Probabilistic Coupling: Mixing Times and Optimality

Roberta Merli

Doctor of Philosophy

UNIVERSITY OF YORK
MATHEMATICS

September 2021

ABSTRACT

The main focus of this thesis is probabilistic coupling. This technique and its connection with the total variation distance will be a common thread through the exploration of the random processes investigated in this thesis.

In Chapter 2, we generalise a recent result on the mixing time of the random walk on $\mathbb{Z}_2^n$ that at each step flips k randomly chosen coordinates. In our work, we let the number of coordinates flipped at each step be a random variable K, and, using a path coupling argument, we establish bounds for the mixing time of this random walk. Furthermore, we show that, under some stricter assumptions on the distribution of K, the random walk has a pre-cutoff.

In Chapter 3, we focus on properties of particular couplings, such as co-adaptedness, maximality, and other types of optimality. We consider the Brownian motion on the circumference of the unit circle that, at times of an independent Poisson process of rate λ, jumps to the opposite point on the circle. We construct a co-adapted coupling for this process and, using excursion theory and Bellman's principle of optimality, we prove that it is mean-optimal in the class of co-adapted couplings, i.e. it minimises the expected coupling time. We describe how this coupling depends upon λ, and show that it is maximal only when $\lambda = 0$. We also give an explicit construction of a maximal coupling for this "jumpy Brownian motion" (for any value of λ) in the case where the two copies of the process begin at opposite sides of the circle.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

As my PhD comes to an end, I would like to thank all the people who have supported and encouraged me throughout this long and exciting journey.

My heartfelt gratitude goes to my supervisor, Dr. Stephen Connor, for the great patience, support, and consideration he's showed me throughout my PhD and during the pandemic. I'm very grateful for his help, I couldn't have asked for a better supervisor.

I would like to thank my family for supporting me even when I decided to move far away from home to do something that they still don't completely understand.

Warm thanks to my precious friend Nayana for the many laughs and conversations, her unconditional support, and for sharing the frustration of mixing multiple languages when speaking.

Many thanks to my friends in Italy, who have always been there for me despite the thousands of kilometres that separate us.

Thanks to everyone in the Department of Mathematics in the University of York who helped make the long working days lighter. Special thanks to Peipei and Dalal for offering me their friendship and sharing the difficulties encountered during the PhD.

Thanks to all the friends I met through Friend International York for warmly welcoming me in York and for broadening my mind by introducing me to so many different people and cultures. Especially, I'd like to thank Helen, Katherine, Julie, Rebecca, and all the friends from Film Night and Friday Night Feast for supporting me in very difficult times and helping make the pandemic more bearable and less lonely.

Finally, I'd like to thank Prof. Pietro Caputo for encouraging me to pursue this objective. Without his help, I'd have never known the wonderful people who have enriched my life in York, and I'd have never lived the extraordinary experiences that life abroad has offered me.

RINGRAZIAMENTI

Con questa tesi il mio dottorato si avvicina alla fine, quindi vorrei ringraziare tutte le persone che mi hanno supportato e incoraggiato durante questo lungo percorso.

La mia più sentita gratitudine va al mio relatore, Dr Stephen Connor, per la grande pazienza, sostegno e riguardo dimostrati nei miei confronti durante il dottorato e la pandemia. Sono molto grata per il suo aiuto, non avrei potuto avere un relatore migliore.

Vorrei ringraziare la mia famiglia per avermi supportato anche nella decisione di trasferirmi così lontano da casa per fare qualcosa che ancora non gli è completamente chiaro.

Un caloroso ringraziamento alla mia preziosa amica Nayana per le tante risate e conversazioni, il suo sostegno incondizionato e per aver condiviso la frustrazione nel parlare mescolando più lingue insieme.

Moltissime grazie ai miei amici in Italia che mi sono sempre stati vicini nonostante le migliaia di chilometri che ci separano.

Grazie a tutti coloro che nel dipartimento di matematica della University of York hanno contribuito a rendere le lunghe giornate lavorative più leggere. Un ringraziamento speciale a Peipei e Dalal per avermi offerto la loro amicizia e aver condiviso le difficoltà affrontate durante il dottorato.

Grazie a tutti gli amici incontrati tramite Friend International York per avermi accolta così calorosamente a York e per aver ampliato la mia mente facendomi conoscere così tante persone e culture diverse. Grazie soprattutto a Helen, Katherine, Julie, Rebecca e tutti gli amici di Film Night e Friday Night Feast per avermi supportato in momenti difficili e aver contribuito a rendere la pandemia più sopportabile e meno solitaria.

Infine vorrei ringraziare il Prof. Pietro Caputo per avermi incoraggiato a perseguire questo obiettivo. Senza il suo aiuto non avrei mai conosciuto tutte le persone meravigliose che hanno arricchito la mia vita a York e non avrei mai vissuto le esperienze che la vita all'estero mi ha offerto.

DECLARATION

I declare that this thesis is my own work and has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

1. INTRODUCTION

In this first part of the thesis, we report a summary of material on Markov chains that are at the base of our project. We define their main properties, and we give an introduction to coupling describing some of its useful applications in the study of the convergence of Markov chains. The material related to Markov chains is taken from "Markov chains and mixing times" by Levin, Peres and Wilmer [21].

1.1 Markov chains and the cutoff phenomenon

Let $X = (X_n)_{n \geq 0}$ be a Markov chain with state space Ω and transition matrix P. A distribution π over Ω satisfying $\pi = \pi P$ is called a stationary distribution, and it represents the long-term limiting distribution of the Markov chain, provided P is irreducible and aperiodic.

The convergence of a Markov chain is studied in terms of distance between the distribution of the chain and the stationary distribution. Several types of distance have been defined, one of them being the *total variation distance*, which is the metric we will use in this thesis.

Definition 1.1. Let P be the transition matrix of a Markov chain with state space Ω. Let π the stationary distribution and P^t the distribution of the Markov chain at time t. Then, we define the total variation distance between the distributions P^t and π as

$$\|P^t(x, \cdot) - \pi\|_{TV} := \max_{A \subset \Omega} |P^t(x, A) - \pi(A)|,$$

i.e. the maximum difference between the two probability distributions over all the possible subsets of the state space Ω.

As it is not always convenient to work with this definition, we can use an alternative characterisation that reduces the maximum over all subsets of Ω to a sum over its states, provided that Ω is a countable space.

Proposition 1.2. *If Ω is a countable space, then*

$$\|P^t(x,\cdot) - \pi\|_{TV} = \frac{1}{2}\sum_{y\in\Omega} |P^t(x,y) - \pi(y)|.$$

Proof. Let $B = \{y : P^t(x,y) \geq \pi(y)\}$ and let $A \subset \Omega$. Then,

$$P^t(x,A) - \pi(A) \leq P^t(x,A\cap B) - \pi(A\cap B) \leq P^t(x,B) - \pi(B),$$

where the first inequality holds because if $y \in A \cap B^c$ then $P^t(x,y) < \pi(y)$, which implies

$$P^t(x,A)-\pi(A) = P^t(x,(A\cap B)\cup(A\cap B^c))-\pi((A\cap B)\cup(A\cap B^c)) \leq P^t(x,A\cap B)-\pi(A\cap B).$$

Similarly, we have

$$P^t(x,B)-\pi(B) = P^t(x,(A\cap B)\cup(A^c\cap B))-\pi((A\cap B)\cup(A^c\cap B)) \geq P^t(x,A\cap B)-\pi(A\cap B).$$

In the same way, it can be showed that

$$\pi(A) - P^t(x,A) \leq \pi(B^c) - P^t(x,B^c).$$

Observe that the upper bounds $P^t(x,B) - \pi(B) = \pi(B^c) - P^t(x,B^c)$, so, if we take $A = B$ or $A = B^c$, then

$$\|P^t(x,\cdot)-\pi\|_{TV} = \frac{1}{2}[P^t(x,B)-\pi(B)+\pi(B^c)-P^t(x,B^c)] = \frac{1}{2}\sum_{y\in\Omega}|P^t(x,y)-\pi(y)|.$$

□

If Ω is a measurable space, then we have the following definition of total variation distance.

Definition 1.3. Let μ and μ' be two probability measures on a measurable space Ω. Then, the total variation distance between μ and μ' is defined as

$$\|\mu - \mu'\|_{TV} := \frac{1}{2} \sup_{\substack{|f|\leq 1 \\ f \text{ measurable}}} \left| \int f d(\mu - \mu') \right|.$$

Lindvall shows in [23] that we have the following alternative characterisation.

Proposition 1.4. *Let μ and μ' be two probability measures on a measurable space Ω. Let $\lambda = \mu + \mu'$ and*

$$g = \frac{d\mu}{d\lambda}, \qquad g' = \frac{d\mu'}{d\lambda}.$$

Then,

$$||\mu - \mu'||_{TV} = 1 - \int g \wedge g' d\lambda.$$

Proof. From Definition 1.3,

$$\begin{aligned}
2||\mu - \mu'||_{TV} &= \sup_{\substack{|f|\leq 1 \\ f \text{ measurable}}} \left| \int f \cdot (g - g')d\lambda \right| \\
&= \int_{g\geq g'} 1 \cdot (g - g')d\lambda + \int_{g<g'} (-1) \cdot (g - g')d\lambda \\
&= \int |g - g'|d\lambda = \int (g - g \wedge g')d\lambda + \int (g' - g \wedge g')d\lambda \\
&= \int g d\lambda + \int g' d\lambda - 2 \int g \wedge g' d\lambda \\
&= \int d\mu + \int d\mu' - 2 \int g \wedge g' d\lambda \\
&= 2 \left(1 - \int g \wedge g' d\lambda \right)
\end{aligned}$$

□

The following result establishes convergence at an exponential rate of any irreducible and aperiodic Markov chain on a finite state space to the stationary distribution.

Theorem 1.5 (Convergence theorem)**.** *Let P be irreducible and aperiodic with finite state space Ω and stationary distribution π. Then, there exist constants $\alpha \in (0, 1)$ and $C > 0$ such that*

$$\max_{x\in\Omega} ||P^t(x, \cdot) - \pi(\cdot)||_{TV} \leq C\alpha^t.$$

There are different proofs of this theorem. The proof given in [21] applies a decomposition of the chain into a combination of the stationary distribution and another Markov chain.

Observe that assuming that a Markov chain is irreducible and aperiodic is essential to have convergence; if one of these two conditions is not satisfied we might have that the chain does not converge, or that the stationary distribution depends on the starting state.
Given Theorem 1.5, one of the aims in the study of the convergence of a Markov chain is to estimate the rate of convergence, finding bounds on the maximal distance between the distribution of the chain and the stationary distribution. One way to do that is estimating the *mixing time*, a parameter that measures the time required for the distance from the stationary distribution to be small.

Definition 1.6. Let

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV}$$

be the *worst-case total variation distance*; observe that $d(t)$ is non-increasing for all $t \in \mathbb{N}$. Let $\varepsilon \in (0, 1)$, we can define the *mixing time* as

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}.$$

It can be showed that, if ℓ is a non-negative integer, then $d(\ell t_{mix}(\varepsilon)) \leq (2\varepsilon)^{\ell}$. To make use of this inequality, we need $\varepsilon < \frac{1}{2}$, but for algebraic convenience, we usually choose $\varepsilon = \frac{1}{4}$ and we use the notation $t_{mix} = t_{mix}\left(\frac{1}{4}\right)$.

Finally, we introduce the property of *transitivity* of Markov chains.

Definition 1.7. A Markov chain is transitive if for each pair $(x, y) \in \Omega \times \Omega$ there exists a bijection $\phi_{(x,y)} : \Omega \to \Omega$ such that

$$\phi_{(x,y)}(x) = y \qquad \text{and} \qquad P(z, w) = P(\phi_{(x,y)}(z), \phi_{(x,y)}(w)) \text{ for all } z, w \in \Omega.$$

In other words, the Markov chain "looks the same" from any point in Ω, and convergence results do not depend on the starting state. In terms of mixing time, this means the total variation distance does not depend on the starting state, so considering the worst-case is not needed.

1.1.1 The cutoff phenomenon

Starting from the notions of mixing time and total variation distance, we can define the cutoff phenomenon. We say that a sequence of random walks exhibits *cutoff* if the total variation distance drops from near 1 to near 0 in a small interval centred at the mixing time. Formally, we have the following definition.

Definition 1.8. Consider a sequence of Markov chains $X^{(n)}$ on the state spaces $\Omega^{(n)}$, each with mixing time $t_{mix}^{(n)}$ and total variation distance $d_n(t)$ at time t. We say that the sequence exhibits *cutoff* with a window of size ω_n if $\omega_n = o(t_n)$ and

$$\lim_{\alpha\to\infty} \liminf_{n\to\infty} d_n(t_{mix}^{(n)} - \alpha\omega_n) = 1,$$

$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} d_n(t_{mix}^{(n)} + \alpha\omega_n) = 0.$$

Figure 1.1 illustrates the graph of the total variation distance of a sequence of Markov chains that exhibits cutoff.
The cutoff phenomenon gives important information about the convergence of the Markov chain, but it is often difficult to prove. Sufficient conditions under which a Markov chain exhibits cutoff have been established by Saloff-Coste and Chen in [6] for the L^p distance for $1 < p \leq \infty$.

Definition 1.9. Let X be a Markov chain with finite state space Ω, transition matrix P and stationary distribution π. The L^p distance between the distribution of the Markov chain and π is defined as

$$||P^t(x,\cdot) - \pi(\cdot)||_p = \begin{cases} \left(\sum_{y\in\Omega} \left|\frac{P^t(x,y)}{\pi(y)} - 1\right|^p \pi(y)\right)^{\frac{1}{p}} & \text{if } 1 \leq p < \infty \\ \max_{y\in\Omega}\left\{\left|\frac{P^t(x,y)}{\pi(y)} - 1\right|\right\} & \text{if } p = \infty. \end{cases}$$

In particular, $||P^t(x,\cdot) - \pi(\cdot)||_1 = 2 \cdot ||P^t(x,\cdot) - \pi(\cdot)||_{TV}$, but for this value of p we do not know sufficient conditions under which a random walk exhibits cutoff, so it has to be shown case by case.

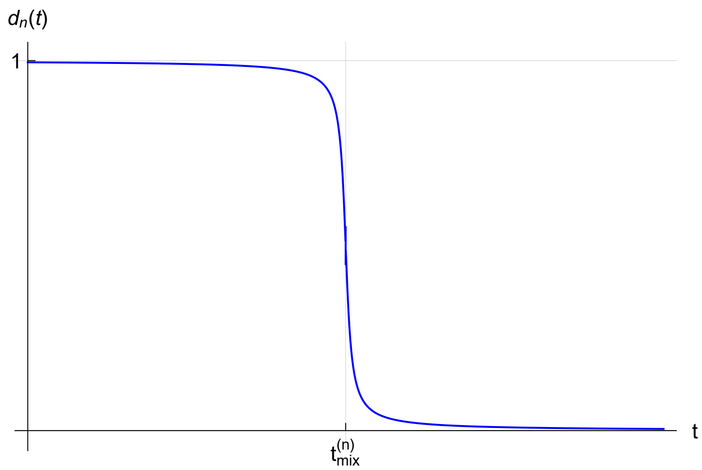


Fig. 1.1: Graph of $d_n(t)$ against time for a Markov chain that exhibits cutoff.

Many results about the cutoff phenomenon have been proved for chains describing card shufflings. In general, if we consider a random walk on a deck of n cards, the state space is the set of all possible arrangements of the deck, which is represented by the set S_n of permutations of n elements.

For a first example, consider the top-to-random shuffling on n cards, a card shuffling consisting of inserting, at each step, the top card randomly in the deck. In [2], Aldous and Diaconis show that the mixing time has upper bound of order $n \log n$. In the same paper, they also prove that there exists a matching lower bound by applying directly the definition of total variation distance: they identify a set of permutations for which the probability under the stationary distribution and the distribution of the random walk differ significantly since the difference between the two probabilities is large. A useful lower bound of the mixing time can be found from Definition 1.1. The proof of the existence of two matching bounds for the mixing time allows Aldous and Diaconis to conclude that the top-to-random shuffling has cutoff at $n \log n$ and window of size n.

We will give more details about the upper bound of this card shuffle in Section 1.2, in which we introduce the coupling method to upper bound the mixing time.

Another card shuffle that exhibits cutoff is the random transposition shuffle, which consists of choosing two cards, independently and uniformly at random, and then transposing them if they are different. In [9], Diaconis and Shahshahani apply representation theory of the symmetric group and

Fourier analysis to find that the mixing time has upper bound at $\frac{1}{2}n \log n$. Using the definition of total variation distance, they also show that a matching lower bound exists, proving that the random transposition shuffle exhibits cutoff at $\frac{1}{2}n \log n$ with a window of order n.
Finally, in [4], Bayer and Diaconis study the riffle shuffle. This is the most common method to shuffle cards, and it consists of cutting the deck into two parts and riffling the two halves together. In their paper, Bayer and Diaconis derive formulas for the probability of seeing any permutation after t shuffles and use those expressions to show that the riffle shuffle on n cards has cutoff at $\frac{3}{2} \log_2 n$.

Another class of Markov chains that has received great attention is that of random walks on the hypercube, for which there exists an abundant literature showing cutoff. A hypercube, or cube of dimension n, is a graph whose vertices are the elements of $\mathbb{Z}_2^n$, i.e. binary strings of length n.
Consider the lazy simple random walk on the hypercube that, at each step, stays at the current position with probability $\frac{1}{2}$ and with probability $\frac{1}{2}$ chooses one of the n coordinates uniformly at random and flips it. In [8], it is shown, by using Fourier analysis and representation theory, that this random walk exhibits cutoff at $\frac{1}{2}n \log n$ with window of size n. In [21], the same result is showed using different tools. The lower bound is found using the Hamming weight, i.e. the number of coordinates equal to 1, at time t to bound the total variation distance between the distribution of the chain at time t and the stationary distribution.
The upper bound is established using coupling, so we will give more details about it in Section 1.2.
In [28], Nestoridi analyses the random walk on the hypercube that flips a fixed number k of coordinates at each step, showing that, if $k = o(n)$, it exhibits cutoff at $\frac{n}{2k} \log n$ with window of order $\frac{n}{2k}$. We give more details about this paper in Chapter 2, where we expand Nestoridi's results by analysing the random walk on the hypercube that flips a *random* number K of coordinates at each step.

More generally, we can view a cube of dimension n as a n-regular graph, i.e. a graph such that each vertex has exactly n neighbours. Lubetzky and Sly, in [25], consider generic d-regular graphs on n vertices. Exploiting the properties of regular graphs, they find some interesting results concerning cutoff. In particular, they prove that for any fixed $d \geq 3$ the simple random walk on a d-regular graph on n vertices has cutoff at $\frac{d}{d-2}\log_{d-1} n$ with window of size $\sqrt{\log n}$. Also, if d tends to infinity with n, then the window decreases to $\sqrt{\frac{\log n}{d \log d}}$.

Finally, in a recent paper [32], Salez considers Markov chains with negative curvature finding a sufficient condition for such chains to exhibit cutoff. He considers two definitions of curvature, the Ollivier-Ricci curvature and the Bakry-Émery curvature, but his result applies to both definitions indifferently. The sufficient condition requires comparing the orders of the mixing time and the relaxation time.

Definition 1.10. Let λ be the second largest eigenvalue in absolute value of the transition matrix, the *relaxation time* is defined as

$$t_{rel} = \frac{1}{1-\lambda}.$$

Theorem. *Consider a sequence of irreducible Markov chains with symmetric transition matrices and non-negative curvature. Suppose that for every* $\varepsilon \in (0,1)$ *we have*

$$\frac{(t_{rel}\log\Delta)^2}{t_{mix}(\varepsilon)} = o(1),$$

where

$$\Delta = \max\left\{\frac{1}{P(x,y)} : dist(x,y) = 1\right\}$$

and

$$dist(x,y) = \min\{k \in \mathbb{Z}_+ : P^k(x,y) > 0\}.$$

Then, the sequence exhibits cutoff.

The proof relies on the *entropic concentration phenomenon*, which ensures control on the variance of the *relative entropy*, a measure of the difference between a probability distribution and a reference probability measure. In the paper, Salez shows that mixing does not occur before the relative entropy is small but, once its level is low enough, mixing happens quickly. Salez also show that having a non-negative curvature is sufficient to establish the entropic concentration phenomenon proving that it is a sufficient condition to have cutoff.

We can also define the pre-cutoff phenomenon.

Definition 1.11. We say that family of Markov chains exhibits *pre-cutoff* if it satisfies the condition

$$\sup_{0<\varepsilon<\frac{1}{2}} \limsup_{n\to\infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} < \infty.$$

In other words, there exist c_0, c_1 such that

$$\liminf_{n\to\infty} d_n(ct_{mix}^{(n)}) = 1,$$

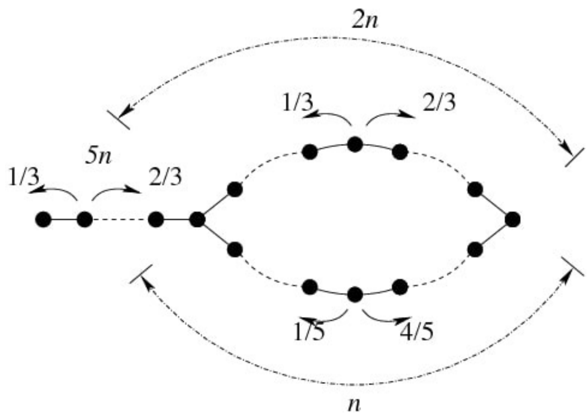
$$\limsup_{n\to\infty} d_n(ct_{mix}^{(n)}) = 0,$$

for $c < c_0$ and $c > c_1$.

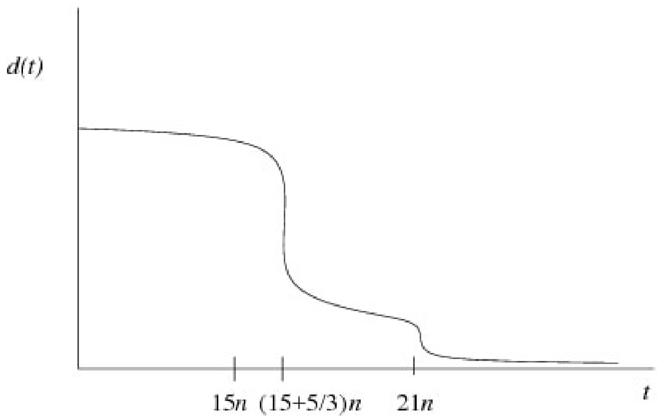
This means that $t_{mix}^{(n)}(\varepsilon)$ and $t_{mix}^{(n)}(1-\varepsilon)$ are comparable and that pre-cutoff is weaker than cutoff, for which the ratio $\frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)}$ tends to 1 as $n \to \infty$ for all $\varepsilon \in (0, \frac{1}{2})$.
The following example helps to provide a clearer idea of what pre-cutoff is and to explain why pre-cutoff is weaker than cutoff.

Example 1.12. At the American Institute of Mathematics (AIM) research workshop "Sharp Thresholds for Mixing Times" organised in December 2004, Aldous constructed a random walk that has pre-cutoff but does not exhibit cutoff. Figure 1.2 shows the transition probabilities and the graph of the total variation distance for this random walk. It can be seen in Figure 1.2a that the stationary distribution has a geometric growth from left to right and the random walk mixes when it reaches the right-most point. If the random walk starts at the left-most point, it takes about $15n$ steps to reach the fork. From there, it takes about $\frac{5}{3}n$ steps to reach the right-most point using the bottom path and about $6n$ steps using the top path. Figure 1.2b shows that after time $\left(15+\frac{5}{3}\right)n$ the total variation distance drops by about $\frac{3}{4}$, while we have to wait time $(15+6)n$ to have the distance drop by the remaining $\frac{1}{4}$. Then, the ratio $\frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)}$ in Definition 1.11 is bounded as $n \to \infty$ implying that we have pre-cutoff, but it does not tend to 1, so this random walk does not exhibit cutoff.

□



(a) Transition probabilities



(b) Total variation distance

Fig. 1.2: The random walk constructed by Aldous in 2004. The images have been taken from [21]

As we already observed, sufficient conditions to prove cutoff using the total variation distance are not known, but necessary conditions exist. A necessary condition is related to pre-cutoff and the relaxation time of a Markov chain. The following proposition, which can be found in [21], gives a necessary condition.

Proposition 1.13. *For a sequence of irreducible aperiodic Markov chains with relaxation times $t_{rel}^{(n)}$ and mixing times $t_{mix}^{(n)}$, if $\frac{t_{mix}^{(n)}}{t_{rel}^{(n)}}$ is bounded above, then there is no pre-cutoff.*

This proposition can be used to prove that a sequence of Markov chains does not exhibit cutoff. For example, consider the lazy random walk on the cycle $\mathbb{Z}_n$ that at each step moves clockwise with probability $\frac{1}{4}$, anti-clockwise with probability $\frac{1}{4}$, and with probability $\frac{1}{2}$ stays in its position. In [8], it is showed that the mixing time is of order n^2, and in [21] a detailed calculation of the eigenvalues of this random walk shows that the relaxation time also has order n^2, so by Proposition 1.13, there is no pre-cutoff, which implies that there is no cutoff.

1.2 Coupling

The coupling method is one of the main probabilistic techniques in the study of convergence of Markov chains, and it consists of comparing two probability measures on a measurable space. To formally define coupling, we refer to Lindvall [23], where a measurable space is denoted as the couple $(E, \mathcal{E})$ of the state space E with its Borel sets $\mathcal{E}$.

Definition 1.14. A *coupling of two probability measures* P and P' on a measurable space $(E, \mathcal{E})$ is a probability measure $\hat{P}$ on $(E^2, \mathcal{E}^2)$ such that

$$P = \hat{P}\pi^{-1} \quad \text{and} \quad P' = \hat{P}\pi'^{-1},$$

where $\pi(x, x') = x$ and $\pi'(x, x') = x'$ for $(x, x') \in E^2$.
In other words, P and P' are marginal distributions of $\hat{P}$.

This definition describes coupling in terms of probability measures but, since in the thesis we are interested in constructing couplings of random processes, we need a definition of coupling that better adapts to working with random elements in the state space.
We can define a random element in $(E, \mathcal{E})$ as a quadruple $(\Omega, \mathcal{F}, \mathbb{P}, X)$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is the underlying probability space and $X \in \mathcal{F}/\mathcal{E}$, i.e. the class of measurable maps from Ω to E. We can define coupling in the following way.

Definition 1.15. A *coupling of two random elements* $(\Omega, \mathcal{F}, \mathbb{P}, X)$ and $(\Omega', \mathcal{F}', \mathbb{P}', X')$ in $(E, \mathcal{E})$ is a random element $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}}, (\hat{X}, \hat{X}'))$ in $(E^2, \mathcal{E}^2)$ such that

$$X \overset{D}{=} \hat{X} \quad \text{and} \quad X' \overset{D}{=} \hat{X}',$$

i.e. X has the same distribution as $\hat{X}$ and X' has the same distribution as $\hat{X}'$.

So, in terms of Definition 1.14, $\hat{\mathbb{P}}(\hat{X},\hat{X}')^{-1}$ is a coupling of $\mathbb{P}X^{-1}$ and $\mathbb{P}'X'^{-1}$.

If we assume that the random variables X and X' are coupled, we can exploit the coupling to produce an upper bound on the total variation distance between the distributions of the two random variables.

Proposition 1.16. *Let X and X' be two random variables with distributions μ and μ' respectively, and assume that X and X' are coupled. Then*

$$||\mu-\mu'||_{TV} \leq \mathbb{P}(X \neq X').$$

In the inequality, we should strictly use $\hat{X}$ and $\hat{X}'$ since the result holds for two coupled random variables, but, by convention, we just write X and X'.

Proof. Let $A \in \mathcal{E}$, then

$$\begin{aligned}\mathbb{P}(X\in A)-\mathbb{P}(X'\in A) &= \mathbb{P}(X\in A, X=X')+\mathbb{P}(X\in A, X\neq X')\\ &\quad -\mathbb{P}(X'\in A, X=X')-\mathbb{P}(X'\in A, X\neq X')\\ &= \mathbb{P}(X\in A, X\neq X')-\mathbb{P}(X'\in A, X\neq X')\\ &\leq \mathbb{P}(X\neq X').\end{aligned}$$

Applying Definition 1.1 completes the proof. □

Now that we have defined coupling of random variables, we can move to random processes. Let $(\hat{X},\hat{X}')$ be a coupling of two random processes $X=(X_t)_{t=0}^{\infty}$ and $X'=(X_t')_{t=0}^{\infty}$ in $(E,\mathcal{E})$, and assume there exists a random time T such that

$$\hat{X}_t=\hat{X}_t', \qquad \text{for} \qquad t\geq T.$$

We call T a *coupling time.* Using the coupling time and observing that $\{\hat{X}_t\neq\hat{X}_t'\}\subseteq\{T>t\}$, Proposition 1.16 can be extended in the following theorem, showed in [23], which states that there exists a convenient relationship between the coupling time and the total variation distance.

Theorem 1.17. *Let $(\hat{X},\hat{X}')$ be a coupling of two random processes $X=(X_t)_{t=0}^{\infty}$ and $X'=(X_t')_{t=0}^{\infty}$ in $(E,\mathcal{E})$. Then*

$$||\mathbb{P}(X_t\in\cdot)-\mathbb{P}(X_t'\in\cdot)||_{TV}\leq\mathbb{P}(T>t). \tag{1.1}$$

Now, let $X = (X_t)_{t=0}^{\infty}$ and $X' = (X'_t)_{t=0}^{\infty}$ be Markov chains with transition matrix P and stationary distribution π such that $X_0 = x$ and $X'_0 \sim \pi$. Then equation (1.1) gives

$$||P^t(x,\cdot) - \pi(\cdot)||_{TV} \leq \mathbb{P}(\tau_{couple} > t).$$

This inequality makes coupling one of the main tools in the estimation of rates of convergence of Markov chains. We also have the following corollary.

Corollary 1.18. *Suppose that for each pair of states $x, x' \in \Omega$ there is a coupling $(\hat{X}, \hat{X}')$ of two Markov chains X and X' with transition matrix P such that $X_0 = x$ and $X'_0 = x'$. Let τ_{couple} be the coupling time. Then*

$$d(t) \leq \max_{x,x' \in \Omega} \mathbb{P}(\tau_{couple} > t).$$

The first time the two chains meet is called the *coupling time*, and it is defined as

$$\tau_{couple} := \min\{t : \hat{X}_t = \hat{X}'_t\}.$$

We can also construct the coupling such that the two chains run together at all times after the coupling time, i.e.

$$\hat{X}_t = \hat{X}'_t, \qquad \text{for all } t \geq \tau_{couple}.$$

Definition 1.19. We say that a coupling is *successful* if the coupling time is finite almost surely, i.e. if

$$\mathbb{P}(\tau_{couple} < \infty) = 1.$$

Example 1.20 (Lazy random walk on the n-cycle)**.** Consider the random walk on the n-cycle $\mathbb{Z}_n = \{0, \ldots, n-1\}$ that, at each step, with probability $\frac{1}{4}$ moves clockwise, with probability $\frac{1}{4}$ moves counterclockwise, and with probability $\frac{1}{2}$ stays in the current state.

Let $X = (X_t)_{t=0}^{\infty}$ be the Markov chain with transition matrix

$$P(j,k) = \begin{cases} \frac{1}{4}, & \text{if } k \equiv j+1 \ (mod\ n), \\ \frac{1}{4}, & \text{if } k \equiv j-1 \ (mod\ n), \\ \frac{1}{2}, & \text{if } k \equiv j \ (mod\ n), \\ 0 & \text{otherwise.} \end{cases}$$

This random walk is irreducible and aperiodic, so we can apply Theorem 1.5 and conclude that the random walk converges to its stationary distribution. In [21], the following coupling is constructed. Consider two lazy random walks on the n-cycle $X = (X_t)_{t=0}^{\infty}$ and $X' = (X'_t)_{t=0}^{\infty}$ with transition matrix P and such that $X = x$ and $X' = x'$. To ensure that one does not jump over the other when they are at unit distance, the two random walks are coupled so as to never move at the same time. At each step, a fair coin is tossed, independently from all previous tosses. If the coin lands heads up, X moves one step left with probability $\frac{1}{2}$ or one step right with probability $\frac{1}{2}$, otherwise X' moves one step left with probability $\frac{1}{2}$ or one step right with probability $\frac{1}{2}$. The two chains move together after the first time they land on the same state.

Let $D = (D_t)_{t=0}^{\infty}$ be the clockwise distance between X and X', then D is a simple random walk on $\{0, \ldots, n\}$ with absorbing states 0 and n. If $\tau = \min\{t \geq 0 : D_t \in \{0, n\}\}$, then τ is a coupling time and, since D models the gambler's ruin problem ([13]), we know that $\mathbb{E}[\tau] = D_0(n - D_0)$.

By Corollary 1.18, using Markov's inequality yields

$$d(t) \leq \max_{x,x' \in \mathbb{Z}_n} \mathbb{P}(\tau > t) \leq \frac{\max_{x,x' \in \mathbb{Z}_n} \mathbb{E}[\tau]}{t} \leq \frac{n^2}{4t}.$$

If $t = n^2$, then $d(t) \leq \frac{1}{4}$, so $t_{mix} \leq n^2$. □

Example 1.21 (Simple random walk on the hypercube)**.** Consider again the simple random walk on $\mathbb{Z}_2^n$. Adjacent vertices on the hypercube differ in exactly one coordinate. Consider the random walk that starts at a vertex $x = (x_1, \ldots, x_n)$ and, at each step, chooses a coordinate $i \in \{0, \ldots, n\}$ uniformly at random and flips it, where flipping coordinate $i = 0$ means that the random walk stays in the current position while flipping coordinate

$i = 1, \dots, n$ represents a step to one of the n neighbours.
In [1], Aldous applies the coupling method to upper bound the mixing time of this random walk. He constructs the following coupling.
Let X and X' be two simple random walks on $\mathbb{Z}_2^n$ started from the states x and x' respectively. Let $X_t(i)$ denote the ith coordinate of X_t. Define

$$U_t = \{i \in \{1, \dots, n\} : X_t(i) \neq X'_t(i)\}$$

the set of unmatched coordinates at time t, and let $L_t = |U_t|$.
At time $t+1$, we choose $i \in \{0, \dots, n\}$ uniformly at random and obtain X_{t+1} by flipping $X_t(i)$. X'_{t+1} is defined according to the following rules.

i. If $L_t = 0$, then we flip $X'_t(i)$

ii. If $L_t = 1$, and $U_t = \{j\}$

   a. if $i = 0$, then we flip $X'_t(j)$
   b. if $i = j$, then we flip $X'_t(0)$
   c. if $X_t(i) = X'_t(i)$, then we flip $X'_t(i)$ to maintain the match at time $t+1$

iii. If $L_t > 1$,

   a. if $i \notin U_t$ or $i = 0$, then we flip $X'_t(i)$
   b. if $i \in U_t$, then we choose $j \in U_t$ with $j \neq i$ and we flip $X'_t(j)$

Intuitively, this coupling ensures that, if $X_t(i) = X'_t(i)$, then the match is maintained at time $t+1$, otherwise we choose another unmatched coordinate of X'_t and we flip it to create two new matches. If the only unmatched coordinate is j, then X and X' are coupled as soon as $i = 0$ or $i = j$. By construction, using this coupling implies that at time $t+1$ we have $L_{t+1} \leq L_t$, and $\tau = \min\{t \geq 0 : L_t = 0\}$ is the coupling time.
Using this coupling and applying Theorem 1.17, Aldous shows that

$$t_{mix}(\varepsilon) \leq \frac{1}{2} n \log n, \qquad \varepsilon \in (0, 1).$$

□

A class of random walks to which coupling has been successfully applied is represented by card shufflings. In [21], the coupling method is applied to the random transposition shuffle. Consider a deck of n cards, we label the cards with numbers $1, \dots, n$. At each step, we choose a card i and, independently, a position $j \in \{1, \dots, n\}$. Then, we transpose card i with the card at position j; if i already occupied position j, the deck is left unchanged. To couple two decks, we use the same choices of i and j. If m_t is the number of matching cards at time t, we have the following possibilities.

- If i is in the same position in both decks and position j is occupied by the same card in both decks, then $m_{t+1} = m_t$.
- If i is in the same position in both decks but j is occupied by different cards, then $m_{t+1} = m_t$.
- If i is in different positions but position j is occupied by the same card in both decks, then applying the shuffling breaks a match but creates a new one, so $m_{t+1} = m_t$.
- If i is in different positions and position j is occupied by different cards, then applying the shuffing creates at least one new match with a maximum of three new matches.

Applying Theorem 1.17 to this coupling yields an upper bound of the mixing time of order n^2. The authors show that this result holds for any initial arrangement of the two decks, so in this particular case it is not necessary to assume that the second deck is uniformly mixed at the start.

In this case coupling is not the most efficient method, and a sharper bound can be found using other tools. Lindvall also observes in [23] that this approach is not always useful in general, and sometimes other methods prove to be more efficient. In some cases, strong uniform times may be involved.

Let X be a random walk with uniform stationary distribution. A *strong uniform time* is a randomised stopping time T such that X_T is distributed uniformly, and X_T and T are independent. In [2], Aldous and Diaconis prove that, if T is a strong uniform time, then the total variation distance between the distribution of the deck at time t and the stationary distribution is bounded by $\mathbb{P}(T > t)$. In their paper, they use this approach to find an upper bound of the mixing time of the top-to-random shuffle on n cards, the random walk that, at each step, inserts the top card randomly into the deck. Aldous and Diaconis define T as the first time when the original top card is inserted into the deck by a shuffle and, showing that T is a strong uniform time, they find an upper bound of order $n \log n$ for the mixing time.

1.2.1 Co-adapted coupling

In this section, we introduce the notions of co-adapted coupling and maximal coupling.

Definition 1.22 (Co-adapted coupling)**.** Let $(X, \hat{X})$ be a coupling of two Markov processes $X = (X_t)_{t=0}^{\infty}$ and $\hat{X} = (\hat{X}_t)_{t=0}^{\infty}$. We say that the coupling is *co-adapted* if there exists a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$ such that X and $\hat{X}$ are Markov processes with respect to $(\mathcal{F}_t)_{t=0}^{\infty}$.

A non co-adapted coupling is a coupling constructed so that the way one of the processes evolves depends on how the other process evolves in the future. In [26], Matthews constructs a non co-adapted coupling for the simple walk on the hypercube. As we will see more in detail in Chapter 2, to find an upper bound of the total variation distance, Matthews constructs a non co-adapted coupling and uses a strong uniform time to prove that the random walk needs $\frac{1}{4}n \log n$ steps to have a small variation distance with upper bound $\frac{1}{2}n \log n$.

Definition 1.23 (Maximal coupling)**.** We say that a coupling is *maximal* if it realises an equality in the coupling inequality (1.1) for all $t \geq 0$, i.e.

$$||\mathbb{P}(X_t \in \cdot) - \mathbb{P}(X'_t \in \cdot)||_{TV} = \mathbb{P}(T > t).$$

The existence of a maximal coupling has been established in several cases. The following theorem ([23]) proves the existence of maximal coupling for probability measures.

Theorem 1.24. *Let μ and μ' be two probability measures on a measurable space $(E, \mathcal{E})$. Then, there exists a coupling $(\hat{Z}, \hat{Z}')$ such that*

(i) $||\mu - \mu'||_{TV} = \mathbb{P}(\hat{Z} \neq \hat{Z}')$

(ii) $\hat{Z}$ and $\hat{Z}'$ are independent conditioned on $\{\hat{Z} \neq \hat{Z}'\}$.

Proof. Let $\lambda = \mu + \mu'$ and

$$g = \frac{d\mu}{d\lambda}, \qquad g' = \frac{d\mu'}{d\lambda}.$$

From Proposition 1.4, we know that

$$||\mu - \mu'||_{TV} = 1 - \int g \wedge g' d\lambda.$$

So, if we can construct a coupling $(\hat{Z}, \hat{Z}')$ such that $\mathbb{P}(\hat{Z} \neq \hat{Z}') = \int g \wedge g' d\lambda$, then this coupling would be maximal. Let Q be the subprobability $dQ = g \wedge g' d\lambda$, and let γ be its total mass. Observe that if $\gamma = 1$, then we would have $\mu = \mu'$, and defining $\hat{Z}' = \hat{Z}$ would be enough to have the result. So, we restrict to the case $\gamma < 1$. We want to construct a coupling $\hat{\mu}$ of μ and μ' such that $\hat{\mu}(\Delta) = \gamma$, where $\Delta = \{(x,x) : x \in E\}$ is the diagonal in E^2. Let $\varphi : E \to E^2$ defined as $\varphi(x) = (x,x)$, and let $\hat{Q} = Q\varphi^{-1}$, then $\hat{Q}$ has mass γ concentrated on Δ. Let

$$\nu = \mu - Q, \qquad \text{and} \qquad \nu' = \mu' - Q,$$

and

$$\hat{\mu} = \frac{\nu \times \nu'}{1-\gamma} + \hat{Q}.$$

Then for all $A, A' \in \mathcal{E}$,

$$\begin{aligned} \hat{\mu}(A \times E) &= \frac{\nu(A) \cdot (\mu'(E) - Q(E))}{1-\gamma} + \hat{Q}(A \times E) \\ &= \frac{\nu(A) \cdot (1-\gamma)}{1-\gamma} + Q\varphi^{-1}(\{(x,x) : x \in A\}) \\ &= \nu(A) + Q(A) = \mu(A), \end{aligned}$$

and in the same way $\hat{\mu}(E \times A') = \mu'(A')$, so $\hat{\mu}$ is a coupling of μ and μ'. So, if we let $(\hat{Z}, \hat{Z}')$ be any pair with distribution $\hat{\mu}$, part (i) follows.
Finally, to show part (ii) we need to prove that for any $A, A' \in \mathcal{E}$

$$\mathbb{P}(\hat{Z} \in A, \hat{Z}' \in A' | \hat{Z} \neq \hat{Z}') = \mathbb{P}(\hat{Z} \in A | \hat{Z} \neq \hat{Z}') \times \mathbb{P}(\hat{Z}' \in A' | \hat{Z} \neq \hat{Z}').$$

Now, $\{\hat{Z} \neq \hat{Z}'\} = \Delta^c$ and from the definition of $\hat{\mu}$ we have

$$\hat{\mu}(A \times A'|\Delta^c) = \frac{\nu(A)}{1-\gamma} \times \frac{\nu'(A')}{1-\gamma},$$

since conditioning on Δ^c implies $\hat{Q}(A \times A') = 0$. □

Analogous results have been found for stochastic processes, but in this case finding a maximal coupling is much more difficult since it must be constructed such that it is maximal for all times.
The existence of a maximal coupling has been established by Griffeath in [15] for homogeneous Markov chains with countable state space.

