad e^{-2\mu_2 (t - s - 2)} c(x, y) e^{-\mu_2 s} \leq c_1(x, y) e^{-\mu_2 t}\,. \end{aligned}$$

We see that (3.9) extends to all $t \geq 0$. Since we have chosen x and y with $\varphi_2(x) \neq \varphi_2(y)$, we conclude that the mirror coupling is efficient in obtuse triangles. $\square$

We defer the proof of Theorem 3.7 (ii) till we have proved several subsidiary lemmas.

Let $\mathcal{D} = \{(x, y) \in \overline{D} \times \overline{D} : x \neq y\}$, $\mathcal{D}(\varepsilon) = \{(x, y) \in \overline{D} \times \overline{D} : |x-y| \geq \varepsilon\}$ and $\hat{\mathcal{D}}(\varepsilon) = \mathcal{D} \setminus \mathcal{D}(\varepsilon)$. In an abuse of notation we use $T(A)$ to denote the hitting time of A for any process, including for example X and (X, Y). Sometimes the notation will record the process as well, as in $T_X(A)$. We work under the hypotheses of Theorem 3.7 (ii); the domain is an acute-angled triangle all of whose angles are different from each other.

Lemma 3.8 *For sufficiently small $a \in (0, diam(D)/10)$ there exist $s, c_1 > 0$, $\mathcal{D}_1 \subset \mathcal{D}(a)$, and a (probability) measure ν on $\mathcal{D}_1$ with $\nu(\mathcal{D}_1) > 0$, such that for all $(x, y) \in \mathcal{D}(a)$ and for every subset A of $\mathcal{D}_1$ we have*

$$\mathbb{P}((X_s, Y_s) \in A \mid X_0 = x, Y_0 = y) \quad > \quad c_1 \nu(A).$$

Proof : Most of the proof is concerned with a description of "all possible" trajectories of (X, Y), before the coupling time. Our description will be partly given in terms of possible motions of the mirror process M and will be partly qualitative in nature. We are interested in all trajectory-related events of positive probability, no matter how small that probability might be.

We will say that a positive measure ν_1 is a component of a (probability) measure ν_2 if $\nu_1(A) \leq \nu_2(A)$ for all A.

Suppose that $X_0 = x, Y_0 = y$, $x, y \in D$, $x \neq y$. Suppose further that B_X and B_Y are non-empty open subsets of D which are mirror-images of each other with respect to M_0 and such that B_X lies totally on the same side of M_0 as X_0. Then it is easy to see that the coupling process (X, Y) may reach $B_X \times B_Y$ without touching the boundary of D, and moreover this can happen in an arbitrarily short time. In particular X, Y can come arbitrarily close to any one of the points of intersection of the initial position M_0 of the mirror with ∂D, before the mirror M has first moved.

Now fix one of the points in $M_0 \cap \partial D$. Call this point h and assume that h is not a vertex of the triangle D. Let $\theta_M(t)$ be the angle between M_t and the side I_1 containing h. We will argue that if X_0 and Y_0 are close to h (if they are not, they can move close to h, by our previous remarks), then the mirror can turn around h in the direction towards the normal (i.e, $\theta_M(t)$ will monotonically move towards $\pi/2$), and for each $t > 0$, the angle between M_t and I_1 is a random variable which has a non-trivial atom at $\pi/2$ and a component with a strictly positive continuous density on $(\theta_M(0), \pi/2)$ (or $(\pi/2, \theta_M(0))$). We will show that all this may happen before X and Y leave a small neighborhood of h, and so the hinge, i.e., the point of intersection of the mirror M with ∂D around which the mirror is turning, will remain fixed at h.

The next part of our argument will be quantitative in nature; note that we actually prove more than is strictly needed in this lemma.

Suppose that $X_0 = x$ and $Y_0 = y$, $|x-y| = \rho$, and at least one of the points x or y is at distance no more than ρ from I_1. We moreover assume that the distance from x to each one of the other sides of ∂D is greater than 10ρ, likewise for y. Consider a polar coordinate system (r, θ) based on h as origin, such that I_1 lies on the line $\{(r, \theta) :\ \text{either } \theta = 0 \text{ or } \theta = \pi\}$. Without loss of generality assume that the distance from x to I_1 is not smaller than that for y, that $\pi/2 \geq \theta_y > 0$, that $\theta_x > \theta_y$, and that $\pi - \theta_x > \theta_y$, where $x = (r_x, \theta_x)$ and $y = (r_y, \theta_y)$. We will write $X_t = (r_X(t), \theta_X(t))$, and use corresponding polar coordinates for Y. We argue geometrically, considering the ray $\{(r, \theta) : \theta = \theta_M(t)\}$ defining the mirror process M for X and Y. We will consider two cases. The first case (C1) is when $\theta_x \leq 3\pi/4$. Let $b_1 > 1$ be defined by

$$b_1^2 \quad = \quad \inf_{\beta \in (0, 3\pi/4)} \frac{\sin(7\beta/12)}{\sin(\beta/2)}.$$

We define a ray segment

$$Q_1 \quad = \quad \{(r, \theta) : r \geq \max(r_x/b_1, r_x - \rho), \theta = 7\theta_x/6\}$$

and the rest of a 4-sided curvilinear domain

$$\begin{aligned} Q_2 \quad = \quad & \{(r, \theta) : r \geq r_x/b_1, \theta = 4\theta_x/5\} \cup \\ & \cup \{(r, \theta) : r = \max(r_x/b_1, r_x - \rho) \text{ or } r = r_x + \rho\}. \end{aligned}$$

By scaling and the effect of the assumption (C1) on possible locations of x, y (and the monotonic effect of reflection in the boundary) it follows that there exists $p_1 > 0$ such that

$$\mathbb{P}(T_X(Q_1) < T_X(Q_2) \mid X_0 = x, Y_0 = y) \quad > \quad p_1.$$

Elementary trigonometry can be used to show that $\theta_M(0) < 4\theta_x/5$, in view of the assumptions that y is closer to I_1 than x, and not further from I_1 than ρ, and that $|x - y| = \rho$. We have

$$\frac{\rho}{2r_x} = \frac{|x-y|/2}{r_x} = \sin((\theta_x - \theta_y)/2).$$

We also have $\theta_M(t) \leq \max(\theta_M(0), 7\theta_x/12) < 4\theta_x/5$ as long as $\theta_X(t) \leq 7\theta_x/6$. This gives control over the coupling time τ: the event $\{T_X(Q_1) < T_X(Q_2)\}$ implies $\{T_X(Q_1) < \tau\}$. The fact that $\theta_X(t)$ reaches the level $7\theta_x/6$ for the first time when $t = T_X(Q_1)$ implies that $\theta_Y(T_X(Q_1)) \leq \theta_y$. Using the fact that Y is a reflection of Y modified by reflection in I_1,

$$\begin{aligned} \frac{|X(T_X(Q_1)) - Y(T_X(Q_1))|/2}{r_X(T_X(Q_1))} \quad &= \quad \sin((\theta_X(T_X(Q_1)) - \theta_Y(T_X(Q_1)))/2) \\ \geq \sin((7\theta_x/6 - (\theta_x + \theta_y)/2) \quad &= \quad \sin(\theta_x/12 + (\theta_x - \theta_y)/2) \\ \geq \sin((\theta_x - \theta_y)/12 + (\theta_x - \theta_y)/2) \quad &= \quad \sin(7(\theta_x - \theta_y)/12). \end{aligned}$$