Theorem (Griffeath)**.** *For any homogeneous Markov chain X with countable state space there exists a maximal coupling $\tilde{X}$.*

Goldstein extended that result to general discrete Markov chains under some condition on tail σ-fields in [14]. Pitman also shows the existence of a maximal coupling in [29] using randomised stopping times to simplify Griffeath's construction.

In [24], Lindvall and Rogers consider multidimensional Brownian motions and construct a coupling between two such processes. As they observe in their paper, when we couple two one-dimensional continuous processes, it is convenient to make them move together after they meet the first time, as two one-dimensional continuous processes cannot pass each other without hitting each other. This is not true in general for multi-dimensional processes, so the construction of the coupling needs to be adapted in order to derive sufficient conditions for successful coupling.
The coupling they consider for two multi-dimensional Brownian motions is the reflection coupling. Consider two multi-dimensional Brownian motions starting from two different states $\mathbf{x}$ and $\mathbf{y}$, we run one of the processes as the reflection of the other with respect to the hyperplane between $\mathbf{x}$ and $\mathbf{y}$. Lindvall and Rogers show that the reflection coupling is a co-adapted maximal coupling of Brownian motions on $\mathbb{R}^n$, and Hsu and Sturm, in [16], prove that it is the unique maximal coupling in the class of co-adapted couplings, but that the uniqueness does not hold if we also consider non co-adapted couplings. The same result has been generalised to Brownian motion on a Riemannian manifold by Kuwada in [20].

Example 1.25 (One-dimensional Brownian motion)**.** In [16], Hsu and Sturm show that the reflection coupling is the unique co-adapted maximal coupling of Brownian motions on $\mathbb{R}^n$. Consider the case $n = 1$.
Let

$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{|x-y|^2}{2t}}$$

be the transition density function of a one-dimensional Brownian motion. Let X_t and Y_t be Brownian motions on $\mathbb{R}$ with starting states $X_0 = 0$ and $Y_0 = y$. Following Hsu and Sturm, we can define the reflection coupling of X_t and Y_t as a process $(B_t, y - B_t)$, where B_t is a standard Brownian motion on $\mathbb{R}$. In particular, due to the symmetry of the density function, $y - B_t$ is also a Brownian motion started at y. Figure 1.3 illustrates an example of reflection coupling $(B_t, y - B_t)$.

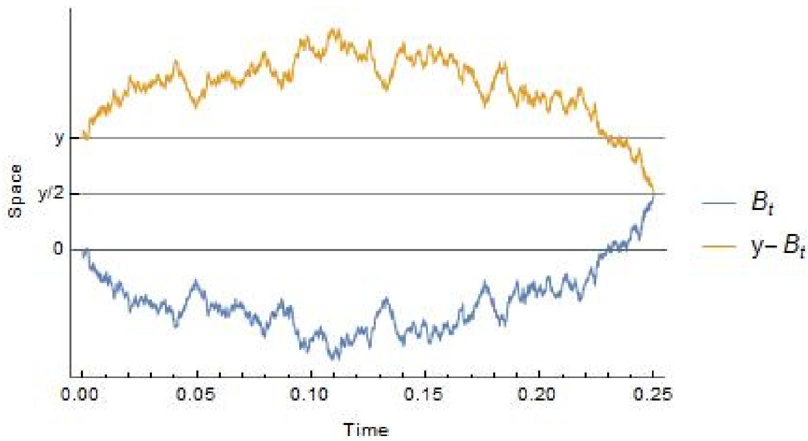


Fig. 1.3: Reflection coupling $(B_t, y - B_t)$

The coupling time is defined as the first time when $B_t = y - B_t$, i.e. as the hitting time of $\frac{y}{2}$

$$\tau_{\frac{y}{2}} = \inf \left\{ t \geq 0 : B_t = \frac{y}{2} \right\}.$$

By Definition 1.23, we need to verify that the total variation distance between X_t and Y_t agrees with $\mathbb{P}(\tau_{\frac{y}{2}} > t)$. We find that

$$||X_t - Y_t||_{TV} = ||B_t - (y - B_t)||_{TV} = \max_{A \subset \mathbb{R}} |\mathbb{P}(B_t \in A) - \mathbb{P}(y - B_t \in A)|$$
$$= \max_{A \subset \mathbb{R}} \frac{1}{\sqrt{2\pi t}} \left| \int_A e^{-\frac{|z|^2}{2t}} - e^{-\frac{|z-y|^2}{2t}} dz \right| = \frac{1}{\sqrt{2\pi t}} \left| \int_{-\infty}^{\frac{y}{2}} e^{-\frac{|z|^2}{2t}} - e^{-\frac{|z-y|^2}{2t}} dz \right|$$

$$= \frac{2}{\sqrt{2\pi t}} \int_0^{\frac{y}{2}} e^{-\frac{z^2}{2t}} dz = \text{Erf}\left(\frac{y}{2\sqrt{2t}}\right),$$

where Erf is the error function that comes out when integrating the standard normal distribution, and it is defined as $\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz$.
To calculate $\mathbb{P}(\tau_{\frac{y}{2}} > t)$, we use the Reflection principle for Brownian motion, which can be found in [10].

Proposition 1.26 (Reflection principle)**.** *Let* $a > 0$ *and let* $\tau_a = \inf\{t : B_t = a\}$*, where* B_t *is a Brownian motion started at* 0*, then*

$$\mathbb{P}(\tau_a \leq t) = 2\mathbb{P}(B_t \geq a).$$

From Proposition 1.26, we have

$$\mathbb{P}(\tau_{\frac{y}{2}} \leq t) = 2\mathbb{P}(B_t \geq \frac{y}{2}) = \frac{2}{\sqrt{2\pi t}} \int_{\frac{y}{2}}^{\infty} e^{-\frac{z^2}{2t}} dz = \frac{1}{\sqrt{2\pi t}} \left(\int_{\frac{y}{2}}^{\infty} e^{-\frac{z^2}{2t}} dz + \int_{-\infty}^{-\frac{y}{2}} e^{-\frac{z^2}{2t}} dz \right)$$
$$= \frac{1}{\sqrt{2\pi t}} \left(\int_{-\infty}^{\infty} e^{-\frac{z^2}{2t}} dz - \int_{-\frac{y}{2}}^{\frac{y}{2}} e^{-\frac{z^2}{2t}} dz \right) = 1 - \frac{2}{\sqrt{2\pi t}} \int_0^{\frac{y}{2}} e^{-\frac{z^2}{2t}} dz.$$

Then, we have

$$\mathbb{P}(\tau_{\frac{y}{2}} > t) = \frac{2}{\sqrt{2\pi t}} \int_0^{\frac{y}{2}} e^{-\frac{z^2}{2t}} dz = \text{Erf}\left(\frac{y}{2\sqrt{2t}}\right),$$

so by Definition 1.23, we can conclude that reflection coupling is a maximal co-adapted coupling. □

In general, a maximal coupling is not co-adapted. Non co-adapted couplings are difficult to study because, as we mentioned above, the evolution of one of the two coupled processes depends on the future of the other. For this reason, we are interested in constructing co-adapted couplings which satisfy other optimality properties.

Definition 1.27 (Optimal coupling)**.** We say that a coupling is *tail-optimal* if it minimises the tail probability of the coupling time $\mathbb{P}(\tau_{couple} > t)$ simultaneously for all $t > 0$.
We say that a coupling is *Laplace-optimal* if it maximises the Laplace transform of the coupling time $\mathbb{E}[e^{-\gamma\tau_{couple}}]$ for all $\gamma > 0$.

We say that a coupling is *mean-optimal* if it minimises the expectation of the coupling time $\mathbb{E}[\tau_{couple}]$.

Connor and Jacka, in [7], consider the symmetric random walk on the hypercube $\mathbb{Z}_2^n$ and construct a co-adapted coupling of two such random walks, which is very similar to the Aldous coupling described in Example 1.21. The difference between the two couplings is that, while Aldous runs the coupling to match the coordinates of the two random walks until they differ only in 1 coordinate, Connor and Jacka first make the random walk run independently until the number of unmatched coordinates is even, then they couple the unmatched coordinates in pairs, so that they obtain two new matches at each step. They show that their coupling is optimal. However, since the coupling time that they obtain is greater than the mixing time, it follows that the coupling is not maximal.
A similar result related to Brownian motion is provided by Kendall in [18]. He studies Brownian motion with its local time at 0, i.e. the time spent by the Brownian motion at 0, and constructs a coupling based on a combination of reflection and synchronised couplings. He describes a co-adapted coupling, and he proves that it is optimal. Kendall also observes that the moment generating functions of the coupling times of his coupling and the maximal coupling are different, so the coupling he constructs is optimal but not maximal.

1.2.2 Path coupling

Other methods have been derived from coupling to help the study of the convergence of Markov chains. One of these methods is path coupling, which was introduced by Bubley and Dyer in [5].
Instead of studying a coupling for all pairs of states, path coupling considers pairs of adjacent states in some path between two arbitrary states. If it is shown that for all pairs of adjacent states in the path, the two Markov chains of the coupling get closer in expectation then, by linearity of expected value and the triangle inequality, the two chains will come closer in expectation on the entire path.
In their paper, Bubley and Dyer apply the path coupling technique to prove two convergence theorems for Markov chain. The statements of the two

theorems are technical, but the proofs offer a good demonstration of the advantage that path coupling can provide in the estimation of the rate of convergence of Markov chains. In this final part of the first chapter, we report only one of them.

Let V and C be finite sets and define $n = |V|$, and let $\Omega = C^V$ be the set of functions from V to C. Let X be a Markov chain with state space Ω and stationary distribution π. For $x \in \Omega$, $v \in V$ and $c \in C$, we denote by $x_{v\to c}$ the state resulting from making the transition at x associated with the pair (v, c), so

$$x_{v\to c}(w) = \begin{cases} c & \text{if } w = v, \\ x(w) & \text{otherwise.} \end{cases}$$

Now, consider two such Markov chains $X = (X_t)_{t=0}^{\infty}$ and $X' = (X'_t)_{t=0}^{\infty}$. Let μ_t be the distribution of X_t and let $X'_0 \sim \pi$.

For $x, x' \in C^V$, the Hamming distance between x and x', denoted by $H(x, x')$, is defined as the number of $v \in V$ such that $x(v) \neq x'(v)$. Let $h = H(x, x')$.

We consider a path between x and x', i.e. a sequence of adjacent states $x = z_0, z_1, \dots, z_h = x'$ in C^V such that $H(z_{a-1}, z_a) = 1$ for all $a = 1, \dots, h$. Now, a coupling $(\hat{X}, \hat{X}')$ is defined at (x, x') by choosing the next state (y, y') according to the following rules.

i. Choose $v \in V$ according to a fixed distribution J, and $c_0 \in C$ according to $K_{z_0,v}$, where $K_{z_i,j}$ is a distribution on C depending only on the current state $z_i \in C^V$ and $j \in V$.

ii For $a = 1, \dots, h$, define $c_a = c_{a-1}$ with probability $\frac{K_{z_a,v}(c_{a-1})}{K_{z_{a-1},v}(c_{a-1})}$, otherwise choose c_a according to the distribution $(K_{z_a,v} - K_{z_{a-1},v})^+$.

iii. Define $y = x_{v\to c_0}$ and $y' = x'_{v\to c_h}$.

Denote $w_a = (z_a)_{v\to c_a}$, then we have $y = w_0$ and $y' = w_h$. Assume that z_{a-1} and z_a differ only at i, then, if J is a fixed distribution on V,

$$\begin{aligned} \mathbb{E}[H(w_a, w_{a-1})] &= 1 - \mathbb{P}(H(w_a, w_{a-1}) = 0) + \mathbb{P}(H(w_a, w_{a-1}) = 2) \\ &= 1 - J(i)\mathbb{P}(c_a = c_{a-1}|v = i) + \sum_{j\neq i} J(j)\mathbb{P}(c_a \neq c_{a-1}|v = j) \end{aligned} \tag{1.2}$$

$$= 1 - J(i)(1 - ||K_{z_a,i} - K_{z_{a-1},i}||_{TV}) + \sum_{j \neq i} J(j)||K_{z_a,j} - K_{z_{a-1},j}||_{TV}$$

where to obtain equality (1.2) we have to observe that, due to the coupling, $H(w_a, w_{a-1}) = 0$ if we choose $v = i$ and $c_a = c_{a-1}$, while to have $H(w_a, w_{a-1}) = 2$ we need $v \neq i$ and $c_a \neq c_{a-1}$.
If we define β such that

$$1 - J(i)(1 - ||K_{z_a,i} - K_{z_{a-1},i}||_{TV}) + \sum_{j \neq i} J(j)||K_{z_a,j} - K_{z_{a-1},j}||_{TV} \leq \beta$$

and consider the whole path between y and y', we can apply the linearity of expectation and the triangle inequality to obtain

$$\mathbb{E}[H(y, y')] \leq \mathbb{E}\left[\sum_{a=1}^{h} H(w_a, w_{a-1})\right] = \sum_{a=1}^{h} \mathbb{E}[H(w_a, w_{a-1})] \leq \sum_{a=1}^{h} \beta = \beta H(x, x').$$

So, after t steps we have

$$\mathbb{E}[H(X_t, X'_t)] \leq \beta^t n,$$

and applying Proposition 1.16, we deduce

$$||\mu_t - \pi||_{TV} \leq \beta^t n.$$

We have the following result.

Theorem 1.28 (General path coupling). *Let $\Omega = C^V$ with $n = |V|$ and let $(\hat{X}, \hat{X}')$ be a coupling of two Markov chains $X = (X_t)_{t=0}^{\infty}$ and $X' = (X'_t)_{t=0}^{\infty}$ with state space Ω and stationary distribution π. Let μ_t be the distribution of X_t and $X'_0 \sim \pi$. Define*

$$\beta = \max_{x,x' \in \Omega, i \in V} \left\{ 1 - J(i) + \sum_{j \in V} J(j)||K_{x,j} - K_{x',j}||_{TV} \,\middle|\, x' = x_{v \to c} \text{ for some } c \in C \text{ and } x' \neq x \right\},$$

i.e. β is an upper bound of the expected distance between adjacent states after one step. Then, if $\beta < 1$ and $t \geq \lceil \frac{\ln(n\varepsilon^{-1})}{\ln(\beta^{-1})} \rceil$, then

$$||\mu_t - \pi||_{TV} \leq \varepsilon.$$

In the theorem, β is an upper bound on the expected distance between adjacent states after one step, so the assumption $\beta < 1$ is essential to have that the two random walks X and X' started at any two states get closer in expectation.

2. MIXING TIME FOR A RANDOM WALK ON THE HYPERCUBE

We consider a hypercube of dimension n as a graph whose vertices are the elements of the set $\mathbb{Z}_2^n = \{(x_1, \dots, x_n) : x_i \in \{0,1\}, i = 1, \dots, n\}$. In particular, adjacent vertices differ in exactly one coordinate. A typical random walk on this structure is the *simple random walk*, that at each step chooses one coordinate uniformly at random and flips it. We know that it is an irreducible and transitive Markov chain, and, if we consider the lazy version, it is also aperiodic. Thus, it converges to its stationary distribution, which is the uniform distribution.

As we have seen in the introduction, to study the convergence of a random walk we aim at bounding the mixing time and, possibly, at showing that the random walk exhibits cutoff. There are many results proving bounds on the mixing time of different random walks on the hypercube using a variety of methods, and for several of these random walks a cutoff has also been proved. For this reason, random walks on the hypercube are widely studied and still receive a great deal of attention as studying these processes could contribute to the development of a stronger literature about cutoff.

2.1 Previous results

As we mentioned at the beginning of this chapter, the simple random walk on $\mathbb{Z}_2^n$ is periodic, so it has to be modified to become aperiodic. In this section, we present some of the studies that have established bounds on the mixing time using two different lazy versions of the simple random walk.
As explained in Chapter 1, Aldous [1] considers the lazy version that moves from the current vertex to one of the n neghbours, each with probability $\frac{1}{n+1}$, and stays still with probability $\frac{1}{n+1}$. He defines a co-adapted coupling and applies equation (1.1) to prove that the mixing time has upper bound

$\frac{n}{2}\log n$. In the same paper, he also gives an exact formula for the total variation distance of this random walk showing that the mixing time is $\frac{1}{4}n\log n$. In [8], Diaconis considers the same lazy version and presents another approach to study the mixing time of random walks that involves the calculation of the eigenvalues of the process using irreducible representations.

In general, a representation of a group G is a homomorphism $\rho : G \to GL(V)$ that assigns to every element of the group an invertible matrix over a vector space V such that $\rho(st) = \rho(s)\rho(t)$ for all $s, t \in G$. If $W \subset V$ is a subspace such that $\rho(s)W \subset W$ for all $s \in G$, then ρ restricted to W is a subrepresentation. We always have two trivial subrepresentations given by the zero space and the subspace $W = V$. We say that the representation ρ is irreducible if it admits no non-trivial subrepresentations. If P is a probability on G, the *Fourier transform* of P at the representation ρ is the matrix

$$\hat{P}(\rho) = \sum_{s} P(s)\rho(s).$$

Diaconis also shows that the Fourier transforms of P at the irreducible representations determine P. In the case of the hypercube, $G = \mathbb{Z}_2^n$ and, since this is an abelian group, all irreducible representations are 1-dimensional representations associated to $x \in \mathbb{Z}_2^n$ given by $\rho_x(y) = 1^{x\cdot y}$ where $y \in \mathbb{Z}_2^n$. Using this representation, Diaconis calculates the Fourier transform of the transition matrix, showing that the mixing time has upper bound of order $\frac{1}{4}n\log n$. Applying Chebychev's inequality, he also shows that a lower bound of the same order can be found.

In the same book, Diaconis also considers another lazy version of the simple random walk that moves to one of the n neighbours, each with probability $\frac{1}{2n}$, and stays still with probability $\frac{1}{2}$. To find bounds of the mixing time, he applies strong uniform time arguments, defining a stopping time and showing that it is a strong uniform time. The stopping time is defined as follows: at each step, we choose a coordinate uniformly at random and, tossing a coin, we decide whether we flip the coordinate or not. We stop when all the coordinates have been chosen at least once. Using this method, he shows that $n\log n + cn$ is an upper bound for the mixing time. Applying Fourier analysis, he also proves that for this random walk we have a lower bound of order $\frac{1}{2}n\log n$.

Another coupling for the same lazy version is presented in [21]: consider two simple random walks on the hypercube started from two different vertices, at each step choose a coordinate uniformly at random and replace it with the same bit in both walks. Let T be the first time when all the coordinates have been chosen at least once, then T has the same distribution as the coupon collector random variable, and it is a coupling time. Applying Theorem 1.17, it is shown that the mixing time has an upper bound $n\log(n)$. This bound does not give the correct mixing time, but in the same book it is defined another coupling, which improves the upper bound and gives the exact mixing time. With this different approach, the study of the random walk on the hypercube is reduced to the study of a lazy version of the Ehrenfest urn chain, which has transition matrix $\frac{1}{2}(I+P)$, where I is the identity and P is given by

$$P(j,k) = \begin{cases} \frac{n-j}{n} & \text{if } k = j+1 \\ \frac{j}{n} & \text{if } k = j-1 \\ 0 & \text{otherwise} \end{cases}$$

Let X_t and Y_t be two lazy Ehrenfest urn chains. At each move, a fair coin is tossed to determine which of the two chains moves; the selected chain makes a step according to the matrix given above, while the other chain remains in its current position. The chains move together once they have met for the first time. Using this coupling, it is shown that the mixing time is bounded above by $\frac{1}{2}n\log n$ and, as shown in the same book, a matching lower bound can be found, proving that this random walk exhibits cutoff at $\frac{1}{2}n\log n$ and window of size n. To find the lower bound, Levin, Peres and Wilmer consider the Hamming weight, i.e. the number of coordinates equal to 1, at time t, and use it to bound the total variation distance between the distribution of the chain at time t and the stationary distribution.

We have seen how coupling is a convenient technique to bound the mixing time, but we do not necessarily have to work with co-adapted couplings. In [26], Matthews uses another approach.

Example 2.1 (Simple random walk on the hypercube)**.** As we have seen in Example 1.21, Aldous defined a co-adapted coupling to establish $\frac{1}{2}n\log(n)$ as an upper bound on the rate of convergence of the simple random walk

on the n-dimensional cube. In [26], Matthews constructs a non co-adapted coupling for the simple random walk on the hypercube based on the coupling proposed by Aldous, finding a sharper upper bound.
Consider two simple random walks $X = (X_t)_{t=0}^{\infty}$ and $Y = (Y_t)_{t=0}^{\infty}$ on the n-dimensional cube $\mathbb{Z}_2^n$ with $X_0 = 0$ and let Y_0 have the uniform distribution on $\mathbb{Z}_2^n$. Create a mythical $(n+1)$st coordinate that flips at the tth step if $Y_t = Y_{t-1}$. Denote this new process on $\mathbb{Z}_2^{n+1}$ by $Y^\star = (Y_t^\star)_{t=0}^{\infty}$ and put

$$Y_0^\star(n+1) = \begin{cases} 1 & \text{if } |Y_0^\star| \text{ is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

Now, if we denote by A the set of coordinates i such that $Y_0^\star(i) = 1$, we define the time T to be the first time t when half of the A-coordinates of $Y_t^\star$ are equal to 0. T is a stopping time for $Y^\star$, and Matthews shows that T is also a coupling time.
Now, we construct a process $X^\star$ on $\mathbb{Z}_2^{n+1}$ (to which X on $\mathbb{Z}_2^n$ will correspond) using the information we have from $Y_T^\star$. We consider the following two sets

$$A_0 = \{\text{coordinates in } A \text{ such that } Y_T^\star = 0\},$$
$$A_1 = \{\text{coordinates in } A \text{ such that } Y_T^\star = 1\},$$

listed in order of increasing coordinate index. Make a list of pairs of coordinates consisting of the first coordinates of A_0 and A_1, the second coordinates of A_0 and A_1, etc.
At step t, if $Y_t^\star$ is obtained by flipping coordinate i, $X_t^\star$ will move according to the following rule.

i. If $t > T$, then $X_t^\star$ flips coordinate i,

ii If $t \leq T$ then we have two possibilities

a. If $i \in A^c$, then $X_t^\star$ flips coordinate i,

b. If $i \in A$ and i and j are paired, then $X_t^\star$ flips coordinate j.

By the construction of the coupling, $X_T^\star = Y_T^\star$, and Matthews shows that the mixing time is bounded above by $\frac{1}{4} n \log(n)$, improving the result found by Aldous by a factor $\frac{1}{2}$. So, this coupling is more efficient but clearly non co-adapted since we first run the process Y up to the coupling time T, then

Author	Method	Upper Bound
Aldous [1]	Coupling inequality	$\frac{n}{2}\log n$
Diaconis [8]	Representation theory	$\frac{n}{4}\log n$
Diaconis [8]	Strong uniform time	$n\log n$
Levin, Peres, Wilmer [21]	Coupon collector	$n\log n$
Levin, Peres, Wilmer [21]	Ehrenfest urn	$\frac{n}{2}\log n$
Matthews [26]	Non co-adapted coupling	$\frac{n}{4}\log n$

Tab. 2.1: Summary of the methods that have been used to find upper bounds for random walks on the hypercube.

we go back in time and we construct the process X according to rules that depend on Y_T. □

Similar results have also been proved by Aldous and Diaconis in [3], in which they define a strong stationary time for the lazy simple random walk on the hypercube, and relate this time to the coupling method to bound the total variation distance.
Table 2.1 illustrates a summary of the results reported in the first part of this section.

In [22], Lim studies an irreversible random walk on the hypercube Q^V, with colour set $Q = \{1, 2, 3\}$ and vertex set $V = \{1, \ldots, n\}$. He considers the following cyclic dynamics. At each step, a vertex is uniformly chosen. Assume that the vertex has colour i, then we reassign colour i with probability $1 - p$ and colour $i + 1$ with probability p, where $0 < p < 1$. Lim shows that this random walk exhibits cutoff at $\frac{1}{3p} n \log(n)$ and window of order n. To find the upper bound, he defines a semi-synchronised coupling, i.e. a synchronised coupling (in which the two chains are as synchronised as possible) that considers different cases depending on the current configuration of the two random walks.

The second lazy version considered by Diaconis in his book is generalised in “A non-local random walk on the hypercube” [28] by Nestoridi. In

her paper, she studies the random walk that at each step stays still with probability $\frac{1}{2}$, else flips a set of k uniformly chosen coordinates, where k is fixed at the beginning of the walk. The number k is chosen in the set of odd numbers in $\{1, \ldots, \frac{n}{2}\}$; choosing k odd avoids parity problems, and she restricts the choice to $\{1, \ldots, \frac{n}{2}\}$ because flipping k coordinates and flipping $n-k$ coordinates lead to the same results.
The main result, showed in [28], is the following theorem.

Theorem 2.2. *For the lazy walk changing $k \leq \frac{n}{2}$ coordinates on the hypercube, the following hold for every $x \in \mathbb{Z}_2^n$.*

(a) For $\ell = \frac{n^2}{2k(n-k)} \log n + c\frac{n^2}{k(n-k)}$, we have that

$$||P_x^\ell - U||_{TV} \leq e^{-c} + 2^{-c},$$

where $c > 0$.

(b) For $\ell = \frac{n}{2k} \log n - c\frac{n}{k}$, where $0 < c \leq \frac{1}{4} \log n$ and for x being the identity element, we have that

$$||P_x^\ell - U||_{TV} \geq 1 - \frac{B}{e^{4c}},$$

for a uniformly bounded constant $B > 0$.

Corollary 2.3. *If $k = o(n)$, then the walk exhibits cutoff at $\frac{n}{2k} \log n$ with window $\frac{n}{2k}$.*

To find the upper bound in the first part of the theorem, Nestoridi introduces a coupling for the Markov chain and applies Theorem 1.17. Then, applying path coupling we described in Section 1.2.2, she proves that the two random walks of the coupling started at any two vertices get closer in expectation. That gives an upper bound for the probability on the right hand side of (1.1), thus an upper bound for the mixing time.
To prove the lower bound, Nestoridi uses representation theory. Starting from the irreducible representations of $\mathbb{Z}_2^n$, she calculates the eigenvalues and eigenfunctions of the transition matrix to show that after ℓ steps the expected value of a particular eigenfunction can get big, while the variance is bounded. Applying Chebyshev's inequality finishes the proof.

2.2 Mixing time for a random walk on $\mathbb{Z}_2^n$ with random step size

In this section, we generalise the results of Theorem 2.2 in the following way. Instead of flipping always the same number k of coordinates, we allow k to be chosen at each step according to some probability distribution. More precisely, let K be a random variable on the set of odd numbers in $\{1, \dots, n\}$. At each step, the random walk first draws a value for K, then it chooses K coordinates uniformly at random and flips them. We will consider the lazy version, so that we have an irreducible and aperiodic Markov chain that converges to the uniform distribution. Also, this random walk is transitive, so the choice of the starting point will not influence our results.
In this section, we prove two main results. The first, summarised in the following theorem, establishes an upper bound for the mixing time.

Theorem 2.4. *Let $\{K_1, K_2, \dots\}$ be i.i.d. random variables supported on the odd integers belonging to the set $\{1, \dots, n\}$. Let $\ell = \lceil \frac{n^2}{\mathbb{E}[K(n-K)]} \log n + c\frac{2n^2}{\mathbb{E}[K(n-K)]} \rceil$, then for the lazy random walk on the hypercube changing K_i coordinates at step i, we have*

$$||P^\ell - U||_{TV} \leq 2^{1-c}.$$

This theorem establishes an upper bound for the mixing time even when we introduce variability in the choice of k. This means that we are able to extend the result found for k fixed when considering a more general version of the random walk on the hypercube.
As it can be seen in the statement of the theorem, if we fixed $k \leq \frac{n}{2}$ instead of using $\{K_1, K_2, \dots\}$, we would not obtain the results showed in Theorem 2.2*(a)* since our upper bound is double the bound found by Nestoridi. As we were working to generalise her path coupling argument, it became evident that there was a mistake in her proof, which Nestoridi confirmed (personal communication), and correcting it gave us an extra factor of 2. We still believe that the upper bound can be tightened to give the correct bound

$$\frac{n^2}{2\mathbb{E}[K(n-K)]} \log n + c\frac{n^2}{\mathbb{E}[K(n-K)]}$$

by using a similar method to the proof in [12] by Nestoridi and Eskenazis. They study the mixing time of the Bernoulli-Laplace urn model with parameters (n, k), which is defined in the following way. Consider two urns, each containing n balls; at each step, we pick k balls uniformly at random from each urn and switch them simultaneously. They show that $\frac{n}{4k}\log n + \frac{3n}{k}\log\log n + O\left(\frac{n}{\varepsilon^4 k}\right)$ is an upper bound for the mixing time for every $\varepsilon \in (0, 1)$. Let X_t and Y_t be two copies of the Bernoulli-Laplace urn random walk, their proof is structured into four steps.

1. Using the eigenvalues and eigenfunctions of the random walk, they show that X_t and Y_t are at distance at most $O(\sqrt{n})$ at time $\frac{n}{4k}\log n$.
2. Then, they run the random walks independently showing that after $O\left(\frac{n}{k}\right)$ X_t and Y_t are at distance $O(\sqrt{k\log n})$.
3. The third step consists of running the random walks according to a coupling, and they prove that with high probability they will be at distance $o(\sqrt{n})$ within $\frac{3n}{k}\log\log n$ steps.
4. Finally, once X_t and Y_t are at distance $o(\sqrt{n})$, they show that the distance becomes $o(1)$ after one step.

Conjecture. Applying a similar argument of [12] used for the Bernoulli-Laplace urn model proves that the right upper bound for the mixing time of the lazy random walk on the hypercube that flips $\{K_1, K_2, \dots\}$ coordinates at each step is

$$\frac{n^2}{2\mathbb{E}[K(n-K)]}\log n + c\frac{n^2}{\mathbb{E}[K(n-K)]}.$$

The following theorem establishes a lower bound for the mixing time.

Theorem 2.5. *Let $\{K_1, K_2, \dots\}$ be i.i.d. random variables supported on the odd integers belonging to the set $\{1, \dots, s_n\}$, where $s_n \leq n/2$ is the maximum possible value assumed by the random variables. Let $C_\varepsilon = \log\left(\frac{n}{4s_n}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)$, we have two cases for the lower bound of $t_{mix}(\varepsilon)$ with $\varepsilon \in (0, 1)$.*

(i) If $s_n = O(n)$ and $\mathbb{E}[K] \sim \frac{n}{D}$ for some constant $D \geq 2$, then the lower bound is $\frac{C_\varepsilon}{2\log\left(\frac{D}{D-1}\right)}$.

(ii) If $\mathbb{E}[K] = o(n)$ then the lower bound is $\frac{n}{2\mathbb{E}[K]}((1-\gamma)\log n + C_\varepsilon)$, where $\gamma \in [0, 1]$ satisfies $s_n = O(n^\gamma)$.

Remark. If the conjecture were true then, under the hypothesis of Theorem 2.5*(ii)* with $\gamma = 0$, the mixing time would have a lower bound at

$$\frac{n}{2\mathbb{E}[K]}(\log n + C_\varepsilon)$$

and an upper bound of order

$$\frac{n}{2\mathbb{E}[K]}\log n.$$

So, the random walk would exhibit a cutoff at $\frac{n}{2\mathbb{E}[K]}\log n$ with window of size $\frac{n}{2\mathbb{E}[K]}$.

2.2.1 Proof of Theorem 2.4

To prove Theorem 2.4, we shall use a coupling argument that is slightly different from the one used in [28], so first we have to construct the coupling that we will use in our proof. At each step we will couple the chains to move some number k of coordinates, so here we define the following measure on $\mathbb{Z}_2^n$ for any fixed odd $k \in \{1, \dots, n\}$.

$$P_k(\mathbf{v}) = \begin{cases} \frac{1}{2}, & \text{if } \mathbf{v} = id = (0, 0, \dots, 0) \\ \frac{1}{2\binom{n}{k}}, & \text{if } \mathbf{v} \in \mathbb{Z}_2^n \text{ has } k \text{ ones and } n-k \text{ zeros} \end{cases} \tag{2.1}$$

and let $P_k(\mathbf{x}, \mathbf{x}+\mathbf{v}) = P_k(\mathbf{v})$ for every $\mathbf{x}, \mathbf{v} \in \mathbb{Z}_2^n$.
Let (X_t) and (X'_t) be two copies of the Markov chain with transition matrix P_k, where X starts at an arbitrary point $\mathbf{x} \in \mathbb{Z}_2^n$, and X' starts at a uniformly chosen point. At time t, let

$$Y_t = \{j \in \{1, \dots, n\} : X_t(j) \neq X'_t(j)\},$$

$$M_t = \{1, \dots, n\} \setminus Y_t,$$

where $X_t(j)$ and $X'_t(j)$ indicate the j-th coordinate of the two chains at time t, and let

$$y_t = |Y_t|, \qquad m_t = |M_t|.$$

Thus, Y_t and M_t are sets of coordinates on which X_t and X'_t differ and agree

respectively.
Suppose the two chains are at time t, and we have to move to step $t+1$. We couple the two chains according to the following rule. We have two cases:

1. If y_t is odd, then take one independent step on each chain according to the probability measure P_k.

2. If y_t is even, both chains stay fixed with probability $\frac{1}{2}$.
   With probability $\frac{1}{2\binom{n}{k}}$, we choose an integer d_{t+1} using the probability distribution

$$\mathbb{P}(d_{t+1} = j \mid M_t) = \frac{\binom{m_t}{j}\binom{y_t}{k-j}}{\binom{n}{k}}, \quad j = 0, \ldots, \min\{k, m_t\}.$$

   Given, d_{t+1}, we uniformly choose two disjoint sets, denoted by D_{t+1} and A_{t+1}, with $|D_{t+1}| = d_{t+1}$ and $|A_{t+1}| = a_{t+1} := k - d_{t+1}$, such that

$$D_{t+1} \sim \text{Unif}(\{\text{sets of size } d_{t+1} \text{ from } M_t\}),$$
$$A_{t+1} \sim \text{Unif}(\{\text{sets of size } a_{t+1} \text{ from } Y_t\}).$$

   Then, put $G_{t+1} = D_{t+1} \cup A_{t+1}$. Note that $|G_{t+1}| = k$.

   Flip $X_{t+1}(i)$ for all $i \in G_{t+1}$, and consider the following rule to move X'_{t+1}:

   (i) If $a_{t+1} > \frac{y_t}{2}$, flip $X'_{t+1}(i)$, for all $i \in G_{t+1}$.

   (ii) If $a_{t+1} \leq \frac{y_t}{2}$, we choose another set of k coordinates $G'_{t+1} = D_{t+1} \cup A'_{t+1}$, where A'_{t+1} is chosen uniformly in the set $Y_t \setminus A_{t+1}$. We have that $|A'_{t+1}| = |A_{t+1}|$ and $A'_{t+1} \cap A_{t+1} = \emptyset$.
   Finally, we flip $X'_{t+1}(j)$ for all $j \in G'_{t+1}$.

We claim that this procedure represents a coupling the two sets G_{t+1} and G'_{t+1} are both uniformly chosen from

$$\mathcal{I}_k = \{\text{subsets of size } k \text{ from } \{1, \ldots, n\}\},$$

so the construction given above really is a coupling of X and X', conditional on each chain moving according to P_k. This is proved in the following two propositions.

Proposition 2.6. *Conditioned on M_t, the set G_{t+1} is distributed uniformly on $\mathcal{I}_k$.*

Proof. For any set $I \in \mathcal{I}_k$,

$$\begin{aligned}
\mathbb{P}(G_{t+1} = I \mid M_t) &= \mathbb{P}(D_{t+1} \cup A_{t+1} = I \mid M_t) \\
&= \mathbb{P}(D_{t+1} \cup A_{t+1} = (I \cap M_t) \cup (I \cap Y_t) \mid M_t) \\
&= \sum_{j=0}^{m_t} \mathbb{P}(D_{t+1} = I \cap M_t, A_{t+1} = I \cap Y_t \mid M_t, d_{t+1} = j)\mathbb{P}(d_{t+1} = j \mid M_t) \\
&= \sum_{j=0}^{m_t} \Bigg[\mathbb{P}(D_{t+1} = I \cap M_t \mid M_t, d_{t+1} = j) \\
&\qquad \cdot \mathbb{P}(A_{t+1} = I \cap Y_t \mid M_t, d_{t+1} = j) \frac{\binom{m_t}{j}\binom{y_t}{k-j}}{\binom{n}{k}} \Bigg] \\
&= \mathbb{P}(D_{t+1} = I \cap M_t \mid M_t, d_{t+1} = |I \cap M_t|) \\
&\quad \cdot \mathbb{P}(A_{t+1} = I \cap Y_t \mid M_t, d_{t+1} = |I \cap M_t|) \frac{\binom{m_t}{|I \cap M_t|}\binom{y_t}{k-|I \cap M_t|}}{\binom{n}{k}} \\
&= \frac{1}{\binom{m_t}{|I \cap M_t|}} \frac{1}{\binom{y_t}{k-|I \cap M_t|}} \frac{\binom{m_t}{|I \cap M_t|}\binom{y_t}{k-|I \cap M_t|}}{\binom{n}{k}} = \frac{1}{\binom{n}{k}}.
\end{aligned}$$

Here, the final line follows from the definitions of D_{t+1} and A_{t+1}. □

Proposition 2.7. *Conditioned on M_t, the set G'_{t+1} is distributed uniformly on $\mathcal{I}$.*

Proof. First, we need to prove that, conditioned on $|A'_{t+1}| = a_{t+1} \leq \frac{y_t}{2}$, G'_{t+1} is distributed uniformly on

$$\mathcal{J} = \{\text{all subsets of } Y_t \text{ of size } a_{t+1}\}.$$

Conditioned on $|A_{t+1}| = |A'_{t+1}| = a_{t+1}$, we have that $\forall J, J' \in \mathcal{J}$ such that $J \cap J' = \emptyset$

$$\mathbb{P}(A_{t+1} = J \mid Y_t, |A_{t+1}| = a_{t+1}) = \frac{1}{\binom{y_t}{a_{t+1}}},$$

$$\mathbb{P}(A'_{t+1} = J' \mid Y_t, A_{t+1} = J, |A_{t+1}| = |A'_{t+1}| = a_{t+1}) = \frac{1}{\binom{y_t - a_{t+1}}{a_{t+1}}}.$$

Therefore,

$$
\begin{aligned}
&\mathbb{P}(A'_{t+1} = J' \mid |A'_{t+1}| = a_{t+1}) \\
&= \sum_{J \in \mathcal{J}} \mathbb{P}(A'_{t+1} = J' \mid Y_t, A_{t+1} = J, |A_{t+1}| = |A'_{t+1}| = a_{t+1})\mathbb{P}(A_{t+1} = J \mid Y_t, |A_{t+1}| = a_{t+1}) \\
&= \frac{1}{\binom{y_t}{a_{t+1}}} \sum_{\substack{J \in \mathcal{J}: \\ J \cap J' = \emptyset}} \mathbb{P}(A'_{t+1} = J' \mid Y_t, A_{t+1} = J, |A_{t+1}| = |A'_{t+1}| = a_{t+1}) \\
&= \frac{1}{\binom{y_t}{a_{t+1}}} \cdot \frac{1}{\binom{y_t - a_{t+1}}{a_{t+1}}} \cdot \binom{y_t - a_{t+1}}{a_{t+1}}.
\end{aligned}
$$

It follows that

$$
\mathbb{P}(A'_{t+1} = J' \mid |A'_{t+1}| = a_{t+1}) = \frac{1}{\binom{y_t}{a_{t+1}}}.
$$

Finally, if we consider $\mathcal{I}_k = \{\text{subsets of size } k \text{ from } \{1, \ldots, n\}\}$ and $I \in \mathcal{I}_k$, then as in the proof of Proposition 2.6,

$$
\begin{aligned}
\mathbb{P}(G'_{t+1} = I | M_t) &= \mathbb{P}(D_{t+1} \cup A'_{t+1} = I \mid M_t) \\
&= \mathbb{P}(D_{t+1} \cup A'_{t+1} = (I \cap M_t) \cup (I \cap Y_t) \mid M_t) \\
&= \mathbb{P}(D_{t+1} = I \cap M_t \mid M_t, d_{t+1} = |I \cap M_t|) \\
&\quad \cdot \mathbb{P}(A'_{t+1} = I \cap Y_t \mid M_t, d_{t+1} = |I \cap M_t|) \frac{\binom{m_t}{|I \cap M_t|}\binom{y_t}{k - |I \cap M_t|}}{\binom{n}{k}} \\
&= \frac{1}{\binom{m_t}{|I \cap M_t|}} \frac{1}{\binom{y_t}{k - |I \cap M_t|}} \frac{\binom{m_t}{|I \cap M_t|}\binom{y_t}{k - |I \cap M_t|}}{\binom{n}{k}} = \frac{1}{\binom{n}{k}}.
\end{aligned}
$$

□

As we already mentioned above, the coupling that we described represents one step of the random walk for a fixed value of k. We can, of course, use this to couple our random walks in which k is randomised: at each step, we draw K from a distribution q_K on the odd integers in $\{1, \ldots, n\}$ and then, conditioned on $K = k$, couple the next step of our random walks using the construction described above.