Using the construction of Q_1 and recalling the definition of b_1,

$$\begin{aligned} |X(T_X(Q_1)) - Y(T_X(Q_1))| &\geq 2r_X(T_X(Q_1))\sin(7(\theta_x - \theta_y)/12) \\ \geq 2(r_x/b_1)\sin(7(\theta_x - \theta_y)/12) &\geq 2r_x b_1 \sin((\theta_x - \theta_y)/2) \\ &\geq 2r_x b_1 \frac{\rho}{2r_x} = b_1\rho\,. \end{aligned}$$

We conclude that in case (C1),

$$\mathbb{P}(T(\mathcal{D}(b_1\rho)) < \tau \mid X_0 = x, Y_0 = y) > p_1.$$

The second case (C2) is when $\theta_x \geq 3\pi/4$. Let

$$\begin{aligned} Q_3 &= \{(r,\theta) : \rho/4 < r < \rho, \theta = \pi\}, \\ Q_4 &= \{(r,\theta) : \theta = \pi/2\} \cup \{(r,\theta) : r = 3\rho\} \cup \\ &\qquad \cup\{(r,\theta) : \theta = \pi, r \leq \rho/4 \text{ or } r \geq \rho\}\,. \end{aligned}$$

It is easy to see that there exists $p_2 > 0$ such that for x and y satisfying the assumptions of (C2),

$$\mathbb{P}(T_X(Q_3) < T_X(Q_4) \mid X_0 = x, Y_0 = y) > p_2.$$

Note that $\{T_X(Q_3) < T_X(Q_4)\}$ implies $\{T_X(Q_3) < \tau\}$. If the former event occurs then the mirror for X and Y will be perpendicular to I_1 after time $T_X(Q_3)$ at least as long as X and Y do not leave the ball of radius 7ρ around h.

Using the strong Markov property and repeated application of properties proved in cases (C1) and (C2), we see that the mirror may turn around h towards the normal position while all the time X and Y may stay in a very small neighborhood of h, without coupling.

Next we will prove that the distribution of $\theta_M(t)$ has a component with a continuous density unless $\theta_M(0) = \pi/2$. Recall our current assumptions that $X_0 = x, Y_0 = y$, $x, y \in D$, $x \neq y$. Suppose without loss of generality that x is not closer to I_1 than y. Let A be a closed disc in D, with non-empty interior, not far from h, on the same side of the mirror as x, and such that $x \notin A$. Assume without loss of generality that I_1 lies on the horizontal axis, $h = (0,0)$, and $\theta_M(0) \in (0, \pi/2)$. Then apply the complex analytic transformation $z \to \log z$ to X and to Y, viewing log as a mapping of the upper half-plane to the strip where the imaginary part is between 0 and π. We can make a single random time-change simultaneously converting each of the processes $\log(X)$ and $\log(Y)$ into reflected Brownian motions $\tilde{X}$ and $\tilde{Y}$ respectively. The same time-change works for both the processes since they are always the same distance from h. For the same reason the processes $\tilde{X}$ and $\tilde{Y}$ have the same real parts and they are related by a mirror coupling. Let $\tilde{X}^2$ and $\tilde{Y}^2$ be the imaginary parts of $\tilde{X}$ and $\tilde{Y}$, and let $\tilde{L}$ measure the local time spent by $\tilde{Y}$ on the real axis. Then

$$\frac{\tilde{Y}_t^2 + \tilde{X}_t^2}{2} - \frac{\tilde{Y}_s^2 + \tilde{X}_s^2}{2} = \frac{\tilde{L}_t - \tilde{L}_s}{2}. \tag{3.10}$$

Let $\tilde{S}$ be the hitting time of $\log(A)$ by $\tilde{X}$. For a fixed t we have that $\tilde{L}_{\tilde{S}} = \tilde{L}_t$, with positive probability. For a fixed t we know that $\tilde{L}_t$ has a continuous density, so it follows from equation (3.10) that $(\tilde{Y}^2(\tilde{S}) + \tilde{X}^2(\tilde{S}))/2$ has a component with a continuous density. If S is the hitting time of A by X_t then $\theta_M(S) = (\tilde{Y}^2(\tilde{S})+\tilde{X}^2(\tilde{S}))/2$. This shows that the distribution of $\theta_M(S)$ has a component with a continuous density. For a fixed s, the event $\{L_S = L_s\}$ has a positive probability, so $\theta_M(s)$ has a component with a continuous density. Furthermore it is now not hard to see that the density is strictly positive on $(\theta_M(0), \pi/2)$.

Our argument so far shows that the mirror can turn at either point of intersection with ∂D towards the normal direction, and the angle where it stops before switching the turning point (hinge) is a random variable with an atom at $\pi/2$ and a component with a continuous positive density on the interval between the starting angle and $\pi/2$. All this can happen with positive probability before the coupling time, and, moreover, any turning with a finite sum of all turning angles can be done in an arbitrarily small time, with positive probability.

We have explicitly and implicitly assumed in our arguments that X_0 and Y_0 belong to D. We will briefly discuss what may happen when X_0 or Y_0 belong to ∂D, including, possibly, one of the vertices. We do not assume that X_0 and Y_0 are necessarily close to the mirror M_0. Since X and Y spend zero time on the boundary of D, then with probability 1, there will be arbitrarily small $s > 0$ with $X_s \in D$ and $Y_s \in D$. Once both processes X and Y are strictly inside D, they can move in the way described earlier in the proof.

Next we will use the above results on the possible movements of the mirror to construct a component of the distribution of (X_1, Y_1), for any starting points $x \neq y$ for X and Y.

Let the sides of the triangle D be called I_1, I_2 and I_3. Let us name the vertices of the triangle as follows: $\{z_1\} = I_1 \cap I_2$, $\{z_2\} = I_2 \cap I_3$ and $\{z_3\} = I_3 \cap I_1$. Let $F_j = (B(z_j, \rho_1) \setminus B(z_j, \rho_1/2)) \cap \partial D$, where $\rho_1 > 0$ is chosen so that F_j's have the following property. For every $z \in F_j$, the line K passing through z and perpendicular to ∂D at this point, crosses only the sides of ∂D which are adjacent to z_j, and the points in $K \cap \partial D$ are at least $2\rho_1$ units from the other vertices. Such a ρ_1 exists in view of the fact that all angles of D are acute.