2.2.2 Upper bound

At time $t = 0$, if y_0 is odd, we let the chains run independently until the first time T_0 when y_{T_0} is even. Then, we claim that, for all times $t > T_0$, we have that y_t is even.This is easily proved by induction. Suppose that $t > T_0$ and y_t is even, we want to prove that y_{t+1} is even. We have two cases:

(i) If $a_{t+1} > \frac{y_t}{2}$, then we flip $X'_{t+1}(i)$, with $i \in G_{t+1}$. Thus, $y_{t+1} = y_t$ and so is even.

(ii) If $a_{t+1} \leq \frac{y_t}{2}$, then we uniformly choose another set A'_{t+1} of size a_{t+1} in $Y_t \setminus A_{t+1}$ and we flip both $X_{t+1}(i)$ and $X'_{t+1}(j)$, where $i \in G_{t+1}$ and $j \in G'_{t+1}$. Thus, we flip D_{t+1} on both, A_{t+1} on X_{t+1}, and A'_{t+1} on X'_{t+1}. In particular, we flip $2a_{t+1}$ mismatching coordinates (those in A_{t+1} and A'_{t+1}), i.e. $y_{t+1} = y_t - 2a_{t+1}$ is even.

To prove the upper bound, we will apply the path coupling technique as explained in Section 1.2.2, where in this context $V = \{1, \ldots, n\}$ is the set of coordinates, $C = \{0, 1\}$, and the distributions $K_{X_t,i}$ are used to assign an element of C to coordinate i in the current state X_t.
The way we defined T_0 implies that it is distributed as a geometric random variable with parameter $\frac{1}{2}$. For any $c > 0$

$$\mathbb{P}\left(T_0 > c\frac{n^2}{\mathbb{E}[K(n-K)]}\right) \leq 2^{-c\frac{n^2}{\mathbb{E}[K(n-K)]}} \leq 2^{-c}, \tag{2.2}$$

since $\frac{n^2}{\mathbb{E}[K(n-K)]} > 1$.

Suppose that at time $t-1$ we have $y_{t-1} = 2$, i.e. just 2 coordinates differ between X_{t-1} and X'_{t-1}. In order to couple at time t (and obtain $y_t = 0$), at least one of the mismatched coordinates must belong to A_t, so $a_t \geq 1$. If $a_t > \frac{y_{t-1}}{2}$, we know that we flip the same coordinates in both chains at time t, so in this case we would have $y_t = y_{t-1} = 2$. Thus, the only way in which we can couple at time t this is if $a_t = 1$. Then, for any fixed value of k,

$$\begin{aligned}\mathbb{P}(X_t = X'_t \mid y_{t-1} = 2, K_t = k) &= \mathbb{P}(a_t = 1 \mid y_{t-1} = 2, K_t = k) \\ &= \frac{1}{2}\frac{\binom{k}{1}\binom{n-k}{1}}{\binom{n}{2}} = \frac{k(n-k)}{n(n-1)}.\end{aligned}$$

In our case, k is chosen according to the distribution q_K:

$$\mathbb{P}(X_t = X'_t \mid y_{t-1} = 2) = \sum_{k=0}^{n} \frac{k(n-k)}{n(n-1)} \cdot q_K(k) = \frac{\mathbb{E}[K(n-K)]}{n(n-1)}. \tag{2.3}$$

Now, let $||X_t - X'_t||$ denote the Hamming distance between X_t and X'_t. Using (2.3), we have

$$\mathbb{E}\left[||X_t - X'_t|| \;\middle|\; X_{t-1}, X'_{t-1}, ||X_{t-1} - X'_{t-1}|| = 2\right] \leq \left[1 - \frac{\mathbb{E}[K(n-K)]}{n(n-1)}\right] ||X_{t-1} - X'_{t-1}||$$

since having $X_t \neq X'_t$ corresponds to $||X_t - X'_t|| = ||X_{t-1} - X'_{t-1}||$. Applying the path coupling, we can extend the previous inequality to

$$\mathbb{E}\left[||X_t - X'_t|| \;\middle|\; X_{T_0}, X'_{T_0}\right] \leq n \left[1 - \frac{\mathbb{E}[K(n-K)]}{n(n-1)}\right]^{t-T_0}. \tag{2.4}$$

Let T be the coupling time, i.e. $T = \min\{t : X_t = X'_t\}$. Then we know that $T > t$ implies $||X_t - X'_t|| \geq 2$. If we define time $\ell = \frac{n^2}{\mathbb{E}[K(n-K)]} \log n + c\frac{2n^2}{\mathbb{E}[K(n-K)]}$ and use (2.2), (2.4), and the Markov inequality, we have

$$\begin{aligned}
\mathbb{P}(T > \ell) &= \mathbb{P}\left(T > \ell \;\middle|\; T_0 \leq c\frac{n^2}{\mathbb{E}[K(n-K)]}\right) \mathbb{P}\left(T_0 \leq c\frac{n^2}{\mathbb{E}[K(n-K)]}\right) \\
&\quad + \mathbb{P}\left(T > \ell \;\middle|\; T_0 > c\frac{n^2}{\mathbb{E}[K(n-K)]}\right) \mathbb{P}\left(T_0 > c\frac{n^2}{\mathbb{E}[K(n-K)]}\right) \\
&\leq \mathbb{P}\left(||X_\ell - X'_\ell|| \geq 1 \;\middle|\; T_0 \leq c\frac{n^2}{\mathbb{E}[K(n-K)]}\right) + \mathbb{P}\left(T_0 > c\frac{n^2}{\mathbb{E}[K(n-K)]}\right) \\
&\leq \mathbb{E}\left[||X_\ell - X'_\ell|| \;\middle|\; T_0 \leq c\frac{n^2}{\mathbb{E}[K(n-K)]}\right] + 2^{-c} \\
&\leq n\left[1 - \frac{\mathbb{E}[K(n-K)]}{n(n-1)}\right]^{\ell - c\frac{n^2}{\mathbb{E}[K(n-K)]}} + 2^{-c} \\
&\leq e^{-c} + 2^{-c}.
\end{aligned}$$

In the final line we make use of the inequality $1 - x \leq e^{-x}$ for $x \in [0, 1]$. The upper bound for the total variation distance derives from the coupling inequality established by Theorem 1.17.

We conclude that an upper bound for the mixing time is given by

$$\ell = \frac{n^2}{\mathbb{E}[K(n-K)]}\log n + c\frac{2n^2}{\mathbb{E}[K(n-K)]},$$

and this concludes the proof of Theorem 2.4.

2.2.3 *A lower bound on the mixing time*

We will prove Theorem 2.5 by using irreducible representations to calculate the eigenvalues and eigenfunctions of the Markov chain. In this section, the random walk will start at the identity $id = (0, 0, \ldots, 0)$ and we will define the set of possible values for K as $\{1, \ldots, s_n\}$, where $s_n \leq \frac{n}{2}$. In particular, we will have to consider two cases: $s_n = O(n)$ and $s_n = o(n)$.

Consider the one-dimensional representations $\rho_{\mathbf{a}}(\mathbf{v}) = (-1)^{\mathbf{a}\cdot\mathbf{v}}$ of $\mathbb{Z}_2^n$ indexed by vectors $\mathbf{a} \in \mathbb{Z}_2^n$, where $\mathbf{v} \in \mathbb{Z}_2^n$ and $\mathbf{a}\cdot\mathbf{v}$ is the dot product of $\mathbf{a}$ and $\mathbf{v}$. Observe that $\rho_{\mathbf{a}}$ is a representation since, given any two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{Z}_2^n$, $\rho_{\mathbf{a}}(\mathbf{v}+\mathbf{w}) = (-1)^{\mathbf{a}\cdot(\mathbf{v}+\mathbf{w})} = (-1)^{\mathbf{a}\cdot\mathbf{v}+\mathbf{a}\cdot\mathbf{w}} = (-1)^{\mathbf{a}\cdot\mathbf{v}}(-1)^{\mathbf{a}\cdot\mathbf{w}} = \rho_{\mathbf{a}}(\mathbf{v})\rho_{\mathbf{a}}(\mathbf{w})$. Moreover, as explained in [8], any one-dimensional representation is irreducible since in that case there are no non-trivial subrepresentations. Thus, the representations $\rho_{\mathbf{a}}(\mathbf{v}) = (-1)^{\mathbf{a}\cdot\mathbf{v}}$ are irreducible.

Let $|\mathbf{a}|$ denote the number of ones in $\mathbf{a}$. Following [8], for any fixed k, the Fourier transform of P_k (2.1) at $\rho_{\mathbf{a}}$ is

$$\widehat{P}_k(\rho_{\mathbf{a}}) = \sum_{\mathbf{v}\in\mathbb{Z}_2^n} \rho_{\mathbf{a}}(\mathbf{v})P_k(\mathbf{v}) = \frac{1}{2} + \frac{1}{2}\sum_{b=0}^{k}(-1)^b \frac{\binom{|\mathbf{a}|}{b}\binom{n-|\mathbf{a}|}{k-b}}{\binom{n}{k}}. \tag{2.5}$$

In the above calculation, we fix $\mathbf{a}$, and we sum over all possible $\mathbf{v} \in \mathbb{Z}_2^n$. From the definition of P_k, if $\mathbf{v} = id$ then $P_k(\mathbf{v}) = 1/2$ and if $|\mathbf{v}| = k$ then $P_k(\mathbf{v}) = \frac{1}{2\binom{n}{k}}$.

The eigenvalues for the transition matrix P_k are the values $\widehat{P}_k(\rho_{\mathbf{a}})$ with corresponding eigenfunctions $f_{\mathbf{a}}(\mathbf{x}) = (-1)^{\mathbf{a}\cdot\mathbf{x}}$. This is true because

$$\begin{aligned} P_k f_{\mathbf{a}}(\mathbf{x}) &= \sum_{\mathbf{v}\in\mathbb{Z}_2^n} P_k(\mathbf{v}) f_{\mathbf{a}}(\mathbf{x}+\mathbf{v}) = \sum_{\mathbf{v}\in\mathbb{Z}_2^n} P_k(\mathbf{v}) f_{\mathbf{a}}(\mathbf{x}) f_{\mathbf{a}}(\mathbf{v}) \\ &= f_{\mathbf{a}}(\mathbf{x})\left[\sum_{\mathbf{v}\in\mathbb{Z}_2^n} P_k(\mathbf{v}) f_{\mathbf{a}}(\mathbf{v})\right] = f_{\mathbf{a}}(\mathbf{x})\widehat{P}_k(\rho_{\mathbf{a}}). \end{aligned}$$

For our random walk of interest, for which we choose a different k for each step according to a generic distribution q_K on $\{1,\dots,s_n\}$, the transition matrix is a linear combination of the matrices P_k defined in (2.1):

$$P(\mathbf{v}) = \sum_{k=1}^{n} q_K(k)P_k(\mathbf{v}).$$

We have that the function $f_{\mathbf{a}}(\mathbf{x}) = (-1)^{\mathbf{a}\cdot\mathbf{x}}$ is once again an eigenfunction for P with corresponding eigenvalue $\widehat{P}(\rho_{\mathbf{a}}) = \sum_k q_K(k)\widehat{P}_k(\rho_{\mathbf{a}})$:

$$\begin{aligned}\widehat{P}(\rho_{\mathbf{a}}) &= \sum_{\mathbf{v}\in\mathbb{Z}_2^n} \rho_{\mathbf{a}}(\mathbf{v})P(\mathbf{v}) = \sum_{\mathbf{v}\in\mathbb{Z}_2^n} \rho_{\mathbf{a}}(\mathbf{v}) \sum_{k=0}^{n} q_K(k)P_k(\mathbf{v}) \\ &= \sum_{k=1}^{n} q_K(k) \sum_{\mathbf{v}\in\mathbb{Z}_2^n} \rho_{\mathbf{a}}(\mathbf{v})P_k(\mathbf{v}) = \sum_{k=1}^{n} q_K(k)\widehat{P}_k(\rho_{\mathbf{a}}).\end{aligned}$$

Notice that all vectors $\mathbf{a} \in \mathbb{Z}_2^n$ that have the same number of ones give the same eigenvalue. Therefore, if we fix a vector $\mathbf{x} \in \mathbb{Z}_2^n$, and we sum the eigenfunctions $f_{\mathbf{a}}(\mathbf{x})$ on all $\mathbf{a}$ such that $|\mathbf{a}| = j$, we obtain the function

$$f_j(\mathbf{x}) = \sum_{\substack{\mathbf{a}\in\mathbb{Z}_2^n:\\ |\mathbf{a}|=j}} f_{\mathbf{a}}(\mathbf{x}) = \sum_{b=0}^{|\mathbf{x}|} (-1)^b \frac{\binom{|\mathbf{x}|}{b}\binom{n-|\mathbf{x}|}{j-b}}{\binom{n}{j}}.$$

Since $f_j(\mathbf{x})$ is a sum of eigenfunctions of P corresponding to the same eigenvalue $\widehat{P}(\rho_{\mathbf{a}})$, it is itself an eigenfunction of P.

To find a lower bound for the mixing time we will use a different approach to that used by Nestoridi. As we mentioned above, to find a lower bound she applied the Chebyshev's inequality using the expectation and variance of an eigenfunction of the random walk. In our case, using the same approach is not useful as the variance we obtain using the eigenfunctions is too big and applying Chebyshev's inequality would be inconclusive. In Section 2.2.4, we will see that introducing assumptions over the size of $\mathbb{E}[K]$ will give us the possibility to bound the variance, and that would make us able to apply Chebyshev's inequality to tighten the bounds of Theorem 2.5.
We will apply the "Wilson's lemma", a method that was introduced by Wilson in his paper "Mixing times of lozenge tilings and card shuffling Markov

chains" [33], in which he employs this result to improve the lower bounds of some Markov chains that had been previously studied.

Theorem 2.8. (Wilson's lemma) *If a function* Φ *on the state space* Ω *of a Markov chain* X *satisfies*

$$\mathbb{E}[\Phi(X_{t+1})|X_t] = (1-\gamma)\Phi(X_t), \tag{2.6}$$

and there exists a constant $R > 0$ *such that*

$$\mathbb{E}[(\Delta\Phi)^2|X_t] \leq R, \tag{2.7}$$

where $\Delta\Phi = \Phi(X_{t+1}) - \Phi(X_t)$*, then, when the number of Markov chain steps* t *is bounded by*

$$t \leq \frac{\log \Phi_{max} + \frac{1}{2}\log(\gamma\varepsilon/(4R))}{-\log(1-\gamma)}$$

where $\Phi_{max} = \sup_{x\in\Omega}\Phi(x)$ *and* $0 < \gamma \leq 2-\sqrt{2}$*, the variation distance from stationarity is at least* $1-\varepsilon < 1$.

The basic idea of Wilson's lemma is that if we can find a function on the state space of a Markov chain for which we can bound the variance then we can apply Chebychev's inequality to show that the distribution of this function is concentrated around its expectation. This function is then used to bound the number of steps of the Markov chain, from which we can deduce a lower bound for the mixing time.
We will use the Fourier transform to find the eigenfunctions of the random walk we study, and any of such eigenfunctions may be used as the function Φ in Wilson's lemma.
To prove the lower bound, we apply the following equivalent formulation ([21]) of Wilson's Lemma.

Theorem 2.9. *Let* (X_t) *be an irreducible, aperiodic Markov chain with state space* Ω *and transition matrix* P*. Let* Φ *be an eigenfunction of* P *with eigenvalue* λ *satisfying* $\frac{1}{2} < \lambda < 1$*. Fix* $0 < \varepsilon < 1$ *and let* $R > 0$ *satisfy*

$$\mathbb{E}_x\left(|\Phi(X_1) - \Phi(x)|^2\right) \leq R$$

for all $x \in \Omega$. Then for any $x \in \Omega$

$$t_{mix}(\varepsilon) \geq \frac{1}{2\log(1/\lambda)}\left[\log\left(\frac{(1-\lambda)\Phi(x)^2}{2R}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right].$$

In our study, the function Φ may be replaced by any of the eigenfunctions of the random walk, and, following Nestoridi [28], we want to consider the eigenfunction with the biggest eigenvalue so that we can have a good bound for the mixing time. For this reason, we will use the eigenfunction $f_j(\mathbf{x})$ for $j = 1$. Once we have the function, we can find R for which the assumption of Theorem 2.9 is satisfied.

In our case, we have $\Omega = \{0,1\}^n$ with transition matrix P. We want to apply Theorem 2.9 using the eigenfunction

$$f_1(\mathbf{x}) = \sum_{b=0}^{1}(-1)^b \frac{\binom{|\mathbf{x}|}{b}\binom{n-|\mathbf{x}|}{1-b}}{\binom{n}{1}} = 1 - \frac{2|\mathbf{x}|}{n}.$$

To find and calculate the corresponding eigenvalue λ, we take $X_0 = \mathbf{0}$:

$$\begin{aligned} Pf_1(X_0) &= \sum_{k=1}^{n} q_K(k) P_k f_1(X_0) = \sum_{k=1}^{n} q_K(k)\left[\frac{1}{2} + \frac{1}{2}\left(1 - \frac{2k}{n}\right)\right] \\ &= \sum_{k=1}^{n} q_K(k)\left(1 - \frac{k}{n}\right) = 1 - \frac{\mathbb{E}[K]}{n} = \left(1 - \frac{\mathbb{E}[K]}{n}\right) f_1(\mathbf{0}), \end{aligned}$$

so we have

$$\lambda = 1 - \frac{\mathbb{E}[K]}{n}$$

which satisfies $\frac{1}{2} \leq \lambda < 1$ due to our assumption that $s_n \leq \frac{n}{2}$.
To find R, we observe that the random walk starts from the identity and that after one step we either have the identity (if the walk stays still) or we have exactly K ones:

$$|f_1(X_1) - f_1(\mathbf{0})| = \left|1 - \frac{2|X_1|}{n} - 1\right| = \frac{2}{n}|X_1|,$$

then,

$$\mathbb{E}\left(|f_1(X_1) - f_1(\mathbf{0})|^2\right) = \frac{4}{n^2}\mathbb{E}[|X_1|^2] = \frac{1}{2}\frac{4}{n^2}\mathbb{E}[K^2] \leq \frac{2}{n^2} s_n \mathbb{E}[K],$$

where the factor $\frac{1}{2}$ derives from the laziness of the random walk (and, if it does not move, then $|X_1| = 0$). We can put $R = \frac{2}{n^2} s_n \mathbb{E}[K]$ and apply Theorem 2.9

$$\begin{aligned} t_{mix}(\varepsilon) &\geq \frac{1}{2\log 1/\lambda}\left[\log\left(\frac{(1-\lambda)f_1(\mathbf{0})^2}{2R}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \\ &= \frac{1}{2\log\left(\frac{n}{n-\mathbb{E}[K]}\right)}\left[\log\left(\frac{\frac{\mathbb{E}[K]}{n}}{\frac{4}{n^2}s_n\mathbb{E}[K]}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \\ &= \frac{1}{2\log\left(1+\frac{\mathbb{E}[K]}{n-\mathbb{E}[K]}\right)}\left[\log\left(\frac{n}{4s_n}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right]. \end{aligned} \tag{2.8}$$

Now, we have two cases:

Case 1 Let $\frac{s_n}{n} \to \frac{1}{B}$ for some constant $B > 1$. To find an asymptotic equivalence for $\log\left(\frac{n}{4s_n}\right)$, we need the following result.

Lemma 2.10. *Suppose* $f(n) \sim g(n)$, *then* $\log(f(n)) \sim \log(g(n))$.

Proof. We know that $f(n) \sim g(n)$, so $\frac{f(n)}{g(n)} \to 1$. Now, if we take logarithms on both sides we have

$$\log(f(n)) - \log(g(n)) = \log\left(\frac{f(n)}{g(n)}\right) \to \log(1) = 0.$$

We now divide by $\log(g(n))$

$$\frac{\log(f(n))}{\log(g(n))} - 1 \to 0.$$

□

Thus, we have that

$$\frac{n}{4s_n} \sim \frac{B}{4} \Rightarrow \log\left(\frac{n}{4s_n}\right) \sim \log\left(\frac{B}{4}\right).$$

We define

$$C_\varepsilon = \log\left(\frac{n}{4s_n}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right),$$

and we have that

$$C_\varepsilon > 0 \Leftrightarrow 0 < \varepsilon < \frac{B}{B+4}.$$

In the setting of *Case 1*, we have two possibilities for $\mathbb{E}[K]$.

(i) If $\mathbb{E}[K] \sim \frac{n}{D}$, with $D > 1$ constant, then from

$$1 + \frac{\mathbb{E}[K]}{n - \mathbb{E}[K]} \sim \frac{D}{D-1}$$

it follows

$$\log\left(1 + \frac{\mathbb{E}[K]}{n - \mathbb{E}[K]}\right) \sim \log\left(\frac{D}{D-1}\right)$$

again by Lemma 2.10. From (2.8), we obtain

$$t_{mix}(\varepsilon) \geq \frac{1}{2\log\left(1 + \frac{\mathbb{E}[K]}{n-\mathbb{E}[K]}\right)}\left[\log\left(\frac{n}{4s_n}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \sim \frac{C_\varepsilon}{2\log\left(\frac{D}{D-1}\right)},$$

and, since $\log\left(\frac{D}{D-1}\right) > 0$ for all $D > 1$, we have a useful lower bound when $C_\varepsilon > 0$, and that happens when $0 < \varepsilon < \frac{B}{B+4}$.

(ii) If $\mathbb{E}[K] = o(n)$, then from $\frac{\mathbb{E}[K]}{n-\mathbb{E}[K]} \xrightarrow[n\to\infty]{} 0$ it follows

$$\log\left(1 + \frac{\mathbb{E}[K]}{n - \mathbb{E}[K]}\right) = \frac{\mathbb{E}[K]}{n - \mathbb{E}[K]} + o\left(\frac{\mathbb{E}[K]}{n - \mathbb{E}[K]}\right).$$

From (2.8),

$$\begin{aligned} t_{mix}(\varepsilon) &\geq \frac{1}{2\log\left(1 + \frac{\mathbb{E}[K]}{n-\mathbb{E}[K]}\right)}\left[\log\left(\frac{n}{4s_n}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \\ &\sim \frac{C_\varepsilon}{2\frac{\mathbb{E}[K]}{n-\mathbb{E}[K]} + o\left(\frac{\mathbb{E}[K]}{n-\mathbb{E}[K]}\right)} \sim \frac{C_\varepsilon n}{2\mathbb{E}[K]}. \end{aligned}$$

As in case *(i)*, we have a useful lower bound when $0 < \varepsilon < \frac{B}{B+4}$.
Comparing what we obtained for the lower bound with our result for the upper bound, we cannot conclude anything about a cutoff when $s_n = O(n)$. This result is not unexpected. Even in [28], where the choice of K is deterministic, Nestoridi had to assume $k = o(n)$ to have cutoff, and the introduction of extra variability in the random walk intuitively makes having cutoff even more unlikely.

Case 2 If $s_n \sim \frac{n^\gamma}{B}$, with $B > 0$ and $\gamma \in [0, 1)$, then $\mathbb{E}[K] = o(n)$.
Again, we have

$$\log\left(1 + \frac{\mathbb{E}[K]}{n - \mathbb{E}[K]}\right) = \frac{\mathbb{E}[K]}{n - \mathbb{E}[K]} + o\left(\frac{\mathbb{E}[K]}{n - \mathbb{E}[K]}\right).$$

From Lemma 2.10, we deduce that

$$\frac{n}{4s_n} \sim \frac{B}{4} n^{1-\gamma} \Rightarrow \log\left(\frac{n}{4s_n}\right) \sim \log\left(\frac{B}{4} n^{1-\gamma}\right) \sim \log(n^{1-\gamma}).$$

For the lower bound of the mixing time,

$$\begin{aligned} t_{mix}(\varepsilon) &\geq \frac{1}{2\log\left(1 + \frac{\mathbb{E}[K]}{n-\mathbb{E}[K]}\right)} \left[\log\left(\frac{n}{4s_n}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \\ &= \frac{1}{2\frac{\mathbb{E}[K]}{n-\mathbb{E}[K]} + o\left(\frac{\mathbb{E}[K]}{n-\mathbb{E}[K]}\right)} \left[\log\left(\frac{n}{4s_n}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \\ &\sim \frac{n}{2\mathbb{E}[K]} \left[\log\left(n^{1-\gamma}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right]. \end{aligned}$$

To have a useful lower bound, we need $\log\left(\frac{B}{4} n^{1-\gamma}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right) > 0$, and we have that when $0 < \varepsilon < \frac{n}{n+4s_n}$.
We observe that in this case the lower bound we found is of order

$$(1-\gamma)\frac{n}{2\mathbb{E}[K]}\log n.$$

From our previous calculations, we have that the upper bound is

$$\frac{n^2}{\mathbb{E}[K(n-K)]}\log n + c \cdot \frac{2n^2}{\mathbb{E}[K(n-K)]},$$

which, assuming $\mathbb{E}[K] = o(n)$ and $s_n = O(n^\gamma)$, with $\gamma \in [0, 1)$, is of order

$$\frac{n}{\mathbb{E}[K]}\log n.$$

As in Definition 1.11, we say that a sequence of Markov chains has a pre-cutoff if it satisfies

$$\sup_{0<\varepsilon<1/2} \limsup_{n\to\infty} \frac{t_n(\varepsilon)}{t_n(1-\varepsilon)} < \infty.$$

From what we said about the bounds for the mixing time we conclude that, under the assumptions of *Case 2*, our random walk has a pre-cutoff

$$\limsup_{n\to\infty} \frac{t_n(\varepsilon)}{t_n(1-\varepsilon)} = \limsup_{n\to\infty} \frac{(1-\gamma)\frac{n}{2\mathbb{E}[K]}\log n}{\frac{n}{\mathbb{E}[K]}\log n} = \frac{1-\gamma}{2} < \infty.$$

However, since the upper and lower bounds do not agree under these assumptions on s_n and $\mathbb{E}[K]$, we are not able to conclude anything about a cutoff.

2.2.4 A tighter lower bound

In some cases, we can improve the lower bound by extending the method used in [28] to our random walk.

Following Nestoridi, we will use the normalized form of $f_j(\mathbf{x})$ for $j = 1$, the eigenfunction $f_1(\mathbf{x}) = \sqrt{n}\left(1 - \frac{2|\mathbf{x}|}{n}\right)$, and the non-normalized form for $j = 2$, the eigenfunction $f_2(\mathbf{x}) = 1 - \frac{4|\mathbf{x}|}{n-1} + \frac{4|\mathbf{x}|^2}{n(n-1)}$.
Let $\mathbf{X}$ be a vector chosen uniformly from $\mathbb{Z}_2^n$, and let $Z = |\mathbf{X}|$ be the number of ones in $\mathbf{X}$. We have that $Z \sim \mathrm{Bin}\left(n, \frac{1}{2}\right)$. Then,

$$\mathbb{E}[f_1(Z)] = 0, \qquad \mathrm{Var}(f_1(Z)) = 1.$$

Let $\mathbf{X}_0 = \mathbf{0}$ and let $Z_\ell = |\mathbf{X}_\ell|$ be the number of ones at time ℓ. We can find the eigenvalues of f_1 and f_2, which will be useful in our calculation.

$$Pf_1(Z_0) = \sum_k q_K(k) P_k f_1(Z_0) = \sum_k q_K(k)\left[\frac{1}{2} + \frac{1}{2}\left(1 - \frac{2k}{n}\right)\right] = \left(1 - \frac{\mathbb{E}[K]}{n}\right) f_1(\mathbf{0}),$$

so $1 - \frac{\mathbb{E}[K]}{n}$ is the eigenvalue corresponding to f_1, and

$$\begin{aligned} Pf_2(Z_0) &= \sum_k q_K(k) P_k f_2(Z_0) = \sum_k q_K(k)\left[\frac{1}{2} + \frac{1}{2}\left(1 - \frac{4k}{n-1} + \frac{4k^2}{n(n-1)}\right)\right] \\ &= \sum_k q_K(k)\left(1 - \frac{2k(n-k)}{n(n-1)}\right) = \left(1 - \frac{2\mathbb{E}[K(n-K)]}{n(n-1)}\right) f_2(\mathbf{0}), \end{aligned}$$

so $1 - \frac{2\mathbb{E}[K(n-K)]}{n(n-1)}$ is the eigenvalue corresponding to f_2. Then (since f_1 is an eigenfunction for P),

$$\mathbb{E}[f_1(Z_\ell)] = \widehat{P}^\ell f_1(Z_0) = \left(\sum_k q_K(k)\widehat{P}_k\right)^\ell f_1(Z_0) = \sqrt{n}\left(\sum_k q_K(k)\left(1-\frac{k}{n}\right)\right)^\ell$$
$$= \sqrt{n}\left(1-\frac{1}{n}\sum_k q_K(k)k\right)^\ell = \sqrt{n}\left(1-\frac{\mathbb{E}[K]}{n}\right)^\ell.$$

Using $f_1^2(\mathbf{x}) = n\left(\frac{1}{n}+\frac{n-1}{n}f_2(\mathbf{x})\right)$ yields

$$\begin{aligned}\mathbb{E}[f_1^2(Z_\ell)] &= n\left(\frac{1}{n}+\frac{n-1}{n}\widehat{P}^\ell f_2(Z_0)\right)\\ &= 1+(n-1)\left(\sum_k q_K(k)\widehat{P}_k\right)^\ell f_2(Z_0)\\ &= 1+(n-1)\left[\sum_k q_K(k)\left(1-\frac{2kn-2k^2}{n(n-1)}\right)\right]^\ell\\ &= 1+(n-1)\left[1-\frac{2\mathbb{E}[K(n-K)]}{n(n-1)}\right]^\ell,\end{aligned}$$

and

$$\mathrm{Var}(f_1(Z_\ell)) = 1+(n-1)\left[1-\frac{2\mathbb{E}[K(n-K)]}{n(n-1)}\right]^\ell - n\left(1-\frac{\mathbb{E}[K]}{n}\right)^{2\ell}.$$

Now that we have an expression for the expectation and variance of f_1, we can prove that, under the hypotheses described in the following Lemma on the distribution of K, the expectation grows with n and the variance is bounded. Then, applying the Chebyshev's inequality gives us a lower bound for the mixing time.

Lemma 2.11. *We have the following two cases.*

Case 1. If $\mathbb{E}[K] = O(n^\varepsilon)$, with $\varepsilon \in (0,1)$, then we have a lower bound for the mixing time at

$$\ell = \frac{n^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c),$$

for $0 < c < \frac{1}{4}\log\left(\frac{n}{n^\varepsilon}\right)$.

Case 2. If $\mathbb{E}[K] = \frac{n}{d}$, with $d > 1$, then we have a lower bound for the mixing *time at*

$$\ell = \frac{1}{2}\log_{\frac{d}{d-1}} n - c$$

when $1 - \frac{1}{2d} \leq \gamma \leq 1$.

Proof. Case 1. Assume $\mathbb{E}[K] = O(n^{\varepsilon})$, with $\varepsilon \in (0,1)$, then $\mathbb{E}[K(n-K)] = O(n^{1+\varepsilon})$. Let $\ell = \frac{n^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c)$ with $c < \frac{1}{2}\log n$ and suppose that $\frac{n\mathbb{E}[K]}{\mathbb{E}[K(n-K)]} \sim 1$. Under these assumptions we claim that

$$\mathbb{E}[f_1(Z_\ell)] \sim e^c \left(1 + \frac{(\mathbb{E}[K])^2}{4\mathbb{E}[K(n-K)]}(\log n - 2c)\right) \geq e^c, \tag{2.9}$$

and

$$\operatorname{Var}(f_1(Z_\ell)) \sim 1 + e^{2c}\left(\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c) - \frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c)\right).$$

The details of the calculations of $\mathbb{E}[f_1(Z_\ell)]$ and $\operatorname{Var}(f_1(Z_\ell))$ can be found in Appendix A. Therefore, for $c < \frac{1}{4}\log\left(\frac{n}{n^{\varepsilon}}\right)$, the expectation in (2.9) can get big and grow with n, and the variance is bounded, so there exists a constant $B > 0$ such that $\operatorname{Var}(f_1(Z_\ell)) < B$. For the set $A_\alpha = \{x : |f(x)| \leq \alpha\}$, using Chebyshev's inequality we have

$$U(A_\alpha) = U(|f_1(Z_\ell)| \leq \alpha) = U\left(|f_1(Z_\ell) - \mathbb{E}[f_1(Z_\ell)]| \leq \alpha\sqrt{\operatorname{Var}(f_1(Z_\ell))}\right) \geq 1 - \frac{1}{\alpha^2}$$

and

$$P^{\ell}(A_\alpha) \leq \mathbb{P}(|f_1(Z_\ell) - \mathbb{E}[f_1(Z_\ell)]| > \mathbb{E}[f_1(Z_\ell)] - \alpha) \leq \frac{\operatorname{Var}(f_1(Z_\ell))}{(\mathbb{E}[f_1(Z_\ell)] - \alpha)^2} \leq \frac{B}{(e^c - \alpha)^2}.$$

If we take $\alpha = \frac{e^c}{2}$, we obtain

$$||P^{\ell}(A_\alpha) - U(A_\alpha)||_{TV} \geq 1 - \frac{1}{\alpha^2} - \frac{B}{(e^c - \alpha)^2} = 1 - \frac{4}{e^{2c}} - \frac{4B}{e^{2c}}.$$

We conclude that, if $\mathbb{E}[K] = O(n^{\varepsilon})$, with $\varepsilon \in (0,1)$, and $\frac{n\mathbb{E}[K]}{\mathbb{E}[K(n-K)]} \sim 1$, we have a lower bound for the mixing time at

$$\ell = \frac{n^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c),$$

for $0 < c < \frac{1}{4}\log\left(\frac{n}{n^{\varepsilon}}\right)$, which improves the bound of Theorem 2.5*(ii)*.
Observe that if we restricted our result to the case when $K = k$ is a deterministic value from the set $\{1, \dots, \frac{n}{2}\}$, then we would obtain the same result for the lower bound shown in [28] by Nestoridi. We would have $\mathbb{E}[K] = k = O(n^{\varepsilon})$, with $\varepsilon \in (0, 1)$, and the lower bound would be

$$\ell = \frac{n^2}{2k(n-k)}(\log n - 2c) \sim \frac{n}{2k}(\log n - 2c),$$

with $c < \frac{1}{4}\log\left(\frac{n}{k}\right) \sim \frac{1}{4}\log\left(\frac{n}{n^{\varepsilon}}\right)$.

Case 2. Assume that $\mathbb{E}[K] = \frac{n}{d}$, with $d > 1$, then we have that $\mathbb{E}[K(n - K)] \sim \gamma\frac{n^2}{d}$, with $\frac{1}{2} \leq \gamma < 1$. Let $\ell = \frac{1}{2}\log_{\frac{d}{d-1}} n - c$, then

$$\mathbb{E}[f_1(Z_\ell)] = \sqrt{n}\left(1 - \frac{1}{d}\right)^{\ell} = \left(\frac{d}{d-1}\right)^{c},$$

and

$$\begin{aligned}
\mathrm{Var}(f_1(Z_\ell)) &\sim 1 + (n-1)\left(1 - \frac{2\gamma}{d}\right)^{\ell} - n\left(1 - \frac{1}{d}\right)^{2\ell} \\
&= 1 + (n-1)\left(\frac{d^2 - 2d\gamma}{d^2}\right)^{\ell} - n\left(\frac{d-1}{d}\right)^{2\ell} \\
&\leq 1 + (n-1)\left(\frac{d-1}{d}\right)^{2\ell} - n\left(\frac{d-1}{d}\right)^{2\ell} \\
&= 1 - \left(\frac{d-1}{d}\right)^{2\ell} = 1 - \frac{1}{n}\left(\frac{d-1}{d}\right)^{2c}.
\end{aligned}$$

To obtain the inequality in the third line, we used $d^2 - 2d\gamma \leq d^2 - 2d + 1$, so this argument holds if and only if $\gamma \geq 1 - \frac{1}{2d}$. If $c < \frac{1}{4}\log_{\frac{d}{d-1}} n$, then the expectation can grow with n while $\mathrm{Var}(f_1(Z_\ell)) < 1$. We can apply Chebyshev's inequality to find that we have a lower bound for the mixing time at

$$\ell = \frac{1}{2}\log_{\frac{d}{d-1}} n - c$$

when $1 - \frac{1}{2d} \leq \gamma \leq 1$.
Also in this case, we can obtain the same result shown in [28] by considering $K = k$, where k is a deterministic value in the set $\{1, \dots, \frac{n}{2}\}$. If we restricted

our result to that choice for K, we would have $\mathbb{E}[K] = k \sim \frac{n}{d}$, $\mathbb{E}[K(n-K)] = k(n-k) \sim \frac{n^2}{d}(1-\frac{1}{d})$ for a constant $d \geq 2$ and lower bound

$$\ell = \frac{1}{2} \log_{\frac{d}{d-1}} n - c,$$

with $c < \frac{1}{4} \log_{\frac{d}{d-1}} n$.

□

3. JUMPY BROWNIAN MOTION ON THE CIRCUMFERENCE OF THE UNIT CIRCLE

Consider a continuous time stochastic process X_t defined as

$$X_t = \frac{1}{2}B_t + \pi N_t \ (\mathrm{mod}\ 2\pi),$$

where B_t is a standard $\mathbb{R}$-valued Brownian motion, N_t is an independent Poisson process of rate $\lambda > 0$, and $\frac{1}{2}$ is introduced for the convenience of calculations.
In other words, X_t can be consider as a Brownian motion $\frac{1}{2}B_t$ on the circumference of the unit circle that, at times given by an independent Poisson process, jumps to the opposite point of the circumference from which it continues diffusing. For this reason, we will refer to the process X_t as a *jumpy Brownian motion*.

The aim of our study is to construct a mean-optimal co-adapted coupling of the jumpy Brownian motion as we met in Definitions 1.22 and 1.27.

Definition 1.22 (Co-adapted coupling)**.** Let $(X, \hat{X})$ be a coupling of two Markov processes $X = (X_t)_{t=0}^{\infty}$ and $\hat{X} = (\hat{X}_t)_{t=0}^{\infty}$. We say that the coupling is *co-adapted* if there exists a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$ such that X and $\hat{X}$ are Markov processes with respect to $(\mathcal{F}_t)_{t=0}^{\infty}$.

Definition 1.27 (Optimal coupling)**.** We say that a coupling is *tail-optimal* if it minimises the tail probability of the coupling time $\mathbb{P}(\tau_{couple} > t)$ simultaneously for all $t > 0$.
We say that a coupling is *Laplace-optimal* if it maximises the Laplace transform of the coupling time $\mathbb{E}[e^{-\gamma\tau_{couple}}]$ for all $\gamma > 0$.
We say that a coupling is *mean-optimal* if it minimises the expectation of the coupling time $\mathbb{E}[\tau_{couple}]$.

As we have mentioned in Chapter 1, a lot of attention has been given to constructing optimal couplings of Brownian motion, and, for different state spaces, the maximality of the reflection coupling has been established. Introducing the possibility for the Brownian motion to jump, we want to investigate if and to what extent the jumps influence the construction of a coupling and its optimality.

3.1 Previous studies

As we have seen in Chapter 1, in [24] Lindvall and Rogers show that for Euclidean Brownian motion on $\mathbb{R}^n$, the reflection coupling is maximal and co-adapted. Consider two Euclidean Brownian motions X_1 and X_2 with different initial states x_1 and x_2, the reflection coupling consists in letting X_2 be the reflection of X_1 with respect to the hyperplane which is the perpendicular bisector of the line $x_1 x_2$; after the first time X_1 and X_2 meet on that hyperplane, they run together.

In [16], Hsu and Sturm start from that result and study the uniqueness of this coupling. In general, the reflection coupling is not the unique maximal coupling of Euclidean Brownian motion, so they need to restrict to the class of Markovian couplings to obtain the uniqueness.

Definition 3.1 (Markovian coupling)**.** Let $X = (X_1, X_2)$ be a coupling of Brownian motions and let $\{\mathcal{F}_t^X\}$ be the filtration generated by X. The coupling X is Markovian if for each $s \geq 0$, conditioned on $\mathcal{F}_s^X$, the process

$$\{(X_1(t+s), X_2(t+s), t \geq 0\}$$

is still a coupling of Brownian motion.

In other words, a Markovian coupling is a coupling that, conditioning on the past, is still a coupling of the process in the future, i.e. the joint process $X = (X_1, X_2)$ is a Markov process. So, a Markovian coupling requires a stronger condition than a co-adapted coupling, for which X_1 and X_2 are required to be Markov processes with respect to the same filtration.

To prove their result, Hsu and Sturm use a martingale argument. They consider the joint process $X = (X_1, X_2)$. From the Markovian hypothesis on the coupling, it follows that X and $X_1 - X_2$ are continuous martingales. Then, by Lévy's decomposition, it is possible to express the process $X_1 -$

X_2 as a Brownian motion W with time increments given by the quadratic variation of $X_1 - X_2$. They then define the following times

$$\tau_1 = \inf\{t \geq 0 : X_1(t) = X_2(t)\}, \qquad \tau_2 = \inf\{t \geq 0 : W(t) = 0\},$$

and they show that the coupling time $T = \tau_1 = \tau_2$, from which it follows that X_2 is the reflection of X_1. Then, the reflection coupling is the only maximal Markovian coupling of n-dimensional Brownian motions.

The study of uniqueness of the reflection coupling is developed by Kuwada who, in [20], extends the result to Brownian motion on Riemannian manifolds. The reflection structure of the coupling is defined in the following way. Let B_1 and B_2 be two Brownian motions on a Riemannian manifold M started from two distinct points $x_1, x_2 \in M$. The reflection structure is defined by the two following properties on M.

(1) There is a continuous map $R : M \to M$, with $R \circ R = id$ and $\mathbb{P}_{x_1} \circ R^{-1} = \mathbb{P}_{x_2}$.

(2) The set of fixed points $H = \{x \in M : R(x) = x\}$ separates M into two disjoint sets M_1 and M_2 with $R(M_1) = M_2$.

From these properties, the reflection coupling can be constructed. Let $\tau = \inf\{t > 0 : B_1(t) \in H\}$ be the hitting time of H, then let

$$B_2(t) = \begin{cases} R(B_1(t)) & t < \tau \\ B_1(t) & t \geq \tau. \end{cases}$$

This construction defines the reflection coupling, and τ is the coupling time. Kuwada shows that the reflection coupling is the unique maximal coupling of Euclidean Brownian motion in the class of Markovian couplings. To prove this result, Kuwada applies the Markovian hypothesis to reduce to a mass transportation problem, which he uses to prove the uniqueness.

In [18], Kendall studies the case of coupling for the two-dimensional process consisting of Brownian motion together with its local time at 0. He defines a reflection/synchronised coupling showing that it is tail-optimal among all co-adapted couplings.

Applying Tanaka's formula (which can be found in [30]), Kendall represents a Brownian motion with local time at 0 as a pair (B, S), where B is a real Brownian motion and $S_t = \max\{L_0^{(0)}, \sup\{B_s, s \leq t\}\}$, where $L_0^{(0)}$ is the local time at 0. He then defines a coupling of two copies (B, S) and $(\widetilde{B}, \widetilde{S})$ of the process started with different initial conditions; the coupling has two stages:

1. Reflection coupling. $\widetilde{B} = -B$ until the first time T_1 that B and $\widetilde{B}$ meet.

Then, if (B, S) and $(\widetilde{B}, \widetilde{S})$ are not already coupled,

2. Synchronised coupling. $\widetilde{B} = B$ until the first time T_2 after T_1 that $B \equiv \widetilde{B}$ hit the level $S_{T_1} \vee \widetilde{S}_0$.

At the end of the second stage of the coupling, we have $B = \widetilde{B}$ and $S = \widetilde{S}$, so the processes are coupled. The reflection/synchronised coupling can be reformulated as $d\widetilde{B} = JdB$, where the control J assumes values ± 1, i.e. it is a "bang-bang" control.

$$J_t = \begin{cases} -1 & t < T_1 \text{ (reflection stage)} \\ 1 & T_1 \leq t \leq T_2 \text{ (synchronised stage)} \end{cases}$$

In his paper, Kendall shows that any optimal co-adapted coupling of (B, S) and $(\widetilde{B}, \widetilde{S})$ can be approximated by a coupling with "bang-bang" control. Using this result, Kendall shows that a coupling consisting of an initial synchronised stage followed by the reflection/synchronised coupling has a positive chance of reducing the probability of success. This implies that the reflection/synchronised coupling is tail-optimal among the co-adapted couplings of Brownian motion together with local time at 0. Finally, Kendall shows that the reflection/synchronised coupling is not maximal by comparing the moment generating function of the coupling times. To find the expression of the moment generating function of the coupling time of the reflection/synchronised coupling, he applies excursion theory adapting the calculations of [31, §VI.56]. We will examine excursion theory in detail in Section 3.5. Then, using the joint density of B and S given in [30], he calculates numerically the moment generating function of the maximal coupling time and compares it with that of the reflection/synchronised coupling, showing that their distributions are different. This implies that the tail-optimal co-adapted coupling cannot be maximal.

3.2 The distribution of X_t

In this section, we first of all explore distributional properties of the jumpy Brownian motion. The way X_t is defined allows us to describe its distribution in terms of a standard real valued Brownian motion and the Poisson process that generates the jump times. From the definition of X_t, we have that, if N_t is even, then $X_t = \frac{1}{2}B_t \pmod{2\pi}$, otherwise we have that X_t is at the opposite point on the circumference, i.e. $X_t = \frac{1}{2}B_t + \pi \pmod{2\pi}$. So, if $A \subset [0, 2\pi]$, we have the following:

$$\begin{aligned}\mathbb{P}(X_t \in A | N_t \text{ even}) &= \sum_{k\in\mathbb{Z}} \mathbb{P}\left(\frac{1}{2}B_t \in A + 2k\pi\right),\\ \mathbb{P}(X_t \in A | N_t \text{ odd}) &= \sum_{k\in\mathbb{Z}} \mathbb{P}\left(\frac{1}{2}B_t \in A + (2k+1)\pi\right).\end{aligned} \tag{3.1}$$

Since N is a Poisson process of rate λ, it follows that

$$\begin{aligned}\mathbb{P}(N_t \text{ even}) &= \sum_{k\geq 0} e^{-\lambda t}\frac{(\lambda t)^{2k}}{(2k)!} = e^{-\lambda t}\cosh(\lambda t),\\ \mathbb{P}(N_t \text{ odd}) &= \sum_{k\geq 0} e^{-\lambda t}\frac{(\lambda t)^{2k+1}}{(2k+1)!} = e^{-\lambda t}\sinh(\lambda t).\end{aligned} \tag{3.2}$$

Putting this together, we see that:

$$\begin{aligned}\mathbb{P}(X_t \in A) &= \mathbb{P}(X_t \in A | N_t \text{ even})\mathbb{P}(N_t \text{ even}) + \mathbb{P}(X_t \in A | N_t \text{ odd})\mathbb{P}(N_t \text{ odd})\\ &= \mathbb{P}(X_t \in A | N_t \text{ even})\frac{1+e^{-2\lambda t}}{2} + \mathbb{P}(X_t \in A | N_t \text{ odd})\frac{1-e^{-2\lambda t}}{2}.\end{aligned} \tag{3.3}$$

Now, let X_t and $\hat{X}_t$ be two independent jumpy processes that start from opposite points on the circumference, and let N_t and $\hat{N}_t$ their respective driving Poisson processes.

Remark. From this point, we will often drop the (mod 2π) and denote the circumference of the unit circle as S^1 and the distance on the circle between

$x, y \in S^1$ as $|x - y|$.

We have the following result.

Lemma 3.2. *Let* $A = \{y \in S^1 : |X_0 - y| < |\hat{X}_0 - y|\}$ *with* X_0 *and* $\hat{X}_0$ *opposite points on* S^1*, then for all* $t \geq 0$

$$\mathbb{P}(X_t \in A) \geq \mathbb{P}(\hat{X}_t \in A). \tag{3.4}$$

Proof. In this calculation, we use (3.1), (3.2) and the fact that, since λt is always a real, positive value, then $\cosh(\lambda t) \geq \sinh(\lambda t)$, which implies $\mathbb{P}(N_t \text{ even}) \geq \mathbb{P}(N_t \text{ odd})$.
We can also observe that, since X_0 and $\hat{X}_0$ are antipodal points, if N_t is even, then $X_t = X_0 + \frac{1}{2}B_t$, while if N_t is odd then $X_t = X_0 + \pi + \frac{1}{2}B_t$. So, from the choice of A, it follows that $\mathbb{P}(X_t \in A|N_t \text{ even}) > \mathbb{P}(X_t \in A|N_t \text{ odd})$. In the same way, $\mathbb{P}(\hat{X}_t \in A|N_t \text{ odd}) > \mathbb{P}(\hat{X}_t \in A|N_t \text{ even})$.

$$\begin{aligned}
&\mathbb{P}(X_t \in A) = \mathbb{P}(X_t \in A|N_t \text{ even})\mathbb{P}(N_t \text{ even}) + \mathbb{P}(X_t \in A|N_t \text{ odd})\mathbb{P}(N_t \text{ odd}) \\
&= \mathbb{P}(X_t \in A|N_t \text{ even})\mathbb{P}(N_t \text{ even}) + \mathbb{P}(X_t \in A|N_t \text{ odd})[1 - \mathbb{P}(N_t \text{ even})] \\
&= \mathbb{P}(X_t \in A|N_t \text{ odd}) + \mathbb{P}(N_t \text{ even})[\mathbb{P}(X_t \in A|N_t \text{ even}) - \mathbb{P}(X_t \in A|N_t \text{ odd})] \\
&\geq \mathbb{P}(X_t \in A|N_t \text{ odd}) + \mathbb{P}(N_t \text{ odd})[\mathbb{P}(X_t \in A|N_t \text{ even}) - \mathbb{P}(X_t \in A|N_t \text{ odd})] \\
&= \mathbb{P}(\hat{X}_t \in A|\hat{N}_t \text{ even}) + \mathbb{P}(\hat{N}_t \text{ odd})[\mathbb{P}(\hat{X}_t \in A|\hat{N}_t \text{ odd}) - \mathbb{P}(\hat{X}_t \in A|\hat{N}_t \text{ even})] \\
&= \mathbb{P}(\hat{X}_t \in A|\hat{N}_t \text{ even})\mathbb{P}(\hat{N}_t \text{ even}) + \mathbb{P}(\hat{X}_t \in A|\hat{N}_t \text{ odd})\mathbb{P}(\hat{N}_t \text{ odd}) \\
&= \mathbb{P}(\hat{X}_t \in A).
\end{aligned}$$

□

Now, for simplicity assume that $X_0 = 0$ and $\hat{X}_0 = \pi$ (still independent). Using (3.1) and (3.3), we want to find an expression for the total variation distance between the two processes.

$$||X_t - \hat{X}_t||_{TV} = \max_{A \subseteq S^1} |\mathbb{P}(X_t \in A) - \mathbb{P}(\hat{X}_t \in A)|.$$

To maximise the total variation distance, we need to maximise $\mathbb{P}(X_t \in A) - \mathbb{P}(\hat{X}_t \in A)$. From 3.2 with $X_0 = 0$ and $\hat{X}_0 = \pi$, it follows that we need to choose $A = \{y \in S^1 : |y| < |\pi - y|\}$, so $A = (-\frac{\pi}{2}, \frac{\pi}{2})$. The following lemma gives an explicit expression of the total variation distance between X_t and $\hat{X}_t$.

Lemma 3.3. *Let* $X_0 = 0$ *and* $\hat{X}_0 = \pi$.

$$||X_t - \hat{X}_t||_{TV} = 2e^{-2\lambda t} \sum_{k\in\mathbb{Z}} \left[\Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right) \right]. \tag{3.5}$$

Proof. Following the discussion above and equation (3.3), we can rewrite the total variation distance as

$$\begin{aligned} ||X_t - \hat{X}_t||_{TV} &= \left| \mathbb{P}\left(X_t \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\right) - \mathbb{P}\left(\hat{X}_t \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\right) \right| \\ &= \frac{1+e^{-2\lambda t}}{2} \sum_{k\in\mathbb{Z}} \left[\mathbb{P}\left(2k\pi - \frac{\pi}{2} \le \mathcal{N}\left(0, \frac{t}{4}\right) \le 2k\pi + \frac{\pi}{2}\right) \right. \\ &\qquad \left. -\mathbb{P}\left(2k\pi - \frac{\pi}{2} \le \mathcal{N}\left(\pi, \frac{t}{4}\right) \le 2k\pi + \frac{\pi}{2}\right) \right] \\ &\quad + \frac{1-e^{-2\lambda t}}{2} \sum_{k\in\mathbb{Z}} \left[\mathbb{P}\left(2k\pi + \frac{\pi}{2} \le \mathcal{N}\left(0, \frac{t}{4}\right) \le 2k\pi + \frac{3\pi}{2}\right) \right. \\ &\qquad \left. -\mathbb{P}\left(2k\pi + \frac{\pi}{2} \le \mathcal{N}\left(\pi, \frac{t}{4}\right) \le 2k\pi + \frac{3\pi}{2}\right) \right], \end{aligned}$$

where $\mathcal{N}(\mu, \sigma^2)$ is a normal distribution with mean μ and variance σ^2. Given the choice of A, we know that

$$\mathbb{P}(X_t \in A) = \mathbb{P}(\hat{X}_t \in A^c).$$

The periodic nature of S^1 enables us to deduce that

$$\mathbb{P}(X_t \in A+2k\pi) = \mathbb{P}(\hat{X}_t \in A^c+2k\pi), \quad \text{and} \quad \mathbb{P}(X_t \in A^c+2k\pi) = \mathbb{P}(\hat{X}_t \in A+2k\pi)$$

for all $k \in \mathbb{Z}$, so we can simplify the previous expression

$$\begin{aligned} ||X_t - \hat{X}_t||_{TV} = e^{-2\lambda t} \sum_{k\in\mathbb{Z}} & \left[\mathbb{P}\left(2k\pi - \frac{\pi}{2} \le \mathcal{N}\left(0, \frac{t}{4}\right) \le 2k\pi + \frac{\pi}{2}\right) \right. \\ & \left. - \mathbb{P}\left(2k\pi - \frac{\pi}{2} \le \mathcal{N}\left(\pi, \frac{t}{4}\right) \le 2k\pi + \frac{\pi}{2}\right) \right]. \end{aligned}$$

Now, we analyse the two probabilities in these series. Let $\Phi(x)$ be the cumulative standard normal distribution.

$$\mathbb{P}\left(2k\pi - \frac{\pi}{2} \le \mathcal{N}\left(0, \frac{t}{4}\right) \le 2k\pi + \frac{\pi}{2}\right) = \mathbb{P}\left(\frac{(4k-1)\pi}{\sqrt{t}} \le \mathcal{N}(0,1) \le \frac{(4k+1)\pi}{\sqrt{t}}\right)$$

$$= \Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right),$$

$$\begin{aligned}\mathbb{P}\left(2k\pi - \frac{\pi}{2} \leq \mathcal{N}\left(\pi, \frac{t}{4}\right) \leq 2k\pi + \frac{\pi}{2}\right) &= \mathbb{P}\left(\frac{(4k-3)\pi}{\sqrt{t}} \leq \mathcal{N}(0,1) \leq \frac{(4k-1)\pi}{\sqrt{t}}\right) \\ &= \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k-3)\pi}{\sqrt{t}}\right),\end{aligned}$$

Then,

$$\begin{aligned}||X_t - \hat{X}_t||_{TV} &= e^{-2\lambda t} \sum_{k\in\mathbb{Z}} \left[\Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right)\right. \\ &\qquad \left. -\Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right) + \Phi\left(\frac{(4k-3)\pi}{\sqrt{t}}\right)\right] \\ &= e^{-2\lambda t} \sum_{k\in\mathbb{Z}} \left[\Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right)\right. \\ &\qquad \left. -\Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) + \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right)\right] \\ &= 2e^{-2\lambda t} \sum_{k\in\mathbb{Z}} \left[\Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right)\right].\end{aligned}$$

□

We could now consider the case when X and $\hat{X}$ start from any two points on S^1. In the discussion above, we assumed that X_t and $\hat{X}_t$ start from 0 and π respectively, and that condition allowed us to find an explicit expression of the total variation distance. We could follow the same idea for any two starting states $X_0, \hat{X}_0 \in S^1$, but the expression we would have in Lemma 3.3 would not be as nice as when X and $\hat{X}$ start at distance π. However, we would still be able to determine the set A that maximises the total variation distance following the same idea as in the case $X_0 = 0$ and $\hat{X}_0 = \pi$. Applying 3.2 with any two starting states X_0 and $\hat{X}_0$, we can define A as the symmetric interval of size π centred at X_0.

3.3 Maximal coupling

Let X_t and $\hat{X}_t$ be two jumpy Brownian motions started from points 0 and π respectively. In Lemma 3.3, we found an explicit expression of the total variation distance between two jumpy Brownian motions and, from Definition 1.23, we know that a maximal coupling is a coupling that realises the equality in the coupling inequality Theorem 1.17.

In this section, we show how to construct a maximal coupling of X and $\hat{X}$.

Lemma 3.4. *Let* $\tau = \min\{t \geq 0 : |B_t| = \pi\}$*, where* B *is a standard Brownian motion on* $\mathbb{R}$*, and* $J \sim \text{Exp}(2\lambda)$*, then*

$$||X_t - \hat{X}_t||_{TV} = \mathbb{P}(\tau > t)\mathbb{P}(J > t) = \mathbb{P}(\min\{\tau, J\} > t). \tag{3.6}$$

Proof. The distribution of J is

$$\mathbb{P}(J > t) = \int_t^\infty 2\lambda e^{-2\lambda x} dx = e^{-2\lambda t}.$$

To find the distribution of τ, we can observe that

$$\mathbb{P}(\tau > t) = \mathbb{P}_0\left(B_s \in (-\pi, \pi), 0 \leq s \leq t\right) = \mathbb{P}_\pi(B_s \in (0, 2\pi), 0 \leq s \leq t).$$

From [27], this probability is given by

$$\begin{aligned}
&\mathbb{P}_\pi(B_s \in (0, 2\pi), 0 \leq s \leq t) = \\
&= \sum_{k=-\infty}^{\infty} \left[\Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right) - \Phi\left(\frac{(4k+3)\pi}{\sqrt{t}}\right) + \Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right)\right] \\
&= \sum_{k=-\infty}^{\infty} \left[2\Phi\left(\frac{(4k+1)\pi}{\sqrt{t}}\right) - 2\Phi\left(\frac{(4k-1)\pi}{\sqrt{t}}\right)\right].
\end{aligned}$$

Thus, from (3.5) we deduce that

$$||X_t - \hat{X}_t||_{TV} = \mathbb{P}(\tau > t)\mathbb{P}(J > t).$$

□

Then, a coupling of two jumpy Brownian motions started from 0 and π is maximal if it satisfies (3.6) for all $t \geq 0$. This observation now allows us to define a maximal coupling for X_t and $\hat{X}_t$ in the following way. We let B be a standard Brownian motion on $\mathbb{R}$, and J an independent $\text{Exp}(2\lambda)$ random variable. Let τ be the first time B hits $\pm\pi$. We have two possibilities.

1. On the event $\{\tau < J\}$, let $X_t = \frac{1}{2}B_t \pmod{2\pi}$ and $\hat{X}_t = \pi - \frac{1}{2}B_t \pmod{2\pi}$ for $t \in [0, \tau]$.
2. On the event $\{J < \tau\}$, let $X_t = \frac{1}{2}B_t \pmod{2\pi}$ and $\hat{X}_t = \pi + \frac{1}{2}B_t \pmod{2\pi}$ for $t \in [0, J)$. At time J, we toss a fair coin
   (i) If we get a head, then set $X_J = \hat{X}_J = \pi + \frac{1}{2}B_J \pmod{2\pi}$
   (ii) If we get a tail, then set $X_J = \hat{X}_J = B_J \pmod{2\pi}$.

   In both cases, the two processes will be coupled as soon as we see the first jump.

In other words, we couple X and $\hat{X}$ with starting distance π in the following way: we first observe the process B up to time τ. We then compare τ and J, the time of the first jump: if $\tau < J$, then reflecting B would make X and $\hat{X}$ couple before time J; if $J < \tau$, then we keep X and $\hat{X}$ at distance π so that at time J one of them jumps to the other and they couple. By construction, this coupling satisfies (3.6), but it is not co-adapted since the choice of the strategy depends on what happens in the future.

Considering a different starting distance would probably compromise the maximality of this coupling. If $|X_0 - \hat{X}_0| < \pi$, then using a jump to couple X and $\hat{X}$ would require them to reach distance π first. So, deciding whether reflecting B or jumping is faster would be more complicated, and we would probably have to combine the two strategies.

3.4 Our candidate mean-optimal coupling

We are now going to construct a co-adapted coupling and prove that it is mean-optimal. To define a coupling, we consider two jumpy Brownian motions

$$X_t = \frac{1}{2}B_t + \pi N_t \pmod{2\pi} \qquad \text{and} \qquad \hat{X}_t = x + \frac{1}{2}\hat{B}_t + \pi \hat{N}_t \pmod{2\pi},$$

where B_t and $\hat{B}_t$ are two standard Brownian motions, N_t and $\hat{N}_t$ are two Poisson processes with parameter $\lambda > 0$, and $x \in [0, \pi]$ is the distance on the circle between X_0 and $\hat{X}_0$. To construct the coupling, we consider the process

$$D_t = |X_t - \hat{X}_t| \in [0, \pi]$$

defined as the distance on the circle between X_t and $\hat{X}_t$. We define the following coupling for X_t and $\hat{X}_t$. First, assume that $x \in (0, \pi)$. We start with reflection coupling, i.e. $\hat{B}_t = -B_t$. We shall generate jump times for X and $\hat{X}$ using a Poisson process $\widetilde{N}$ of rate 2λ, and censoring the events of that process. If we see a jump in $\widetilde{N}$ at time t, then we have two possibilities:

i) if $D_{t-} > \frac{\pi}{2}$, then we toss a fair coin.

   a. If we get a head, then X_t jumps, i.e. $X_t = X_{t-} + \pi$.

   b. If we get a tail, then $\hat{X}_t$ jumps, i.e. $\hat{X}_t = \hat{X}_{t-} + \pi$.

   In both cases, D_t jumps downwards, reflecting over $\frac{\pi}{2}$. That is $D_t = \pi - D_{t-} < \frac{\pi}{2}$.

ii) If $D_{t-} \leq \frac{\pi}{2}$, then we toss a fair coin.

   a. If we get a head, then both X_t and $\hat{X}_t$ jump.

   b. If we get a tail, neither of them jumps.

   In both cases, D_t is unchanged, i.e. $D_t = D_{t-}$.

With this construction, D diffuses like a standard Brownian motion in the interval $[0, \pi]$ (that is why we used a factor $\frac{1}{2}$ in our original definition of X) with downward jumps which reflect the process over $\frac{\pi}{2}$, occurring at rate 2λ. If D hits zero then the two jumpy Brownian motions meet; but if D hits π (and so X and $\hat{X}$ arrive at opposite points of S^1) then we consider two possible ways in which to continue coupling them from that point.

Definition 3.5 (Synchronised coupling)**.** Let X and $\hat{X}$ be at distance π. We let them diffuse synchronously, i.e. $B_t = \hat{B}_t$, until we see a jump on $\widetilde{N}$ and the two processes meet. That means that the process D stays at π and, as soon as we see a jump, D jumps to 0.

Definition 3.6 (Reflection coupling)**.** Let X and $\hat{X}$ be at distance π. Under the reflection coupling we follow the same strategy as for $x < \pi$, with $B_t = -\hat{B}_t$, until X and $\hat{X}$ meet.

The coupling time is the time taken by D_t to hit 0

$$T_x = \min\{t : D_t = 0 \text{ with } D_0 = x\}.$$

Figures 3.1 and 3.2 show a simulation of the reflection and synchronised couplings respectively, with starting distance $D_0 = \frac{3\pi}{4}$ and jump rate $\lambda = 0.5$. As we can see in the simulations, the two couplings differ in the behaviour of D at π. Figure 3.1, when D hits π, it reflects and continues diffusing in $[0, \pi]$, under the synchronised coupling illustrated in Figure 3.2, once D hits π, it stays there until the next jump, that will bring the process directly to 0.

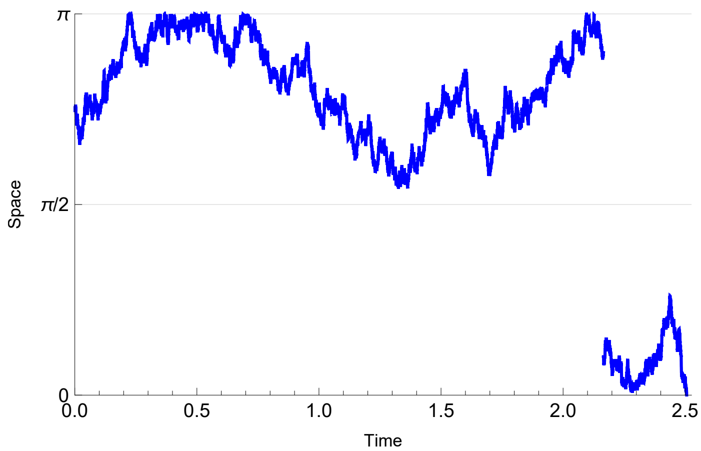


Fig. 3.1: Simulation of the reflection coupling with starting distance $\frac{3\pi}{4}$ and $\lambda = 0.5$. We can see a jump in D at about time 2.15 and the coupling time is approximately 2.5.

The choice we made for our coupling strategy when $x = \pi$ is justified intuitively by the following result.

Lemma 3.7. *Let B_t be a standard Brownian motion in the interval $(-a, a)$ and J be the first interarrival time of a Poisson process of parameter 2λ. Let $\tau = \min\{t \geq 0 : B_t \notin (-a, a)\}$. Then*

$$\mathbb{P}(\tau > J) = 1 - \frac{1}{\cosh(2a\sqrt{\lambda})}.$$

Proof. From [11], we know that for a standard Brownian motion B_t in the

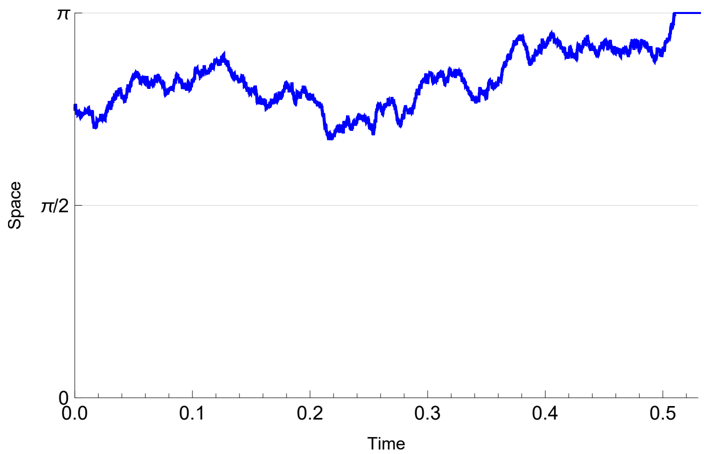


Fig. 3.2: Simulation of the synchronised coupling with starting distance $\frac{3\pi}{4}$ and $\lambda = 0.5$. We can see a jump in D at about time 0.53, which is also the coupling time.

interval $(-a, a)$ we have $\mathbb{E}_0[e^{-\gamma\tau}] = \frac{1}{\cosh(a\sqrt{2\gamma})}$. Since $J \sim \text{Exp}(2\lambda)$,

$$\begin{aligned}\mathbb{P}(\tau > J) &= \mathbb{E}[\mathbb{P}(\tau > J|\tau)] = 1 - \mathbb{E}[e^{-2\lambda\tau}] \\ &= 1 - \frac{1}{\cosh(2a\sqrt{\lambda})}.\end{aligned}$$

□

Now, we have that

$$\mathbb{P}(\tau > J) = 1 - \frac{1}{\cosh(2a\sqrt{\lambda})} \to \begin{cases} 1 & \text{as } \lambda \to 0 \\ 0 & \text{as } \lambda \to \infty \end{cases}$$

Interpreting this result in terms of the jumpy Brownian motion, if $\lambda \to 0$, i.e. if the jumps happen more rarely, then the process D_t hits 0 by reflecting before we see the first jump, so it is reasonable to expect that the reflection coupling is faster. On the other hand, if $\lambda \to \infty$, i.e. if we often see jumps, then it is convenient to keep D_t at π because, as soon as we see the next jump, D_t jumps directly to 0.

The main aim of this chapter is proving the following theorem.

Theorem 3.8. *Let X and $\hat{X}$ be two jumpy Brownian motions with jumps* *occurring at rate λ. Let $\lambda^\star$ be the unique solution to the equation*

$$\frac{\pi}{2\cosh(\pi\sqrt{\lambda})-1}=\frac{\operatorname{cosech}(\pi\sqrt{\lambda})}{2\sqrt{\lambda}},$$

i.e. $\lambda^\star = 0.08337$. *Then,*

(i) If $\lambda < \lambda^\star$, then the reflection coupling of Definition 3.6 is the unique mean-optimal coupling in the class of co-adapted couplings.

(ii) If $\lambda > \lambda^\star$, then the synchronised coupling of Definition 3.5 is the unique mean-optimal coupling in the class of co-adapted couplings.

To prove the mean-optimality of our coupling, we follow the idea of Kendall [18] of using excursion theory to find an expression of the moment generating function of the coupling times for the reflection and synchronised couplings defined above. In the next section, we introduce excursion theory, which is essential in Section 3.6 to calculate the expressions of the Laplace transform of the coupling times. In Section 3.8, we use the Laplace transform to calculate the expectation of the coupling times, which gives us the condition to find the value of $\lambda^\star$ mentioned in the statement of Theorem 3.8. Finally, in Section 3.9, we apply Bellman's principle of optimality to complete the proof of Theorem 3.8 and establish the mean-optimality of the reflection and synchronised couplings.

3.5 Excursion theory

Excursion theory was introduced by K. Itô in [17], and it is one of the most useful techniques in the study of the behaviour of Markov processes. Excursion theory consists in breaking time into intervals and analysing the excursions of the process on those intervals. Itô discovered that the excursions over different intervals are independent and identically distributed. This makes excursion theory a method that simplifies the study of the diffusion.

Excursion theory is very efficient when, given a random variable T, we want to study the behaviour of the process at time T. Consider the interval $[0,T]$ and a recurrent state a for the Markov process. We can split $[0,T]$ into smaller time intervals at the times when the process comes back at a. By the strong Markov property, we have that the excursions of the process over those intervals are independent. Finally, we mark the excursions with an independent Poisson process. Using the distribution of the marks, we can then use the law of the excursions to calculate the Laplace transform of the random variable T.

In this section, we give an overview of Excursion Theory, describing its application to standard Brownian motion: this will be helpful when we come to apply it to the more complicated jumpy Brownian motion. We start with defining the excursion point process.

Definition 3.9 (Inverse local time)**.** Let $B = (B_t)_{t\geq 0}$ be a one-dimensional Brownian motion started at recurrent state a. We denote by L_t^a the *local time* at a at time t, i.e. the time spent by B at the state a by time t. The corresponding *inverse local time* is defined as

$$\gamma_t^a = \inf\{u > 0 : L_u^a > t\}.$$

Definition 3.10 (Excursion point process)**.** Let $B = (B_t)_{t\geq 0}$ be a one-dimensional Brownian motion started at recurrent state a. The *excursion point process* from a is defined as

$$\Pi = \{(t, e_t) : \gamma_t \neq \gamma_{t-}\},$$

where

$$e_t(s) = \begin{cases} B(\gamma_{t-} + s) & \text{for } 0 \leq s < \gamma_t - \gamma_{t-}, \\ a & \text{for } s \geq \gamma_t - \gamma_{t-}. \end{cases}$$

We call e_t the *excursion* at local time t.

In other words, we can split the path at the times when the Brownian motion comes back to state a. An excursion of the Brownian motion is the portion of its path in one of those intervals.

The following result, showed in detail in [31, §VI.55], describes the measure of the excursions.

Theorem 3.11. *Let $W = (W_t)_{t\geq 0}$ be a reflecting Brownian motion on $\mathbb{R}^+$. Let f be an excursion of W, i.e. a continuous function from $\mathbb{R}^+$ to $\mathbb{R}^+$ such that $f(0) = 0$ and*

$$f(t) > 0, \quad 0 < t < \zeta \qquad \textit{and} \qquad f(t) = 0, \quad t \geq \zeta,$$

for some $\zeta > 0$. *Let* $S(f) = \sup_t f(t)$ *for the excursion* f, *and let* $n(\cdot)$ *denote the excursion measure, then*

(i) $n(S(f) > x) = x^{-1}$ *for* $x > 0$

(ii) Each excursion f *can be split at its maximum* $S(f)$ *into two pieces representing two independent Bessel processes of dimension* 3 *started at* 0.

Recall that a Bessel process is defined as follows.

Definition 3.12 (Bessel process)**.** Let $B = (B_t)_{t\geq 0}$ be an n-dimensional Brownian motion started at x. An n-dimensional *Bessel process* started at x, denoted by $\mathrm{BES}^x(n)$, is the one-dimensional process $R = (R_t)_{t\geq 0}$ defined as

$$R_t = ||B_t||,$$

where $||\cdot||$ denotes the Euclidean norm.

To illustrate how Excursion Theory can be used to solve problems related to Brownian motion, we report, as an example, the calculations showed in [31, §VI.56]. We will use the same ideas in Section 3.6, adapting these calculations to the jumpy Brownian process.

Example 3.13. Let $B = (B_t)_{t\geq 0}$ be a one-dimensional Brownian motion started at 0. Define

$$T = \inf\{t \geq 0 : |B_t| = 1\}, \qquad \sigma = \sup\{t < T : B_t = 0\},$$

i.e. T is the hitting time of ± 1, and σ is the last time when B_t is equal to 0 before hitting level ± 1. We also define

$$A_t^+ = \int_0^t \mathbb{1}_{[0,\infty)}(B_s)ds, \qquad A_t^- = t - A_t^+,$$

i.e. A_t^+ is the time spent by B in the interval $[0, \infty)$ up to time t. We want to find the joint law of A_σ^+, A_σ^-, and T. To do that, we define

$$\xi = \mathbb{E}[e^{-\alpha A_\sigma^+ - \beta A_\sigma^- - \gamma(T-\sigma)}], \tag{3.7}$$

where $\alpha, \beta, \gamma > 0$. We want to interpret ξ using excursions of B from 0. We define the following disjoint sets:

$$U_+ = \{\text{upward excursions from 0 that do not hit 1}\}$$
$$U_- = \{\text{downward excursions from 0 that do not hit } -1\}$$
$$U_1 = \{\text{excursions from 0 that hit } +1 \text{ or } -1\}.$$

Following this classification, we split the path of B into excursions, and we mark them with independent Poisson processes of different rates independent from B. We mark the excursions in U_+ at rate α, the excursions in U_- at rate β, and the excursions in U_1 at rate γ. Then, we define the following sets.

$$U_+^\star = \{\text{excursions in } U_+ \text{ that contain a mark}\}$$
$$U_-^\star = \{\text{excursions in } U_- \text{ that contain a mark}\}$$
$$U_1^\star = \{\text{excursions in } U_1 \text{ that contain a mark before reaching 1}\}$$
$$U_1^0 = U_1 \setminus U_1^\star.$$

Then, we can observe that the expectation in (3.7) can be rewritten in terms of a combination of the independent Poisson processes that produce the marks on the excursions.

$$\begin{aligned}\xi &= \mathbb{P}(\text{the first excursion in } U_+^\star \cup U_-^\star \cup U_1 \text{ lies in } U_1^0) \\ &= \frac{n(U_1^0)}{n(U_+^\star) + n(U_-^\star) + n(U_1)},\end{aligned} \tag{3.8}$$

where recall that $n(\cdot)$ denotes the excursion measure.
Now, we need to calculate the excursion measures in equation (3.8).

$$\begin{aligned}n(U_1^0) &= n(U_1)\mathbb{P}(\text{excursion which escapes from } [-1,1] \text{ has no marks before it leaves } [-1,1]) \\ &= n(U_1)\mathbb{E}[e^{-\gamma H^3(1)}],\end{aligned}$$

where $H^3(x) = \inf\{t : R_t = x\}$, and $R = (R_t)_{t\geq 0}$ is a $\text{BES}^0(3)$ process. This last passage comes from Theorem 3.11(ii), while from 3.11(i) we have that $n(U_1) = 1$.

To calculate the expected hitting time of R at level x, we want to use the following lemma.

Lemma 3.14. *Let R_t be a* $\mathrm{BES}^0(3)$*, and let* $H^3(x) = \inf\{t : R_t = x\}$*. Then for* $\lambda > 0$*,*

$$\mathbb{E}[e^{-\lambda H^3(x)}] = x\sqrt{2\lambda}\,\mathrm{cosech}(x\sqrt{2\lambda}).$$

Proof. Let $M_t = e^{-\lambda t} R_t^{-1} \sinh(\sqrt{2\lambda} R_t)$ be a local martingale, and let $T = H^3(x)$. From Doob's optional stopping theorem, we know that $\mathbb{E}[M_T] = \mathbb{E}[M_0]$. Moreover,

$$M_0 = \lim_{t\to 0} \frac{e^{-\lambda t}\sinh(\sqrt{2\lambda}R_t)}{R_t} = \lim_{t\to 0} \frac{\sinh(\sqrt{2\lambda}R_t)}{R_t} = \sqrt{2\lambda}.$$

Then

$$\mathbb{E}[M_T] = \mathbb{E}\left[\frac{e^{-\lambda T}\sinh(\sqrt{2\lambda}R_T)}{R_T}\right] = \frac{\sinh(x\sqrt{2\lambda})}{x}\mathbb{E}[e^{-\lambda T}] = \sqrt{2\lambda},$$

and rearranging yields

$$\mathbb{E}[e^{-\lambda T}] = x\sqrt{2\lambda}\,\mathrm{cosech}(x\sqrt{2\lambda}).$$

□

Applying Lemma 3.14 to our expression for $n(U_1^0)$ gives

$$\mathbb{E}[e^{-\gamma H^3(1)}] = \sqrt{2\gamma}\,\mathrm{cosech}(\sqrt{2\gamma}),$$

so we obtain

$$n(U_1^0) = \sqrt{2\gamma}\,\mathrm{cosech}(\sqrt{2\gamma}).$$

To calculate $n(U_+^\star)$, we apply Theorem 3.11(*ii*). Each excursion in $U_+^\star$ has a maximum in $(0,1)$. Given that the excursion has maximum x, we can split it into two pieces, one before the maximum is reached, the other one after the maximum is reached. From Theorem 3.11(*ii*), we have that these two pieces are independent and each distributed as a $\mathrm{BES}^0(3)$ process.

Since we are now interested in the event that there is at least one mark of rate α in the excursions that belong to $U_+^\star$, from Lemma 3.14, we have,

$$n(U_+^\star) = \int_0^1 \frac{1}{2}x^{-2}[1-(x\sqrt{2\alpha}\,\mathrm{cosech}(x\sqrt{2\alpha}))^2]dx = \frac{1}{2}[\sqrt{2\alpha}\coth(\sqrt{2\alpha})-1].$$

Applying the same ideas to $U_-^\star$ yields

$$n(U_-^\star) = \frac{1}{2}[\sqrt{2\beta}\coth(\sqrt{2\beta}) - 1].$$

Putting all of these expressions into equation (3.8), we arrive at the formula stated in [31]:

$$\xi = \mathbb{E}[e^{-\alpha A_\sigma^+ - \beta A_\sigma^- - \gamma(T-\sigma)}] = \frac{2\sqrt{2\gamma}\,\text{cosech}(\sqrt{2\gamma})}{\sqrt{2\alpha}\coth(\sqrt{2\alpha}) + \sqrt{2\beta}\coth(\sqrt{2\beta})}.$$

To apply Excursion Theory to the jumpy Brownian motion, we will need to use the ideas illustrated in Example 3.13. The following lemma extends those results to a generic interval $(-c, d)$, with $c, d > 0$.

Lemma 3.15. *Let B_t be a Brownian motion in the interval $(-c, d)$ starting from 0, where $c, d > 0$. Let $T_0(\{-c, d\}) = \inf\{t : B_t \in \{-c, d\}\}$, $T_0(d) = \inf\{t : B_t = d\}$ and $T_0(c) = \inf\{t : B_t = -c\}$. Then*

$$(i) \quad \mathbb{E}[e^{-\gamma T_0(\{-c,d\})}] = \frac{\text{cosech}(d\sqrt{2\gamma}) + \text{cosech}(c\sqrt{2\gamma})}{\coth(d\sqrt{2\gamma}) + \coth(c\sqrt{2\gamma})}.$$

$$(ii) \quad \mathbb{E}[e^{-\gamma T_0(\{d\})}] = \frac{\text{cosech}(d\sqrt{2\gamma})}{\coth(d\sqrt{2\gamma}) + \coth(c\sqrt{2\gamma})}.$$

$$(iii) \quad \mathbb{E}[e^{-\gamma T_0(\{-c\})}] = \frac{\text{cosech}(c\sqrt{2\gamma})}{\coth(d\sqrt{2\gamma}) + \coth(c\sqrt{2\gamma})}.$$

Observe that if $c = d = 1$, then for (i) we would get $\text{sech}(\sqrt{2\gamma})$, which agrees with ξ above when $\alpha = \beta = \gamma$.