Suppose that M_0 does not pass through any vertex and let h_t^1 and h_t^2 be the points of intersection of M_t with ∂D. Note that the hinge h_t, i.e., the point about which the mirror M_t is turning, is sometimes equal to h_t^1 and sometimes to h_t^2. We can and will choose h_t^1 and h_t^2 so that the functions $t \to h_t^j$ are continuous up to the time when one of the hinges jumps to the third line segment. Suppose that $h_0^1 \in I_1$ and $h_0^2 \in I_2$, and both points are close to z_1; if they are not, the argument requires only minor modifications. It is possible that h_t^1 will not move until time t_1 when M_{t_1} is perpendicular to ∂D at $h_{t_1}^1$. Then h_t^1 will start moving while h_t^2 will remain fixed at the position $h_{t_1}^2$ until time t_2 when M_{t_2} is perpendicular to ∂D at $h_{t_1}^2 = h_{t_2}^2$. The effect of these motions is that $h_{t_2}^1$ is at a greater distance from z_1 than h_0^1. If the same type of motions are repeated again, then h_t^1 will move away from z_1 by even greater distance. Since this process cannot be continued indefinitely, either h_t^1 or h_t^2 must hit one of the vertices z_2 or z_3. Before this happens, either h^1_t or h^2_t must reach F_2 or F_3. Suppose, for example, that h^1_t hits F_3 first. Then it follows from the definition of F_3 that h^2_t may slide along I_2 and reach F_2 while h^1_t remains fixed at a point of F_3. Hence we may have $h^1_t \in F_3$ and $h^2_t \in F_2$, for some t, with positive probability.

Next we may suppose that h^1_t does not move until the mirror is perpendicular to ∂D at this point. The other point, h^2_t, will then move to the side I_3. By repeating the process discussed in the previous paragraph, the points h^1_t and h^2_t may move so that $h^1_t \in F_1$ and $h^2_t \in F_2$. At some future times we may have, in succession, $h^1_t \in F_1$ and $h^2_t \in F_3$, and then, $h^1_t \in F_2$ and $h^2_t \in F_3$.

We now discuss the case when M_0 passes through a vertex. For example, assume that $h^1_0 = z_1$. If the mirror is not perpendicular to the opposite side I_3, then it can turn around the point h^2_0 and it will no longer pass through a vertex. Suppose that the mirror is perpendicular to I_3. We have assumed in Theorem 3.7 (ii) that all angles of D are different, so the angles formed by M_0 with I_1 and I_2 are not equal. It follows that when X_t or Y_t reflect on I_1 or I_2, the mirror will turn around $h^1_0 = z_1$ and it will no longer be perpendicular to I_3.

Consider any point $z \in F_1 \cap I_1$ and an orthonormal coordinate system $CS(z)$ in which I_1 lies on the horizontal axis, z is the origin, and D lies in the upper half-plane. Then choose $\rho_3 > 0$ such that for all $z \in F_1 \cap I_1$ we have (viewed in the $CS(z)$ coordinate system) $\hat{B}(z) = B((3\rho_3, 3\rho_3), 2\rho_3) \subset D$ and $B((-3\rho_3, 3\rho_3), 2\rho_3) \subset D$.

Let $\hat{\theta}_M(t)$ be the angle formed by M_t and the line containing I_1; let $\hat{h}_t$ be the intersection point of M_t and I_1, with the convention that $\hat{h}_t = z_1$ if $M_t \cap I_1 = \emptyset$; and let $\hat{X}_t$ be the position of X_t expressed in terms of the $CS(\hat{h}_t)$ coordinate system.

Our argument has shown that for any starting points x and y for X and Y there is positive probability that at time $t = 1$ the mirror M_1 passes through $F_1 \cap I_1$ and is perpendicular to I_1, and $\hat{X}_1 \in \hat{B}(\hat{h}_1)$. Hence, the event $A_1 = \{\hat{\theta}_M(1) = \pi/2\}$ has a positive probability. Moreover, given the event A_1, $\hat{h}_1$ has a strictly positive density on $F_1 \cap I_1$. Given A_1 and $\hat{h}_1$, the density of $\hat{X}_1$ is strictly positive on $\hat{B}(\hat{h}_1)$.

We sketch a proof, using compactness, that there exist lower and upper bounds for the densities of $\hat{h}_1$ and $\hat{X}_1$, uniform in $x, y \in \mathcal{D}(a)$ for any fixed small $a > 0$. Let $\psi^{x,y}(v)$ be the density of $\hat{h}_1$ restricted to $F_1 \cap I_1$ (the proof for the density of $\hat{X}_1$ is analogous and so it is omitted). Suppose that there exist $v_0 \in F_1 \cap I_1$ and a sequence $(x_k, y_k) \in \mathcal{D}(a)$ such that $\psi^{x_k,y_k}(v_0) \to 0$ as $k \to \infty$. By compactness, we may suppose that $x_k \to x_\infty$ and $y_k \to y_\infty$. Note that we must have $(x_\infty, y_\infty) \in \mathcal{D}(a)$. Now going back to our argument, it is not hard to see that the infimum of $\psi^{x,y}(v_0)$ taken over (x, y) in a neighborhood of (x_∞, y_∞) must be strictly positive. The crucial observation here is that the distance between x_∞ and y_∞ is strictly positive. This gives us the desired contradiction.

We fix some small $a > 0$ and $x_0, y_0 \in \mathcal{D}(a)$ and take ν to be the restriction of the distribution of (X_1, Y_1) to the event $A_1 \cap \{\hat{h}_1 \in F_1 \cap I_1\}$, given $(X_0, Y_0) = (x_0, y_0)$. Chosen in this way, ν satisfies the condition in the lemma. $\square$

Lemma 3.9

(i) There exist $a, \alpha, c_1 > 0$ such that for any $\varepsilon \in (0, a)$ and $(x, y) \in \mathcal{D}(\varepsilon)$, we have

$$\mathbb{P}\left[T(\mathcal{D}(a)) < \tau \mid X_0 = x, Y_0 = y\right] \quad \geq \quad c_1 \varepsilon^{\alpha} .$$

(ii) There exists $c_2 < \infty$ such that for all $\varepsilon > 0$ and $(x, y) \in \hat{\mathcal{D}}(\varepsilon)$,

$$\mathbb{E}\left[T(\hat{\mathcal{D}}^c(\varepsilon)) \mid X_0 = x, Y_0 = y\right] \quad \leq \quad c_2 \varepsilon^2 .$$

Proof :

(i) Let ρ_0 be so small that any disc of radius $100\rho_0$ can intersect at most 2 sides of the triangle D. Let $A(\rho)$ be the event that the process (X, Y) will hit $\mathcal{D}(2\rho)$ before exiting $\mathcal{D}$ and, moreover, this will happen before X or Y move more than 4ρ away from their starting points. For a fixed $\rho \leq \rho_0$, let $p = p(\rho)$ be the infimum of $\mathbb{P}(A(\rho) \mid X_0 = x, Y_0 = y)$, evaluated over all x and y with $|x - y| = \rho$. By the arguments presented in the proof of Lemma 3.8, we know that $p(\rho) > 0$. We will argue that $p(\rho) = p(\rho_1)$ for some $\rho_1 > 0$ and all $\rho < \rho_1$.

Consider a vertex z_1 of the triangle D. Then

$$p_1(\rho) = \inf_{\substack{|x-y|=\rho \\ |x-z_1|\leq 10\rho}} \mathbb{P}(A(\rho) \mid X_0 = x, Y_0 = y)$$

depends only on the angle at the vertex z_1, because neither process X nor Y can hit the side of D opposite to z_1 before moving more than 4ρ units away from its starting point. By scaling, we obtain $p_1(\rho) = p(\rho_1)$ for some $\rho_1 > 0$ and all $\rho < \rho_1$. The same argument applies to the neighborhoods of the other two vertices, and to the points of D more than 9ρ units away from any vertex.

By repeatedly applying the strong Markov property at the hitting times of $\mathcal{D}(2^j \rho)$, we see that

$$\mathbb{P}(T(\mathcal{D}(a)) < \tau \mid X_0 = x, Y_0 = y) \geq c_3 p^k,$$

for $(x, y) \in \mathcal{D}(a/2^k)$. This can be easily rewritten as the estimate in part (i) of the lemma.

(ii) Recall the process $\rho = |X - Y|$ from the proof of Theorem 3.7 (i). It is the sum of Brownian motion (with variance twice the standard one) and a non-increasing process. Let $T_\rho(0)$ be the hitting time of 0 for ρ_t. By comparing ρ with the Brownian motion with diffusion rate 2, we obtain the following estimate. There exists $p > 0$ such that for all $\varepsilon > 0$ and $(x, y) \in \hat{\mathcal{D}}(\varepsilon)$,

$$\mathbb{P}(T_\rho(0) < \varepsilon^2 \mid X_t = x, Y_t = y) > p,$$

which clearly implies

$$\mathbb{P}(T(\hat{\mathcal{D}}^c(\varepsilon)) \geq \varepsilon^2 \mid X_t = x, Y_t = y) \leq 1 - p.$$

By the Markov property,

$$\mathbb{P}(T(\hat{\mathcal{D}}^c(\varepsilon)) \geq k\varepsilon^2 \mid X_t = x, Y_t = y) \leq (1 - p)^k,$$

and so

$$\mathbb{E}(T(\hat{\mathcal{D}}^c(\varepsilon)) \mid X_t = x, Y_t = y) \leq \sum_{k \geq 1} k\varepsilon^2 (1 - p)^k = c_2 \varepsilon^2.$$

□

The following lemma is almost the same as [4, Lemma 5.1]. We reproduce that result here as many details of the original proof have to be changed. The intuitive meaning of the lemma is that if we condition X and Y on not coupling before time s, then the processes are likely to move apart for a considerable distance at time s. Hence, Lemma 3.10 below is a version of the parabolic boundary Harnack principle for the process (X, Y).

Recall the mirror coupling (X, Y). It will be convenient to write $Z = (X, Y)$ as the separate components of Z will play no role in Lemma 3.10. The state space of Z is $\overline{D} \times \overline{D}$. Recall that $\mathcal{D} = \{(x, y) \in \overline{D} \times \overline{D} : x \neq y\}$, $\mathcal{D}(\varepsilon) = \{(x, y) \in \overline{D} \times \overline{D} : |x - y| \geq \varepsilon\}$ and $\hat{\mathcal{D}}(\varepsilon) = \mathcal{D} \setminus \mathcal{D}(\varepsilon)$.

In the following lemma, $\mathbb{P}^z$ and $\mathbb{E}^z$ will denote the distribution of Z starting from z and the corresponding expectation. Conditioning by a harmonic function h will be reflected in the notation by writing $\mathbb{P}^z_h$ and $\mathbb{E}^z_h$. See [20] for the discussion of conditioned Brownian motion and [44] for conditioning of general Markov processes.

We will denote the space-time counterpart of Z by V. More precisely, if Z has law $\mathbb{P}^z$, then the law of the space-time process V=$\{V_t = (Z_t, s - t), t \geq 0\}$ will be denoted $\mathbb{P}^{z,s}$. The distribution of space-time process conditioned by a parabolic function g will be denoted $P^{z,s}_g$. By abuse of notation, $T(A)$ will denote the first hitting time of A for V as well as for Z.

Lemma 3.10 *There exist $\varepsilon, c, u > 0$ such that for all $z \in \mathcal{D}$,*

$$\mathbb{P}^z(Z_u \in \mathcal{D}(\varepsilon), \tau > u) \quad \geq \quad c\,\mathbb{P}^z(\tau > u).$$

Proof : Fix some small $\varepsilon_0 > 0$ such that $M = \mathcal{D}(\varepsilon_0)$ contains a non-empty open ball and let $\mathcal{D}_1 = \mathcal{D} \setminus M$. Let $h(z) = \mathbb{P}^z(T(M) < \tau)$ and $U_k = \{x \in D_1 : h(x) \in [2^{k-1}, 2^k]\}$ for integer k.

By Lemma 3.9 (i), $U_k \subset \hat{\mathcal{D}}(c_1 2^{k/\alpha})$, for some $c_1, \alpha > 0$. Then Lemma 3.9 (ii) shows that $\sup_{z \in U_k} \mathbb{E}^z(T(U_k^c)) \le c_2 2^{2k/\alpha}$. It follows that

$$\sum_{k=0}^{\infty} \sup_{z \in U_{-k}} \mathbb{E}^z(T(U_{-k}^c)) < \infty.$$

An argument of Chung [13] (see also [4]) shows that for suitable c_3,

$$c_3 \sum_{k=0}^{\infty} \sup_{z \in U_{-k}} \mathbb{E}^z(T(U_{-k}^c))$$

is an upper bound for $\mathbb{E}_h^z(T(\mathcal{D}_1^c))$. It follows that for a suitable $u > 0$ and every $z \in \mathcal{D}$,

$$\mathbb{P}_h^z(T(\mathcal{D}_1^c)) < u/4) > 1/2. \tag{3.11}$$

Recall the discussion of space-time processes before the statement of the lemma. The function

$$(z,t) \mapsto g(z,t) = \mathbb{P}^z(\tau > t)$$

is parabolic in $\mathcal{D} \times [0, \infty)$ with boundary values 1 on $\mathcal{D} \times \{0\}$ and 0 otherwise.

Let g_1 be a parabolic function in $\mathcal{D} \times [0, \infty)$ which has the same boundary values as g except that $g_1(z, 0)$ is changed from 1 to δ for $z \in \mathcal{D}_1$, where $\delta \in (0,1)$ will be chosen later. Now we will estimate g_1 on $\mathcal{D} \times [u/2, u]$.