Proof. To prove this lemma, we want to apply excursion theory to the interval $(-c, d)$ adapting the calculations of Example 3.13. We distinguish four different types of excursions:

$$\begin{aligned}
U_+ &= \{\text{upward excursions that do not hit } d\},\\
U_- &= \{\text{downward excursions that do not hit } -c\},\\
U_d &= \{\text{upward excursions that hit } d\},\\
U_c &= \{\text{downward excursions that hit } -c\}.
\end{aligned}$$

We mark the excursions at rate γ, and we consider the following sets

$$U_+^\star = \{\text{marked upward excursions that do not hit } d\},$$

$$U_-^\star = \{\text{marked downward excursions that do not hit } -c\},$$
$$U_d^0 = \{\text{upward excursions with no marks before hitting } d\},$$
$$U_c^0 = \{\text{downward excursions with no marks before hitting } -c\}.$$

Starting from (i), we can express the Laplace transform of $T_0(\{-c,d\})$ as the probability that we have an excursion with no marks that hits d or $-c$ before an excursion with at least one mark:

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_0(\{-c,d\})}] &= \mathbb{P}_0(\text{the first excursion in } U_+^\star \cup U_-^\star \cup U_d \cup U_c \text{ lies in } U_d^0 \cup U_c^0) \\ &= \frac{n(U_d^0)+n(U_c^0)}{n(U_+^*)+n(U_-^*)+n(U_d)+n(U_c)}. \end{aligned} \tag{3.9}$$

From the definition of U_d, we have that since we are considering only upward excursions to $+d$ (not $-d$) then $n(U_d) = \frac{1}{2d}$. In the same way, $n(U_c) = \frac{1}{2c}$. As before,

$$\begin{aligned} n(U_d^0) &= n(U_d)\mathbb{P}_0(\text{excursion that hits } d \text{ has no marks before hitting } d) \\ &= \frac{1}{2d}\mathbb{E}[e^{-\gamma H^3(d)}]. \end{aligned}$$

From Lemma 3.14, we obtain

$$\mathbb{E}[e^{-\gamma H^3(d)}] = d\sqrt{2\gamma}\,\mathrm{cosech}(d\sqrt{2\gamma}),$$

which implies

$$n(U_d^0) = \frac{\sqrt{2\gamma}}{2}\,\mathrm{cosech}(d\sqrt{2\gamma}).$$

Doing the same for the downward excursions implies

$$n(U_c^0) = \frac{\sqrt{2\gamma}}{2}\,\mathrm{cosech}(c\sqrt{2\gamma}).$$

To calculate the measure of $U_+^\star$, we consider an upward excursion with maximum y, and we split the excursion into two pieces, one before y is reached and the other after y. The two sections of the excursion behave like independent 3-dimensional Bessel processes started from 0, and we want at least one mark on their union, so

$$n(U_+^*) = \int_0^d \frac{1}{2}y^{-2}[1-(y\sqrt{2\gamma}\,\mathrm{cosech}(y\sqrt{2\gamma}))^2]dy = \frac{1}{2}\left(\sqrt{2\gamma}\coth(d\sqrt{2\gamma}) - \frac{1}{d}\right).$$

We do the same for $U_-^\star$, and we obtain

$$n(U_-^*) = \frac{1}{2}\left(\sqrt{2\gamma}\coth(c\sqrt{2\gamma}) - \frac{1}{c}\right).$$

Then, from (3.9),

$$\mathbb{E}[e^{-\gamma T_0(\{-c,d\})}] = \frac{\operatorname{cosech}(d\sqrt{2\gamma}) + \operatorname{cosech}(c\sqrt{2\gamma})}{\coth(d\sqrt{2\gamma}) + \coth(c\sqrt{2\gamma})},$$

which completes the proof of part (i). To show parts (ii) and (iii), we just observe that

$$\mathbb{E}[e^{-\gamma T_0(\{d\})}] = \frac{n(U_d^0)}{n(U_+^*) + n(U_-^*) + n(U_d) + n(U_c)} = \frac{\operatorname{cosech}(d\sqrt{2\gamma})}{\coth(d\sqrt{2\gamma}) + \coth(c\sqrt{2\gamma})},$$

and

$$\mathbb{E}[e^{-\gamma T_0(\{-c\})}] = \frac{n(U_c^0)}{n(U_+^*) + n(U_-^*) + n(U_d) + n(U_c)} = \frac{\operatorname{cosech}(c\sqrt{2\gamma})}{\coth(d\sqrt{2\gamma}) + \coth(c\sqrt{2\gamma})}.$$

□

3.6 Laplace transform of the coupling time

In this section, we show how to find an expression for the Laplace transform of the coupling time for the jumpy Brownian motion. The expressions of the Laplace transform are given by the following lemmas for the two coupling strategies defined in Definitions 3.6 and 3.5.

Lemma 3.16. *Let T_x^r be the coupling time for the jumpy Brownian motion with $D_0 = x$ under the reflection coupling. Let $\alpha = \sqrt{2(2\lambda+\gamma)}$ and $\beta = \sqrt{2\gamma}$. Then, we have the following formulas for the Laplace transform of T_x^r.*

1. *If $x \in (0, \frac{\pi}{2})$,*

$$\mathbb{E}[e^{-\gamma T_x^r}] = \frac{\operatorname{cosech}(x\beta)[2 - \operatorname{sech}(\alpha\frac{\pi}{2})\operatorname{sech}(\beta\frac{\pi}{2})]}{[2 - \operatorname{sech}(\alpha\frac{\pi}{2})\operatorname{sech}(\beta\frac{\pi}{2})][\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)]} - \frac{\operatorname{cosech}((\frac{\pi}{2} - x)\beta)[-2\operatorname{sech}(\beta\frac{\pi}{2}) + \operatorname{sech}(\alpha\frac{\pi}{2})]}{[2 - \operatorname{sech}(\alpha\frac{\pi}{2})\operatorname{sech}(\beta\frac{\pi}{2})][\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)]}.$$

2. *If* $x \in (\frac{\pi}{2}, \pi)$

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_x^r}] = &\frac{\operatorname{cosech}((\pi - x)\beta) + \operatorname{cosech}((x - \frac{\pi}{2})\beta)\operatorname{sech}(\beta\frac{\pi}{2})}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} \\ &- \frac{\operatorname{cosech}((x - \frac{\pi}{2})\beta)\operatorname{sech}(\alpha\frac{\pi}{2})\tanh^2(\beta\frac{\pi}{2})}{[2 - \operatorname{sech}(\alpha\frac{\pi}{2})\operatorname{sech}(\beta\frac{\pi}{2})][\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)]} \\ &- \frac{2\frac{\beta}{\alpha}\operatorname{cosech}((\pi - x)\alpha)\tanh(\alpha\frac{\pi}{2})\tanh(\beta\frac{\pi}{2})}{[2 - \operatorname{sech}(\alpha\frac{\pi}{2})\operatorname{sech}(\beta\frac{\pi}{2})][\coth((\pi - x)\alpha) + \coth((x - \frac{\pi}{2})\alpha)]}.\end{aligned}$$

3. *If* $x = \frac{\pi}{2}$,

$$\mathbb{E}[e^{-\gamma T_{\frac{\pi}{2}}^r}] = \frac{2\cosh(\frac{\pi}{2}\alpha) - \cosh(\frac{\pi}{2}\beta)}{2\cosh(\frac{\pi}{2}\beta)\cosh(\frac{\pi}{2}\alpha) - 1}.$$

4. *If* $x = \pi$,

$$\mathbb{E}[e^{-\gamma T_\pi^r}] = 1 - \frac{\frac{\beta}{\alpha}\tanh\left(\frac{\pi}{2}\alpha\right)\tanh\left(\frac{\pi}{2}\beta\right)}{1 - \frac{1}{2}\operatorname{sech}\left(\frac{\pi}{2}\beta\right)\operatorname{sech}\left(\frac{\pi}{2}\alpha\right)}.$$

Lemma 3.17. *Let* T_x^s *be the coupling time for the jumpy Brownian motion with* $D_0 = x$ *under the synchronised coupling. Let* $\alpha = \sqrt{2(2\lambda + \gamma)}$ *and* $\beta = \sqrt{2\gamma}$. *Then, we have the following formulas for the Laplace transform of* T_x^s.

1. *If* $x \in (0, \frac{\pi}{2})$,

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_x^s}] = &\frac{\operatorname{cosech}(x\beta) + \operatorname{cosech}((\frac{\pi}{2} - x)\beta)\operatorname{sech}(\beta\frac{\pi}{2})}{\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)} \\ &- \frac{\beta}{2\alpha}\frac{\operatorname{cosech}((\frac{\pi}{2} - x)\beta)\tanh(\beta\frac{\pi}{2})\operatorname{cosech}(\alpha\frac{\pi}{2})}{\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)}.\end{aligned}$$

2. *If* $x \in (\frac{\pi}{2}, \pi)$

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_x^s}] = &\frac{\operatorname{cosech}((\pi - x)\beta) + \operatorname{cosech}((x - \frac{\pi}{2})\beta)\operatorname{sech}(\beta\frac{\pi}{2})}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} \\ &- \frac{\beta}{2\alpha}\frac{\operatorname{cosech}((x - \frac{\pi}{2})\beta)\tanh(\beta\frac{\pi}{2})\operatorname{cosech}(\alpha\frac{\pi}{2})}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} \\ &- \frac{\beta^2\operatorname{cosech}((\pi - x)\alpha)}{\alpha^2[\coth((\pi - x)\alpha) + \coth((x - \frac{\pi}{2})\alpha)]}.\end{aligned}$$

3. If $x = \frac{\pi}{2}$,

$$\mathbb{E}\left[e^{-\gamma T^s_{\frac{\pi}{2}}}\right] = \frac{1}{\cosh(\frac{\pi}{2}\beta)} - \frac{1}{2}\frac{\beta}{\alpha}\tanh\left(\beta\frac{\pi}{2}\right)\operatorname{cosech}\left(\frac{\pi}{2}\alpha\right).$$

4. If $x = \pi$

$$\mathbb{E}[e^{-\gamma T^s_{\pi}}] = \frac{2\lambda}{\gamma + 2\lambda}.$$

The remainder of Section 3.6 is dedicated to the proof of Lemmas 3.16 and 3.17. We shall split our analysis into three cases, depending on the value of D_0.

I. $D_0 = \frac{\pi}{2}$

II. $D_0 = \pi$

III. $D_0 \neq \frac{\pi}{2}, \pi$.

In the remainder of the chapter, we will use the following notations:

- T^r_x for any $x \in [0, \pi]$ as the coupling time with $D_0 = x$ under the reflection coupling,
- T^s_x for any $x \in [0, \pi]$ as the coupling time with $D_0 = x$ under the synchronised coupling.

3.6.1 Case I. $D_0 = \frac{\pi}{2}$

Consider the jumpy Brownian motion D_t starting from $\frac{\pi}{2}$. In this case, the process behaves identically for both the coupling strategies until D hits the set $E = \{0, \pi\}$. This means that, if $T_{\frac{\pi}{2}}(E)$ denotes the hitting time of the set E with $D_0 = \frac{\pi}{2}$, $T_{\frac{\pi}{2}}(E)$ does not depend on the coupling. We can write

$$T^r_{\frac{\pi}{2}} = T_{\frac{\pi}{2}}(E) + \mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)} = \pi\}} T^r_{\pi},$$

and similarly

$$T^s_{\frac{\pi}{2}} = T_{\frac{\pi}{2}}(E) + \mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)} = \pi\}} T^s_{\pi},$$

where T^r_π and T^s_π are independent from $T_{\frac{\pi}{2}}(E)$. Then, we can express the Laplace transform of $T^r_{\frac{\pi}{2}}$ as

$$\begin{aligned}\mathbb{E}\left[e^{-\gamma T^r_{\frac{\pi}{2}}}\right] &= \mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=0\}} + e^{-\gamma(T_{\frac{\pi}{2}}(E)+T^r_\pi)}\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=\pi\}}\right] \\ &= \mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}(1-\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=\pi\}}) + e^{-\gamma(T_{\frac{\pi}{2}}(E)+T^r_\pi)}\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=\pi\}}\right] \\ &= \mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\right] + \mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=\pi\}}\right]\left(\mathbb{E}\left[e^{-\gamma T^r_\pi}\right]-1\right).\end{aligned} \tag{3.10}$$

To determine the first term in 3.10, we can observe that the time for D to hit E is the same as for a Brownian motion to hit $\pm\frac{\pi}{2}$. So, we can rewrite the Laplace transform of $T_{\frac{\pi}{2}}(E)$ for the process D_t started at $\frac{\pi}{2}$ as the Laplace transform of the hitting time $T_0(\{-\frac{\pi}{2},\frac{\pi}{2}\})$ for a Brownian motion started at 0:

$$\mathbb{E}[e^{-\gamma T_{\frac{\pi}{2}}(E)}] = \mathbb{E}[e^{-\gamma T_0(\{-\frac{\pi}{2},\frac{\pi}{2}\})}] = \frac{1}{\cosh(\frac{\pi}{2}\sqrt{2\gamma})}.$$

To find an expression for the second term in (3.10), we apply excursion theory as in Section 3.5. We distinguish the following types of excursions for a Brownian motion started at $\frac{\pi}{2}$:

1. upward excursions that do not hit π, and which have no marks
2. upward excursions that do not hit π, with at least one mark
3. downward excursions that do not hit 0
4. upward excursions that hit π, and which have no marks
5. upward excursions that hit π, with at least one mark
6. downward excursions that hit 0.

Let $\sigma = \sup\{t < T_{\frac{\pi}{2}}(E) : D_{T_{\frac{\pi}{2}}(E)} = \frac{\pi}{2}\}$ be the last time the process leaves $\frac{\pi}{2}$ before hitting E. First, we rewrite the expectation we are considering in terms of σ and $T_{\frac{\pi}{2}}(E)$:

$$\begin{aligned}\mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=\pi\}}\right] &= \mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\mathbb{E}\left[\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=\pi\}}\middle|\sigma, T_{\frac{\pi}{2}}(E)\right]\right] \\ &= \mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\mathbb{P}\left(D_{T_{\frac{\pi}{2}}(E)}=\pi\middle|\sigma, T_{\frac{\pi}{2}}(E)\right)\right]\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\frac{1}{2}\mathbb{P}\left(\text{no jumps in } \widetilde{N} \text{ in time } [\sigma, T_{\frac{\pi}{2}}(E)]\middle|\sigma, T_{\frac{\pi}{2}}(E)\right)\right] \\
&= \frac{1}{2}\mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)-2\lambda(T_{\frac{\pi}{2}}(E)-\sigma)}\right] \\
&= \frac{1}{2}\mathbb{E}\left[e^{-(2\lambda+\gamma)(T_{\frac{\pi}{2}}(E)-\sigma)-\gamma\sigma}\right]. \qquad (3.11)
\end{aligned}
$$

We will use the notations $\alpha = \sqrt{2(2\lambda+\gamma)}$ and $\beta = \sqrt{2\gamma}$. Now, we define the following sets for a Brownian motion started at 0:

$$
\begin{aligned}
U_+ &= \{\text{positive excursions that do not hit } \frac{\pi}{2}\} \\
U_- &= \{\text{negative excursions that do not hit } -\frac{\pi}{2}\} \\
U_{\frac{\pi}{2}} &= \left\{\text{positive and negative excursions that hit } \left\{-\frac{\pi}{2}, \frac{\pi}{2}\right\}\right\}
\end{aligned}
$$

We mark the excursions in $U_\pm$ at rate γ and the excursions in $U_{\frac{\pi}{2}}$ at rate $2\lambda+\gamma$, and we consider the sets

$$
\begin{aligned}
U^\star_\pm &= \{\text{excursions in } U^\pm \text{ with at least one mark}\} \\
U^\star_{\frac{\pi}{2}} &= \{\text{excursions in } U_{\frac{\pi}{2}} \text{ with at least one mark}\} \\
U^0_{\frac{\pi}{2}} &= \{\text{excursions in } U_{\frac{\pi}{2}} \text{ with no marks}\}.
\end{aligned}
$$

We can now rewrite the following in terms of the measures of these sets:

$$
\begin{aligned}
\mathbb{E}\left[e^{-(2\lambda+\gamma)(T_{\frac{\pi}{2}}(E)-\sigma)-\gamma\sigma}\right] &= \mathbb{P}(\text{the first excursion in } U_{\frac{\pi}{2}} \cup U^\star_\pm \text{ is in } U^0_{\frac{\pi}{2}}) \\
&= \frac{n(U^0_{\frac{\pi}{2}})}{n(U^\star_\pm) + n(U_{\frac{\pi}{2}})}.
\end{aligned}
$$

First, we calculate the measure in the numerator.

$$
\begin{aligned}
n(U^0_{\frac{\pi}{2}}) &= n(U_{\frac{\pi}{2}})\mathbb{P}\left(\text{excursion that hits } \left\{-\frac{\pi}{2}, \frac{\pi}{2}\right\} \text{ has no marks before hitting it}\right) \\
&= \frac{2}{\pi}\mathbb{E}\left[e^{-(2\lambda+\gamma)H^3(\frac{\pi}{2})}\right],
\end{aligned}
$$

where $n(U_{\frac{\pi}{2}})$ is calculated using 3.11*(i)*, and $H^3(\frac{\pi}{2})$ is the time for a 3-dimensional Bessel process to hit $\frac{\pi}{2}$.

To find an expression for the expectation in the previous paragraph, we apply Lemma 3.14 with $x = \frac{\pi}{2}$ and $\alpha = 2\lambda + \gamma$ to obtain

$$\mathbb{E}\left[e^{-(2\lambda+\gamma)H^3(\frac{\pi}{2})}\right] = \frac{\pi}{2}\alpha \operatorname{cosech}\left(\alpha\frac{\pi}{2}\right)$$

and therefore

$$n(U^0_{\frac{\pi}{2}}) = \alpha \operatorname{cosech}\left(\alpha\frac{\pi}{2}\right).$$

To calculate $n(U^\star_+)$, given x is the maximum of the excursion, we can split it into two parts, the part before x and the part after x. Both of these two parts of the excursion are independent and can be interpreted as a 3-dimensional Bessel process started from 0. We want at least one mark of rate γ in the union these two parts, so

$$n(U^\star_\pm) = 2n(U^\star_+) = \int_0^{\frac{\pi}{2}} x^{-2}[1 - (\beta x \operatorname{cosech}(\beta x))^2]dx = \beta \coth\left(\beta\frac{\pi}{2}\right) - \frac{2}{\pi}.$$

Then,

$$\mathbb{E}\left[e^{-(2\lambda+\gamma)(T_{\frac{\pi}{2}}(E)-\sigma)-\gamma\sigma}\right] = \frac{\alpha \operatorname{cosech}(\frac{\pi}{2}\alpha)}{\beta\coth(\beta\frac{\pi}{2}) - \frac{2}{\pi} + \frac{2}{\pi}} = \frac{\alpha}{\beta}\tanh\left(\beta\frac{\pi}{2}\right)\operatorname{cosech}\left(\alpha\frac{\pi}{2}\right)$$

and from (3.11)

$$\mathbb{E}\left[e^{-\gamma T_{\frac{\pi}{2}}(E)}\mathbb{1}_{\{D_{T_{\frac{\pi}{2}}(E)}=\pi\}}\right] = \frac{1}{2}\frac{\alpha}{\beta}\tanh\left(\beta\frac{\pi}{2}\right)\operatorname{cosech}\left(\frac{\pi}{2}\alpha\right).$$

From (3.10), we obtain

$$\mathbb{E}[e^{-\gamma T^r_{\frac{\pi}{2}}}] = \frac{1}{\cosh(\frac{\pi}{2}\beta)} + \frac{1}{2}\frac{\alpha}{\beta}\tanh\left(\beta\frac{\pi}{2}\right)\operatorname{cosech}\left(\frac{\pi}{2}\alpha\right)\left(\mathbb{E}\left[e^{-\gamma T^r_\pi}\right] - 1\right), \tag{3.12}$$

with the same equation holding for $T^s_{\frac{\pi}{2}}$ in terms of T^s_π.

At this point, we need to find expressions for $\mathbb{E}\left[e^{-\gamma T^r_\pi}\right]$ and $\mathbb{E}\left[e^{-\gamma T^s_\pi}\right]$.

3.6.2 Case II. $D_0 = \pi$

In this section, we calculate the Laplace transforms of T^s_π and T^r_π, the coupling times respectively for the synchronised coupling and the reflection coupling with $D_0 = \pi$. We continue to write $\alpha = \sqrt{2(2\lambda+\gamma)}$ and $\beta = \sqrt{2\gamma}$.

Synchronised process

In this case, $D_t = \pi$ until the first jump is generated by the Poisson process $\widetilde{N}_t$. The coupling time is the time taken to generate a jump at rate 2λ, so $T^s_\pi \sim \text{Exp}(2\lambda)$, and

$$\mathbb{E}[e^{-\gamma T^s_\pi}] = \int_0^\infty e^{-\gamma x} 2\lambda e^{-2\lambda x} dx = 2\lambda \int_0^\infty e^{-(\gamma+2\lambda)x} dx = \frac{2\lambda}{\gamma + 2\lambda}.$$

From (3.12), we obtain the following formula

$$\begin{aligned}\mathbb{E}\left[e^{-\gamma T^s_{\frac{\pi}{2}}}\right] &= \frac{1}{\cosh(\frac{\pi}{2}\beta)} + \frac{1}{2}\frac{\alpha}{\beta}\tanh\left(\beta\frac{\pi}{2}\right)\operatorname{cosech}\left(\frac{\pi}{2}\alpha\right)\left(\frac{2\lambda}{\gamma+2\lambda} - 1\right) \\ &= \frac{1}{\cosh(\frac{\pi}{2}\beta)} - \frac{1}{2}\frac{\beta}{\alpha}\tanh\left(\beta\frac{\pi}{2}\right)\operatorname{cosech}\left(\frac{\pi}{2}\alpha\right).\end{aligned}$$

This completes the proof of parts *3* and *4* of Lemma 3.17.

Reflected process

We want to apply again excursion theory to calculate the Laplace transform of the coupling time T^r_π. We consider two types of excursions of D from π:

$$\begin{aligned} U_\pi &= \left\{\text{downward excursions that return to } \pi \text{ without hitting } \frac{\pi}{2}\right\} \\ U_{\frac{\pi}{2}} &= \left\{\text{downward excursions that hit } \frac{\pi}{2}\right\}\end{aligned}$$

We mark the excursions with marks at rate γ and, independently, with jumps at rate 2λ. We define the following sets:

$$\begin{aligned} U^m_\pi &= \{\text{excursions in } U_\pi \text{ with at least one mark}\} \\ U^{j, no\ m}_\pi &= \{\text{excursions in } U_\pi \text{ with at least one jump and no marks}\} \\ U^{j,m}_\pi &= \{\text{excursions in } U_\pi \text{ with at least one jump and one mark}\} \\ U^{no\ j, no\ m}_\pi &= \{\text{excursions in } U_\pi \text{ with no jumps and no marks}\} \\ U^{no\ m}_{\frac{\pi}{2}} &= \{\text{excursions in } U_{\frac{\pi}{2}} \text{ with no marks}\}\end{aligned}$$

Let $\theta = \inf\{t > J_1 : D_t = \pi\}$, where J_1 is the first jump time of D, and η the time D, under the reflection coupling, takes to hit $\frac{\pi}{2}$. We can write

$$T^r_\pi = \theta \mathbb{1}_{\theta<\eta} + (T_{\frac{\pi}{2}} + \eta)\mathbb{1}_{\eta<\theta},$$

where $T^r_{\frac{\pi}{2}}$ is the coupling time of the process started from $\frac{\pi}{2}$, and

$$\begin{aligned}\mathbb{E}[e^{-\gamma T^r_\pi}] &= \mathbb{E}[e^{-\gamma\theta}\mathbb{1}_{\theta<\eta}] + \mathbb{E}[e^{-\gamma(T^r_{\frac{\pi}{2}}+\eta)}\mathbb{1}_{\eta<\theta}] \\ &= \mathbb{E}[e^{-\gamma\theta}\mathbb{1}_{\theta<\eta}] + \mathbb{E}[e^{-\gamma\eta}\mathbb{1}_{\eta<\theta}]\mathbb{E}[e^{-\gamma T^r_{\frac{\pi}{2}}}], \end{aligned} \tag{3.13}$$

where the second equality is justified by the fact that η and $T_{\frac{\pi}{2}}$ are independent.

The first term in (3.13) is the probability of seeing an excursion with a jump but no marks that goes back to π, before seeing an excursion that hits $\frac{\pi}{2}$ or an excursion with at least one mark that goes back to π. That is,

$$\begin{aligned}\mathbb{E}[e^{-\gamma T^r_\pi}\mathbb{1}_{\theta<\eta}] &= \mathbb{P}(\text{the first excursion in } U^{j,no\ m}_\pi \cup U^m_\pi \cup U_{\frac{\pi}{2}} \text{ lies in } U^{j,no\ m}_\pi) \\ &= \frac{n(U^{j,no\ m}_\pi)}{n(U_{\frac{\pi}{2}}) + n(U^m_\pi) + n(U^{j,no\ m}_\pi)}.\end{aligned}$$

The measure of the set of excursions hitting $\frac{\pi}{2}$ is, as before,

$$n(U_{\frac{\pi}{2}}) = \frac{2}{\pi}.$$

To calculate $n(U^m_\pi)$, we proceed as before, splitting the excursion around its maximum value:

$$n(U^m_\pi) = \int_0^{\frac{\pi}{2}} x^{-2}\left[1 - (x\beta\,\mathrm{cosech}(x\beta))^2\right] dx = \beta\coth\left(\frac{\pi}{2}\beta\right) - \frac{2}{\pi}.$$

To calculate $n(U^{j,no\ m}_\pi)$, we rewrite $U^{j,no\ m}_\pi = U^{no\ m}_\pi \setminus U^{no\ j,no\ m}_\pi$, so if we split again the excursion into two pieces around its maximum, if the excursion is in $U^{no\ m}_\pi$ we want no marks in both pieces, while if the excursion is in $U^{no\ j,no\ m}_\pi$ we want no marks and no jumps in both pieces.

$$\begin{aligned}n(U^{j,no\ m}_\pi) &= n(U^{no\ m}_\pi) - n(U^{no\ j,no\ m}_\pi) \\ &= \int_0^{\frac{\pi}{2}} x^{-2}\left[(x\beta\,\mathrm{cosech}(x\beta))^2 - (x\alpha\,\mathrm{cosech}(x\alpha))^2\right] dx \\ &= -\beta\coth\left(\frac{\pi}{2}\beta\right) + \alpha\coth\left(\frac{\pi}{2}\alpha\right).\end{aligned}$$

Then we have

$$\mathbb{E}[e^{-\gamma T_\pi^r}\mathbb{1}_{\theta<\eta}] = \frac{\alpha\coth(\frac{\pi}{2}\alpha) - \beta\coth(\frac{\pi}{2}\beta)}{\alpha\coth(\frac{\pi}{2}\alpha)} = 1 - \frac{\beta\coth(\frac{\pi}{2}\beta)}{\alpha\coth(\frac{\pi}{2}\alpha)}.$$

Now, we calculate the second term in (3.13). It can be interpreted as the probability of having an excursion with no marks that goes back to π, before an excursion with at least one jump and one mark that goes back to π or an excursion that hits $\frac{\pi}{2}$:

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_\pi^r}\mathbb{1}_{\eta<\theta}] &= \mathbb{P}(\text{the first excursion in } U_\pi^{j,m} \cup U_{\frac{\pi}{2}}^{no\ m} \cup U_{\frac{\pi}{2}} \text{ lies in } U_\pi^{no\ m})\\ &= \frac{n(U_{\frac{\pi}{2}}^{no\ m})}{n(U_\pi^{j,m}) + n(U_{\frac{\pi}{2}})}.\end{aligned}$$

Applying again excursion theory and Lemma 3.14,

$$\begin{aligned}n(U_{\frac{\pi}{2}}^{no\ m}) &= n(U_{\frac{\pi}{2}})\mathbb{P}(\text{excursion that hits } \frac{\pi}{2} \text{ has no marks})\\ &= \frac{2}{\pi}\mathbb{E}[e^{-\gamma H^3(\frac{\pi}{2})}] = \frac{2}{\pi}\frac{\pi}{2}\beta\,\mathrm{cosech}\left(\frac{\pi}{2}\beta\right),\end{aligned}$$

while if we consider the set of excursions with at least one jump and one mark

$$n(U_\pi^{j,m}) = \int_0^{\frac{\pi}{2}} x^{-2}[1 - (x\alpha\,\mathrm{cosech}(x\alpha))^2]dx = \alpha\coth\left(\frac{\pi}{2}\alpha\right) - \frac{\pi}{2}.$$

Thus we have

$$\mathbb{E}[e^{-\gamma T_\pi^r}\mathbb{1}_{\eta<\theta}] = \frac{\beta}{\alpha}\frac{\mathrm{cosech}(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)}.$$

In conclusion, we can express (3.13) as

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_\pi^r}] &= \mathbb{E}[e^{-\gamma\theta}\mathbb{1}_{\theta<\eta}] + \mathbb{E}[e^{-\gamma\eta}\mathbb{1}_{\eta<\theta}]\mathbb{E}[e^{-\gamma T_{\pi/2}^r}] =\\ &= 1 - \frac{\beta}{\alpha}\frac{\coth\left(\frac{\pi}{2}\beta\right)}{\coth\left(\frac{\pi}{2}\alpha\right)} + \frac{\beta}{\alpha}\frac{\mathrm{cosech}\left(\frac{\pi}{2}\beta\right)}{\coth\left(\frac{\pi}{2}\alpha\right)}\cdot\mathbb{E}[e^{-\gamma T_{\pi/2}^r}]. \qquad (3.14)\end{aligned}$$

To find a complete expression for $\mathbb{E}[e^{-\gamma T_\pi^r}]$ we combine (3.12) and (3.14) together:

$$\mathbb{E}[e^{-\gamma T_\pi^r}] = 1 - \frac{\beta}{\alpha}\frac{\coth(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)} + \frac{\beta}{\alpha}\frac{\mathrm{cosech}(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)}$$

$$\cdot \left[\frac{1}{\cosh(\frac{\pi}{2}\beta)} + \frac{\alpha}{2\beta} \tanh\left(\beta\frac{\pi}{2}\right) \cdot \operatorname{cosech}\left(\frac{\pi}{2}\alpha\right) \left(\mathbb{E}\left[e^{-\gamma T_\pi^r}\right] - 1\right) \right],$$

and rearranging we obtain

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_\pi^r}] &= \frac{1 - \frac{\beta}{\alpha}\frac{\coth(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)}}{1 - \frac{1}{2}\frac{\operatorname{cosech}(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)} \tanh\left(\frac{\pi}{2}\beta\right) \cdot \operatorname{cosech}\left(\frac{\pi}{2}\alpha\right)} + \\
&+ \frac{\frac{\beta}{\alpha}\frac{\operatorname{cosech}(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)}\left[\frac{1}{\cosh(\frac{\pi}{2}\beta)} - \frac{\alpha}{2\beta}\tanh\left(\frac{\pi}{2}\beta\right) \cdot \operatorname{cosech}\left(\frac{\pi}{2}\alpha\right)\right]}{1 - \frac{1}{2}\frac{\operatorname{cosech}(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)} \tanh\left(\frac{\pi}{2}\beta\right) \cdot \operatorname{cosech}\left(\frac{\pi}{2}\alpha\right)} \\
&= 1 - \frac{\frac{\beta}{\alpha}\tanh\left(\frac{\pi}{2}\alpha\right)\tanh\left(\frac{\pi}{2}\beta\right)}{1 - \frac{1}{2}\operatorname{sech}\left(\frac{\pi}{2}\beta\right)\operatorname{sech}\left(\frac{\pi}{2}\alpha\right)}. \qquad (3.15)
\end{aligned}$$

Combining (3.12) and (3.14), we can also obtain an explicit expression for the coupling time starting from $\frac{\pi}{2}$:

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_{\frac{\pi}{2}}^r}] &= \frac{1}{\cosh(\frac{\pi}{2}\beta)} + \frac{\alpha}{2\beta}\tanh\left(\frac{\pi}{2}\beta\right)\operatorname{cosech}\left(\frac{\pi}{2}\alpha\right)\left(\mathbb{E}\left[e^{-\gamma T_\pi^r}\right] - 1\right) \\
&= \frac{1}{\cosh(\frac{\pi}{2}\beta)} + \frac{\alpha}{2\beta}\tanh\left(\frac{\pi}{2}\beta\right)\operatorname{cosech}\left(\frac{\pi}{2}\alpha\right) \\
&\left(-\frac{\beta}{\alpha}\frac{\coth(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)} + \frac{\beta}{\alpha}\frac{\operatorname{cosech}(\frac{\pi}{2}\beta)}{\coth(\frac{\pi}{2}\alpha)} \cdot \mathbb{E}[e^{-\gamma T_{\pi/2}^r}]\right).
\end{aligned}$$

Rearranging this expression we obtain the following formula for $T_{\frac{\pi}{2}}^r$

$$\mathbb{E}[e^{-\gamma T_{\frac{\pi}{2}}^r}] = \frac{2\cosh(\frac{\pi}{2}\alpha) - \cosh(\frac{\pi}{2}\beta)}{2\cosh(\frac{\pi}{2}\beta)\cosh(\frac{\pi}{2}\alpha) - 1}. \qquad (3.16)$$

This completes the proof of parts *3* and *4* of Lemma 3.16.

3.6.3 Case III. $D_0 \neq \frac{\pi}{2}, \pi$

Finally, consider the coupled processes starting from any point $D_0 = x \in (0, \pi)\setminus\{\frac{\pi}{2}\}$. First, we find the expressions of the Laplace transforms without distinguishing the coupling strategy, so we denote by T_x the coupling time. Let $F = \{0, \frac{\pi}{2}, \pi\}$, we denote by $T_x(F)$ the hitting time of D of the set F

with $D_0 = x$. We can express T_x as

$$T_x = T_x(F) + T_\pi \mathbb{1}_{\{D_{T_x(F)}=\pi\}} + T_{\pi/2}\mathbb{1}_{\{D_{T_x(F)}=\frac{\pi}{2}\}}.$$

To find the Laplace transform of T_x, we rewrite it in the following way

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_x}] &= \mathbb{E}[e^{-\gamma T_x(F)}\mathbb{1}_{\{D_{T_x(F)}=0\}}] + \mathbb{E}[e^{-\gamma(T_x(F)+T_\pi)}\mathbb{1}_{\{D(T_x(F))=\pi\}}] \\
&\quad + \mathbb{E}[e^{-\gamma(T_x(F)+T_{\frac{\pi}{2}})}\mathbb{1}_{\{D_{T_x(F)}=\frac{\pi}{2}\}}] \\
&= \mathbb{E}[e^{-\gamma T_x(F)}] + \mathbb{E}[e^{-\gamma T_x(F)}\mathbb{1}_{\{D_{T_x(F)}=\frac{\pi}{2}\}}](\mathbb{E}[e^{-\gamma T_{\frac{\pi}{2}}}] - 1) \\
&\quad + \mathbb{E}[e^{-\gamma T_x(F)}\mathbb{1}_{\{D(T_x(F))=\pi\}}](\mathbb{E}[e^{-\gamma T_\pi}] - 1). \qquad (3.17)
\end{aligned}$$

To find expressions for the three terms in (3.17) we use excursion theory again. Due to the definition of the coupling, the process D_t behaves differently depending on whether the starting point x is smaller or greater than $\frac{\pi}{2}$. That means that we need to study the two cases $x < \frac{\pi}{2}$ and $x > \frac{\pi}{2}$ separately and, in each interval, we restrict to considering the hitting time of the extrema of the interval in which D_t starts.
To make the calculations easier we can apply Lemma 3.15 and we again use the notations $\alpha = \sqrt{2(2\lambda + \gamma)}$ and $\beta = \sqrt{2\gamma}$.
Since the Brownian motion B is shift-invariant, we can think of B as starting from 0 and diffusing in an interval $(-c, d)$, where $d, c > 0$. In other words, with $D_0 = x$,

$$\begin{aligned}
&1.\ \text{if } x < \frac{\pi}{2},\ \text{we shift to the interval } (-c, d) = \left(-x, \frac{\pi}{2} - x\right), \\
&2.\ \text{if } x > \frac{\pi}{2},\ \text{we shift to the interval } (-c, d) = \left(\frac{\pi}{2} - x, \pi - x\right).
\end{aligned} \qquad (3.18)$$

First term of (3.17)

To find a formula for the first term $\mathbb{E}[e^{-\gamma T_x(F)}]$, we can observe that in this case the hitting time of F is not affected by the jumps since we are not considering which point of F is hit by D. This means that the calculations do not depend on whether x is bigger or smaller than $\frac{\pi}{2}$ and that, when applying Lemma 3.15(i), we only mark the excursions at rate γ.
We can conclude that

$$\mathbb{E}[e^{-\gamma T_x(F)}] = \mathbb{E}[e^{-\gamma T_0(\{-c,d\})}] = \frac{\operatorname{cosech}(d\beta) + \operatorname{cosech}(c\beta)}{\coth(d\beta) + \coth(c\beta)}, \qquad (3.19)$$

where c and d are functions of x, as defined in (3.18).

Second term of (3.17)

To calculate $\mathbb{E}[e^{-\gamma T_x(F)}\mathbb{1}_{\{D_{T_x(F)}=\frac{\pi}{2}\}}]$, we need to consider the case in which D_t hits $\frac{\pi}{2}$. Since we always use a reflection coupling in the interval $(0,\pi)$, D hits $\frac{\pi}{2}$ independently of the jumps in $\widetilde{N}$: this happens because, if $D \leq \frac{\pi}{2}$ at the time of a jump, then its position is left unchanged; if $D > \frac{\pi}{2}$, the jump would reflect D downwards over $\frac{\pi}{2}$, and the process would continue diffusing upwards hitting $\frac{\pi}{2}$. That means that in both intervals, the time taken by D to hit F when starting from x is the same as the time taken by B to hit the boundary points of an interval $(-c,d)$ and, when we apply Lemma 3.15, we mark the excursions at rate γ.

1. Let $x \in (0,\frac{\pi}{2})$. To calculate the hitting time of $\frac{\pi}{2}$, we need to consider the times when B hits $d=\frac{\pi}{2}-x$, so from Lemma 3.15(*ii*)

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_x(F)}\mathbb{1}_{\{D_{T_x(F)}=\frac{\pi}{2}\}}] &= \mathbb{E}[e^{-\gamma T_0(\{-x,\frac{\pi}{2}-x\})}\mathbb{1}_{\{B_{T_0(\{-x,\frac{\pi}{2}-x\})}=\frac{\pi}{2}-x\}}] \\
&= \frac{\operatorname{cosech}(d\beta)}{\coth(d\beta)+\coth(c\beta)} \\
&= \frac{\operatorname{cosech}((\frac{\pi}{2}-x)\beta)}{\coth((\frac{\pi}{2}-x)\beta)+\coth(x\sqrt{2\gamma})}
\end{aligned}$$

2. Let $x \in (\frac{\pi}{2},\pi)$. To have the hitting time of $\frac{\pi}{2}$, we consider the times when B hits $-c=\frac{\pi}{2}-x$. From Lemma 3.15(*iii*),

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_x(F)}\mathbb{1}_{\{D_{T_x(F)}=\frac{\pi}{2}\}}] &= \mathbb{E}[e^{-\gamma T_0(\{\frac{\pi}{2}-x,\pi-x\})}\mathbb{1}_{\{B_{T_0(\{\frac{\pi}{2}-x,\pi-x\})}=\frac{\pi}{2}-x\}}] \\
&= \frac{\operatorname{cosech}(c\beta)}{\coth(d\beta)+\coth(c\beta)} \\
&= \frac{\operatorname{cosech}((x-\frac{\pi}{2})\beta)}{\coth((x-\frac{\pi}{2})\beta)+\coth((\pi-x)\beta)}
\end{aligned}$$

Third term of (3.17)

Finally, we calculate $\mathbb{E}[e^{-\gamma T_x(F)}\mathbb{1}_{\{D(T_x(F))=\pi\}}]$. As we did for the second term, we need to study the two cases $x<\frac{\pi}{2}$ and $x>\frac{\pi}{2}$ separately.

1. If $x \in (0, \frac{\pi}{2})$, it is not possible for D_t to hit π without hitting $\frac{\pi}{2}$ first, so

$$\mathbb{E}[e^{-\gamma T_x(F)} \mathbb{1}_{\{D_{T_x(F)}=\pi\}}] = 0.$$

2. Let $x \in (\frac{\pi}{2}, \pi)$. To have the hitting time of π, we apply Lemma 3.15, but this time the jumps will influence the calculations.
As usual, we consider excursions of B in an interval $(-c, d) = (\frac{\pi}{2} - x, \pi - x)$. To have B hit $d = \pi - x$, we need that to happen before seeing a jump, so we mark the excursions at rate $2\lambda + \gamma$. From Lemma 3.15(ii):

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_x(F)} \mathbb{1}_{\{D_{T_x(F)}=\pi\}}] &= \mathbb{E}[e^{-\gamma T_0(\{\frac{\pi}{2}-x, \pi-x\})} \mathbb{1}_{\{B_{T_0(\{\frac{\pi}{2}-x,\pi-x\})}=\pi-x\}}] \\
&= \frac{\operatorname{cosech}(d\alpha)}{\coth(d\alpha) + \coth(c\alpha)} \\
&= \frac{\operatorname{cosech}((\pi - x)\alpha)}{\coth((\pi - x)\alpha) + \coth((x - \frac{\pi}{2})\alpha)}.
\end{aligned} \tag{3.20}$$

This completes the proof of parts *1* and *2* of Lemmas 3.16 and 3.17. The calculations of the expressions of the Laplace tranform as established in those lemmas are showed in Appendix B.

3.7 Laplace transform

In this section, we show the graphs of some simulations of the Laplace transforms and we compare them to the formulas obtained in the previous section. As the graphs illustrate, the formulas we calculated agree with the simulations of the jumpy Brownian motion under the reflection and synchronised couplings, so we have also a numerical confirmation that the expressions given in Lemmas 3.16 and 3.17 are correct.

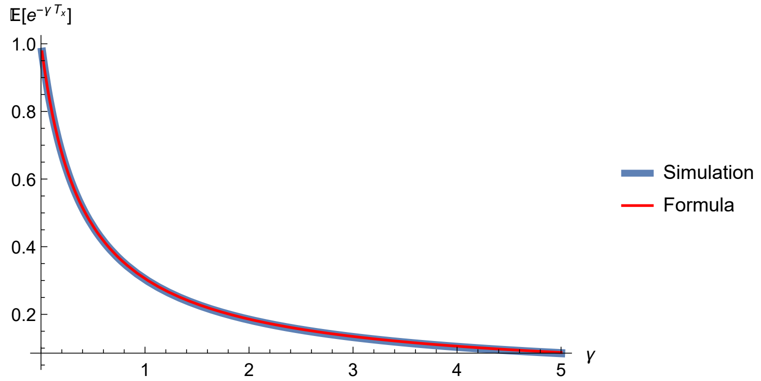
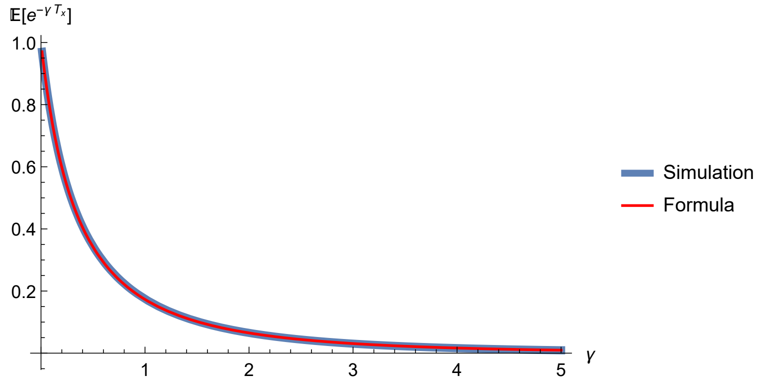
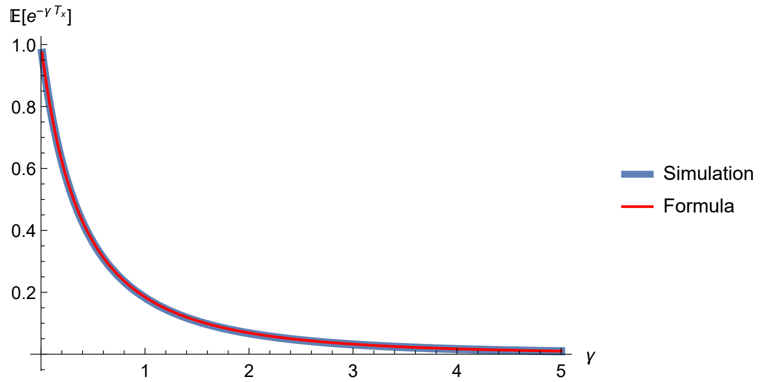


Fig. 3.3: Comparison of the simulation of the Laplace tranform for $D_0 = \pi$ and $\lambda = 0.5$ and the formulas obtained in Section 3.6.2 under the reflection coupling.

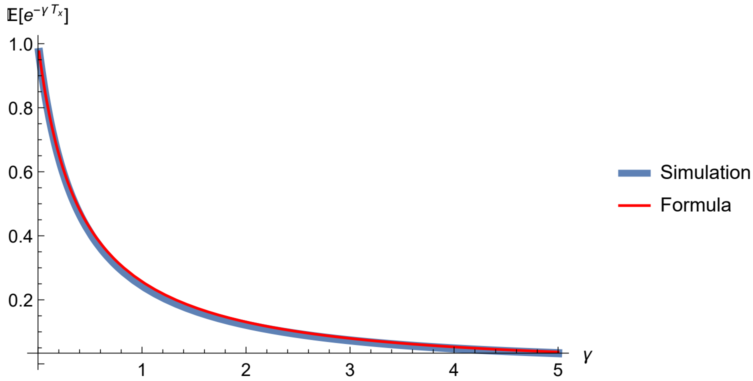


(a) Comparison of the simulation and the formulas of the Laplace tranform for $D_0 = \frac{\pi}{2}$ and $\lambda = 0.5$ under the reflection coupling.

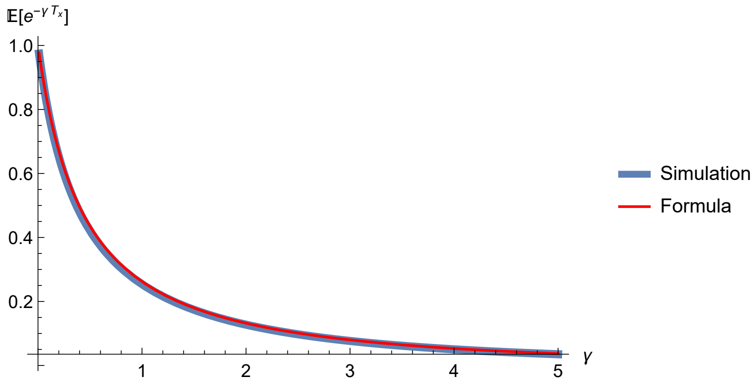


(b) Comparison of the simulation and the formulas of the Laplace tranform for $D_0 = \frac{\pi}{2}$ and $\lambda = 0.5$ under the synchronised coupling.

Fig. 3.4: Comparison of the simulation of the Laplace tranform for $D_0 = \frac{\pi}{2}$ and the formulas obtained in Section 3.6.1.

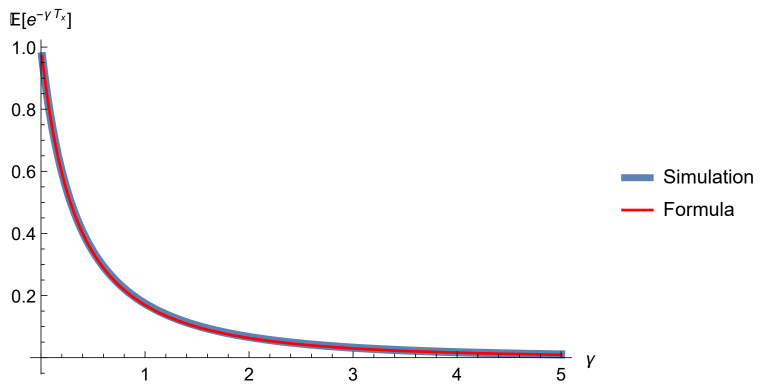


(a) Comparison of the simulation and the formulas of the Laplace tranform for $D_0 = x < \frac{\pi}{2}$ and $\lambda = 0.5$ under the reflection coupling.

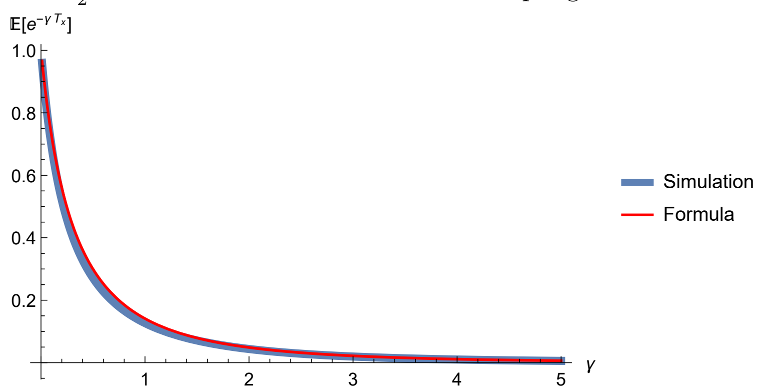


(b) Comparison of the simulation and the formulas of the Laplace tranform for $D_0 = x < \frac{\pi}{2}$ and $\lambda = 0.5$ under the synchronised coupling.

Fig. 3.5: Comparison of the simulation of the Laplace tranform for $D_0 = x < \frac{\pi}{2}$ and the formulas obtained in Section 3.6.3.



(a) Comparison of the simulation and the formulas of the Laplace tranform for $D_0 = x > \frac{\pi}{2}$ and $\lambda = 0.5$ under the reflection coupling.



(b) Comparison of the simulation and the formulas of the Laplace tranform for $D_0 = x > \frac{\pi}{2}$ and $\lambda = 0.5$ under the synchronised coupling.

Fig. 3.6: Comparison of the simulation of the Laplace tranform for $D_0 = x > \frac{\pi}{2}$ and the formulas obtained in Section 3.6.3.

At this point, it is interesting to compare the Laplace transform of the coupling time for the two coupling strategies on which we are focusing. Figure 3.7 shows the graph of the difference $\mathbb{E}[e^{-\gamma T_\pi^r}] - \mathbb{E}[e^{-\gamma T_\pi^s}]$ of the Laplace transforms evaluated in Section 3.6.2 for two jumpy Brownian motions started at distance π under our two coupling strategies. As in Definition 1.27, we say that a coupling is Laplace-optimal if it maximizes the Laplace transform of the coupling time for all $\gamma > 0$. From the relation between the expectation and Laplace transform of a random variable, we can see that the Laplace-optimality is a stronger property than the mean-optimality. Finally, observe that the tail-optimality is the strongest type of optimality among the three of Definition 1.27, and it corresponds to finding the co-adapted coupling that realises the best upper bound for the total variation distance in the coupling inequality (1.1) simultaneously for all $t \geq 0$. As we can see from Figure 3.7a, if $\lambda > \lambda^\star$, then the Laplace transform under the synchronised coupling is bigger than that under the reflection coupling for all $\gamma > 0$. Thus, we can conjecture that the synchronised coupling is Laplace-optimal for all $\lambda > \lambda^\star$. Unfortunately, we cannot expect a similar result in the case $\lambda < \lambda^\star$. As Figure 3.7b illustrates, when γ is small, $\mathbb{E}[e^{-\gamma T_\pi^r}] - \mathbb{E}[e^{-\gamma T_\pi^s}] > 0$, but as γ grows, the graph decreases until it becomes negative. This indicates that neither of the two couplings under consideration is Laplace-optimal when $\lambda < \lambda^\star$.

Remark. If $\lambda = 0$, then the jumpy Brownian motion is a one-dimensional Brownian motion with no jumps, for which we know that the mean-optimal co-adapted coupling is given by our reflection coupling, which is clearly maximal.

Now that we have graphically seen how the mean-optimal co-adapted coupling compares to the maximal coupling, we can show that the mean-optimal coupling is not maximal for all $\lambda > 0$ directly from the explicit formulas of the expected coupling time. Let $\tau^\star, T_\pi^s$, and T_π^r be the coupling time for two jumpy Brownian motions started at distance π under the maximal, synchronised, and reflection couplings respectively. We know that

$$\begin{aligned}
\mathbb{E}[\tau^\star] &= \frac{1 - \operatorname{sech}(2\pi\sqrt{\lambda})}{2\lambda}, \\
\mathbb{E}[T_\pi^s] &= \frac{1}{2\lambda}, \\
\mathbb{E}[T_\pi^r] &= \frac{\pi \sinh(\pi\sqrt{\lambda})}{\sqrt{\lambda}(2\cosh(\pi\sqrt{\lambda}) - 1)}.
\end{aligned}$$

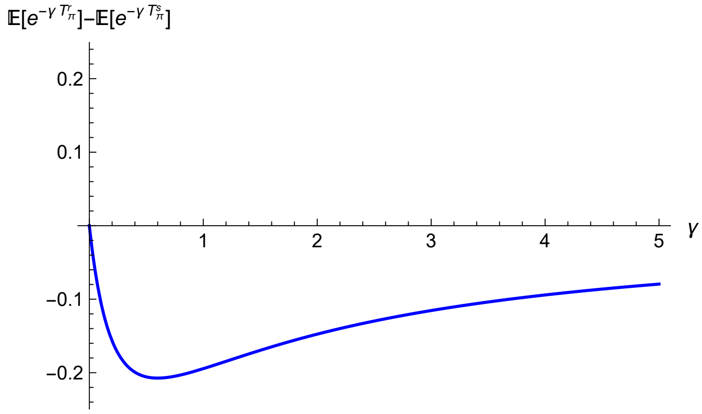
We want to show that, for all $\lambda \geq 0$,

1. $\mathbb{E}[\tau^\star] \leq \mathbb{E}[T_\pi^s]$,
2. $\mathbb{E}[\tau^\star] \leq \mathbb{E}[T_\pi^r]$.

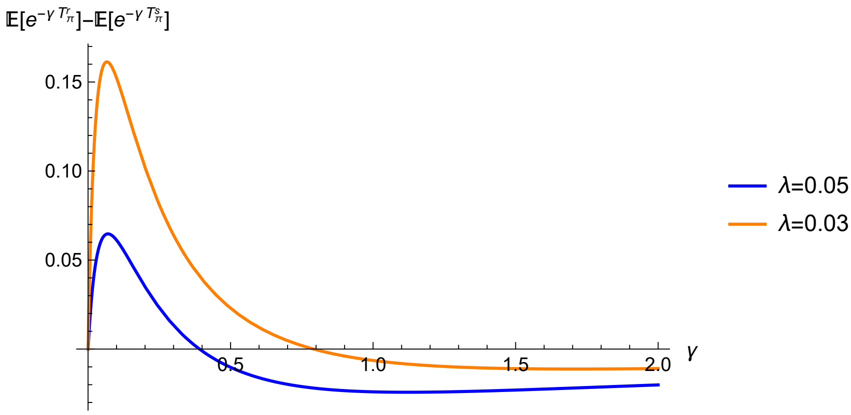
1. Since $\operatorname{sech}(2\pi\sqrt{\lambda}) \geq 0$ for all $\lambda \geq 0$, then

$$\mathbb{E}[\tau^\star] = \frac{1 - \operatorname{sech}(2\pi\sqrt{\lambda})}{2\lambda} \leq \frac{1}{2\lambda} = \mathbb{E}[T_\pi^s],$$

for all $\lambda \geq 0$.



(a) Graph of $\mathbb{E}[e^{-\gamma T_\pi^r}] - \mathbb{E}[e^{-\gamma T_\pi^s}]$ with $D_0 = \pi$ for $\lambda = 0.5 > \lambda^\star$.



(b) Graph of $\mathbb{E}[e^{-\gamma T_\pi^r}] - \mathbb{E}[e^{-\gamma T_\pi^s}]$ with $D_0 = \pi$ for two different values of $\lambda < \lambda^\star$.

Fig. 3.7: Graph of $\mathbb{E}[e^{-\gamma T_\pi^r}] - \mathbb{E}[e^{-\gamma T_\pi^s}]$ with $D_0 = \pi$.

2. We distinguish two cases based on the value of λ.

- If $\lambda \geq \lambda^\star$, then $\mathbb{E}[T_\pi^r] \geq \mathbb{E}[T_\pi^s] \geq \mathbb{E}[\tau^\star]$.
- If $\lambda < \lambda^\star$, we want to show that $\frac{\mathbb{E}[\tau^\star]}{\mathbb{E}[T_\pi^r]} \leq 1$. Let $x = \pi\sqrt{\lambda}$, from the formulas above,

$$\frac{\mathbb{E}[\tau^\star]}{\mathbb{E}[T_\pi^r]} = \frac{(1 - \operatorname{sech}(2x))(2\cosh(x) - 1)}{2x\sinh(x)}.$$

Then, using the exponential forms $\sinh(z) = \frac{e^z - e^{-z}}{2}$ and $\cosh(z) =$

$\frac{e^z+e^{-z}}{2}$ yields

$$\begin{aligned}
\frac{\mathbb{E}[\tau^\star]}{\mathbb{E}[T^r_\pi]} &= \frac{(1-\operatorname{sech}(2x))(2\cosh(x)-1)}{2x\sinh(x)} = \frac{\operatorname{sech}(2x)\sinh(x)(2\cosh(x)-1)}{x} \\
&= \frac{(e^{2x}-e^{-2x})-(e^{x}-e^{-x})}{x(e^{2x}+e^{-2x})} = \frac{\int_x^{2x}(e^y+e^{-y})dy}{x(e^{2x}+e^{-2x})} \\
&\leq \frac{x\max_{x\leq y\leq 2x}(e^y+e^{-y})}{x(e^{2x}+e^{-2x})} = 1,
\end{aligned}$$

where the last equality derives from the fact that $e^y + e^{-y}$ is an increasing function.

In conclusion, the mean-optimal coupling is not maximal for all $\lambda > 0$.

3.8 Expectation of T_x

Using the Laplace transforms calculated in Section 3.6, we can derive the expectation of the coupling time T_x for our two coupling strategies. As usual, the calculations depend on the starting point x and on the coupling strategy adopted at π.

Lemma 3.18. *For the reflection coupling of Definition 3.6 starting from $D_0 = x$, the expected coupling time satisfies*

$$\mathbb{E}[T^r_x] = \begin{cases} x(\pi-x) + \frac{\pi(\pi-x)}{2\cosh(\sqrt{\lambda}\pi)-1} + \frac{\pi\sinh(\sqrt{\lambda}(2x-\pi))}{\sqrt{\lambda}(2\cosh(\sqrt{\lambda}\pi)-1)} & \text{if } x \geq \frac{\pi}{2} \\ \\ x(\pi-x) + \frac{\pi x}{2\cosh(\sqrt{\lambda}\pi)-1} & \text{if } x \leq \frac{\pi}{2}. \end{cases} \tag{3.21}$$

For the synchronised coupling of Definition 3.5, the expectation of the coupling time satisfies

$$\mathbb{E}[T^s_x] = \begin{cases} x(\pi-x) + \frac{(\pi-x)\operatorname{cosech}(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}} + \frac{\operatorname{cosech}(\sqrt{\lambda}\pi)\sinh(\sqrt{\lambda}(2x-\pi))}{2\lambda} & \text{if } x \geq \frac{\pi}{2} \\ \\ x(\pi-x) + \frac{x\operatorname{cosech}(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}} & \text{if } x \leq \frac{\pi}{2}. \end{cases} \tag{3.22}$$

Proof. To find the expectation of the coupling time, we take the expressions of the Laplace transform $\mathbb{E}[e^{-\gamma T_x}]$ found in Section 3.6, we differentiate them with respect to γ, and finally we take the limit as $\gamma \to 0$. $\square$

Once we have the expectation, we can study how it behaves when λ varies.
Looking first at the reflection coupling and the expectation given by (3.21), we have the following:

$$\lim_{\lambda\to\infty} \mathbb{E}[T_x^r] = x(\pi - x), \qquad \lim_{\lambda\to 0} \mathbb{E}[T_x^r] = x(2\pi - x)$$

for all $x \in [0, \pi]$. Combining (3.21) and the limits, we can easily verify that $\mathbb{E}[T_x^r]$ is a continuous function of x for all $\lambda \geq 0$ with a maximum at $x = \pi$ when $\lambda < \lambda^\star$ and a maximum at $x = \frac{\pi}{2}$ when $\lambda > \lambda^\star$. The limits we obtained are what we would have expected intuitively, as we now explain.
When $\lambda \to \infty$, the process D_t jumps instantly. If $x \in (0, \pi)$, the process diffuses in the interval $\left(0, \frac{\pi}{2}\right]$ because every time it gets above $\frac{\pi}{2}$ it instantly jumps back below $\frac{\pi}{2}$. So, D_t behaves in the same way as a standard Brownian motion in $\left(0, \frac{\pi}{2}\right]$ with reflection at $\frac{\pi}{2}$. Thus, the time taken by D to hit 0 has the same distribution as the time taken by a standard Brownian motion started at x to hit $\{0, \pi\}$. Then, $\mathbb{E}[T_x^r] = x(\pi - x)$. In particular, if $x = \frac{\pi}{2}$ the mean time is $\frac{\pi^2}{4}$. If $x = \pi$, we instantly jump to 0, so the expected coupling time is 0.
If $\lambda \to 0$, there are no jumps, so D_t diffuses like a standard Brownian motion in $[0, \pi]$ with reflection at π, so the time to hit 0 is equivalent in distribution to the time for a standard Brownian motion started at x to hit $\{0, 2\pi\}$. Then, we know that $\mathbb{E}[T_x^r] = x(2\pi - x)$.

As we did for the reflected process, we now analyse the behaviour of the expectation of the synchronised coupling time described in (3.22).

$$\lim_{\lambda\to\infty} \mathbb{E}[T_x^s] = x(\pi - x) \quad \text{for all } x \in [0, \pi], \qquad \lim_{\lambda\to 0} \mathbb{E}[T_x^s] = \begin{cases} +\infty & \text{for all } x \in (0, \pi] \\ 0 & \text{if } x = 0. \end{cases}$$

From (3.22) and the limits, we can conclude that $\mathbb{E}[T_x^s]$ is again a continuous function of x for all λ except at $x = 0$ when $\lambda = 0$. As a function of x, $\mathbb{E}[T_x^s]$ has a maximum at $x = \pi$ if $\lambda < \lambda^\star$, while it has a maximum at $x = \frac{\pi}{2}$ and a minimum at $x = \pi$ if $\lambda > \lambda^\star$. That follows by the fact that $\mathbb{E}[T_0^s] = 0$, while from the limits above we have $\lim_{\lambda\to 0} \mathbb{E}[T_x^s] = +\infty$ for all $x \in (0, \pi]$. As we did for the reflecting strategy, we can intuitively justify the limits we obtained.

If $\lambda \to \infty$, then the probability of hitting π tends to 0. This means that we can repeat the same argument as for the reflection coupling, and $\mathbb{E}[T^s_x] = x(\pi - x)$ for all $x \in (0, \pi)$. If $x = \pi$, D_t instantly jumps to 0, so $\mathbb{E}[T^s_\pi] = 0$. When $\lambda \to 0$, the process does not jump. This implies that, if $x = \pi$, the process will never hit 0, so $\mathbb{E}[T^s_x] = \infty$. If $x \in (0, \pi)$, the process starts performing reflection, so it has a positive probability of hitting π before getting to 0. Thus, there is a positive chance of the coupling being unsuccessful, and so $\mathbb{E}[T^s_x] = \infty$.

Now, looking at the expectation of the coupling time of the two coupling strategies in Lemma 3.18 we notice that they only differ by a function that depends on λ. This means that we can rewrite

$$\mathbb{E}[T_x] = \begin{cases} x(\pi - x) + (\pi - x)C(\lambda) + \frac{\sinh(\sqrt{\lambda}(2x-\pi))}{\sqrt{\lambda}} C(\lambda) & \text{if } x \geq \frac{\pi}{2} \\ \\ x(\pi - x) + xC(\lambda) & \text{if } x \leq \frac{\pi}{2}, \end{cases}$$

where

$$C(\lambda) = \begin{cases} C^r(\lambda) = \frac{\pi}{2\cosh(\pi\sqrt{\lambda})-1} & \text{for the reflection coupling,} \\ C^s(\lambda) = \frac{\operatorname{cosech}(\pi\sqrt{\lambda})}{2\sqrt{\lambda}} & \text{for the synchronised coupling.} \end{cases}$$

Comparing the formulas for $C(\lambda)$ for the two couplings, as illustrated in Figure 3.8, we see that there exists a unique value $\lambda^\star = 0.083$ of λ such that $C(\lambda)$ assumes the same value in both strategies, i.e. $C^r(\lambda) = C^s(\lambda)$. This is the value of $\lambda^\star$ given in Theorem 3.8. Comparing Figures 3.8 and 3.9, we also deduce that if $\lambda > \lambda^*$, then $C^s(\lambda) < C^r(\lambda)$ and the synchronised process is faster on average than the reflected process. The opposite happens if $\lambda < \lambda^*$.

Figures 3.10 and 3.11 confirm this conclusion. Figure 3.10 illustrates a comparison between the expected coupling times of the two processes with fixed λ (with $\lambda > \lambda^\star$ in Figure 3.10a and $\lambda < \lambda^\star$ in Figure 3.10b) and $x \in [0, \pi]$. Figure 3.11 shows a comparison between the expectation of the coupling time under the two coupling strategies when λ varies for two fixed values of x. As we see from the graphs, the expectations agree at $\lambda^\star$.

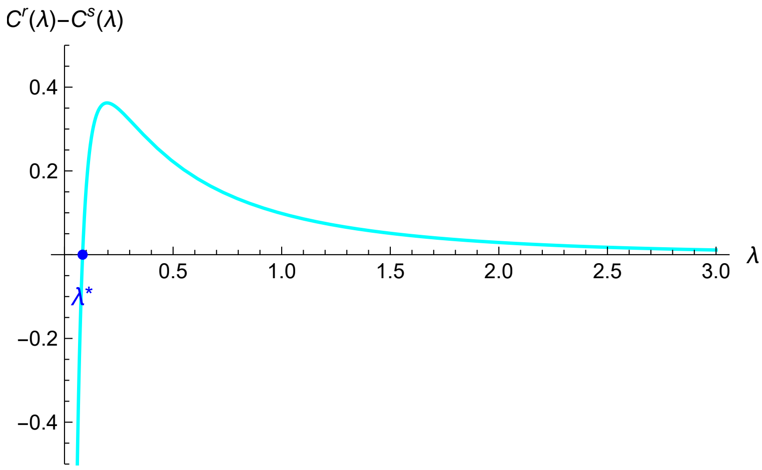


Fig. 3.8: Comparison of the constant $C(\lambda)$ under the two couplings.

3.9 Optimal coupling

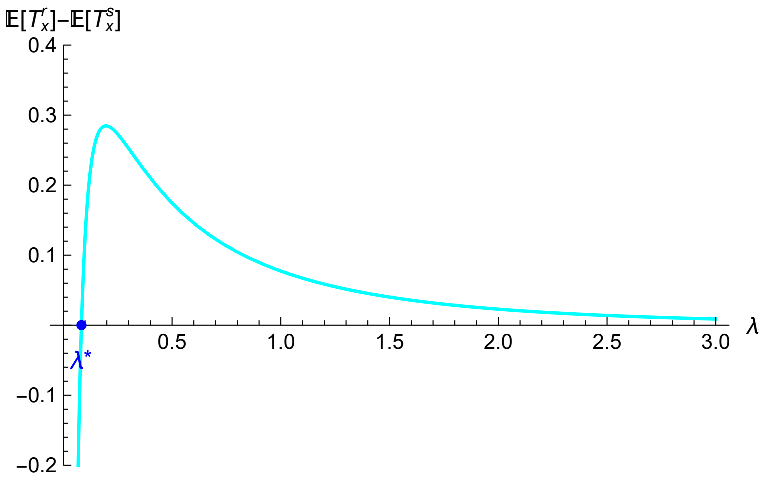
Now that we have calculated the expectations of the coupling time under the reflection and synchronised couplings and established which strategy is better depending on the value of λ, we complete the proof of Theorem 3.8 showing that the couplings defined in Section 3.4 are mean-optimal in the class of co-adapted couplings. To prove the mean-optimality of our coupling, we use Bellman's principle of optimality, which can be found in [19]. We define the value function as

$$v(x) = \inf_{c} \mathbb{E}[T_x^c], \tag{3.23}$$

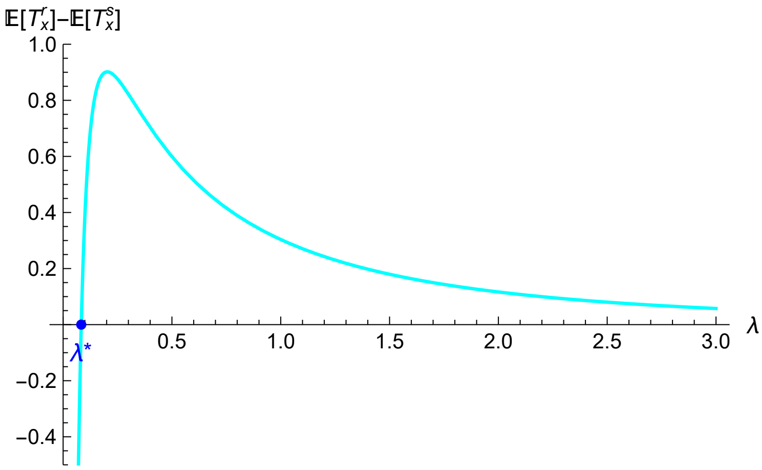
the infimum over all co-adapted couplings c of $\mathbb{E}[T_x^c]$ for a pair of jumpy Brownian motions started at distance x under the coupling c. To establish the optimality of our couplings, we need to show that they attain the equality of (3.23). To do that, we define the process

$$V_x^c(\delta) = \int_0^\delta \mathbb{1}_{\{D_s^c>0\}} ds + \hat{v}(D_\delta^c), \tag{3.24}$$

for any coupling c in the class of co-adapted couplings, where $\hat{v}(D_\delta^c)$ is the value function under our candidate optimal couplings $\hat{c}$ described in Section



(a) Comparison of the expectation of the coupling time under the two couplings as a function of λ and $x = \frac{\pi}{4}$.

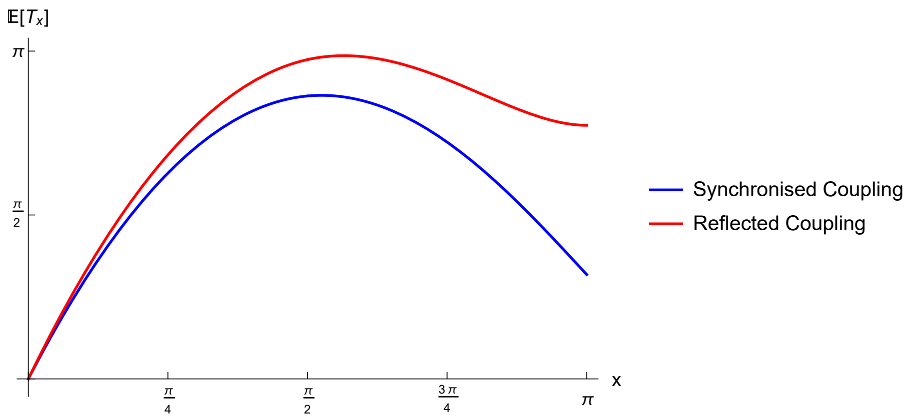


(b) Comparison of the expectation of the coupling time under the two couplings as a function of λ and $x = \frac{3\pi}{4}$

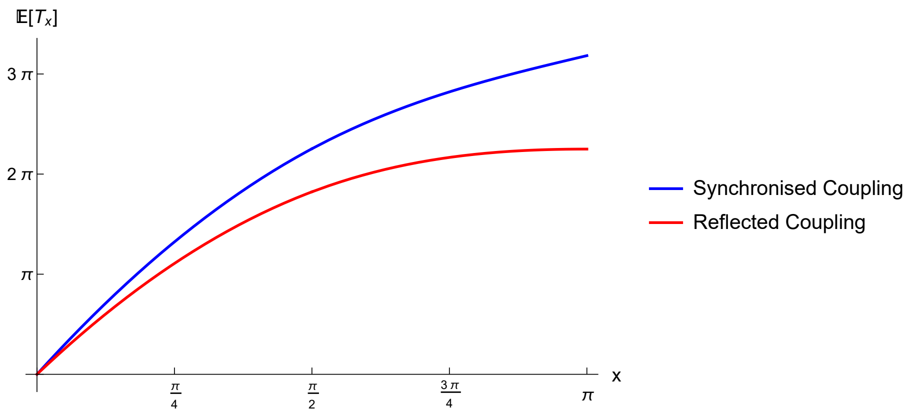
Fig. 3.9: Comparison of the expectation of the coupling time under the two couplings as a function of λ.

3.4:

$$\hat{v}(D_\delta) = D_\delta(\pi - D_\delta) + C(\lambda)\min\{D_\delta, \pi - D_\delta\} + C(\lambda)\frac{\sinh(\sqrt{\lambda}(2D_\delta - \pi))}{\sqrt{\lambda}}\mathbb{1}_{D_\delta > \pi/2}, \tag{3.25}$$



(a) Expectation of the coupling time as a function of x for the two couplings for a fixed value $\lambda = 0.5 > \lambda^{\star}$



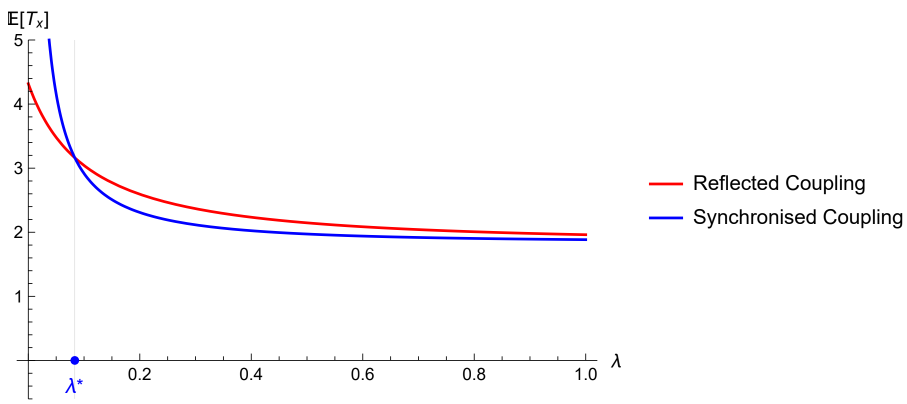
(b) Expectation of the coupling time as a function of x for the two couplings for a fixed value $\lambda = 0.05 < \lambda^{\star}$

Fig. 3.10: Comparison of the expectation of the coupling time for the reflected and the synchronised couplings for a fixed λ.

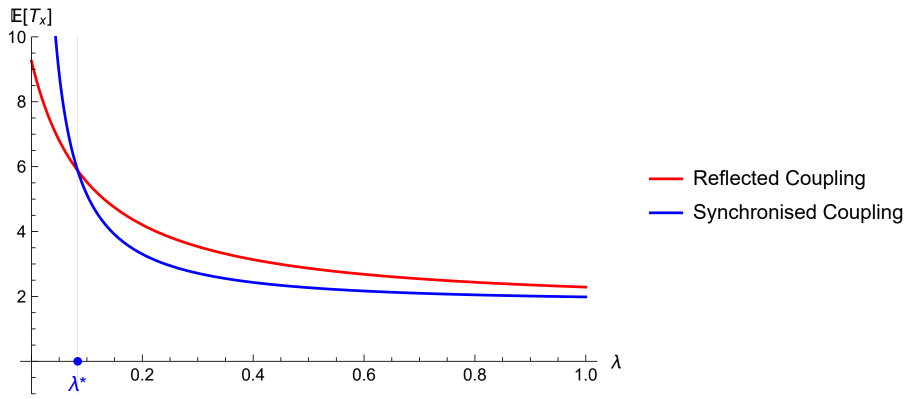
where

$$C(\lambda) = \begin{cases} \frac{\pi}{2\cosh(\sqrt{\lambda}\pi)-1} & \text{if we apply the reflection coupling} \\ \frac{\operatorname{cosech}(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}} & \text{if we apply the synchronised coupling.} \end{cases} \tag{3.26}$$

In other words, we run the jumpy Brownian motion using any co-adapted coupling c until time δ, and then switch to our candidate optimal coupling at time δ. We denote by $V_x^{\hat{c}}$ the function in (3.24) corresponding to using our couplings since time 0. Bellman's principle of optimality states that if



(a) Comparison of the expectation of the coupling time for the two couplings for $x = \frac{\pi}{4}$.



(b) Comparison of the expectation of the coupling time for the two couplings for $x = \frac{3\pi}{4}$.

Fig. 3.11: Comparison of the expectation of the coupling time under the two strategies for two fixed values of x.

V_x^c is a submartingale for all co-adapted couplings c, i.e.

$$\lim_{\delta \to 0} \frac{\mathbb{E}[V_x^c(\delta) - V_x^c(0)]}{\delta} \geq 0,$$

and $V_x^{\hat{c}}$ is a martingale, then our candidate coupling $\hat{c}$ is mean-optimal. In fact, since V_x^c is a submartingale,

$$\hat{v}(x) = V_x^c(0) \leq \mathbb{E}[V_x^c(\delta)]$$

for all $\delta \geq 0$ and for all co-adapted couplings c. Then, from (3.24)

$$\begin{aligned}\lim_{\delta\to\infty} \mathbb{E}[V_x^c(\delta)] &= \lim_{\delta\to\infty} \mathbb{E}\left[\int_0^\delta \mathbb{1}_{\{D_s^c>0\}} ds + \hat{v}(D_\delta^c)\right] \\ &= \mathbb{E}\left[\int_0^{+\infty} \mathbb{1}_{\{D_s^c>0\}} ds\right] + 0 = \mathbb{E}[T_x^c]\end{aligned}$$

for all successful co-adapted coupling c. So, $\hat{v}(x) \leq \inf_c \mathbb{E}[T_x^c] = v(x)$. Since $V_x^{\hat{c}}$ is a martingale, the equality is satisfied by our candidate coupling $\hat{c}$, which is therefore mean-optimal.

Now, in order to define a generic co-adapted coupling of two jumpy Brownian motions X and X', we need to consider all the possible ways to couple their Brownian motion and Poisson process components. Let $J_k \sim \text{Exp}(2\lambda)$ for $k \geq 1$ be i.i.d. random variables that represent the potential jump times, and let $Y_k \sim \text{Ber}\left(\frac{1}{2}\right)$ be i.i.d. that represent tossing a fair coin to decide whether the jumpy Brownian motion jumps at time J_k. Let B_δ and B'_δ are two independent standard Brownian motions, and consider two jumpy Brownian motions on $\mathbb{R}$

$$\begin{aligned}X_\delta &= \frac{1}{2}B_\delta + \pi\sum_{k=1}^{\infty} \mathbb{1}_{\{J_k<\delta\}} Y_k \\ X'_\delta &= x + \frac{\theta}{2}B_\delta + \frac{\sqrt{1-\theta^2}}{2}B'_\delta + \pi\sum_{k=1}^{\infty} \mathbb{1}_{\{J_k<\delta\}}((1-p)Y_k + p(1-Y_k)),\end{aligned} \tag{3.27}$$

where $\theta \in [-1, 1]$ and $p \in [0, 1]$ are the control parameters of the Brownian motion and the Poisson process components respectively. So, the diffusion of X' is defined as a combination of two independent Brownian motions: if $\theta = -1$, it is defined as $x - \frac{1}{2}B_\delta$, which means that X_δ and X'_δ are coupled under the reflection strategy; while if $\theta = 1$, we get $x + \frac{1}{2}B_\delta$, which corresponds to the synchronised coupling. To decide whether X_δ jumps at time J_k, we toss a fair coin Y_k. As to the jumps of X'_δ, we introduce another control parameter p: if $p = 0$, then the two processes always jump simultaneously, while if $p = 1$, they always jump independently. Observe that the coupling defined in (3.27) is constructed so that X' does not depend on how X evolves in the future, so the coupling is co-adapted.

As we did previously, we could obtain the jumpy Brownian motions on the circle by taking X and X' (mod 2π), but that would cause issues with the calculations required in this section. For that reason, we first define the distance between X_δ and X'_δ as the real-valued process

$$\begin{aligned} Z_\delta = X_\delta - X'_\delta &= x + \frac{1-\theta}{2} B_\delta - \frac{\sqrt{1-\theta^2}}{2} B'_t \\ &\quad + \pi \sum_{k=1}^{\infty} \mathbb{1}_{\{J_k < \delta\}} [(1-(1-p))Y_k - p(1-Y_k)] \\ &\stackrel{d}{=} x + \sqrt{\frac{1-\theta}{2}} \tilde{B}_\delta + \pi \sum_{k=1}^{\infty} \mathbb{1}_{\{J_k < \delta\}} p(2Y_k - 1), \end{aligned} \tag{3.28}$$

where $\tilde{B}_\delta$ is a standard Brownian motion. Finally, we define $D_\delta = R(Z_\delta)$, where

$$R(x) = \begin{cases} x - 2k\pi & x \in (2k\pi, (2k+1)\pi) \text{ for any integer } k \geq 0 \\ & \textit{or } x \in ((2k+1)\pi, (2k+2)\pi) \text{ for any integer } k \leq -1 \\ 2\pi - (x - 2k\pi) & x \in ((2k+1)\pi, (2k+2)\pi) \text{ for any integer } k \geq 0 \\ & \textit{or } x \in (2k\pi, (2k+1)\pi) \text{ for any integer } k \leq -1, \end{cases}$$

is a function that projects any point of $\mathbb{R}$ onto a point on the interval $[0, \pi]$. In other words, R is a periodic function of period 2π: it is the identity on the interval $[0, \pi]$, and it reflects any point in $[\pi, 2\pi]$ over π. Any other point on $\mathbb{R}$ is either translated directly to a point in the interval $[0, \pi]$, or it is first translated to $[\pi, 2\pi]$ and then reflected with respect to π, so that any point in $\mathbb{R}$ is mapped into a value in the interval $[0, \pi]$. Figure 3.12 shows the graph of this function. So, $R(Z)$ has the same effect of taking $Z \pmod{2\pi}$, but it allows us to avoid the complications that would arise from the use of $\pmod{2\pi}$ in the calculations exposed in the rest of this section.

From Definition 3.24, we can rewrite $V^c_{D_0}$ as

$$V^c_{D_0}(\delta) = \begin{cases} \delta + \hat{v}(D_\delta) & \text{if } \delta < T^c_{D_0} \\ T^c_{D_0} & \text{otherwise} \end{cases}$$

where T^c_x is the coupling time of the process started from x under a coupling c. Observe that if $\delta > T^c_{D_0}$, then X_δ and X'_δ have already coupled by time δ, so it follows that $\hat{v}(D_\delta) = 0$. We want to calculate the following expectation.

$$\mathbb{E}[V^c_{D_0}(\delta) - V^c_{D_0}(0)] = \mathbb{E}[(\delta + \hat{v}(D^c_\delta))\mathbb{1}_{\{T^c_{D_0} > \delta\}} + (T^c_{D_0} + \hat{v}(D^c_\delta))\mathbb{1}_{\{T^c_{D_0} < \delta\}} - V^c_{D_0}(0)]$$

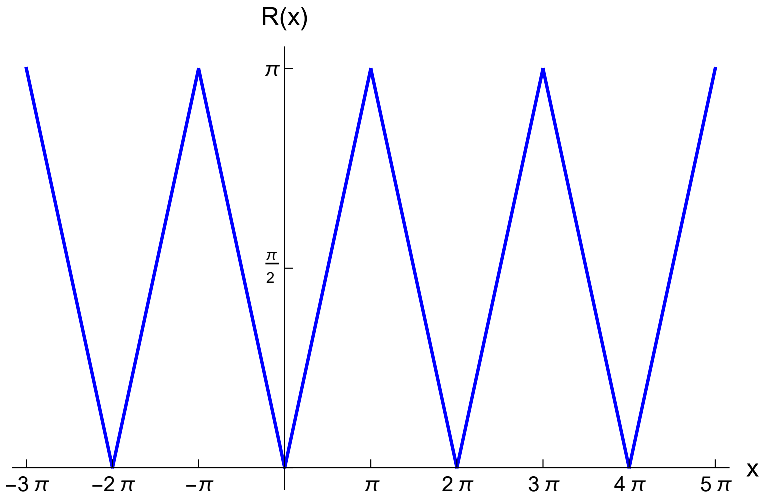


Fig. 3.12: Graph of the function $R(x)$.

$$
\begin{aligned}
&= \mathbb{E}[\delta \mathbb{1}_{\{T^c_{D_0}>\delta\}} + \hat{v}(D^c_\delta) + T^c_{D_0}\mathbb{1}_{\{T^c_{D_0}<\delta\}} - V^c_{D_0}(0)] \\
&\geq \delta \mathbb{P}(T^c_{D_0} > \delta) + \mathbb{E}[\hat{v}(D^c_\delta)] - V^c_{D_0}(0) \\
&= \delta(1 - o(\delta)) + \mathbb{E}[\hat{v}(D^c_\delta)] - V^c_{D_0}(0), \qquad (3.29)
\end{aligned}
$$

where the last equality comes from the following observation. Let $J \sim \text{Exp}(2\lambda p)$ be the random variable, independent from $\tilde{B}_\delta$, that represents the time of the first jump in D^c, then

$$
\mathbb{P}(T^c_{D_0} > \delta) \geq \mathbb{P}(\text{both } J \text{ and the hitting time of the diffusion at } \pm D_0 \text{ are greater then } \delta)
$$
$$
= \mathbb{P}(J > \delta)\mathbb{P}\left(\left|\sqrt{\frac{1-\theta}{2}}\tilde{B}_\delta\right| < D_0\right) = (1 - e^{-2\lambda p\delta})\left(1 - \text{Erfc}\left(\frac{D_0}{\sqrt{(1-\theta)\delta}}\right)\right),
$$

where $e^{-2\lambda p\delta} = o(\delta)$ and $\text{Erfc}\left(\frac{D_0}{\sqrt{(1-\theta)\delta}}\right) = o(\delta)$ for all $p \in [0,1]$ and $\theta \in [-1,1]$. So,

$$
\mathbb{P}(T^c_{D_0} > \delta) \geq 1 - o(\delta)
$$

for all $p \in [0,1]$ and $\theta \in [-1,1]$.

From the discussion above, it is now sufficient to show that

$$
\lim_{\delta \to 0} \frac{\delta + \mathbb{E}[\hat{v}(D^c_\delta)] - V^c_{D_0}(0)}{\delta} \geq 0
$$

for all co-adapted couplings c and that we have the equality if and only if we use our candidate optimal coupling since time 0. This is what we shall do in the remainder of this section.

To make the reading easier, we use the following notations for the error functions that will appear in the calculations in the following sections:

$$\begin{aligned}\operatorname{Erf}(\delta,\theta,j,k) &= \operatorname{Erf}\left(\frac{j}{2\sqrt{\delta(1-\theta)}} + k\sqrt{\delta(1-\theta)}\right)\\ \operatorname{Erfc}(\delta,\theta,j,k) &= 1-\operatorname{Erf}(\delta,\theta,j,k),\end{aligned}$$

and the following property

$$\operatorname{Erf}(-z) = -\operatorname{Erf}(z).$$

From (3.29) and (3.25), we need to find an expression

$$\mathbb{E}[\hat{v}(D_\delta)] = \mathbb{E}\left[D_\delta(\pi - D_\delta) + C(\lambda)\min\{D_\delta, \pi - D_\delta\} + C(\lambda)\frac{\sinh(\sqrt{\lambda}(2D_\delta - \pi))}{\sqrt{\lambda}}\mathbb{1}_{D_\delta > \pi/2}\right], \tag{3.30}$$

where $C(\lambda)$ depends on the coupling strategy we apply. The formula of the value function depends on the value of D_δ, so we distinguish the following cases:

1. $D_0 \in \left(0, \frac{\pi}{2}\right)$
2. $D_0 \in \left(\frac{\pi}{2}, \pi\right)$
3. $D_0 = \frac{\pi}{2}$
4. $D_0 = \pi$.

Since the value functions $\hat{v}(D_\delta)$ for our two candidate optimal couplings differ only by the constant $C(\lambda)$, we will first find lower bounds of the functions that depend on D_δ in (3.30). We will then combine all the terms together distinguishing between the two expressions of $C(\lambda)$ as seen in (3.26), and we will use the lower bounds previously found to establish the mean-optimality of our candidate couplings. We will proceed in this way only for cases 1, 2, and 3. Case 4 will require a separate discussion.

First term of (3.30)

The first term of (3.30) can be bounded in the same way for all D_0. Since both the function $D_\delta(\pi - D_\delta)$ and the jumps are symmetric with respect to $\frac{\pi}{2}$, we have that $D_\delta(\pi - D_\delta)$ is unchanged whether we see a jump before time δ or not. From the definition of the function R, we know that $R(Z_\delta)(\pi - R(Z_\delta)) \geq Z_\delta(\pi - Z_\delta)$, as showed in Figure 3.13.

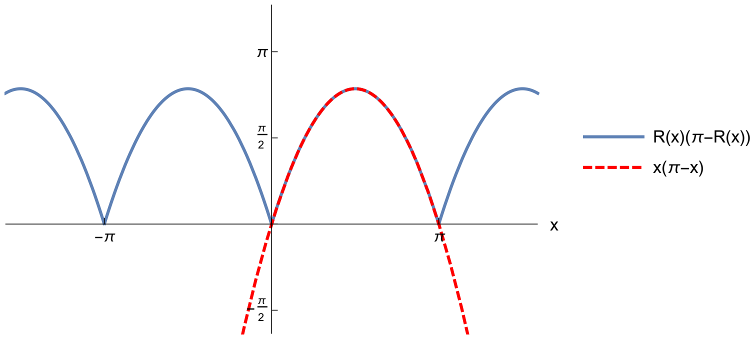


Fig. 3.13: Graph of the functions $R(x)(\pi - R(x))$ and $x(\pi - x)$.

We can therefore find a lower bound in the following way:

$$\begin{aligned}\mathbb{E}[D_\delta(\pi - D_\delta)] &= \mathbb{E}[R(Z_\delta)(\pi - R(Z_\delta))] \geq \mathbb{E}[Z_\delta(\pi - Z_\delta)] \\ &= \mathbb{E}\left[\left(D_0 + \sqrt{\frac{1-\theta}{2}}\tilde{B}_\delta\right)\left(\pi - D_0 - \sqrt{\frac{1-\theta}{2}}\tilde{B}_\delta\right)\right] \\ &= D_0(\pi - D_0) - \frac{1-\theta}{2}\delta. \qquad (3.31)\end{aligned}$$

Second term of (3.30)

To calculate the second term, we observe that $\min\{D_\delta, \pi - D_\delta\}$ is symmetric with respect to $\frac{\pi}{2}$. Since the jumps are also symmetric with respect to the same point, the value of $\min\{D_\delta, \pi - D_\delta\}$ is once again not influenced by the jumps. Again from the definition of the function R, we deduce that $\min\{R(x), \pi - R(x)\} \geq \min\{x, \pi - x\}$, as Figure 3.14 shows.

We rewrite

$$\min\{D_\delta, \pi - D_\delta\} = \frac{\pi}{2} - \frac{|2D_\delta - \pi|}{2},$$

and we distinguish the three cases depending on D_0.

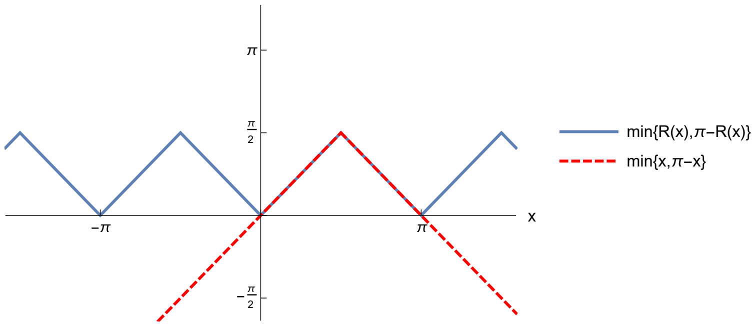


Fig. 3.14: Graph of the functions $\min\{R(x), \pi - R(x)\}$ and $\min\{x, \pi - x\}$.

1. Let $D_0 \in \left(0, \frac{\pi}{2}\right)$. Then,

$$\begin{aligned}
\mathbb{E}[\min\{D_\delta, \pi - D_\delta\}] &= \mathbb{E}\left[\frac{\pi}{2} - \frac{|2D_\delta - \pi|}{2}\right] = \frac{\pi - \mathbb{E}[|2D_\delta - \pi|]}{2} \\
&= \frac{1}{2}\left(\pi - \mathbb{E}\left[\left|2D_0 - \pi + 2\sqrt{\frac{1-\theta}{2}}\tilde{B}_\delta\right|\right]\right) \\
&= \frac{1}{2}\left(\pi - 2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} - (\pi - 2D_0)\,\mathrm{Erf}(\delta, \theta, \pi - 2D_0, 0)\right) \\
&= \frac{1}{2}\left(2D_0 - 2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} + (\pi - 2D_0)\,\mathrm{Erfc}(\delta, \theta, \pi - 2D_0, 0)\right),
\end{aligned} \tag{3.32}$$

where we use $\mathrm{Erf}(z) = 1 - \mathrm{Erfc}(z)$.

2. Let $D_0 \in (\frac{\pi}{2}, \pi)$. Then,

$$\begin{aligned}
\mathbb{E}[\min\{D_\delta, \pi - D_\delta\}] &= \mathbb{E}\left[\left(\frac{\pi}{2} - \frac{|2D_\delta - \pi|}{2}\right)\right] = \frac{\pi - \mathbb{E}[|2D_\delta - \pi|]}{2} \\
&= \frac{1}{2}\left(\pi - \mathbb{E}[|2D_0 - \pi + 2\sqrt{\frac{1-\theta}{2}}B_\delta|]\right) \\
&= \frac{1}{2}\left(\pi - 2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} - (\pi - 2D_0)\,\mathrm{Erf}(\delta, \theta, \pi - 2D_0, 0)\right) \\
&= \frac{1}{2}\left(\pi - 2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} - (2D_0 - \pi)\,\mathrm{Erf}(\delta, \theta, 2D_0 - \pi, 0)\right)
\end{aligned}$$

$$= \frac{1}{2}\left(2\pi - 2x - 2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} + (2D_0-\pi)\operatorname{Erfc}(\delta,\theta,2D_0-\pi,0)\right), \tag{3.33}$$

where we use $\operatorname{Erf}(-z) = -\operatorname{Erf}(z)$ and $\operatorname{Erf}(z) = 1 - \operatorname{Erfc}(z)$.

3. Let $D_0 = \frac{\pi}{2}$. In this case, we have

$$Z_\delta = \frac{\pi}{2} + \sqrt{\frac{1-\theta}{2}}\tilde{B}_\delta.$$

As we observed above, $\min\{R(x), \pi - R(x)\} \geq \min\{x, \pi - x\}$, so

$$\begin{aligned}\mathbb{E}[\min\{D_\delta, \pi - D_\delta\}] &= \mathbb{E}[\min\{R(Z_\delta), \pi - R(Z_\delta)\}] \geq \mathbb{E}[\min\{Z_\delta, \pi - Z_\delta\}] \\ &= \frac{\pi}{2} - \frac{\mathbb{E}[|2Z_\delta - \pi|]}{2} = \frac{\pi}{2} - \frac{\mathbb{E}[|\sqrt{2(1-\theta)}\tilde{B}_\delta|]}{2} \\ &= \frac{\pi}{2} - \sqrt{2(1-\theta)}\frac{\mathbb{E}[|\tilde{B}_\delta|]}{2} = \frac{\pi}{2} - \frac{\sqrt{2(1-\theta)}}{2}\sqrt{\frac{2\delta}{\pi}},\end{aligned} \tag{3.34}$$

where we used

$$\mathbb{E}[|\tilde{B}_\delta|] = 2\int_0^\infty \frac{x}{\sqrt{2\delta\pi}}e^{-\frac{x^2}{2\delta}}\,dx = \sqrt{\frac{2\delta}{\pi}}.$$

Third term of (3.30)

To find a lower bound for the third term of (3.30), we use the process Z and an independent variable $J \sim \operatorname{Exp}(2\lambda p)$ that represents the time of the first jump in X or X'.

1. Let $D_0 \in \left(0, \frac{\pi}{2}\right)$. Restricting to the case when $J < \delta$, we obtain:

$$\begin{aligned}\mathbb{E}[\sinh(\sqrt{\lambda}(2D_\delta - \pi))\mathbb{1}_{\{D_\delta > \frac{\pi}{2}\}}] &\geq \mathbb{E}[\sinh(\sqrt{\lambda}(2D_\delta - \pi))\mathbb{1}_{\{D_\delta > \frac{\pi}{2}\}}\mathbb{1}_{\{J<\delta\}}] \\ &= \mathbb{E}[\sinh(\sqrt{\lambda}(2R(Z_\delta) - \pi))\mathbb{1}_{\{\pi - D_0 + \sqrt{(1-\theta)/2} > \frac{\pi}{2}\}}]\mathbb{P}(J < \delta) \\ &\geq 2\lambda p\delta \int_{\frac{\pi}{2}}^{\pi} \frac{1}{\sqrt{\pi\delta(1-\theta)}}\sinh(\sqrt{\lambda}(2y-\pi))e^{-\frac{(\pi-D_0-y)^2}{(1-\theta)\delta}}\,dy \\ &= 2\lambda p\delta\frac{1}{4}e^{\frac{\sqrt{\lambda}(2(1+\theta)\pi + (\theta-1)^2\sqrt{\lambda}\delta + 2D_0(1+\theta))}{1-\theta}}\end{aligned}$$

$$\left(e^{-\frac{\sqrt{\lambda}(\pi+3\pi\theta+4D_0)}{1-\theta}}\left(-\operatorname{Erf}(\delta,\theta,-2D_0,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi-2D_0,\sqrt{\lambda})\right)\right.$$
$$\left.+\,e^{-\frac{\sqrt{\lambda}(3\pi+\pi\theta+4\theta D_0)}{1-\theta}}\left(-2+\operatorname{Erfc}(\delta,\theta,2D_0,\sqrt{\lambda})+\operatorname{Erfc}(\delta,\theta,\pi-2D_0,-\sqrt{\lambda})\right)\right). \tag{3.35}$$

2. Let $D_0 \in \left(\frac{\pi}{2},\pi\right)$. Restricting to the case when $J > \delta$ yields

$$\begin{aligned}
&\mathbb{E}[\sinh(\sqrt{\lambda}(2D_\delta-\pi))\mathbb{1}_{\{D_\delta>\frac{\pi}{2}\}}] \geq \mathbb{E}[\sinh(\sqrt{\lambda}(2D_\delta-\pi))\mathbb{1}_{\{D_\delta>\frac{\pi}{2}\}}\mathbb{1}_{\{J>\delta\}}]\\
&\quad\geq \mathbb{E}[\sinh(\sqrt{\lambda}(2R(Z_\delta)-\pi))\mathbb{1}_{\{D_0+\sqrt{\frac{(1-\theta)}{2}}\tilde{B}_\delta>\frac{\pi}{2}\}}]\mathbb{P}(J>\delta)\\
&\quad\geq (1-2\lambda p\delta)\int_{\frac{\pi}{2}}^{\pi}\frac{1}{\sqrt{\pi\delta(1-\theta)}}\sinh(\sqrt{\lambda}(2y-\pi))e^{-\frac{(D_0-y)^2}{(1-\theta)\delta}}\,dy\\
&\quad= (1-2\lambda p\delta)\frac{1}{4}e^{\lambda\delta(1-\theta)+\sqrt{\lambda}(2D_0-\pi)}\left(-1+\operatorname{Erf}(\delta,\theta,2\pi-2D_0,-\sqrt{\lambda})\right.\\
&\qquad+e^{-2\sqrt{\lambda}(2D_0-\pi)}\left(-\operatorname{Erf}(\delta,\theta,2\pi-2D_0,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi-2D_0,\sqrt{\lambda})\right)\\
&\qquad\left.+\operatorname{Erfc}(\delta,\theta,\pi-2D_0,-\sqrt{\lambda})\right).
\end{aligned}\tag{3.36}$$

3. Let $D_0 = \frac{\pi}{2}$. We again use Z and the following observation: $\mathbb{1}_{\{R(Z_\delta)>\frac{\pi}{2}\}} = 1$ if $\tilde{B}_\delta > 0$ and $J > \delta$, or if $\tilde{B}_\delta < 0$ and $J < \delta$.

$$\begin{aligned}
&\mathbb{E}[\sinh(\sqrt{\lambda}(2D_\delta-\pi))\mathbb{1}_{\{D_\delta>\frac{\pi}{2}\}}] = \mathbb{E}[\sinh(\sqrt{\lambda}(2R(Z_\delta)-\pi))\mathbb{1}_{\{R(Z_\delta)>\frac{\pi}{2}\}}]\\
&= \mathbb{E}\left[\sinh(\sqrt{\lambda}(2R(Z_\delta)-\pi))\left(\mathbb{1}_{\{\tilde{B}_\delta>0\}}\mathbb{1}_{\{J>\delta\}}+\mathbb{1}_{\{\tilde{B}_\delta<0\}}\mathbb{1}_{\{J<\delta\}}\right)\right]\\
&\geq \int_{\frac{\pi}{2}}^{\pi}\frac{1}{\sqrt{\pi\delta(1-\theta)}}\sinh(\sqrt{\lambda}(2y-\pi))e^{-\frac{(\frac{\pi}{2}-y)^2}{(1-\theta)\delta}}\,dy\\
&= \frac{1}{4}e^{(1-\alpha)\lambda\delta}\left(2\operatorname{Erf}(\delta,\theta,0,\sqrt{\lambda})-\operatorname{Erf}(\delta,\theta,\pi,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi,-\sqrt{\lambda})\right).
\end{aligned}\tag{3.37}$$

3.9.1 Proof of Theorem 3.8

In this section, we prove that our candidate coupling is the unique mean-optimal coupling if $\lambda < \lambda^\star$ and $\lambda > \lambda^\star$ completing the proof of Theorem 3.8. To do that, we bound $\mathbb{E}[V^c_{D_0}(\delta) - V^c_{D_0}(0)]$ as in (3.29) combining the lower bounds found in the previous paragraphs in (3.25) with $C(\lambda)$ defined as

$$C(\lambda) = \begin{cases} \frac{\pi}{2\cosh(\sqrt{\lambda}\pi)-1} & \text{if we apply the reflection coupling}\\ \frac{\operatorname{cosech}(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}} & \text{if we apply the synchronised coupling.}\end{cases}$$

Case 1. Since $D_0 \in \left(0, \frac{\pi}{2}\right)$, the value function at time 0 is

$$V^c_{D_0}(0) = D_0(\pi - D_0) + D_0 C(\lambda).$$

Using equations (3.31), (3.32), and (3.35), we obtain

$$\begin{aligned}
&\mathbb{E}[V^c_{D_0}(\delta) - V^c_{D_0}(0)] \geq \delta(1 - o(\delta)) + D_0(\pi - D_0) - \frac{1-\theta}{2}\delta \\
&\quad + \frac{C(\lambda)}{2}\left(2D_0 - 2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} + (\pi - 2D_0)\operatorname{Erfc}(\delta, \theta, \pi - 2D_0, 0)\right) \\
&\quad + \frac{2\lambda p\delta C(\lambda)}{4\sqrt{\lambda}} e^{\frac{\sqrt{\lambda}(2(1+\theta)\pi + (\theta-1)^2\sqrt{\lambda}\delta + 2D_0(1+\theta))}{1-\theta}} \\
&\quad \left[e^{-\frac{\sqrt{\lambda}(\pi+3\pi\theta+4D_0)}{1-\theta}}\left(-\operatorname{Erf}(\delta, \theta, -2D_0, \sqrt{\lambda}) + \operatorname{Erf}(\delta, \theta, \pi - 2D_0, \sqrt{\lambda})\right)\right. \\
&\qquad \left. + e^{-\frac{\sqrt{\lambda}(3\pi+\pi\theta+4\theta D_0)}{1-\theta}}\left(-2 + \operatorname{Erfc}(\delta, \theta, 2D_0, \sqrt{\lambda}) + \operatorname{Erf}(\delta, \theta, \pi - 2D_0, -\sqrt{\lambda})\right)\right] \\
&\quad - D_0(\pi - D_0) - D_0 C(\lambda) \\
&= \delta(1 - o(\delta)) - \frac{1-\theta}{2}\delta \\
&\quad + \frac{C(\lambda)}{2}\left(-2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} + (\pi - 2D_0)\operatorname{Erfc}(\delta, \theta, \pi - 2D_0, 0)\right) \\
&\quad + \frac{2\lambda p\delta C(\lambda)}{4\sqrt{\lambda}} e^{\frac{\sqrt{\lambda}(2(1+\theta)\pi + (\theta-1)^2\sqrt{\lambda}\delta + 2D_0(1+\theta))}{1-\theta}} \\
&\qquad \left[e^{-\frac{\sqrt{\lambda}(\pi+3\pi\theta+4D_0)}{1-\theta}}\left(-\operatorname{Erf}(\delta, \theta, -2D_0, \sqrt{\lambda}) + \operatorname{Erf}(\delta, \theta, \pi - 2D_0, \sqrt{\lambda})\right)\right. \\
&\quad \left. + e^{-\frac{\sqrt{\lambda}(3\pi+\pi\theta+4\theta D_0)}{1-\theta}}\left(-2 + \operatorname{Erfc}(\delta, \theta, 2D_0, \sqrt{\lambda}) + \operatorname{Erfc}(\delta, \theta, \pi - 2D_0, -\sqrt{\lambda})\right)\right].
\end{aligned}$$

Dividing by δ, we obtain

$$\begin{aligned}
&\frac{\mathbb{E}[V^c_{D_0}(\delta) - V^c_{D_0}(0)]}{\delta} \geq 1 - o(\delta) - \frac{1-\theta}{2} \\
&\quad + \frac{C(\lambda)}{2}\left(-2\sqrt{\frac{1-\theta}{\pi\delta}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} + \frac{(\pi - 2D_0)}{\delta}\operatorname{Erfc}(\delta, \theta, \pi - 2D_0, 0)\right) \\
&\quad + \frac{2\lambda p C(\lambda)}{4\sqrt{\lambda}} e^{\frac{\sqrt{\lambda}(2(1+\theta)\pi + (\theta-1)^2\sqrt{\lambda}\delta + 2D_0(1+\theta))}{1-\theta}} \\
&\qquad \left[e^{-\frac{\sqrt{\lambda}(\pi+3\pi\theta+4D_0)}{1-\theta}}\left(-\operatorname{Erf}(\delta, \theta, -2D_0, \sqrt{\lambda}) + \operatorname{Erf}(\delta, \theta, \pi - 2D_0, \sqrt{\lambda})\right)\right.
\end{aligned}$$

$$+ e^{-\frac{\sqrt{\lambda}(3\pi+\pi\theta+4\theta D_0)}{1-\theta}} \left(-2 + \operatorname{Erfc}(\delta, \theta, 2D_0, \sqrt{\lambda}) + \operatorname{Erfc}(\delta, \theta, \pi - 2D_0, -\sqrt{\lambda})\right)\Bigg].$$

Finally, we take the limit as $\delta \to 0$. The limit of the second line is

$$\lim_{\delta\to 0}\left(-2\sqrt{\frac{1-\theta}{\pi\delta}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} + \frac{(\pi-2D_0)}{\delta}\operatorname{Erfc}(\delta,\theta,\pi-2D_0,0)\right) = 0,$$

while in the last two lines we have

$$\lim_{\delta\to 0}\operatorname{Erf}(\delta,\theta,-2D_0,\sqrt{\lambda}) = -1, \qquad \lim_{\delta\to 0}\operatorname{Erf}(\delta,\theta,\pi-2D_0,\sqrt{\lambda}) = 1,$$
$$\lim_{\delta\to 0}\operatorname{Erfc}(\delta,\theta,2D_0,\sqrt{\lambda}) = 0, \qquad \lim_{\delta\to 0}\operatorname{Erfc}(\delta,\theta,\pi-2D_0,-\sqrt{\lambda}) = 0.$$

Thus, after taking the limit and combining the remaining exponential functions, we obtain

$$\begin{aligned}\lim_{\delta\to 0}\frac{\mathbb{E}[V^c_{D_0}(\delta) - V^c_{D_0}(0)]}{\delta} &\geq \frac{1+\theta}{2} + \frac{2\lambda p C(\lambda)}{2\sqrt{\lambda}}e^{-\sqrt{\lambda}(\pi-2D_0)}(e^{2\sqrt{\lambda}(\pi-2D_0)} - 1) \\ &= \frac{1+\theta}{2} + \frac{2\lambda p C(\lambda)}{\sqrt{\lambda}}\sinh(\sqrt{\lambda}(\pi-2D_0)).\end{aligned}$$

We now observe that this lower bound is non-negative for all $\theta \in [-1,1]$ and $p \in [0,1]$, and the equality is attained if and only if $\theta = -1$ and $p = 0$. Then, if $D_0 \in (0, \frac{\pi}{2})$, $V^c_{D_0}$ is a submartingale for any co-adapted coupling c, and it is a martingale if and only if we use our candidate mean-optimal coupling, i.e. the processes X and X' reflect their Brownian motion components ($\theta = -1$) and they jump at the same time ($p = 0$).

Case 2. Since $D_0 \in (\frac{\pi}{2}, \pi)$, the value function at time 0 is

$$V^c_{D_0}(0) = D_0(\pi - D_0) + (\pi - D_0)C(\lambda) + \frac{\sinh(\sqrt{\lambda}(2D_0-\pi))}{\sqrt{\lambda}}C(\lambda).$$

We then use equations (3.31), (3.33), and (3.36) to obtain

$$\begin{aligned}&\mathbb{E}[V^c_{D_0}(\delta) - V^c_{D_0}(0)] \geq \delta(1 - o(\delta)) + D_0(\pi - D_0) - \frac{1-\theta}{2}\delta \\ &\quad + \frac{C(\lambda)}{2}\left(2\pi - 2D_0 - 2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}} + (2D_0-\pi)\operatorname{Erfc}(\delta,\theta,2D_0-\pi,0)\right)\end{aligned}$$

$$+\frac{(1-2\lambda p\delta)C(\lambda)}{4\sqrt{\lambda}}e^{\lambda\delta(1-\theta)+\sqrt{\lambda}(2D_0-\pi)}\Big[-1+\operatorname{Erf}(\delta,\theta,2\pi-2D_0,-\sqrt{\lambda})$$
$$+e^{-2\sqrt{\lambda}(2D_0-\pi)}\Big(-\operatorname{Erf}(\delta,\theta,2\pi-2D_0,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi-2D_0,\sqrt{\lambda})\Big)$$
$$+\operatorname{Erfc}(\delta,\theta,\pi-2D_0,-\sqrt{\lambda})\Big]-D_0(\pi-D_0)-(\pi-D_0)C(\lambda)-\frac{C(\lambda)\sinh(\sqrt{\lambda}(2D_0-\pi))}{\sqrt{\lambda}}$$
$$=\delta(1-o(\delta))-\frac{1-\theta}{2}\delta$$
$$+\frac{C(\lambda)}{2}\left(-2\sqrt{\delta\frac{1-\theta}{\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}}+(2D_0-\pi)\operatorname{Erfc}(\delta,\theta,2D_0-\pi,0)\right)$$
$$+\frac{(1-2\lambda p\delta)C(\lambda)}{4\sqrt{\lambda}}e^{\lambda\delta(1-\theta)+\sqrt{\lambda}(2D_0-\pi)}\Big[-1+\operatorname{Erf}(\delta,\theta,2\pi-2D_0,-\sqrt{\lambda})$$
$$+e^{-2\sqrt{\lambda}(2D_0-\pi)}\Big(-\operatorname{Erf}(\delta,\theta,2\pi-2D_0,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi-2D_0,\sqrt{\lambda})\Big)$$
$$+\operatorname{Erfc}(\delta,\theta,\pi-2D_0,-\sqrt{\lambda})\Big]-\frac{C(\lambda)\sinh(\sqrt{\lambda}(2D_0-\pi))}{\sqrt{\lambda}}.$$

We divide by δ,

$$\frac{\mathbb{E}[V^c_{D_0}(\delta)-V^c_{D_0}(0)]}{\delta}\geq(1-o(\delta))-\frac{1-\theta}{2}$$
$$+\frac{C(\lambda)}{2}\left(-2\sqrt{\frac{1-\theta}{\delta\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}}+\frac{(2D_0-\pi)}{\delta}\operatorname{Erfc}(\delta,\theta,2D_0-\pi,0)\right)$$
$$+\frac{1}{\delta}\Bigg[\frac{(1-2\lambda p\delta)C(\lambda)}{4\sqrt{\lambda}}e^{\lambda\delta(1-\theta)+\sqrt{\lambda}(2D_0-\pi)}\Big(-1+\operatorname{Erf}(\delta,\theta,2\pi-2D_0,-\sqrt{\lambda})$$
$$+e^{-2\sqrt{\lambda}(2D_0-\pi)}\Big(-\operatorname{Erf}(\delta,\theta,2\pi-2D_0,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi-2D_0,\sqrt{\lambda})\Big)$$
$$+\operatorname{Erfc}(\delta,\theta,\pi-2D_0,-\sqrt{\lambda})\Big)-\frac{C(\lambda)\sinh(\sqrt{\lambda}(2D_0-\pi))}{\sqrt{\lambda}}\Bigg],$$

and we take the limit as $\delta\to 0$. The limit of the second line is

$$\lim_{\delta\to 0}\left(-2\sqrt{\frac{1-\theta}{\delta\pi}}e^{-\frac{(\pi-2D_0)^2}{4(1-\theta)\delta}}+\frac{(2D_0-\pi)}{\delta}\operatorname{Erfc}(\delta,\theta,2D_0-\pi,0)\right)=0,$$

and

$$\lim_{\delta\to 0}\Bigg[\frac{(1-2\lambda p\delta)C(\lambda)}{4\delta\sqrt{\lambda}}e^{\lambda\delta(1-\theta)+\sqrt{\lambda}(2D_0-\pi)}\Big(-1+\operatorname{Erf}(\delta,\theta,2\pi-2D_0,-\sqrt{\lambda})$$
$$+e^{-2\sqrt{\lambda}(2D_0-\pi)}\Big(-\operatorname{Erf}(\delta,\theta,2\pi-2D_0,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi-2D_0,\sqrt{\lambda})\Big)$$

$$+ \operatorname{Erfc}(\delta, \theta, \pi - 2D_0, -\sqrt{\lambda})\Big) - \frac{C(\lambda)\sinh(\sqrt{\lambda}(2D_0 - \pi))}{\sqrt{\lambda}}\Bigg]$$
$$= \frac{1}{2}e^{\sqrt{\lambda}(2D_0-\pi)}(-1 + e^{-2\sqrt{\lambda}(2D_0-\pi)})\lambda(-1 + \theta + 2p).$$

Thus,

$$\begin{aligned}\lim_{\delta\to 0}\frac{\mathbb{E}[V^c_{D_0}(\delta) - V^c_{D_0}(0)]}{\delta} &\geq 1 - \frac{1-\theta}{2} + \frac{C(\lambda)}{2\sqrt{\lambda}}e^{\sqrt{\lambda}(2D_0-\pi)} \\ &\quad (-1 + e^{-2\sqrt{\lambda}(2D_0-\pi)})\lambda(-1+\theta+2p) \\ &= \frac{1+\theta}{2} + \frac{C(\lambda)}{2\sqrt{\lambda}}(-e^{\sqrt{\lambda}(2D_0-\pi)} + e^{-\sqrt{\lambda}(2D_0-\pi)})\lambda(-1+\theta+2p) \\ &= \frac{1+\theta}{2} + \frac{C(\lambda)}{\sqrt{\lambda}}(-\sinh(\sqrt{\lambda}(2D_0-\pi)))\lambda(-1+\theta+2p) \\ &= \frac{1+\theta}{2} + \frac{C(\lambda)\sinh(\sqrt{\lambda}(2D_0-\pi))}{\sqrt{\lambda}}\lambda(1-\theta-2p).\end{aligned} \tag{3.38}$$

To show that our coupling is optimal, we need to prove that the expression (3.38) is non-negative for all $\theta \in [-1, 1]$ and $p \in [0, 1]$ and equal to 0 if and only if $\theta = -1$ and $p = 1$. That comes from the fact that, as we defined our candidate optimal coupling, we want X and X' to reflect their Brownian motion components ($\theta = -1$) and to jump independently ($p = 1$).
In the second term of the formula, the factor $1-\theta-2p$ is not always positive, so we need to distinguish two cases.

i. If $1-\theta-2p \geq 0$, then both terms of the sum in (3.38) are non-negative and are both zero if and only if $\theta = -1$ and $p = 1$.

ii. Alternatively, suppose that $1 - \theta - 2p < 0$. We distinguish two cases depending on the value of λ.

   (a) Let $\lambda < \lambda^\star$, then

$$\frac{\pi\sinh(\sqrt{\lambda}(\pi))}{\sqrt{\lambda}(2\cosh(\sqrt{\lambda}\pi) - 1)} < \frac{1}{2\lambda}.$$

   Since $D_0 \in \left(\frac{\pi}{2}, \pi\right)$, we have that $\sinh(\sqrt{\lambda}(2D_0-\pi)) < \sinh(\sqrt{\lambda}\pi)$, so

$$\lim_{\delta\to 0}\frac{\mathbb{E}[V^c_{D_0}(\delta)-V^c_{D_0}(0)]}{\delta}\geq\frac{1+\theta}{2}+\frac{\pi\sinh(\sqrt{\lambda}(2D_0-\pi))}{\sqrt{\lambda}(2\cosh(\sqrt{\lambda}\pi)-1)}\lambda(1-\theta-2p)$$
$$\geq\frac{1+\theta}{2}+\frac{1-\theta-2p}{2}=1-p\geq 0.$$

(b) Let $\lambda>\lambda^\star$. The initial assumption $D_0\in(\frac{\pi}{2},\pi)$ implies that $2D_0-\pi<\pi$, so we have

$$\lim_{\delta\to 0}\frac{\mathbb{E}[V^c_{D_0}(\delta)-V^c_{D_0}(0)]}{\delta}\geq\frac{1+\theta}{2}+\frac{\operatorname{cosech}(\sqrt{\lambda}\pi)\sinh(\sqrt{\lambda}\pi)}{2\lambda}\lambda(1-\theta-2p)$$
$$=\frac{1+\theta}{2}+\frac{1}{2\lambda}\lambda(1-\theta-2p)=\frac{1+\theta}{2}+\frac{1-\theta-2p}{2}$$
$$=1-p\geq 0.$$

In both cases, also when $1-\theta-2p<0$, the formula is non-negative.

In conclusion, we have that $\lim_{\delta\to 0}\frac{\mathbb{E}[V^c_{D_0}(\delta)-V^c_{D_0}(0)]}{\delta}\geq 0$ for all $\theta\in[-1,1]$ and $p\in[0,1]$ and it is equal to 0 if and only if $\theta=-1$ and $p=1$, i.e. if and only if we use our candidate mean-optimal coupling.

Case 3. Let $D_0=\frac{\pi}{2}$, then

$$V^c_{\frac{\pi}{2}}(0)=\frac{\pi^2}{4}+\frac{\pi}{2}C(\lambda).$$

Using (3.31), (3.34), and (3.37) yields

$$\mathbb{E}[V^c_{\frac{\pi}{2}}(\delta)-V^c_{\frac{\pi}{2}}(0)]\geq\delta(1-o(\delta))+\frac{\pi^2}{4}-\frac{1-\theta}{2}\delta+C(\lambda)\left(\frac{\pi}{2}-\frac{\sqrt{2(1-\theta)}}{2}\sqrt{\frac{2\delta}{\pi}}\right)$$
$$+\frac{C(\lambda)e^{(1-\theta)\lambda\delta}}{4\sqrt{\lambda}}\left(2\operatorname{Erf}(\delta,\theta,0,\sqrt{\lambda})-\operatorname{Erf}(\delta,\theta,\pi,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi,-\sqrt{\lambda})\right)$$
$$-\frac{\pi^2}{4}-\frac{\pi C(\lambda)}{2}$$
$$=\delta(1-o(\delta))-\frac{1-\theta}{2}\delta+C(\lambda)\left(-\sqrt{\frac{(1-\theta)\delta}{\pi}}\right)$$
$$+\frac{C(\lambda)e^{(1-\theta)\lambda\delta}}{4\sqrt{\lambda}}\left(2\operatorname{Erf}(\delta,\theta,0,\sqrt{\lambda})-\operatorname{Erf}(\delta,\theta,\pi,\sqrt{\lambda})+\operatorname{Erf}(\delta,\theta,\pi,-\sqrt{\lambda})\right).$$

We divide by δ,

$$\frac{\mathbb{E}[V^c_{\frac{\pi}{2}}(\delta) - V^c_{\frac{\pi}{2}}(0)]}{\delta} \geq (1 - o(\delta)) - \frac{1-\theta}{2} + \frac{C(\lambda)}{\delta\sqrt{\lambda}}$$
$$\cdot \left[\frac{e^{(1-\theta)\lambda\delta}}{4} \Big(2\operatorname{Erf}(\delta, \theta, 0, \sqrt{\lambda}) - \operatorname{Erf}(\delta, \theta, \pi, \sqrt{\lambda}) + \operatorname{Erf}(\delta, \theta, \pi, -\sqrt{\lambda}) \Big) - \sqrt{\frac{(1-\theta)\lambda\delta}{\pi}} \right],$$

and we take the limit as δ tends to 0,

$$\lim_{\delta \to 0} \frac{\mathbb{E}[V^c_{\frac{\pi}{2}}(\delta) - V^c_{\frac{\pi}{2}}(0)]}{\delta} \geq 1 - \frac{1-\theta}{2} = \frac{1+\theta}{2},$$

which is non-negative for all $\theta \in [-1, 1]$ and it is equal to 0 if and only if $\theta = -1$. So, our candidate mean-optimal coupling is the unique mean-optimal coupling strategy.

Case 4. Finally, we consider the case $D_0 = \pi$. We have

$$\mathbb{E}[V^c_\pi(\delta) - V^c_\pi(0) | D_0 = \pi] = \mathbb{E}\left[\int_0^\delta \mathbb{1}_{\{T^c_\pi > y\}} dy + \hat{v}(D^c_\delta) \right] - \hat{v}(\pi). \qquad (3.39)$$

The following results will help us find a lower bound of (3.39).
Let $\phi(\pi, \theta, \delta; y)$ be the density of a Brownian motion with mean π and variance $\frac{(1-\theta)\delta}{2}$ at $y \in \mathbb{R}$. We define

$$I_1 := \int_{-\infty}^{\pi} (y(\pi - y) + C(\lambda)\min\{y, \pi - y\}) 2\phi(\pi, \theta, \delta; y) dy$$
$$= -\frac{(1-\theta)\delta}{2} + \sqrt{\frac{(1-\theta)\delta}{\pi}} [C(\lambda) + \pi - 2C(\lambda)e^{-\frac{\pi^2}{(1-\theta)\delta}}] + C(\lambda)\pi \operatorname{Erfc}(\delta, \theta, \pi, 0),$$

and

$$I_2 := \int_{-\infty}^{\pi} \frac{C(\lambda)}{\sqrt{\lambda}} \sinh(\sqrt{\lambda}(2y - \pi)) 2\phi(\pi, \theta, \delta; y) dy$$
$$= e^{(1-\theta)\delta\lambda} \frac{C(\lambda)}{\sqrt{\lambda}} [\sinh(\sqrt{\lambda}\pi) - \cosh(\sqrt{\lambda}\pi) \operatorname{Erf}(\delta, \theta, 0, \sqrt{\lambda})].$$

Now, let $J \sim \operatorname{Exp}(2\lambda p)$ be the first time X or X' jumps, then

$$\mathbb{E}[\hat{v}(D_\delta)] \geq \mathbb{E}[\hat{v}(D_\delta) | J > \delta] e^{-2\lambda p\delta} \geq (I_1 + I_2) e^{-2\lambda p\delta},$$

where $e^{-2\lambda p\delta} = \mathbb{P}(J > \delta)$. So,

$$\begin{aligned}
&\mathbb{E}[\hat{v}(D_\delta)] - \hat{v}(\pi) \geq (I_1 + I_2)e^{-2\lambda p\delta} - \hat{v}(\pi) \\
&= e^{-2\lambda p\delta}\left[-\frac{(1-\theta)\delta}{2} + \sqrt{\frac{(1-\theta)\delta}{\pi}}[C(\lambda) + \pi - 2C(\lambda)e^{-\frac{\pi^2}{(1-\theta)\delta}}] + C(\lambda)\pi \operatorname{Erfc}(\delta,\theta,\pi,0)\right. \\
&\quad \left. + e^{(1-\theta)\delta\lambda}\frac{C(\lambda)}{\sqrt{\lambda}}[\sinh(\sqrt{\lambda}\pi) - \cosh(\sqrt{\lambda}\pi)\operatorname{Erf}(\delta,\theta,0,\sqrt{\lambda})]\right] - \hat{v}(\pi) \\
&= e^{-2\lambda p\delta}\left[-\frac{(1-\theta)\delta}{2} - \sqrt{\frac{(1-\theta)\delta}{\pi}}2C(\lambda)e^{-\frac{\pi^2}{(1-\theta)\delta}} + C(\lambda)\pi \operatorname{Erfc}(\delta,\theta,\pi,0)\right] \\
&\quad - \hat{v}(\pi)[1 - e^{-2\lambda p\delta + (1-\theta)\delta\lambda}] \\
&\quad + e^{-2\lambda p\delta}\left[\sqrt{\frac{(1-\theta)\delta}{\pi}}(C(\lambda) + \pi) - e^{(1-\theta)\delta\lambda}\frac{C(\lambda)}{\sqrt{\lambda}}\cosh(\sqrt{\lambda}\pi)\operatorname{Erf}(\delta,\theta,0,\sqrt{\lambda})\right].
\end{aligned} \tag{3.40}$$

We take the Taylor series around point 0 of the last term

$$\begin{aligned}
&\sqrt{\frac{(1-\theta)\delta}{\pi}}(C(\lambda) + \pi) - e^{(1-\theta)\delta\lambda}\frac{C(\lambda)}{\sqrt{\lambda}}\cosh(\sqrt{\lambda}\pi)\operatorname{Erf}(\delta,\theta,0,\sqrt{\lambda}) \\
&\sim (C(\lambda) + \pi)\frac{1-\theta}{\pi}\sqrt{\delta} - \frac{C(\lambda)}{\sqrt{\lambda}}\cosh(\sqrt{\lambda}\pi)2\frac{(1-\theta)\lambda}{\pi}\sqrt{\delta} + O(\delta^{\frac{3}{2}}) \\
&= \frac{(1-\theta)\delta}{\pi}(C(\lambda) + \pi - 2C(\lambda)\cosh(\sqrt{\lambda}\pi)) + O(\delta^{\frac{3}{2}}) = O(\delta^{\frac{3}{2}}).
\end{aligned}$$

So, we get

$$\begin{aligned}
\mathbb{E}[\hat{v}(D_\delta)] - \hat{v}(\pi) \geq e^{-2\lambda p\delta}&\left[-\frac{(1-\theta)\delta}{2} - \sqrt{\frac{(1-\theta)\delta}{\pi}}2C(\lambda)e^{-\frac{\pi^2}{(1-\theta)\delta}} + C(\lambda)\pi \operatorname{Erfc}(\delta,\theta,\pi,0)\right] \\
&- \hat{v}(\pi)[1 - e^{-2\lambda p\delta + (1-\theta)\delta\lambda}] + O(\delta^{\frac{3}{2}}).
\end{aligned}$$

Now, we need to distinguish the following two cases.

- Let $\lambda < \lambda^\star$. Then,

$$\begin{aligned}
\mathbb{E}[V_\pi^c(\delta) - V_\pi^c(0)|D_0 = \pi] \geq \delta + e^{-2\lambda p\delta}&\left[-\frac{(1-\theta)\delta}{2} - \sqrt{\frac{(1-\theta)\delta}{\pi}}2C(\lambda)e^{-\frac{\pi^2}{(1-\theta)\delta}}\right. \\
&\left. + C(\lambda)\pi \operatorname{Erfc}(\delta,\theta,\pi,0)\right] - \hat{v}(\pi)[1 - e^{-2\lambda p\delta + (1-\theta)\delta\lambda}] + O(\delta^{\frac{3}{2}}).
\end{aligned}$$

We divide by δ and let $\delta \to 0$

$$\lim_{\delta\to 0} \frac{\mathbb{E}[V^c_\pi(\delta) - V^c_\pi(0)|D_0 = \pi]}{\delta} \geq 1 - \frac{1-\theta}{2} - \hat{v}(\pi)(2\lambda p - (1-\theta)\lambda)$$
$$= \left(\frac{1}{2} + \lambda\hat{v}(\pi)\right) - p(2\lambda\hat{v}(\pi)) + \theta\left(\frac{1}{2} - \lambda\hat{v}(\pi)\right).$$

Since $\lambda < \lambda^\star$, $\hat{v}(\pi) \leq \frac{1}{2\lambda}$, the last term is non-negative for all $\theta \in [-1,1]$ and $p \in [0,1]$, and the lower bound is minimised when $p = 1$ and $\theta = -1$. Finally, when $p = 1$ and $\theta = -1$, it is equal to 0, as required. So, the reflection coupling is the unique mean-optimal coupling strategy.

This completes the proof of Theorem 3.8*(i)*.

- Let $\lambda > \lambda^\star$. Observe that

$$\begin{aligned} C(\lambda) + \pi - 2C(\lambda)\cosh(\sqrt{\lambda}\pi) &= \frac{\operatorname{cosech}(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}} + \pi - 2\cosh(\sqrt{\lambda}\pi)\frac{\operatorname{cosech}(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}} \\ &= \frac{1}{2\sqrt{\lambda}\sinh(\sqrt{\lambda}\pi)} + \pi - \frac{2\cosh(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}\sinh(\sqrt{\lambda}\pi)} \\ &= \frac{1 + 2\pi\sqrt{\lambda}\sinh(\sqrt{\lambda}\pi) - 2\cosh(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}\sinh(\sqrt{\lambda}\pi)} \\ &> \frac{1 + 2\cosh(\sqrt{\lambda}\pi) - 1 - 2\cosh(\sqrt{\lambda}\pi)}{2\sqrt{\lambda}\sinh(\sqrt{\lambda}\pi)}, \end{aligned}$$

where the last inequality comes from the assumption $\lambda > \lambda^\star$, so

$$\frac{\pi\sinh(\sqrt{\lambda}\pi)}{\sqrt{\lambda}(2\cosh(\sqrt{\lambda}\pi) - 1)} > \frac{1}{2\lambda},$$

which is equivalent to

$$2\pi\sqrt{\lambda}\sinh(\sqrt{\lambda}\pi) > 2\cosh(\sqrt{\lambda}\pi) - 1.$$

So we have that $C(\lambda) + \pi - 2C(\lambda)\cosh(\sqrt{\lambda}\pi) \geq 0$ for all $\theta \in [-1,1]$ and $p \in [0,1]$, then

$$\frac{(1-\theta)\delta}{\pi}(C(\lambda) + \pi - 2C(\lambda)\cosh(\sqrt{\lambda}\pi)),$$

is non-negative for all $\theta \in [-1,1]$ and $p \in [0,1]$ and it is equal to 0 if and only if $\theta = 1$, i.e. if and only if X and X' are synchronised. So, if we let $\theta = 1$, we have that the third term of (3.40) cancels, and, since $\mathrm{Erfc}(\delta, \theta, \pi, 0) \to_{\theta \to 1} 0$, also the first term disappears. So,

$$\begin{aligned}\mathbb{E}[\hat{v}(D^c_\delta)] - \hat{v}(\pi) &\geq -\hat{v}(\pi)[1 - e^{-2\lambda p \delta}] + O(\delta^{\frac{3}{2}}) \\ &= -\frac{1}{2\lambda}[1 - e^{-2\lambda p \delta}] + O(\delta^{\frac{3}{2}}).\end{aligned}$$

Thus, from (3.39),

$$\mathbb{E}[V^c_\pi(\delta) - V^c_\pi(0) | D_0 = \pi] \geq \delta - \frac{1}{2\lambda}[1 - e^{-2\lambda p \delta}] + O(\delta^{\frac{3}{2}}).$$

Dividing by δ and taking the limit as $\delta \to 0$ yield

$$\frac{\mathbb{E}[V^c_\pi(\delta) - V^c_\pi(0) | D_0 = \pi]}{\delta} \geq 1 - \frac{1}{2\lambda} \frac{1 - e^{-2\lambda p \delta}}{\delta} + O(\delta^{\frac{1}{2}}) \xrightarrow[\delta \to 0]{} 1 - \frac{1}{2\lambda} 2\lambda p = 1 - p.$$

We know that $1 - p \geq 0$ for all $p \in [0, 1]$, while $1 - p = 0$ if and only if $p = 1$, as required. So, the synchronised coupling is the unique mean-optimal coupling strategy.

This completes the proof of Theorem 3.8*(ii)*.

3.10 Further thoughts

In this chapter, we showed for any value of $\lambda \neq \lambda^\star$ there exists a unique mean-optimal co-adapted coupling: when $\lambda < \lambda^\star$, this is the reflection coupling of Definition 3.6, and when $\lambda > \lambda^\star$, this is the synchronised coupling of Definition 3.5. In general, a co-adapted coupling can be maximal; an example of that is given in Chapter 1, where we observed that the reflection coupling is maximal and co-adapted for the one-dimensional Brownian motion. To decide whether our mean-optimal coupling is maximal, we can compare it to a maximal coupling.

First, we find the Laplace transform of the maximal coupling time $\tau^\star$ when the two jumpy Brownian motions start at distance π.

$$\mathbb{E}[e^{-\gamma\tau^\star}] = \int_0^\infty e^{-\gamma t}\mathbb{P}(\tau^\star \in dt) = \int_0^\infty e^{-\gamma t}\left(-\frac{d}{dt}\mathbb{P}(\tau^\star > t)\right)dt.$$

As we have seen in Section 3.3,

$$\mathbb{P}(\tau^\star > t) = \mathbb{P}(\tau > t)\mathbb{P}(J > t) = \mathbb{P}(\tau > t)e^{-2\lambda t},$$

so, where $J \sim \text{Exp}(2\lambda)$ is the time of the first jump and τ is the first hitting time of a standard Brownian motion at level $\pm\pi$.

$$\begin{aligned}
\mathbb{E}[e^{-\gamma\tau^\star}] &= \int_0^\infty e^{-\gamma t}\left(2\lambda e^{-2\lambda t}\mathbb{P}(\tau > t) - e^{-2\lambda t}\frac{d}{dt}\mathbb{P}(\tau > t)\right)dt \\
&= \frac{2\lambda}{2\lambda+\gamma}\int_0^\infty (2\lambda+\gamma)e^{-(2\lambda+\gamma)t}\mathbb{P}(\tau > t)dt + \int_0^\infty e^{-(2\lambda+\gamma)t}\mathbb{P}(\tau \in dt) \\
&= \frac{2\lambda}{2\lambda+\gamma}\mathbb{P}(\tau > \text{Exp}(2\lambda+\gamma)) + \mathbb{E}[e^{-(2\lambda+\gamma)\tau}],
\end{aligned}$$

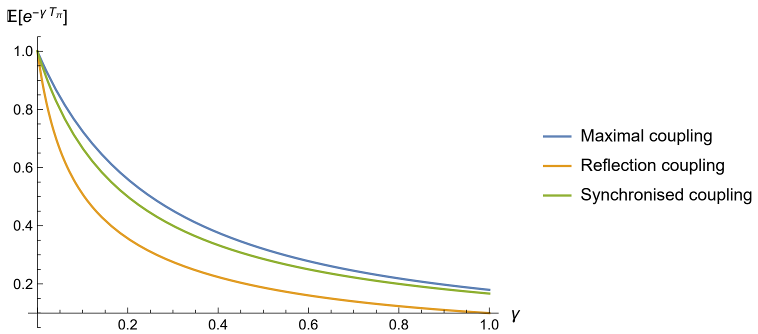
where the probability in the first term is given by Lemma 3.7, and the second term is the Laplace transform of the hitting time of a standard Brownian motion at level $\pm\pi$. Thus,

$$\mathbb{E}[e^{-\gamma\tau^\star}] = \frac{2\lambda}{2\lambda+\gamma} + \frac{\gamma}{2\lambda+\gamma}\frac{1}{\cosh(\pi\sqrt{2(2\lambda+\gamma)})}.$$

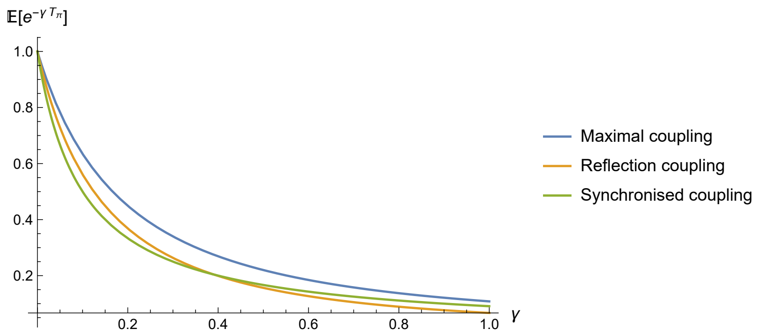
To calculate the expectation of $\tau^\star$, we differentiate the Laplace transform with respect to γ and we take the limit as $\gamma \to 0$:

$$\mathbb{E}[\tau^\star] = \frac{1 - \operatorname{sech}(2\pi\sqrt{\lambda})}{2\lambda}.$$

Figures 3.16 and 3.15 show a comparison between the expectation and Laplace transform of the coupling time under our mean-optimal strategies and under the maximal coupling defined in Section 3.3 for two jumpy Brownian motions started at distance π. From the graphs of the expectation and Laplace transforms, we can conclude that these coupling times do not have the same distribution and that the mean-optimal co-adapted coupling is not maximal.



(a) Laplace transform of the coupling time with $\lambda = 0.2 > \lambda^\star$ for two copies of the jumpy Brownian motion started at distance π under the reflection and synchronised couplings and under a maximal coupling.



(b) Laplace transform of the coupling time with $\lambda = 0.05 < \lambda^\star$ for two copies of the jumpy Brownian motion started at distance π under the reflection and synchronised couplings and under a maximal coupling.

Fig. 3.15: Laplace transform of the coupling time for two copies of the jumpy Brownian motion started at distance π under the reflection and synchronised couplings and under a maximal coupling.

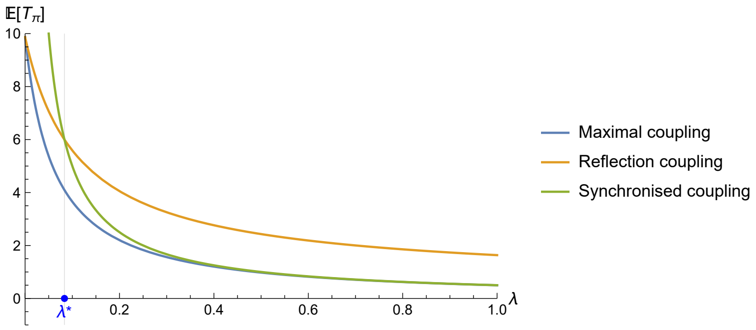


Fig. 3.16: Expectation of the coupling time for two copies of the jumpy Brownian motion started at distance π under the reflection and synchronised couplings and under the maximal coupling.

An interesting extension of the study presented in this chapter is the possibility for the Brownian motion to jump at rate λ to any point on the circumference. As an easier case, we can start with the Brownian motion that at rate λ jumps from the current point x to $x \pm \frac{2\pi}{3} \pmod{2\pi}$. Figure 3.17 shows a simulation of this jumpy Brownian motion.

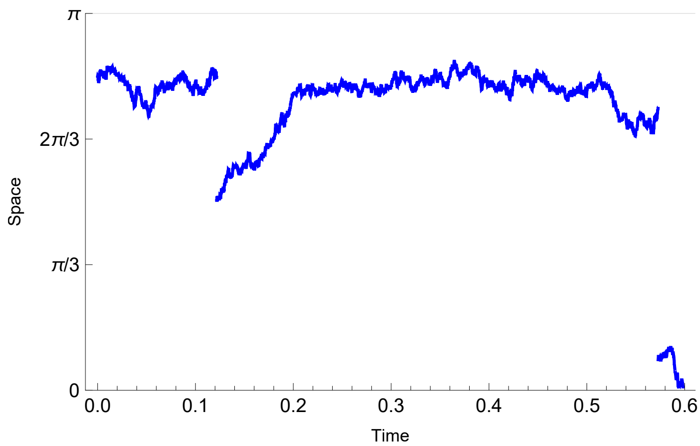


Fig. 3.17: Simulation of the jumpy Brownian motion started at $\frac{5\pi}{6}$ that jumps from x to $x \pm \frac{2\pi}{3} \pmod{2\pi}$ with $\lambda = 0.3$.

As we did before, we can consider the process D_t as the distance on the circumference between two copies X and X' of the jumpy Brownian motion, so that D diffuses in the interval $[0, \pi]$ with potential jumps of rate 2λ. We can split the interval $[0, \pi]$ in which D diffuses into three intervals, $\left[0, \frac{\pi}{3}\right]$, $\left(\frac{\pi}{3}, \frac{2\pi}{3}\right)$, and $\left[\frac{2\pi}{3}, \pi\right]$. As we did for our original jumpy Brownian motion problem, it is intuitively reasonable to reflect the Brownian motion components of X and X', so we start with a reflection coupling. We then need to determine a sensible strategy to couple the jump components in order to minimise the expected coupling time. When we see a jump we have different ways of coupling X and X': we could let only the corresponding Brownian motion jump and keep the other one still; or we could synchronise the jumps and let the two processes jump at the same time, and we could also choose of letting them jump in the same direction or in the opposite. Analysing numerically some of the possible couplings, we think that the following rules might be needed to construct an optimal coupling:

1. If $D_{t-} \in \left[0, \frac{\pi}{3}\right]$, we synchronise both the jump times and directions, so that X and X' jump synchronously and $D_t = D_{t-}$ is unchanged.

2. If $D_{t-} \in \left[\frac{2\pi}{3}, \pi\right]$, we let the processes jump independently. To decide the direction of the jump, we toss a coin and if we get heads, the process will jump anticlockwise, otherwise it will jump clockwise.

It is still not clear to us what the most efficient strategy is when $D_{t-} \in \left(\frac{\pi}{3}, \frac{2\pi}{3}\right)$. At the moment, we considered the following strategy: we synchronise the jumps to leave the distance on the circumference unchanged if $D_{t-} \leq x^\star$, and we let the processes jump independently otherwise. The question then becomes making a sensible choice for $x^\star$.

Figure 3.18 shows a simulation of the expectation of the coupling time τ_{couple} under the strategy described above for $x^\star = \frac{\pi}{3}$ and $x^\star = \frac{\pi}{2}$. Based on the simulation, it looks like setting $x^\star = \frac{\pi}{3}$ would be more efficient. However, this is a numerical evaluation under only one of the possible couplings and using only two of the possible values of $x^\star$, and other co-adapted couplings might reveal to be even more efficient. Also, as we have seen above, the optimal coupling strategy, assuming one exists, will depend upon λ, and so further investigation would be required before we are able to arrive at a conjectured optimal coupling under these altered dynamics.

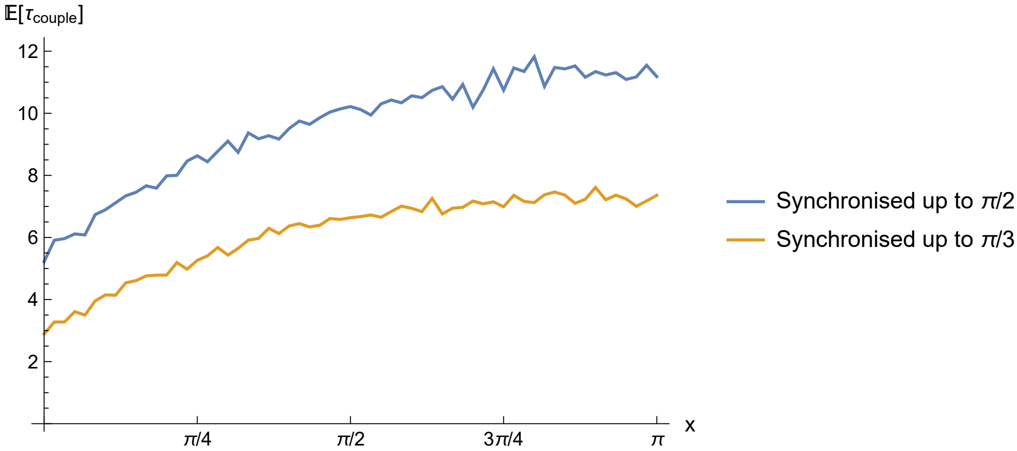


Fig. 3.18: Simulation of the expectation of the coupling time with $\lambda = 0.2$ for $x^\star = \frac{\pi}{2}$ and $x^\star = \frac{\pi}{3}$.

4. CONCLUSION

The coupling method has played an essential role in this thesis. Coupling has a strong connection with the total variation distance and can be an effective technique to help study the convergence of random walks and, possibly, establish cutoff.

In chapter 2, we generalised Nestoridi's paper on the mixing time for the random walk on $\mathbb{Z}_2^n$ that at each step flips k randomly chosen coordinates. We let the number of coordinates flipped at each step be a random variable K, and we established bounds for the mixing time of this random walk. To find an upper bound, we employed a similar path coupling argument of Nestoridi, correcting her result and establishing an upper bound for the mixing time in terms of $\mathbb{E}[K(n-K)]$. The upper bound we found is double what we think the correct mixing time is, but we suggested another method that might give a tighter bound. Then, using representation theory and Wilson's lemma, we found a lower bound that in general does not match the upper bound. However, we showed that in some cases it can be improved, and we proved that, under the assumption $\mathbb{E}[K] = o(n)$, the random walk exhibits a pre-cutoff.

In Chapter 3, we focused on another aspect of coupling, namely mean-optimality. The process under examination was the Brownian motion on the circumference of the unit circle that, at times given by an independent Poisson process of rate λ, jumps to the opposite point on the circle. After exploring some distributional properties of this process, we constructed two co-adapted couplings of two jumpy Brownian motions X and X' using a third process D defined as the distance on the circle between X and X'. These two couplings, that we called the reflection coupling and the synchronised coupling, differ only in the strategy adopted when D hits π. Since we were interested in studying the optimality of these couplings, we needed to know more about the distribution of the coupling time. The presence of the jumps in the diffusion of the jumpy Brownian motion suggested that we should apply excursion theory to find an explicit expression of the Laplace transform of the coupling time, which we then used to derive its expectation. As we saw in Section 3.8, the expected coupling time under the two strategies differs only by a function $C(\lambda)$, and there exists a unique point $\lambda^\star$ at which $C(\lambda)$ assumes the same value in the two couplings. Then, using Bellman's principle of optimality, we showed that for any $\lambda < \lambda^\star$ the reflection coupling is the unique mean-optimal co-adapted coupling, while if $\lambda > \lambda^\star$ the synchronised coupling is the unique mean-optimal co-adapted coupling. Finally, we compared the reflection and synchronised couplings to a maximal coupling. Using the exact formulas of the Laplace transform and expectation of the maximal coupling time, we graphically showed that the distribution of the coupling times under the mean-optimal co-adapted coupling differs from the distribution of the maximal coupling time, therefore the reflection and synchronised couplings are mean-optimal but not maximal.

APPENDIX

A. EXPECTATION AND VARIANCE FROM SECTION 2.2.4

In Section 2.2.4, we used the eigenfunction $f_1(\mathbf{x}) = \sqrt{n}\left(1 - \frac{2|\mathbf{x}|}{n}\right)$ to find tighter lower bounds for the mixing time of the random walk on the hypercube than the lower bound found in Section 2.2.3. In *Case 1*, we assumed $\mathbb{E}[K] = O(n^{\varepsilon})$ with $\varepsilon \in (0,1)$, and we claimed that if $\ell = \frac{n^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c)$ with $c < \frac{1}{2}\log n$ and $\frac{n\mathbb{E}[K]}{\mathbb{E}[K(n-K)]} \sim 1$, then

$$\mathbb{E}[f_1(Z_\ell)] \sim e^c\left(1 + \frac{(\mathbb{E}[K])^2}{4\mathbb{E}[K(n-K)]}(\log n - 2c)\right) \geq e^c,$$

and

$$\mathrm{Var}(f_1(Z_\ell)) \sim 1 + e^{2c}\left(\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c) - \frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c)\right).$$

Using the eigenvalue corresponding to f_1, we obtain

$$\begin{aligned}
&\mathbb{E}[f_1(Z_\ell)] = \sqrt{n}\left(1 - \frac{\mathbb{E}[K]}{n}\right)^\ell = \sqrt{n}e^{\ell\log\left(1 - \frac{\mathbb{E}[K]}{n}\right)} \\
&= \sqrt{n}\exp\left\{-\frac{n\mathbb{E}[K]}{2\mathbb{E}[K(n-K)]}(\log n - 2c) + \frac{(\mathbb{E}[K])^2}{4\mathbb{E}[K(n-K)]}(\log n - 2c)\right. \\
&\qquad\qquad \left. + o\left(\frac{(\mathbb{E}[K])^2}{\mathbb{E}[K(n-K)]}\log n\right)\right\} \\
&\sim \exp\left\{c + \frac{(\mathbb{E}[K])^2}{4\mathbb{E}[K(n-K)]}(\log n - 2c) + o\left(\frac{(\mathbb{E}[K])^2}{\mathbb{E}[K(n-K)]}\log n\right)\right\}.
\end{aligned}$$

The second line follows from the Taylor expansion

$$\log\left(1 - \frac{\mathbb{E}[K]}{n}\right) = -\frac{\mathbb{E}[K]}{n} + \frac{(\mathbb{E}[K])^2}{2n^2} + o\left(\frac{(\mathbb{E}[K])^2}{n^2}\right).$$

Also, from $\frac{n\mathbb{E}[K]}{\mathbb{E}[K(n-K)]} \sim 1$, it follows

$$\exp\left\{-\frac{n\mathbb{E}[K]}{2\mathbb{E}[K(n-K)]}(\log n - 2c)\right\} \sim e^{-\frac{1}{2}(\log n - 2c)}.$$

Since
$$\frac{(\mathbb{E}[K])^2}{4\mathbb{E}[K(n-K)]}(\log n - 2c) = O(n^{\varepsilon-1}\log n) \to 0,$$
the expectation
$$\mathbb{E}[f_1(Z_\ell)] \sim e^c\left(1 + \frac{(\mathbb{E}[K])^2}{4\mathbb{E}[K(n-K)]}(\log n - 2c)\right) \geq e^c.$$

For the variance, we apply the Taylor expansion
$$\log\left(1 - \frac{2\mathbb{E}[K(n-K)]}{n(n-1)}\right) = -\frac{2\mathbb{E}[K(n-K)]}{n(n-1)} + \frac{2(\mathbb{E}[K(n-K)])^2}{n^2(n-1)^2} + o\left(\frac{(\mathbb{E}[K(n-K)])^2}{n^2(n-1)^2}\right).$$

We can observe that
$$n\left(1 - \frac{\mathbb{E}[K]}{n}\right)^{2\ell} = \exp\left\{2c + \frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c) + o\left(\frac{(\mathbb{E}[K])^2}{\mathbb{E}[K(n-K)]}\log n\right)\right\}$$
and that, from $\frac{n}{n-1} \sim 1$, it follows
$$\exp\left\{-\frac{n}{n-1}(\log n - 2c)\right\} \sim e^{-(\log n - 2c)}.$$

Then,
$$\begin{aligned}
&(n-1)\left[1 - \frac{2\mathbb{E}[K(n-K)]}{n(n-1)}\right]^\ell \\
&= (n-1)\exp\left\{\ell \log\left(1 - \frac{2\mathbb{E}[K(n-K)]}{n(n-1)}\right)\right\} \\
&= (n-1)\exp\left\{-\frac{n}{n-1}(\log n - 2c) + \frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c)\right. \\
&\qquad\qquad \left. + o\left(\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}\log n\right)\right\} \\
&\sim \exp\left\{2c + \frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c) + o\left(\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}\log n\right)\right\}.
\end{aligned}$$

Then,
$$\begin{aligned}
\mathrm{Var}(f_1(Z_\ell)) \sim 1 + e^{2c}&\left[\exp\left\{\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c) + o\left(\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}\log n\right)\right\}\right. \\
&\left. - \exp\left\{\frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c) + o\left(\frac{(\mathbb{E}[K])^2}{\mathbb{E}[K(n-K)]}\log n\right)\right\}\right].
\end{aligned}$$

We have that

$$\frac{\mathbb{E}[K(n-K)]}{(n-1)^2} = O(n^{\varepsilon-1}) \to 0 \qquad \text{and} \qquad \frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]} = O(n^{\varepsilon-1}) \to 0$$

. Then,

$$\begin{aligned}
&\exp\left\{\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c)\right\} - \exp\left\{\frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c)\right\} \\
&\quad \sim \frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c) - \frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c) \\
&\quad = O(n^{\varepsilon-1}\log n) \to 0,
\end{aligned}$$

and

$$\mathrm{Var}(f_1(Z_\ell)) \sim 1+e^{2c}\left(\frac{\mathbb{E}[K(n-K)]}{(n-1)^2}(\log n - 2c) - \frac{(\mathbb{E}[K])^2}{2\mathbb{E}[K(n-K)]}(\log n - 2c)\right).$$

B. FORMULAS OF THE LAPLACE TRANSFORM FROM SECTION 3.6.3

We can give an explicit expression for (3.17) for any $x \in (0, \pi)$. Since the formulas depend on the coupling, we denote by T_x^r and T_x^s the coupling time for, respectively, the reflection coupling and the synchronised coupling with $D_0 = x$.

1. If $x \in (0, \frac{\pi}{2})$,

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_x}] &= \mathbb{E}[e^{-\gamma T_x(F)}] + \mathbb{E}[e^{-\gamma T_x(F)} \mathbb{1}_{\{D(T_x(F))=\frac{\pi}{2}\}}](\mathbb{E}[e^{-\gamma T_{\frac{\pi}{2}}}] - 1) \\ &= \frac{\operatorname{cosech}((\frac{\pi}{2} - x)\beta) + \operatorname{cosech}(x\beta)}{\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)} + \frac{\operatorname{cosech}((\frac{\pi}{2} - x)\beta)}{\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)} \cdot \\ &\quad \cdot \left[\frac{1}{\cosh(\frac{\pi}{2}\beta)} + \frac{\alpha}{2\beta} \tanh\left(\frac{\pi}{2}\beta\right) \cdot \operatorname{cosech}\left(\frac{\pi}{2}\alpha\right) \left(\mathbb{E}\left[e^{-\gamma T_\pi}\right] - 1\right) - 1 \right],\end{aligned}$$

where the expression for $\mathbb{E}\left[e^{-\gamma T_\pi}\right]$ depends on the strategy adopted when starting from π.

A. Reflected process.

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_x^r}] = &\frac{\operatorname{cosech}(x\beta)[2 - \operatorname{sech}(\frac{\pi}{2}\alpha)\operatorname{sech}(\frac{\pi}{2}\beta)]}{[2 - \operatorname{sech}(\frac{\pi}{2}\alpha)\operatorname{sech}(\frac{\pi}{2}\beta)][\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)]} \\ &- \frac{\operatorname{cosech}((\frac{\pi}{2} - x)\beta)[-2\operatorname{sech}(\frac{\pi}{2}\beta) + \operatorname{sech}(\frac{\pi}{2}\alpha)]}{[2 - \operatorname{sech}(\frac{\pi}{2}\alpha)\operatorname{sech}(\frac{\pi}{2}\beta)][\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)]}\end{aligned}$$

B. Synchronised process.

$$\begin{aligned}\mathbb{E}[e^{-\gamma T_x^s}] = &\frac{\operatorname{cosech}(x\beta) + \operatorname{cosech}((\frac{\pi}{2} - x)\beta)\operatorname{sech}(\frac{\pi}{2}\beta)}{\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)} \\ &- \frac{\beta}{2\alpha} \frac{\operatorname{cosech}((\frac{\pi}{2} - x)\beta)\tanh(\frac{\pi}{2}\beta)\operatorname{cosech}(\frac{\pi}{2}\alpha)}{\coth((\frac{\pi}{2} - x)\beta) + \coth(x\beta)}\end{aligned}$$

2. If $x \in (\frac{\pi}{2}, \pi)$,

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_x}] &= \mathbb{E}[e^{-\gamma T_x(F)}] + \mathbb{E}[e^{-\gamma T_x(F)} \mathbb{1}_{\{D(T_x(F))=\frac{\pi}{2}\}}](\mathbb{E}[e^{-\gamma T_{\frac{\pi}{2}}}] - 1) \\
&\quad + \mathbb{E}[e^{-\gamma T_x(F)} \mathbb{1}_{\{D(T_x(F))=\pi\}}](\mathbb{E}[e^{-\gamma T_\pi}] - 1) \\
&= \frac{\operatorname{cosech}((\pi - x)\beta) + \operatorname{cosech}((x - \frac{\pi}{2})\beta)}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} + \frac{\operatorname{cosech}((\pi - x)\beta)}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} \cdot \\
&\quad \cdot \left[\frac{1}{\cosh(\frac{\pi}{2}\beta)} + \frac{\alpha}{2\beta} \tanh\left(\frac{\pi}{2}\beta\right) \cdot \operatorname{cosech}\left(\frac{\pi}{2}\alpha\right) \left(\mathbb{E}\left[e^{-\gamma T_\pi}\right] - 1\right) - 1\right] \\
&\quad + \frac{\operatorname{cosech}((\pi - x)\alpha)}{\coth((\pi - x)\alpha) + \coth((x - \frac{\pi}{2})\alpha)} (\mathbb{E}[e^{-\gamma T_\pi}] - 1),
\end{aligned}$$

where the expression for $\mathbb{E}\left[e^{-\gamma T_\pi}\right]$ depends on the strategy adopted when starting from π.

A. Reflected process.

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_x^r}] &= \frac{\operatorname{cosech}((\pi - x)\beta) + \operatorname{cosech}((x - \frac{\pi}{2})\beta) \operatorname{sech}(\frac{\pi}{2}\beta)}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} \\
&\quad - \frac{\operatorname{cosech}((x - \frac{\pi}{2})\beta) \operatorname{sech}(\frac{\pi}{2}\alpha) \tanh^2(\frac{\pi}{2}\beta)}{[2 - \operatorname{sech}(\frac{\pi}{2}\alpha) \operatorname{sech}(\frac{\pi}{2}\beta)][\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)]} \\
&\quad - \frac{\frac{2\beta}{\alpha} \operatorname{cosech}((\pi - x)\alpha) \tanh(\frac{\pi}{2}\alpha) \tanh(\frac{\pi}{2}\beta)}{[2 - \operatorname{sech}(\frac{\pi}{2}\alpha) \operatorname{sech}(\frac{\pi}{2}\beta)][\coth((\pi - x)\alpha) + \coth((x - \frac{\pi}{2})\alpha)]}
\end{aligned}$$

B. Synchronised process.

$$\begin{aligned}
\mathbb{E}[e^{-\gamma T_x^s}] &= \frac{\operatorname{cosech}((\pi - x)\beta) + \operatorname{cosech}((x - \frac{\pi}{2})\beta) \operatorname{sech}(\frac{\pi}{2}\beta)}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} \\
&\quad - \frac{\beta}{2\alpha} \frac{\operatorname{cosech}((x - \frac{\pi}{2})\beta) \tanh(\frac{\pi}{2}\beta) \operatorname{cosech}(\frac{\pi}{2}\alpha)}{\coth((\pi - x)\beta) + \coth((x - \frac{\pi}{2})\beta)} \\
&\quad - \frac{\beta^2 \operatorname{cosech}((\pi - x)\alpha)}{\alpha^2[\coth((\pi - x)\alpha) + \coth((x - \frac{\pi}{2})\alpha)]}
\end{aligned}$$

BIBLIOGRAPHY

[1] David Aldous. Random walks on finite groups and rapidly mixing markov chains. In *Séminaire de Probabilités XVII 1981/82*, pages 243–297. Springer, 1983.

[2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.

[3] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, 8(1):69–97, 1987.

[4] Dave Bayer, Persi Diaconis, et al. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2):294–313, 1992.

[5] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 223–231. IEEE, 1997.

[6] Guan-Yu Chen, Laurent Saloff-Coste, et al. The cutoff phenomenon for ergodic markov processes. *Electronic Journal of Probability*, 13:26–78, 2008.

[7] Stephen Connor and Saul Jacka. Optimal co-adapted coupling for the symmetric random walk on the hypercube. *Journal of applied probability*, 45(3):703–713, 2008.

[8] Persi Diaconis. Group representations in probability and statistics. *Lecture notes-monograph series*, 11:i–192, 1988.

[9] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(2):159–179, 1981.

[10] R Durrett. Probability: Theory and examples. cambridge university press, 2010.

[11] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[12] Alexandros Eskenazis and Evita Nestoridi. Cutoff for the bernoulli–laplace urn model with $o(n)$ swaps. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 2621–2639. Institut Henri Poincaré, 2020.

[13] W Feller. Introduction to probability theory third edition, vol. 1, 1968.

[14] Sheldon Goldstein. Maximal coupling. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 46(2):193–204, 1979.

[15] David Griffeath. A maximal coupling for markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31(2):95–106, 1975.

[16] Elton P Hsu and Karl-Theodor Sturm. Maximal coupling of euclidean brownian motions. *Communications in Mathematics and Statistics*, 1(1):93–104, 2013.

[17] Kiyosi Itô. Poisson point processes attached to markov processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, volume 3, pages 225–239, 1972.

[18] Wilfrid S Kendall et al. Coupling, local times, immersions. *Bernoulli*, 21(2):1014–1046, 2015.

[19] Nikolaj Vladimirovič Krylov. *Controlled diffusion processes*, volume 14. Springer Science & Business Media, 2008.

[20] Kazumasa Kuwada. On uniqueness of maximal coupling for diffusion processes with a reflection. *Journal of Theoretical Probability*, 20(4):935–957, 2007.

[21] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times.* American Mathematical Soc., 2009.

[22] Keunwoo Lim. Cutoff phenomenon for cyclic dynamics on hypercube. *arXiv preprint arXiv:2010.01756*, 2020.

[23] Torgny Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002.

[24] Torgny Lindvall, L Cris G Rogers, et al. Coupling of multidimensional diffusions by reflection. *The Annals of Probability*, 14(3):860–872, 1986.

[25] Eyal Lubetzky, Allan Sly, et al. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010.

[26] Peter Matthews. Mixing rates for a random walk on the cube. *SIAM Journal on Algebraic Discrete Methods*, 8(4):746–752, 1987.

[27] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[28] Evita Nestoridi. A non-local random walk on the hypercube. *Advances in Applied Probability*, 49(4):1288–1299, 2017.

[29] JW Pitman. On coupling of markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35(4):315–322, 1976.

[30] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[31] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales. Volume 2: Itô Calculus*, volume 2. Cambridge university press, 2000.

[32] Justin Salez. Cutoff for non-negatively curved markov chains. *arXiv preprint arXiv:2102.05597*, 2021.

[33] David Bruce Wilson et al. Mixing times of lozenge tiling and card shuffling markov chains. *The Annals of Applied Probability*, 14(1):274–325, 2004.