It is easy to see that $g_1(z,s) > c_4$ for all $z \in M$ and $s \in [u/4, u]$. We obviously have $h(y) \le 1$ for all y. Let $h(x,s) = h(x)$. For $x \in \mathcal{D}_1$ and $s \ge u/2$ we have, by (3.11),

$$\begin{aligned}
g_1(x,s) &\ge \int_{\substack{t \in [u/4,u] \\ y \in \partial \mathcal{D}_1}} g_1(y,t)\, \mathbb{P}^{x,s}(T(\mathcal{D}_1^c) \in dt, X(T(\mathcal{D}_1^c)) \in dy) \\
&= \int_{\substack{t \in [u/4,u] \\ y \in \partial \mathcal{D}_1}} \frac{h(x,s)}{h(y,t)} \frac{h(y,t)}{h(x,s)} g_1(y,t)\, \mathbb{P}^{x,s}(T(\mathcal{D}_1^c) \in dt, X(T(\mathcal{D}_1^c)) \in dy) \\
&= \int_{\substack{t \in [u/4,u] \\ y \in \partial \mathcal{D}_1}} \frac{h(x,s)}{h(y,t)} g_1(y,t)\, \mathbb{P}_h^{x,s}(T(\mathcal{D}_1^c) \in dt, X(T(\mathcal{D}_1^c)) \in dy) \\
&\ge \int_{\substack{t \in [u/4,u] \\ y \in \partial \mathcal{D}_1}} h(x,s) c_4\, \mathbb{P}_h^{x,s}(T(\mathcal{D}_1^c) \in dt, X(T(\mathcal{D}_1^c)) \in dy) \\
&= h(x,s) c_4\, \mathbb{P}_h^{x,s}(T(\mathcal{D}_1^c) \in [u/4, s]) \\
&\ge h(x,s) c_4/2 \quad = \quad c_5 h(x,s) \quad = \quad c_5 h(x)\,.
\end{aligned}$$

Let

$$W_k = \{(z,s) : g_1(z,s) \in [2^k, 2^{k+1}], s \in [u/2, u]\},$$

$$W = \bigcup_{k=-\infty}^{k_1} W_k,$$

where $k_1 < 0$ will be chosen later. If $2^{-m} < c_5$ then $W_k \subset U_{k+m} \times [u/2, u]$. Using the estimate of Chung [13] we obtain for small k_1 and all $z \in \mathcal{D}$,

$$\begin{aligned}
\mathbb{E}_{g_1}^{z,u}(T(W^c)) &\leq c_6 \sum_{k=-\infty}^{k_1} \sup_{(y,s)\in W_k} \mathbb{E}^{y,s}\, T(W_k^c) \\
&\leq c_6 \sum_{k=-\infty}^{k_1} \sup_{(y,s)\in U_{k+m}} \mathbb{E}^{y,s}\, T(U_{k+m}^c) < \infty.
\end{aligned}$$

Choose k_1 so small that for any $z \in \mathcal{D}$,

$$\mathbb{E}_{g_1}^{z,u}\, T(W^c) < u/8\,. \tag{3.12}$$

Let

$$Q = \{(x,s) : g_1(x,s) \geq 2^{k_1}, s \in [u/2, u]\}\,.$$

Since the g_1-process cannot exit $\mathcal{D} \times [0,\infty)$ through $\partial\mathcal{D} \times [0,\infty)$, (3.12) implies

$$P_{g_1}^{z,u}(T(Q) > u/4) < 1/2\,. \tag{3.13}$$

Now let $\delta = 2^{k_1 - 1}$. Since $0 \leq g_1 \leq 1$, the process $g_1(V_t)$ is a martingale under $\mathbb{P}^{z,s}$, and $g_1(z,s) \geq 2^{k_1}$ for $(z,s) \in Q$, we see that there is at least $2^{k_1-1}/2$ chance that V under $\mathbb{P}^{z,s}$ will hit $M \times \{0\}$ before hitting any other part of $\partial(\mathcal{D} \times [0,\infty))$. Thus we have for $(z,s) \in Q$,

$$\begin{aligned}
&\mathbb{P}_{g_1}^{z,s}[V_s \in M \times \{0\}] = \\
&\int_M (g_1(y,0)/g_1(z,s))\, \mathbb{P}^{z,s}[V_s \in dy, T(\partial(\mathcal{D} \times [0,\infty))) = s] \\
&\geq \int_M \mathbb{P}^{z,s}[V_s \in dy, T(\partial(\mathcal{D} \times [0,\infty))) = s] \\
&\geq 2^{k_1-1}/2.
\end{aligned}$$

This and (3.13) yield, by the strong Markov property, for all $z \in \mathcal{D}$,

$$\mathbb{P}_{g_1}^{z,u}[V_u \in M \times \{0\}] \geq c_7 > 0.$$

The ratio of g and g_1 is bounded away from 0 and ∞ on the accessible boundary of $\mathcal{D} \times [0,\infty)$, so

$$P_g^{z,u}[V_u \in M \times \{0\}] \geq c_8 > 0$$

for all $z \in \mathcal{D}$. This is equivalent to the statement in the lemma. □

Proof of Theorem 3.7 (ii): We start by constructing a coupling of couplings. More precisely, we will construct processes (X,Y) and $(\tilde{X},\tilde{Y})$ such that each one of them is a mirror coupling of reflected Brownian motions in D. Hence, each of these processes is Markov. The two processes will also form a coupling, but the combined process $((X,Y),(\tilde{X},\tilde{Y}))$ will *not* be Markov since the coupling will fail to have the co-adapted property.

Let τ and $\tilde{\tau}$ denote the coupling times for (X,Y) and $(\tilde{X},\tilde{Y})$, respectively.

Fix some $a_1, c_1, u_1 > 0$ which satisfy Lemma 3.10 in place of ε, c and u. Find $a_2 \in (0, a_1)$, $c_2 > 0$, a set $\mathcal{D}_1 \subset \mathcal{D}(a_2)$ and a measure ν supported by $\mathcal{D}_1$ which satisfy Lemma 3.8. By Lemmas 3.8 and 3.10, for $A \subset \mathcal{D}_1$ and any $x, y \in \overline{D}$,

$$\begin{aligned}
&\mathbb{P}((X_{u_1+1}, Y_{u_1+1}) \in A \mid X_0 = x, Y_0 = y) \\
&= \int_{\mathcal{D}} \mathbb{P}((X_1, Y_1) \in A \mid X_0 = x', Y_0 = y') \times \\
&\qquad \times \mathbb{P}((X_{u_1}, Y_{u_1}) \in (dx', dy') \mid X_0 = x, Y_0 = y) \\
&\geq \int_{\mathcal{D}_1} \mathbb{P}((X_1, Y_1) \in A \mid X_0 = x', Y_0 = y') \times \\
&\qquad \times \mathbb{P}((X_{u_1}, Y_{u_1}) \in (dx', dy') \mid X_0 = x, Y_0 = y) \\
&\geq c_2 \int_{\mathcal{D}_1} \nu(A)\, \mathbb{P}((X_{u_1}, Y_{u_1}) \in (dx', dy') \mid X_0 = x, Y_0 = y) \\
&\geq c_1 c_2 \nu(A)\, \mathbb{P}(\tau > u_1 \mid X_0 = x, Y_0 = y) \\
&\geq c_1 c_2 \nu(A)\, \mathbb{P}(\tau > u_1 + 1 \mid X_0 = x, Y_0 = y).
\end{aligned}$$

Let $u_2 = u_1 + 1$. The last formula implies that

$$\mathbb{P}((X_{u_2}, Y_{u_2}) \in A \mid \tau > u_2, X_0 = x, Y_0 = y) \quad \geq \quad \nu_1(A) = c_1 c_2 \nu(A)\,. \tag{3.14}$$

Consider any $x, y, \tilde{x}$ and $\tilde{y}$ in $\overline{D}$. On a single probability space, we will construct processes (X_t, Y_t) and $(\tilde{X}_t, \tilde{Y}_t)$ starting from (x,y) and $(\tilde{x},\tilde{y})$ respectively. In view of (3.14) we may construct random vectors $\mathcal{V} = (\mathcal{V}_1, \mathcal{V}_2)$ and $\tilde{\mathcal{V}} = (\tilde{\mathcal{V}}_1, \tilde{\mathcal{V}}_2)$ such that for $A \subset \mathbb{R}^4$,

$$\mathbb{P}(\mathcal{V} \in A) \quad = \quad \mathbb{P}((X_{u_2}, Y_{u_2}) \in A \mid X_0 = x, Y_0 = y)\,,$$

and

$$\mathbb{P}(\tilde{\mathcal{V}} \in A) \quad = \quad \mathbb{P}((\tilde{X}_{u_2}, \tilde{Y}_{u_2}) \in A \mid \tilde{X}_0 = \tilde{x}, \tilde{Y}_0 = \tilde{y})\,;$$

moreover, for $A \subset \mathcal{D}_1$,

$$\mathbb{P}(\mathcal{V} = \tilde{\mathcal{V}} \in A \mid \mathcal{V}_1 \neq \mathcal{V}_2, \tilde{\mathcal{V}}_1 \neq \tilde{\mathcal{V}}_2) \quad \geq \quad \nu_1(A)\,. \tag{3.15}$$

Now we set

$$\begin{aligned}
(X_0, Y_0) = (x, y)\,, \qquad & (\tilde{X}_0, \tilde{Y}_0) = (\tilde{x}, \tilde{y})\,, \\
(X_{u_2}, Y_{u_2}) \quad &= \quad \mathcal{V}, \\
(\tilde{X}_{u_2}, \tilde{Y}_{u_2}) \quad &= \quad \tilde{\mathcal{V}}.
\end{aligned}$$

Next we construct $\{(X_t, Y_t), t \in [0, u_2]\}$ and $\{(\tilde{X}_t, \tilde{Y}_t), t \in [0, u_2]\}$ by adding bridges between the endpoints of the trajectories in such a way that each of these processes is a mirror coupling of reflected Brownian motions in D. Let $Q^{x,y,\tilde{x},\tilde{y}}_{u_2}$ denote the distribution of $\{\mathcal{Z}_t = ((X_t, Y_t), (\tilde{X}_t, \tilde{Y}_t)), t \in [0, u_2]\}$. We inductively define $Q^{x,y,\tilde{x},\tilde{y}}_{(k+1)u_2}$ for $k = 1, 2, 3, \dots$, by the following "Markov-like property" formula (the Markov property does not extend to other times besides ku_2),

$$\begin{aligned} Q^{x,y,\tilde{x},\tilde{y}}_{(k+1)u_2} & (\{\mathcal{Z}_t, t \in [0, ku_2]\} \in A_1, \{\mathcal{Z}_t, t \in [ku_2, (k+1)u_2]\} \in A_2) \\ &= \int Q^{x,y,\tilde{x},\tilde{y}}_{ku_2}(\{\mathcal{Z}_t, t \in [0, ku_2]\} \in A_1, \mathcal{Z}_{ku_2} \in (dv, dz, d\tilde{v}, d\tilde{z})) \\ &\qquad \times Q^{v,z,\tilde{v},\tilde{z}}_{u_2}(\{\mathcal{Z}_t, t \in [0, u_2]\} \in A_2^{ku_2}), \end{aligned} \tag{3.16}$$

for all $A_1 \subset C([0, ku_2], \mathbb{R}^8)$ and $A_2 \subset C([ku_2, (k+1)u_2], \mathbb{R}^8)$. Here $A_2^{ku_2}$ is the family of functions in A_2, shifted to the left by ku_2 units. Note that the measure $Q^{x,y,\tilde{x},\tilde{y}}_{(k+1)u_2}$ is uniquely defined if we specify its value on cylinders of the form $A_1 \times A_2$, as in (3.16). We define $Q^{x,y,\tilde{x},\tilde{y}}_{\infty}$, i.e., the distribution of $\{((X_t, Y_t), (\tilde{X}_t, \tilde{Y}_t)), t \in [0, \infty)\}$ using the measures $Q^{x,y,\tilde{x},\tilde{y}}_{ku_2}$ and consistency, in the obvious way.

Let $\tau^* = \inf\{t : (X_t, Y_t) = (\tilde{X}_t, \tilde{Y}_t)\}$. It follows from (3.15) that

$$Q^{x,y,\tilde{x},\tilde{y}}_{\infty}(\tau^* > u_2 \mid \tau > u_2, \tilde{\tau} > u_2) \quad \leq \quad p_1 = 1 - \nu_1(\mathcal{D}_1) < 1,$$

for all $x, y, \tilde{x}, \tilde{y} \in \overline{D}$. By the "Markov-like property" (3.16),

$$Q^{x,y,\tilde{x},\tilde{y}}_{\infty}(\tau^* > ku_2 \mid \tau > ku_2, \tilde{\tau} > ku_2) \quad \leq \quad p_1^k.$$

Hence, for some $c_3 > 0$, all $t \geq u_2$ and all $x, y, \tilde{x}, \tilde{y} \in \overline{D}$,

$$Q^{x,y,\tilde{x},\tilde{y}}_{\infty}(\tau^* > t \mid \tau > t, \tilde{\tau} > t) \quad \leq \quad e^{-c_3 t}. \tag{3.17}$$

The rest of the proof is very similar to the end of the proof of Theorem 2.6 (i). Consider $x, y \in \overline{D}$ with $\varphi_2(x) \neq \varphi_2(y)$. Recall (3.5);

$$\mathbb{P}(\tau > t \mid X_0 = x, Y_0 = y) \quad \leq \quad c(x, y) e^{-\mu t}, \qquad t \geq 0,$$

for some $\mu > 0$. In the following we will use the generic notation $\mathbb{P}$ for probability but we will assume that the process $\{((X_t, Y_t), (\tilde{X}_t, \tilde{Y}_t)), t \in [0, \infty)\}$ has the distribution $Q^{x,y,y,x}_{\infty}$. Hence, $X_0 = \tilde{Y}_0 = x$ and $Y_0 = \tilde{X}_0 = y$. Then we can use (3.17) to show, for $A \subset D$, and large t,

$$\begin{aligned} &\left| \int_A p(t, x, z) dz - \int_A p(t, y, z) dz \right| \\ &= |\mathbb{P}(X_t \in A \mid X_0 = x) - \mathbb{P}(Y_t \in A \mid Y_0 = y)| \\ &= |\mathbb{P}(X_t \in A, t < \tau \mid X_0 = x, Y_0 = y) - \mathbb{P}(Y_t \in A, t < \tau \mid X_0 = x, Y_0 = y)| \\ &= \big| \mathbb{P}(X_t \in A \mid t < \tau, X_0 = x, Y_0 = y) \, \mathbb{P}(t < \tau \mid X_0 = x, Y_0 = y) \end{aligned}$$

$$
\begin{aligned}
&\quad -\mathbb{P}(Y_t \in A \mid t < \tau, X_0 = x, Y_0 = y)\,\mathbb{P}(t < \tau \mid X_0 = x, Y_0 = y)\big| \\
&\le |\mathbb{P}(X_t \in A \mid t < \tau, X_0 = x, Y_0 = y) - \mathbb{P}(Y_t \in A \mid t < \tau, X_0 = x, Y_0 = y)| \times \\
&\qquad \times c(x,y)e^{-\mu t} \\
&= \Big|\mathbb{P}(X_t \in A \mid t < \tau, X_0 = x, Y_0 = y) - \mathbb{P}(\tilde{X}_t \in A \mid t < \tilde{\tau}, \tilde{X}_0 = y, \tilde{Y}_0 = x)\Big| \\
&\qquad \times ce^{-\mu t} \\
&\le \big|\,\mathbb{P}(X_t \in A, t < \tau^* \mid t < \tilde{\tau}, t < \tau, X_0 = x, Y_0 = y) \\
&\quad -\mathbb{P}(\tilde{X}_t \in A, t < \tau^* \mid t < \tilde{\tau}, t < \tau, \tilde{X}_0 = y, \tilde{Y}_0 = x)\big| ce^{-\mu t} \\
&\le \mathbb{P}(t < \tau^* \mid t < \tilde{\tau}, t < \tau, X_0 = x, \tilde{X}_0 = y)| ce^{-\mu t} \quad \le \quad e^{-c_3 t} ce^{-\mu t}\,.
\end{aligned}
$$

This and (3.6) show that $\mu_2 \ge \mu + c_3 > \mu$. Thus the mirror coupling for reflected Brownian motion in a triangle with distinct acute angles is not efficient. □

4 Some further examples

This section contains a selection of rather informal examples, giving some indication of how far the results on mirror couplings in triangles, presented in the last section, can be generalized to mirror couplings in other planar sets. The motivation for our efforts comes from two different sources. First, mirror couplings have been used to estimate the spectral gap for some diffusions [48] so it is a natural question how sharp those estimates are. This we cannot say but our examples may be a source of inspiration for future research in this direction. Second, the mirror coupling techniques developed for this project have already been applied [3, 6, 5] to a problem on "hot spots." Hence, the techniques seem to have some interest beyond the efficiency of Markovian couplings.

The reader might have noticed that the assumption that D is a triangle, adopted in Section 3, does not play a major role in the arguments. The following example makes this point explicit.

Example 4.1 *Fig. 4 shows a convex polygonal domain D_1 whose boundary is naturally divided into "upper" and "lower" parts. The angles between line segments in the upper part of ∂D_1 and those in the lower part are less than $\pi/2$. The arguments presented in Section 3 carry over to this case and it is easy to see that both synchronous and mirror couplings for reflected Brownian motion in D_1 are efficient.*

[Figure 4 about here.]

On the other hand, Theorem 3.7 (ii) can be also generalized to some other domains besides triangles. The domain D_2 illustrated in Fig. 5 is an acute triangle whose corners have been cut. The mirror coupling can be proved to be inefficient in D_2 just as in the case of a triangle with acute angles.

[Figure 5 about here.]

What about non-convex domains? Convexity is used in Theorem 3.7 (i) to show that the distance between the processes X and Y is a Brownian motion plus a process which always pushes X and Y towards each other. This is true only in convex domains. However it can be circumvented, at a price of obscure conditions and tedious details.

Example 4.2 *It is easy to check that the mirror will never turn more than the angle π in some non-convex domains, for example in the domain D_3 in Fig. 6. The angles between any two line segments in ∂D_3 are less than $\pi/2$. We believe that the mirror coupling is efficient in D_3 but the proof of Theorem 3.7 (i) does not completely apply in this case because of the lack of convexity as indicated above. However we believe that one can circumvent the need for convexity for this kind of example, albeit with the need for more involved arguments.*

[Figure 6 about here.]

Our next example addresses the question of what happens if the domain has smooth boundary, rather than polygonal boundary. The construction of a mirror coupling in such a domain has to proceed along different lines than that presented in Section 3, which works only for polygonal domains. However, the construction does not present major problems; an example in a similar context is to be found in [29].

Example 4.3 *The domain D_4 in Fig. 7 has a piecewise smooth boundary. If we consider two tangent lines to ∂D_4, one to the upper part of ∂D_4 and the other tangent to the lower part then they form an angle less than $\pi/2$. In a domain D_4, the hinge will be the point of the intersection of the mirror and the tangent line to ∂D_4 at the point where one of the processes is reflecting from the boundary. Hence, the hinge will move not only by jumps but also in a continuous fashion. The general qualitative behavior of the mirror movement does not change in a fundamental way, however, from the polygonal domain case. Hence, the mirror cannot turn more than π in D_4. This is all we have to know to prove that the mirror coupling is efficient in D_4.*

[Figure 7 about here.]

It is also possible to make some progress if a symmetry is present:

Example 4.4 *In [3] it is proved that the mirror cannot turn more than π in a convex domain D_5 if we assume in addition that*

(1) D_5 has a line of symmetry S which intersects ∂D_5 at x and y,

(2) $|x-z| \vee |y-z| < |x-y|$ for all $z \in \overline{D}_5 \setminus \{x,y\}$, and

(3) for all $r>0$ the sets $\partial B(x,r) \cup D$ and $\partial B(y,r) \cup D$ are connected.

Just as in Example 4.2, we believe that the mirror coupling is efficient in domains D_5 satisfying these assumptions but the proof given in Section 3 would have to be modified. In the present case, we cannot claim that $Y_t^1 - X_t^1 > \alpha|Y_t^1 - X_t^1|$ for some $\alpha > 0$. This property holds if we impose some more assumptions on the slope of ∂D_5.

It should be noted that our methods cannot decide whether the mirror coupling is efficient for reflected Brownian motion in the domains considered in [3, Theorem 1.3 (A1)]. Those domains are assumed to have two perpendicular axes of symmetry. The proof of that result is based on the behavior of the mirror coupling for the reflected Brownian motion with absorption on one of the axes of symmetry. Hence, the technique does not directly apply to reflected Brownian motion in the whole domain.

Finally we will provide some details of the mirror coupling behavior in the case when the domain D is a disc. This highly symmetric case makes the analysis especially easy and complete.

Example 4.5 *Recall from Example 4.3 that the hinge h_t lies at the intersection of the mirror and the line tangent to the circle ∂D where one of the processes X or Y is reflecting. A quick look at Fig. 8 should convince the reader that the mirror M must move towards the center of the disc (i.e., its distance from the center can only decrease). Moreover, the points of intersection of the mirror with ∂D can only move upwards in Fig. 8. These remarks follow from the fact that the effect of reflection is the counterclockwise motion of the mirror M around the (instantaneous) hinge position h. If the mirror passes through the center at some time s, it will never change its position after time s, because after that time, the processes X and Y will reflect at the boundary of D at the same time, until their coupling time. Hence, the mirror can never intersect ∂D inside the (smaller) arc between a and b and likewise not between a_1 and b_1 (the antipodal points to a and b). The process Y can not start reflecting on ∂D before the time when the mirror passes through the center. Hence, the mirror must hit the center of the disc before or at the same time when Y hits the smaller arc between a_1 and b_1. Given these properties of the mirror coupling in a disc, it is not hard to prove that it is efficient.*

[Figure 8 about here.]

References

[1] D. Aldous and J.A. Fill. Reversible Markov chains and random walks on graphs. (monograph in preparation), 1996. URL: `www.stat.Berkeley.EDU/users/aldous/book.html`.

[2] D.J. Aldous and P. Diaconis. Strong uniform times and finite random walk. *Advances in Applied Mathematics*, 8(1):69–97, 1987.

[3] R. Bañuelos and K. Burdzy. On the 'hot spots' conjecture of J. Rauch. *Journal of Functional Analysis*, 164:1–33, 1999.

[4] R. Bass and K. Burdzy. Lifetimes of conditioned diffusions. *Probability Theory and Related Fields*, 91:405–443, 1992.

[5] R. Bass and K. Burdzy. Fiber Brownian motion and the "hot spots" problem. Research report, Department of Mathematics, University of Washington, 1999.

[6] K. Burdzy and W. Werner. A counterexample to the "hot spots" conjecture. *Annals of Mathematics*, 149:309–317, 1999.

[7] H. Cai. A note on an exact sampling algorithm and Metropolis Markov chains. Technical report, University of Missouri-St. Louis, 1997. URL: `ftp://ftp.umsl.edu/pub/cai/convergence.ps`.

[8] M.-F. Chen. *From Markov chains to non-equilibrium particle systems.* World Scientific, Singapore, 1992.

[9] M.-F. Chen. Optimal Markovian couplings. *Progr. Natur. Sci. (English Ed.)*, 4:364–366, 1994.

[10] M.-F. Chen. Optimal Markovian couplings and applications. *Acta Math. Sinica, New Series*, 10:260–275, 1994.

[11] M.-F. Chen. Trilogy of couplings and general formulas for lower bound of spectral gap. In L. Accardi and C. Heyde, editors, *Proceedings of the Symposium on Probability Towards the Year 2000*, pages 123–136. Springer-Verlag, New York, 1998.

[12] M.-F. Chen and F.-Y. Wang. Application of coupling method to the first eigenvalue on manifold. *Science in China (Ser. A)*, 37:1–14, 1994.

[13] K.L. Chung. The lifetime of conditioned Brownian motion in the plane. *Ann. Inst. Henri Poincaré*, 20:349–351, 1984.

[14] D.R. Cox and H.D. Miller. *The Theory of Stochastic Processes.* Methuen, London, 1965.

[15] M. Cranston and Y. Le Jan. On the noncoalescence of a two point Brownian motion reflecting on a circle. *Ann. Inst. H. Poincaré, Prob. Stat.*, 25:99–107, 1989.

[16] M. Cranston and Y. Le Jan. Noncoalescence for the Skorohod equation in a convex domain of $\mathbb{R}^2$. *Probability Theory and Related Fields*, 87:241–252, 1990.

[17] P. Diaconis and J. Fill. Strong stationary times via a new form of duality. *Annals of Probability*, 18:1483–1522, 1990.

[18] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Annals of Applied Probability*, 6(3):695–750, 1996.

[19] R. Dobrušin. Markov processes with a large number of locally interacting components (in Russian). *Problemy Peredači Informacii*, 7(2):70–87, 1971.

[20] J.L. Doob. *Classical Potential Theory and Its Probabilistic Counterpart.* Springer-Verlag, New York, 1984.

[21] R. Durrett. *Probability: Theory and Examples.* Wadsworth, Pacific Grove, 1991.

[22] J. Fill. An interruptible algorithm for exact sampling via Markov Chains. *Annals of Applied Probability*, 8:131–162, 1998.

[23] G.B. Folland. *Introduction to Partial Differential Equations.* Princeton Univ. Press, Princeton, 1976.

[24] D. Griffeath. A maximal coupling for Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verve Gebiete*, 31(2):95–106, 1975.

[25] D. Griffeath. Coupling methods for Markov processes. In *Studies in probability and ergodic theory; Advances in Mathematics, Supplementary Studies*, volume 2, pages 1–43. Academic Press, 1978.

[26] S. Ingrassia. Geometric approaches to the estimation of the spectral gap of reversible Markov chains. *Combinatorics, Probability and Computing*, 2:301–323, 1993.

[27] K. Itô and H.P. McKean. *Diffusion Processes and Their Sample Paths.* Springer-Verlag, New York, second edition, 1974.

[28] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York, 1988.

[29] W.S. Kendall. Coupled Brownian motions and partial domain monotonicity for the Neumann heat kernel. *Journal of Functional Analysis*, 86:226–236, 1989.

[30] W.S. Kendall. Stochastic calculus in *Mathematica*: software and examples. Research Report 333, Department of Statistics, University of Warwick, 1998. To appear in *CDROM for Mathematics 1998.*

[31] T.M. Liggett. *Interacting Particle Systems.* Springer-Verlag, New York, 1985.

[32] T. Lindvall. *Lectures on the coupling method.* John Wiley & Sons, New York, 1992.

[33] J.S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.

[34] R.B. Lund, S.P. Meyn, and R.L. Tweedie. Computable exponential convergence rates for stochastically ordered Markov processes. *Ann. Appl. Probab.*, 6:218–237, 1996.

[35] P. Matthews. Strong stationary times and eigenvalues. *Journal of Applied Probability*, 29(1):228–233, 1992.

[36] Ch. Meise. Lower bounds for the spectral gap of Markov chains. Preprintreihe Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik" 96-019, Universität Bielefeld, 1996. URL: `www.mathematik.uni-bielefeld.de/sfb343/preprints/index96.html`.

[37] T.S. Mountford and M. Cranston. Efficient coupling on the circle. Unpublished manuscript, 1999.

[38] M.A. Pinsky. The eigenvalues of an equilateral triangle. *SIAM journal of Mathematics*, 11(5):819–827, 1980.

[39] M.A. Pinsky. Completeness of the eigenfunctions of the equilateral triangle. *SIAM J. Math. Anal.*, 16(4):848–851, 1985.

[40] J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.

[41] G.O. Roberts and J.S. Rosenthal. Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society*, B61(3):643–660, 1999.

[42] J.S. Rosenthal. Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Statistics and Computing*, 6(3):269–275, 1996.

[43] L. Saloff-Coste. Simple examples of the use of Nash inequalities for finite Markov chains. In O.E. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, pages 365–400. Chapman and Hall, London, 1998.

[44] M. Sharpe. *General Theory of Markov Processes*. Academic Press, New York, 1989.

[45] R.L. Smith and L. Tierney. Exact transition probabilities for the Independence sampler. Technical report, University of North Carolina, 1996. URL: `www.unc.edu/depts/statistics/postscript/rs/exact.ps`.

[46] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1728, 1994.

[47] S.R.S. Varadhan and R.J. Williams. Brownian motion in a wedge with oblique reflection. *Comm. Pure Appl. Math.*, 38:405–443, 1985.

[48] F.-Y. Wang. Application of coupling methods to the Neumann eigenvalue problem. *Probability Theory and Related Fields*, 98:299–306, 1994.

Krzysztof Burdzy
Department of Mathematics
University of Washington
Box 354350
Seattle, WA 98195-4350, USA
burdzy@math.washington.edu

Wilfrid Kendall
Department of Statistics
University of Warwick
Coventry
United Kingdom
w.s.kendall@warwick.ac.uk

List of Figures

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6:

Figure 7:

Figure 8: