Unbiased Shifts of Stochastic Processes

Redl, Istvan

Award date:
2016

Awarding institution:
University of Bath

Unbiased Shifts of Stochastic Processes

Department of Mathematical Sciences, University of Bath

A thesis submitted for the degree of Doctor of Philosophy

Istvan Redl

May 2016

Abstract

Unbiased shifts of stochastic processes can be regarded as the general theoretical framework for the fundamental *extra head problem*. In the first part of the thesis, we aim to give an overview of the rich mathematical landscape related to the main theme. This exploration is made gradually from relatively simple concepts to more complex ones. The second and third part contain original results with a focus on optimal unbiased shifts. Though some insight from the second part proved to be useful in the third part, new ideas and some involved technicalities were also necessary to establish the main result of the third part.

Anyunak, Karcsinak es örök barátomnak, Sáfrán Péternek
To my Mum, my brother, Karcsi and my dearest friend, Péter Sáfrán

Acknowledgements

I am genuinely grateful to my supervisor prof Peter Mörters for introducing me to the problem that became the backbone of this work, for the illuminating mathematical discussions and for his generous and unconditional support. I feel privileged to have had him as a supervisor and have had the opportunity to work with him.

At this point, I would also like to thank all the other brilliant mathematicians I have met during my studies, who inspired me; among them two names stand out: Dr Bálint Vető and prof Bálint Tóth.

I would also like to thank the Prob L@b Group and my fellow PhD students in Bath for providing a stimulating environment during my research.

Last, but not least I am thankful to my brother Karcsi and my family and friends beyond measure, for their unconditional love and for standing by me through thick and thin.

Contents

Chapter 1

Introduction - Exploring the fundamental problem and related areas

The concept of *unbiased shift* was first introduced in the Brownian motion context, in [23], and it has essentially three main characteristics. The underlying stochastic process is *doubly infinite* or *two sided*, that is, indexed by $\mathbb{Z}$ or $\mathbb{R}$ in the discrete or continuous time setting, respectively. The second key ingredient is some kind of *invariance* under time shift. This means that in the Markov process case the transition probabilities are preserved after the shift. Thirdly, initial distributions at the origins - before and after shift - are imposed. In general, the canonical problem can be formulated as follows. *Given initial distributions at the origin, how do we shift a doubly-infinite Markov process in such a way that the (doubly-infinite) shifted process has the same transition probabilities?* This - admittedly - informal description will be made precise and illustrated in various examples later on.

There are some typical questions attached to this problem, which people are interested in.

- *Existence*: Do such shifts always exist? If the answer is positive, how do they look like? How can we characterize them?
- *Moment properties*: As these shifts are random variables, it is natural to investigate their moment properties.
- *Optimality*: What is the optimal tail behaviour of these shifts? Which shifts are better on average, than others? In other words, optimality of shifts can be studied also in expected value terms.

These questions will be addressed in detail in this thesis.

A remarkable feature of this problem - as the key characteristics may suggest - is that it links together many different areas of mathematics, which are well developed and recently gained significant attention. *Couplings, allocation and optimal transport problems, (Skorokhod) embeddings*; to name the main ones.

In this introductory chapter, our aim will be to demonstrate these aspects throughout examples and to explore some first details in different kind of related problems. The results in this chapter will be given mostly without proofs. If there is an exception to this, we present the proof mainly because we think it contains interesting steps or notable arguments.

As chapter 2 and 3 contain all the new results in a concise way, we omit listing them separately here.

1.1 Fundamental examples

1.1.1 The extra head problem

Consider the following problem: Given an i.i.d. sequence of coin tosses indexed by $\mathbb{Z}$, find a random coin that shows head, such that the sequences to its left and its right remain i.i.d. Assume the coins X_i are such that for some $0 < p < 1$, $\mathbb{P}(X_i = \texttt{head}) = p = 1 - \mathbb{P}(X_i = \texttt{tail})$, for all $i \in \mathbb{Z}$. This problem was first introduced by Liggett [25], who also provided an explicit solution for certain values of p, and subsequently studied in [18], and [20].

To see why this problem might not be as innocent as it seems, let us check the following attempt at a solution. Wait until we first see a head. In other words, reveal the coin at the origin, if it shows head, then it is the found coin; if the coin at the origin shows tail, then check the coin to its right and keep going, until the first coin showing head. Assume that the coins are fair, that is $p = 1/2$. It is a straightforward calculation that the probability that the coin to the left of the found head, is a tail, is $3/4$. Indeed, the coin at the origin shows head with probability a half, in which case the coin to its left shows tail with probability a half. If the coin at the origin is a tail, then we know for certain that the coin to the left of the found head will show a tail. Therefore this cannot be a solution, as not just the fairness of a coin has been distorted, but also independence property of the entire sequence is lost.

1.1.2 Extra head for Poisson process

The Poisson counterpart of the previous problem reads as follows. Let Π be a homogeneous Poisson process on $\mathbb{R}^d$, with unit intensity. Find a site $X \in \mathbb{R}^d$ such that $\Pi(X) = 1$, which means that there is a point at site X, and the shifted process $\Pi(X + \cdot)$ is, again, a Poisson process on $\mathbb{R}^d$ with unit intensity. In other words, given a d-dimensional Poisson process, we are looking for a point, from where the process still looks like a Poisson process.

Let us demonstrate this problem's non-trivial nature in $d = 1$, through a tempting guess. Namely, take the first point X of Π to the right of the origin. It is readily apparent that the distance between the new origin X and the first point to its left will be a random variable that is distributed as the sum of two exponential random variables with unit intensity. Hence, this cannot be a solution, since this distance should be a unit exponential random variable.

What if we tweak this approach by choosing randomly the k-th point to the right of the origin? To this end, we introduce an integer valued random variable Z that tells us, which point to be the new origin X. Let us denote the distance between between the $k-1$-th and k-th point to the left of X by D_k, for $k = 1, 2 \ldots$, where D_1 is the distance between X and its first left neighbour point. By definition, D_k's are independent and conditioned on $Z = k$,

$$\mathbb{P}(D_i > t | Z = k) = \begin{cases} e^{-t} & \text{if} \quad i \neq k, \\ (1+t)e^{-t} & \text{if} \quad i = k, \end{cases}$$

where the $i = k$ case comes from the fact, we have already seen above, that is D_k is the sum of two independent, unit exponential random variables. This implies

$$\mathbb{P}(D_i > t) = e^{-t} + te^{-t}\mathbb{P}(Z = k).$$

Therefore it is impossible to guarantee that $D_i \sim Exp(1)$ for all $i = 1, 2, \ldots$.

1.2 Shift coupling

The first solution to the two problems described above was given by the powerful machinery of *shift (or transformation) coupling*, see [31] and chapter 7 in [32]. Following the latter reference we give an overview of this tool. Though there exists an established theory for shift couplings in 'one-sided' context, see e.g. chapter 5 in [32], we only present here the 'two-sided' case, relevant for us.

Informally, coupling is a simultaneous construction of two (or more) stochastic processes on the same probability space, with the aim to study certain properties of a stochastic process, e.g. convergence to invariant measure.

Definition 1. *Given an index set $\mathbb{I}$, for all $i \in \mathbb{I}$, let X_i,be a random variable, taking values in a measurable space $(E_i, \mathcal{E}_i)$, and defined on probability space $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$. A family of random variables $(\hat{X}_i : i \in \mathbb{I})$ defined on a common probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ is a coupling of X_i, $i \in \mathbb{I}$, if*

$$\hat{X}_i \stackrel{D}{=} X_i,$$

for all $i \in \mathbb{I}$, and where $\stackrel{D}{=}$ denotes equality in distribution.

Remark 1 The definition does not require X_i's to be defined on a common probability space, that is their joint distribution does not necessarily exist. The important point here is that the collection of copies $(\hat{X}_i : i \in \mathbb{I})$ is defined on a common probability, so they have a join distribution.

A first toy example is the independence coupling that consist of independent copies of X_i's. This already indicates that the dependence structure plays a crucial role, as only the marginal distributions need to match with those of X_i's.

We consider two stochastic processes $X = (X_t)_{t \in \mathbb{R}^d}$ and $X' = (X'_t)_{t \in \mathbb{R}^d}$, $d \geq 1$. An element of the index set $\mathbb{R}^d$ will be called *site*. In our setting these processes have a general state space and path or sample space, denoted by $(E, \mathcal{E})$ and $(\Omega, \mathcal{F})$, respectively. The notation $(E, \mathcal{E})$ is the usual one for a measurable space, where E is the state space and $\mathcal{E}$ is a suitable σ-algebra on E. A standard choice is Polish space, that is a complete and separable metric space, for E and right-continuous processes for the path space Ω. We note that usually the reason behind the choice of Polish space is that in this setting Kolmogorov extension theorem - a fundamental tool in probability theory - is readily available, see e.g. p. 86 in [32] or for a classic application, Theorem 12. 1.2 in [12] in the Brownian motion context.

Define the shift maps T^{-u}, $u \in \mathbb{R}^d$, by $T^{-u}\omega = (\omega_{u+t})_{t\in\mathbb{R}^d}$, where $\omega \in \Omega$. The processes X and X' are assumed to be shift-measurable. This means that $T^{-u}\Omega = \Omega$, for all $u \in \mathbb{R}^d$ and the map $(\omega, u) \in \Omega \times \mathcal{B}(\mathbb{R}^d) \mapsto T^{-u}\omega \in \Omega$ is $\mathcal{F} \times \mathcal{B}(\mathbb{R}^d)/\mathcal{F}$ measurable; where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra on $\mathbb{R}^d$.

Definition 2. *(Non-distributional)* $(\hat{X}, \hat{X}', U, U', C)$ *is a shift-coupling of* X *and* X'*, if* $(\hat{X}, \hat{X}')$ *is a coupling of* X *and* X'*,* U *and* U' *are random sites and* C *is an event such that*

$$T^{-U}\hat{X} = T^{-U'}\hat{X}' \qquad \text{on} \quad C.$$

(Distributional) $(\hat{X}, \hat{X}', U, U', C, C')$ *is a distributional shift-coupling of* X *and* X'*, if* $(\hat{X}, \hat{X}')$ *is a coupling of* X *and* X'*,* U *and* U' *are random sites, and* C *and* C' *are events such that*

$$\mathbb{P}(T^{-U}\hat{X} \in \cdot, C) = \mathbb{P}(T^{-U'}\hat{X}' \in \cdot, C'),$$

therefore $\mathbb{P}(C) = \mathbb{P}(C')$.

A shift-coupling is successful if $\mathbb{P}(C) = 1$*, in which case we omit the explicit reference to the events* C *and* C'*.*

One of the advantages of this setup is that the shift maps form a group, which is partly due to the choice of the index set $\mathbb{R}^d$. If the origin is shifted by some site $u \in \mathbb{R}^d$, then it can always be shifted back and so no information on the path is lost. This implies that distributional shift couplings can be turned into non-distributional ones, without the assumption of Polish space and right continuity of paths, see Theorem 3.2, in Chapter 7 in [32].

Another implication of the group property is that the above definitions can be rewritten. By introducing the *shift* $S := U - U'$, they become

$$T^{-S}\hat{X} = \hat{X}' \qquad \text{on} \quad C,$$

in the non-distributional case, and

$$\mathbb{P}(T^{-S}\hat{X} \in \cdot, C) = \mathbb{P}(\hat{X}' \in \cdot, C'),$$

in the distributional case. The notation reduces to $(\hat{X}, \hat{X}', S, C)$ and $(\hat{X}, \hat{X}', S, C, C')$. This gives a simple, but important insight, namely the two stochastic processes, shift coupled this way, are the same, except only that their origins are different.

The *shift-coupling inequality* is a standard result and can be seen as the counterpart of the usual coupling inequality, see section 5 of chapter 1 in [32], that has applications mainly in proofs of convergence of measures.

Recall that, given two probability measures μ and ν on some measurable space $(E, \mathcal{E})$, the total variation distance is given by

$$\|\mu - \nu\| := 2 \sup_{A\in\mathcal{E}} |\mu(A) - \nu(A)|.$$

Let λ denote the Lebesgue measure on $\mathcal{B}(\mathbb{R}^d)$ and for all $B \in \mathcal{B}(\mathbb{R}^d)$, with $0 < \lambda(B) < \infty$, define the uniform distribution $\lambda(\cdot|B)$ on B as

$$\lambda(A|B) := \frac{\lambda(A \cap B)}{\lambda(B)}, \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

The shift-coupling inequality reads as follows. For all $B \in \mathcal{B}(\mathbb{R}^d)$, such that $0 < \lambda(B) < \infty$,

$$\|\mathbb{P}(T^{-U_B}X \in \cdot) - \mathbb{P}(T^{-U_B}X' \in \cdot)\| \leq 2 - 2\mathbb{E}[\lambda(S + B|B); C], \tag{1.2.1}$$

where C is an event as in Definition 2, $U_B \sim \lambda(\cdot|B)$, i.e. U_B is uniform on B, and independent of X and X', and recall that S is the shift as introduced above.

This inequality can be extended to a general class of averaging sets. A family $B_h \in \mathcal{B}(\mathbb{R}^d)$, $0 < h < \infty$, is called *Folner averaging sets*, if

$$0 < \lambda(B_h) < \infty,$$

$$\lambda(t + B_h|B_h) \to 1 \text{ as } h \to \infty,\ t \in \mathbb{R}^d.$$

If $\mathbb{P}(C) = 1$, using (1.2.1) with $B = B_h$, the above conditions imply the total variation convergence

$$\|\mathbb{P}(T^{-U_{B_h}}X \in \cdot) - \mathbb{P}(T^{-U_{B_h}}X' \in \cdot)\| \to 0, \qquad \text{as} \quad h \to \infty. \tag{1.2.2}$$

Define the *invariant σ-algebra* by $\mathcal{I} = \{A \in \mathcal{F} : T^{-u}A = A, u \in \mathbb{R}^d\} = \{A \in \mathcal{F} : T^{+u}A = A, u \in \mathbb{R}^d\}$, where the second claim follows simply from the group property of the shift T^{-u}, $u \in \mathbb{R}$.

The main result on shift-coupling is due to Thorisson and was originally appeared in [31].

Theorem 1. *The following claims are equivalent:*

(a) There exists a successful distributional shift-coupling of X and X'.

(a') There exists a random site U such that $T^{-U}X \stackrel{D}{=} X'$.

(b) For some Folner averaging sets B_h, $0 < h < \infty$, (1.2.2) holds.

(b') For all Folner averaging sets B_h, $0 < h < \infty$, (1.2.2) holds.

(c) $\mathbb{P}(X \in \cdot|\mathcal{I}) = \mathbb{P}(X' \in \cdot|\mathcal{I})$.

Remark 2 When X' is stationary, i.e. $T^{-u}X' \stackrel{D}{=} X'$, for all $u \in \mathbb{R}^d$, then (1.2.2) becomes

$$T^{-U_{B_h}}X \to X',$$

in total variation, as $h \to \infty$. The equivalence of (c) and (b) implies that two stationary stochastic processes agree in distribution on $\mathcal{I}$ if and only if they are identically distributed.

This theorem is a collection of more general results and its proof can be found in section 6 and 7 in chapter 7 in [32] or indeed as Theorem 1 and 2 in [31]. In the general setting one can take a group G, which consists of measurable mappings from $(E, \mathcal{E})$ to itself. An element $\Gamma \in G$ acts on a random element X defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and ΓX is a random element in $(E, \mathcal{E})$. In our example the group was given by the group of the shift maps $(T^{-u} : u \in \mathbb{R}^d)$. Although this is not immediate to see, but this result can be applied to the fundamental examples and one can obtain existence results. The reason for applicability is that the index sets $\mathbb{Z}$ and $\mathbb{R}^d$ and also the state spaces we deal with are sufficiently nice; for further details see [31].

These exciting developments in the coupling literature provided the first answer to the problems presented in the previous subsection. Though they provided existence of solutions, it was unclear how these solutions would look like. The first explicit solution to the extra head problems was described by Liggett in [25], which was followed by a complete characterization of the solutions to these problems, see [20]. Before we discuss the aspects of this characterization in detail, we present a result concerning the moment properties of solutions. The reason we do this is twofold; the result itself may seem remarkable at first sight and it uses the shift-coupling inequality introduced above.

The following result is Theorem 3.1 in [25] and the proof below is based on the same reference.

Theorem 2. *Let $d = 1$, and Y be a solution to the extra head problem described in section 1.1.1 , i.e. Y is a random site in $\mathbb{Z}$, such that $X_Y =$ **head** almost surely and the shifted sequence $\{X_{Y+z}\}_{z\in\mathbb{Z}}$ is an i.i.d. sequence with parameter p. Then Y satisfies,*

$$\liminf_{t\to\infty} \frac{\mathbb{E}[|pY \wedge t|]}{\sqrt{t}} \geq \sqrt{\frac{1-p}{2\pi}},$$

in addition any solution Y to the (Poisson) extra head problem described in section 1.1.2 satisfies

$$\liminf_{t\to\infty} \frac{\mathbb{E}[|Y| \wedge t]}{\sqrt{t}} \geq \sqrt{\frac{1}{2\pi}}.$$

Furthermore, in both cases, $|Y|$ has infinite square root moment, i.e. $\mathbb{E}|Y|^{1/2} = \infty$.

Proof. As already indicated above, the core of the proof is the shift-coupling inequality. Consider the first part of the assertion. Recall that the shift map T^{-u}, $u \in \mathbb{Z}$ on $\{0,1\}^{\mathbb{Z}}$ is defined as $(T^{-u}X)(k) = X(u + k)$. The shift acts on measures as $(T^{-u}\mu)(A) = \mu(T^{+u}A)$. The total variation norm is defined by

$$\|\mu\| = \sup\left\{\int f d\mu : |f| \leq 1\right\}.$$

Let μ_1 and μ_2 be probability measures on $\{0,1\}^{\mathbb{Z}}$. Assume that X, a random element in $\{0,1\}^{\mathbb{Z}}$ is distributed according to μ_1 and the site $Y \in \mathbb{Z}$ is chosen such that $T^{-Y}X$ is distributed according to μ_2. With these choices, the shift-coupling inequality can be written as

$$\left\| \frac{1}{n}\sum_{k=1}^{n} T^{-k}\mu_1 - \frac{1}{n}\sum_{k=1}^{n} T^{-k}\mu_2 \right\| \le \frac{2}{n}\mathbb{E}[|Y| \wedge n], \tag{1.2.3}$$

for $n \ge 1$. In order to see this, take f with $|f| \le 1$ and write

$$\begin{aligned}\left| \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}f(T^{-k}X) - \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}f(T^{-(Y+k)}X) \right| &= \frac{1}{n}\left| \mathbb{E}\sum_{l} f(T^{-l}X)\left[\mathbb{1}_{[1,n]}(l) - \mathbb{1}_{[Y+1,Y+n]}(l)\right] \right| \\ &\le \frac{1}{n}\mathbb{E}\sum_{l}\left|\mathbb{1}_{[1,n]}(l) - \mathbb{1}_{[Y+1,Y+n]}(l)\right| \\ &= \frac{2}{n}\mathbb{E}[|Y| \wedge n].\end{aligned}$$

Taking the supremum over the set of f with $|f| \le 1$ yields inequality (1.2.3).

We now compute the left hand side of (1.2.3) in our case, that is μ_1 is the distribution of X or in other words the i.i.d. sequence $(X(i) : i \in \mathbb{Z})$ with parameter p, and μ_2 is the conditional distribution $\mu_1(\cdot|X(0) = 1)$. Since μ_1 is shift invariant and $T^{-k}\mu_2$ is absolutely continuous with respect to μ_1 for each k, the left hand side of (1.2.3) becomes

$$\int \left| 1 - \frac{1}{n}\sum_{k=1}^{n} \frac{d(T^{-k}\mu_2)}{d\mu_1} \right| d\mu_1.$$

Observe that $T^{-k}\mu_2(\cdot) = \mu_1(\cdot|X(k) = 1)$,

$$\frac{d(T^{-k}\mu_2)}{d\mu_1}(X) = p^{-1}X(k).$$

Using this, we can rewrite the integral above as

$$\mathbb{E}\left| 1 - p^{-1}\frac{S_n}{n} \right|,$$

where $S_n \sim Bin(n,p)$. By the Central Limit Theorem, the last display has asymptotic $\sqrt{2(1-p)/(\pi p n)}$ as $n \to \infty$. Combined with (1.2.3), this yields

$$\mathbb{E}\left[|Y| \wedge n\right] \ge C\sqrt{n}, \tag{1.2.4}$$

for some constant $C > 0$.

The (Poisson) extra head can be treated similarly. In this case the shift-coupling inequality takes the form

$$\left\| \frac{1}{t}\int_0^t (T^{-s}\mu_1)ds - \frac{1}{t}\int_0^t (T^{-s}\mu_2)ds \right\| \le \frac{2}{t}\mathbb{E}[|Y| \wedge t],$$

where μ_1 and μ_2 now are distributions of Poisson processes, with unit intensity, on $\mathbb{R}$. Now just follow and modify the steps of the argument presented for the i.i.d. case. In detail, μ_1 is the distribution of a Poisson process with unit intensity, which is shift invariant and $\mu_2 = \mu_1(\cdot|\Pi(0) = 1)$, i.e. the distribution of a Poisson process conditioned on having a point at the origin. Note that $T^{-s}\mu_2$ is absolutely continuous with respect to μ_1 for all s. Using these consideration, we can evaluate the left hand side above, to get

$$\mathbb{E}\left|1 - \frac{\Pi[0,t]}{t}\right|,$$

which, again by the Central Limit Theorem, has asymptotic $\sqrt{2/(\pi t)}$ as $t \to \infty$.

We finish the proof with showing that $\mathbb{E}|Y|^{1/2} = \infty$, by using a neat argument. By contradiction assume that $\mathbb{E}|Y|^{1/2} < \infty$. Note that

$$\frac{|Y| \wedge n}{\sqrt{n}} \leq |Y|^{\frac{1}{2}},$$

and we can apply dominated convergence theorem to the left hand side, which yields

$$\mathbb{E}\frac{|Y| \wedge n}{\sqrt{n}} \to 0,$$

as $n \to \infty$, but this contradicts (1.2.4). □

1.3 Palm distributions

Both extra head problems can be studied from a Palm theory perspective, which is - again - a well established theory, closely related to various notions of stationarity. Palm theory has applications in various areas e.g. queueing theory, perfect simulation, disintegration of measures, exchangeable sequences and ergodic theory linked to entropy. Here, we only aim to give a brief and mostly informal, more heuristic introduction to the basic notions. However, for a rigorous treatment of this topic, with some explicit applications, we refer either to chapter 11 in [22] or to the self-contained chapter 8 in [32]. The following material is mainly based on the latter reference.

Recall that a stochastic process is *stationary* if its distribution is invariant under deterministic time shifts. *Cycle stationarity* means that the stochastic process contains cycles that form a stationary sequence; that is to say distributional invariance in this case is meant under shifts from one cycle to another. Recurrent Markov chains are good examples for cycle stationary processes. Indeed, cycles are given by successive visits to a given state and by the Markov property they form an i.i.d. sequence. It is known that if the Markov chain is positive recurrent, then its invariant distribution is a probability measure, hence the process has a stationary version as well. Other examples for cycle stationary chains are renewal processes.

Two dualities describe the relationship between stationarity and cycle stationarity. The first one asserts that the stationary process can be obtained from the cycle stationary process by placing the origin uniformly at random in a cycle after 'length-biasing' the cycle length. Conversely, the cycle stationary process can be obtained from the stationary process by shifting the origin to the right end point of the cycle straddling the origin after 'length-debiasing' the cycle length. The interpretation of this result is that *the cycle-stationary dual, i.e. the cycle-stationary process associated with a given stationary process, looks like the stationary process conditioned on having a point at the origin.* We emphasize that this is indeed just a loose description, because if we take a stationary (two-sided) Poisson process on $\mathbb{R}$, then the probability that it has a point at the origin is zero.

The second duality is similar in nature. The only difference to the first one is that the length-biasing and length debiasing is done under conditioning on the invariant σ-algebra. The interpretation of this result is that *the cycle stationary dual looks like the stationary process with origin shifted to a uniformly chosen point.* Conversely the stationary dual looks like the cycle stationary process with origin shifted to a time chosen uniformly in $\mathbb{R}$. This can also be regarded as taking a *randomized origin*.

We make this rather informal description mathematically more precise and present the first duality formally from a measure theoretic point of view. Let Z_0 denote the length of the cycle straddling the origin, that is the distance between random times S_0 and S_{-1}, the starting points of two consecutive cycles, such that $S_{-1} < 0 < S_0$. If X is a two-sided (irreducible and recurrent) Markov chain, then cycles are given by

$$C_n := (X_{S_{n-1}+s})_{s\in[0,Z_n)}, \qquad n \in \mathbb{Z},$$

where Z_n is the n-th cycle length, that is $Z_n = S_n - S_{n-1}$. A classic example for cycles are excursions and the random times S_n are the return times to the same state. Furthermore, let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$ such that it satisfies

$$\mathbb{E}[1/Z_0] < \infty. \tag{1.3.1}$$

Taking the Radon Nikodym derivative, we can define a new probability measure $\mathbb{P}^\circ$ on $(\Omega, \mathcal{F})$ by the density $d\mathbb{P}^\circ/d\mathbb{P} := 1/(Z_0\mathbb{E}[1/Z_0])$ with respect to $\mathbb{P}$, that is

$$d\mathbb{P}^\circ = \frac{1}{Z_0\mathbb{E}[1/Z_0]}d\mathbb{P}. \tag{1.3.2}$$

This is called 'length-debiasing' $\mathbb{P}$. From this, we get

$$\mathbb{E}^\circ[Z_0] = \frac{1}{\mathbb{E}[1/Z_0]}. \tag{1.3.3}$$

Observe that $0 < Z_0 < \infty$, which implies that $\mathbb{E}[1/Z_0] > 0$ and thus

$$\mathbb{E}^\circ[Z_0] < \infty. \tag{1.3.4}$$

This allows us to rewrite (1.3.2) as

$$d\mathbb{P} = \frac{Z_0}{\mathbb{E}^\circ[Z_0]}d\mathbb{P}^\circ, \tag{1.3.5}$$

which is called 'length-biasing' $\mathbb{P}^\circ$.

Conversely, given a probability measure $\mathbb{P}^\circ$ on $(\Omega, \mathcal{F})$ that satisfies (1.3.4), we can define a new probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ by (1.3.5). From this, we get

$$\mathbb{E}[1/Z_0] = \frac{1}{\mathbb{E}^\circ[Z_0]}. \tag{1.3.6}$$

Observe that $0 < Z_0 < \infty$, which implies $\mathbb{E}^\circ[Z_0]$ and thus using (1.3.6) we conclude that (1.3.1) holds. Therefore (1.3.5) can be rewritten as (1.3.2).

To sum up, we showed the connection between length-debiasing in (1.3.2) and length-biasing in (1.3.5). This provides a duality between the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, with property (1.3.1), and the probability measure $\mathbb{P}^\circ$ on the same measurable space $(\Omega, \mathcal{F})$, with property (1.3.4).

Before we formulate the first duality result, we need a few more ingredients. Let $X = (X_t)_{t\in\mathbb{R}}$ be a stochastic process and $S = (S_k)_{k\in\mathbb{Z}}$ be a (doubly infinite) ordered sequence of random times, both supported on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. S is indexed according to the convention that the time origin $t = 0$ is such that $S_{-1} < 0 \le S_0$. The pair (X, S) is stationary, under $\mathbb{P}$, if its distribution is invariant under time shifts, that is

$$T^{-u}(X, S) \stackrel{D}{=} (X, S), \qquad u \in \mathbb{R},$$

under $\mathbb{P}$. Recall that time shifts were defined as $T^{-u}X = (T^{-u}X_t)_{t\in\mathbb{R}} = (X_{t+u})_{t\in\mathbb{R}}$, for all $u \in \mathbb{R}$. These time shifts are defined on S as $T^{-u}S = (T^{-u}S_k)_{k\in\mathbb{Z}} = (S_{n_{-u}+k} + u)_{k\in\mathbb{Z}}$, for all $u \in \mathbb{R}$, where $n_{-u} = n$ if and only if $u \in (s_{n-1}, s_n]$. Note that the indices of S are compensated in order to preserve the convention that the time origin satisfies $T^{-u}S_{-1} < 0 \le T^{-u}S_0$.

Define a cycle by the random path segments $C_n = (X_{S_{n-1}+s})_{s\in[0,Z_n)}$, for all $n \in \mathbb{Z}$ and where $Z_n = S_n - S_{n-1}$ is the length of the n-th cycle. The cycle C_0 straddling the origin has length Z_0.

Let $\mathbb{P}^\circ$ be another probability measure, supported on the same measurable space $(\Omega, \mathcal{F})$ and assume that the pair (X°, S°) lives on the probability space $(\Omega, \mathcal{F}, \mathbb{P}^\circ)$. We say that (X°, S°) is cycle stationary, under $\mathbb{P}^\circ$, if the sequence of cycles is stationary, that is

$$(\dots, C_{n-1}, C_n, C_{n+1}, \dots) \stackrel{D}{=} (\dots, C_{-1}, C_0, C_1, \dots), \qquad n \in \mathbb{Z}.$$

We note that there is a one-to-one measurable correspondence between $T^{-S_n}(X, S)$ and $(\dots, C_{n-1}, C_n, C_{n+1}, \dots)$, for all $n \in \mathbb{Z}$. This implies that (X°, S°) is cycle stationary if and only if $T^{-S_n}(X, S) \stackrel{D}{=} (X^\circ, S^\circ)$, for all $n \in \mathbb{Z}$.

The first duality that states this equivalence between the two notions of stationarity, is given by

Theorem 3. *Let $(\Omega, \mathcal{F})$ be a measurable space supporting (X, S) and $((X^\circ, S^\circ), U)$, where X and X° are two-sided shift measurable processes, S and S° are two-sided sequences of times increasing strictly from $-\infty$ to ∞ with $S_{-1} < 0 \le S_0$ and $S_0^\circ \equiv 0$, and U is a $(0, 1]$ valued* *variable. Let* (X, S) *and* $((X^\circ, S^\circ), U)$ *be linked by*

$$(X^\circ, S^\circ) = T^{-S_0}(X, S) \qquad \text{and} \qquad U = -S_{-1}/Z_0$$

or, equivalently, by

$$(X, S) = T^{(1-U)Z_0^\circ}(X^\circ, S^\circ),$$

thus $Z_0 \equiv Z_0^\circ$. *Let* $\mathbb{P}$ *and* $\mathbb{P}^\circ$ *be probability measures on* $(\Omega, \mathcal{F})$ *satisfying* (1.3.1) *and* (1.3.4) *and linked by* (1.3.2) *or, equivalently, by* (1.3.5). *Then the following statements are equivalent*

(i) (X, S) *is stationary under* $\mathbb{P}$,

(ii) (X°, S°) *is cycle stationary under* $\mathbb{P}^\circ$ *and* U *is uniform on* $(0, 1]$ *and independent of* (X°, S°).

Remark 3 U is uniform on $(0, 1]$ and independent of (X°, S°) under $\mathbb{P}$ if and only if it is so under $\mathbb{P}^\circ$.

This is Theorem 4.1. in Chapter 8 in [32], where the proof can also be found. This result is the *first* or 'point-at-zero' duality.

A natural way to think about this first Palm duality is in terms of distributions. In other words, given a stationary distribution $\mathbb{P}$ we can apply Theorem 3 to some (X, S) with distribution $\mathbb{P}$ to obtain $\mathbb{P}^\circ$, which is the distribution of the cycle stationary dual (X°, S°). Conversely, given a cycle stationary distribution $\mathbb{P}^\circ$, the same theorem yields that we can obtain a stationary dual from some $((X^\circ, S^\circ), U)$, where (X°, S°) is distributed according to $\mathbb{P}^\circ$ and $U \sim Uni((0, 1])$ is independent of (X°, S°).

We conclude this section with the *second* or 'randomized-origin' duality result that makes the link to shift-coupling even more explicit. It can be proved, that under conditions, similar to that of Theorem 3, the following two statements are equivalent,

(a) (X, S) is stationary under $\mathbb{P}$,

(b) (X°, S°) is cycle stationary under $\mathbb{P}^\circ$ and $U \sim Uni((0, 1])$ is independent of (X°, S°);

for the precise statement and proof, see Theorem 8.1 in chapter 8 in [32]. The difference - which is apparent only in the conditions omitted here - between the last statement and that of Theorem 3 is that the length-biasing and length-debiasing is done under the condition of invariant σ-algebra.

Theorem 4. *Assume the equivalent claims* (a) *and* (b) *described above hold. Then the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *can be extended to support a random integer* K *such that*

$$\mathbb{P}(T^{-S_K}(X, S) \in \cdot) = \mathbb{P}^\circ((X^\circ, S^\circ) \in \cdot).$$

Conversely, the probability space $(\Omega, \mathcal{F}, \mathbb{P}^\circ)$ *can be extended to support a random time* Y *such that*

$$\mathbb{P}^\circ(T^{-Y}(X^\circ, S^\circ) \in \cdot) = \mathbb{P}((X, S) \in \cdot).$$

This is Theorem 9.1 in Chapter 8 in [32] and the proof can also be found therein.

1.4 Allocation and transport problems

In their excellent paper [20], Holroyd and Peres gave a complete characterization of the solutions to the fundamental extra head problems. Thus offering not just an insight to the nature of solutions, but also a link to the vast area of allocation and transport problems. This result acted as a catalyst for a wealth of subsequent research in this direction, see e.g. [19, 24] and perhaps most notably [8]. First, we describe this remarkable characterization in its full generality and next we intend to explore some of the classical allocation and transport problems and results. The following introduction is based on [20] and for convenience we borrow the notation from the same reference.

We start with the discrete extra head problem from section 1.1.1, but instead of $\mathbb{Z}$, we take an infinite countable group G, with identity $\mathbf{i}$, which clearly contains $\mathbb{Z}^d$ for $d \geq 1$, as an example. In this setting, we take the product measure μ on $\{0,1\}^G$. We refer to an element $\gamma \in \{0,1\}^G$ as a *configuration*. Γ will be a random configuration with law μ. We call an element $g \in G$ *site*. We say that the site g is *occupied* if $\Gamma(g) = 1$ and *unoccupied*, if $\Gamma(g) = 0$. Taking the left multiplication as the usual group operation, we define $g : x \mapsto gx$, for $g, x \in G$. This implies that a site $g \in G$ acts on a configuration $\gamma \in \Omega$ via $(g\gamma)(x) = \gamma(g^{-1}x)$, on measurable functions $f : \{0,1\}^G \to \mathbb{R}$ via $(gf)(\gamma) = f(g^{-1}\gamma)$, on events $A \subset \{0,1\}^G$ via $gA = \{g\gamma : \gamma \in A\}$, therefore $\mathbb{1}\{gA\} = g\mathbb{1}\{A\}$, and on measures via $(g\mu)(f) = \mu(g^{-1}f)$. At this stage perhaps we point out that in case $G = \mathbb{Z}^d$, for $d \geq 1$, the group action is the usual shift map $T^{-u}\gamma = (\gamma_{u+t})_{t\in\mathbb{Z}^d}$, for $u \in \mathbb{Z}^d$, we already defined in the $\mathbb{R}^d$ setting above, see section 1.2. This is a useful example, which we will frequently return to in the sequel in order to gain a better insight.

Suppose μ is invariant and ergodic under the action of G. Let p denote the marginal probability

$$p = \mu(\Gamma(\mathbf{i}) = 1),$$

and p is assumed to be $0 < p < 1$. We denote by μ^* the conditional law of Γ given $\Gamma(\mathbf{i}) = 1$,

$$\mu^*(\cdot) = \mu(\Gamma \in \cdot|\Gamma(\mathbf{i}) = 1).$$

Let X be a G-valued random variable on some joint probability space with Γ. Here we use $\mathbb{P}$ for the notation of the probability distribution of X and $\mathbb{E}$ for the expectation with respect to $\mathbb{P}$. X is a *discrete extra head scheme* for μ, if $X^{-1}\Gamma$ has law μ^* under $\mathbb{P}$. Observe that in the example $G = \mathbb{Z}$, $X^{-1}\Gamma$ will be the two-sided sequence of coin tosses with an extra head at the origin.

A *discrete transport rule* is a measurable function $\theta : \{0,1\}^G \times G \times G \to \mathbb{R}_+$ satisfying

$$\sum_{y\in G} \theta_\gamma(x,y) = 1 \qquad \text{for all } x \in G \text{ and } \mu\text{-almost every } \gamma. \tag{1.4.1}$$

For sets $A, B \subset G$ we define

$$\theta_\gamma(A,B) := \sum_{x\in A, y\in B} \theta_\gamma(x,y).$$

Furthermore, θ is G-invariant in the sense that

$$\theta_{g\gamma}(gx, gy) = \theta_{\gamma}(x, y), \tag{1.4.2}$$

for all γ and all $x, y \in G$. $\theta_{\gamma}(x, y)$ is interpreted as the amount of *mass* sent or *transported* from x to y. In this view, equation (1.4.1) tells us that each site sends out unit mass in total.

A transport rule θ is *balancing*, if

$$\theta_{\Gamma}(G, y) = p^{-1}\Gamma(y), \tag{1.4.3}$$

holds μ-almost every $y \in G$. In other words, every occupied site receives exactly p^{-1} mass and every unoccupied site gets no mass.

Given μ, let θ be a transport rule and X be a G-valued random variable, we assume that

$$\theta_{\Gamma}(\mathbf{i}, x) = \mathbb{P}(X = x|\Gamma), \tag{1.4.4}$$

for μ-almost every Γ and all $x \in G$. Observe that (1.4.1) can be obtained by summing (1.4.4) over all x and using (1.4.2). Hence, for any X, (1.4.4) determines θ uniquely up to a $\mathbb{P}$-null event, and the converse also holds, i.e. for any θ, (1.4.4) uniquely determines the joint law of (Γ, X).

Now, we are ready to state the characterization result.

Theorem 5. *Suppose X and θ are related by* (1.4.4)*. Then X is an extra head scheme if and only if θ is balanced.*

The proof can be found in [20] as the proof of Theorem 10. This result inspired some of the results in this thesis, see Theorem 9.

Allocation rules are defined as certain types of transport rules. More precisely, an *allocation rule* for μ is a measurable map $\tau : \{0,1\}^G \times G \to G$ that μ-almost surely assigns every site $x \in G$ to a site $\tau_{\gamma}(x)$. Indeed, allocation rules can be expressed in terms of transport rules as

$$\theta_{\gamma}(x, y) = \mathbb{1}\{\tau_{\gamma}(x) = y\} \tag{1.4.5}$$

Moreover, allocation rules satisfy the properties (1.4.1) and (1.4.2), with representation (1.4.5). Assume for now that $G = \mathbb{Z}^d$, for some $d \geq 1$ and the marginal probability p is a reciprocal of an integer. In this case, condition (1.4.1) becomes $|(\tau_{\gamma}^{-1})(y)| = p^{-1}\gamma(y)$ for μ-almost every γ and every $y \in \mathbb{Z}^d$. Condition (1.4.2) takes the following form; if $\tau_{\gamma}(x) = y$, then $\tau_{T^{-u}\gamma}(T^{-u}x) = T^{-u}y$, for all $u \in \mathbb{Z}^d$ and where T^{-u} is the shift operator as seen above. This representation gives rise to a corollary of the previous result.

Proposition 1.4.1. *Let Γ have law μ, and suppose p is the reciprocal of an integer. If τ is a balanced allocation rule for μ, then the random variable X given by*

$$X = \tau_{\Gamma}(0) \tag{1.4.6}$$

is a nonrandomized extra head scheme for μ. Conversely, if X is a nonrandomized extra head scheme, then there exists a μ-almost everywhere unique balanced allocation rule τ satisfying (1.4.6).

For the proof we refer to the proof of Proposition 7 in [20].

It is not so surprising that there is an analogous result in the continuous case, i.e. for the Poisson extra head problem from section 1.1.2. Recall that Π is a translation invariant, simple point process of unit intensity on $\mathbb{R}^d$, for $d \geq 1$, whose law will now be denoted by Λ. Let Π^* be the Palm version of Π, that is Π^* is again a translation invariant, simple point process of unit intensity on $\mathbb{R}^d$ with an added point at the origin. The law of Π^* is denoted by Λ^*. Similarly to the discrete case, an element $x \in \mathbb{R}^d$ is called site and a realization π of the random process Π is called configuration, in other words π is an integer valued Borel measure on $\mathbb{R}^d$. For all $u \in \mathbb{R}^d$, T^{-u} is the usual shift operator, as already seen above, $T^{-u}\pi(\cdot) = \pi(\cdot + u)$. A *continuum extra head scheme* for Π is an $\mathbb{R}^d$-valued random variable Y such that $T^{-Y}\Pi$ has law Π^*.

The *continuum* counterpart of the discrete *allocation rule* can be defined as a measurable function Ψ that assigns to Λ-almost every configuration π and every site $x \in \mathbb{R}^d$ a site $\Psi_\pi(x)$, and Ψ is translation invariant in the sense that if $\Psi_\pi(x) = y$, then $\Psi_{T^{-u}\pi}(T^{-u}x) = T^{-u}y$. Let $[\Pi] := \{x \in \mathbb{R}^d : \Pi(\{x\}) = 1\}$ be the support of Π, that is $[\Pi]$ is a random collection of sites. An allocation rule is *balanced*, if $\Psi_\Pi^{-1}(y)$ has Lebesgue measure 1 for all $y \in [\Pi]$ and at the same time the set $\Psi_\Pi^{-1}(\mathbb{R}^d \setminus [\Pi])$ has Lebesgue measure 0.

Theorem 6. *Let Ψ be an allocation rule for Π. The random variable $Y = \Psi_\Pi(0)$ is a nonrandomized extra head scheme for Π if and only if Ψ is balanced.*

This theorem is proved in [20] as Theorem 13.

We saw above that extra head schemes typically have bad moment properties in dimension one and two, see Theorem 2; the two dimensional result can be found in [20]. Somewhat surprisingly in $d \geq 3$ there is a striking contrast.

Theorem 7. *Let μ be a product measure with parameter p on $\mathbb{Z}^d$. If $d \geq 3$, then there exists a balanced discrete transport rule θ satisfying*

$$\mathbb{E}\exp(C\|\theta_\Gamma(0)\|^d) < \infty$$

for some $C = C(d,p) > 0$.

If we consider that an extra head scheme is at least as far as the first occupied site from the origin, meaning that $\mathbb{P}(\|X\|^d \geq x) > \exp(-C_1 x^d)$, for some $C_1(d,p) > 0$, then we also see that the above result is the best we may hope for.

Theorem 8. *Let Π be a Poisson process of unit intensity on $\mathbb{R}^d$. If $d \geq 3$, then there exists a continuum extra head scheme for Π satisfying*

$$\mathbb{E}\exp(C\|Y\|^d) < \infty$$

for some $C = C(d) > 0$.

The above theorems are proved in Section 8 [20]. Though these results are surprising, at this stage they do not give a clue how such schemes might look like explicitly. The first explicit construction was offered by the gravitational allocation rules - a truly original result appeared in [8].

1.4.1 Further themes for allocations

In the Poisson setting above, we saw that to every Poisson point $x \in \mathbb{R}^d$ there was a set $S(x) \subset \mathbb{R}^d$ assigned by some allocation map Ψ such that these sets had the same (unit) area. More precisely, for any two distinct Poisson points $x, y \in \mathbb{R}^d$, there were given sets $S(x), S(y) \subset \mathbb{R}^d$ such that $S(x) \cap S(y) = \emptyset$ and $\mathcal{L}(S(x)) = \mathcal{L}(S(y)) = 1$, where $\mathcal{L}$ denotes the Lebesgue measure in $\mathbb{R}^d$. In other words, the Poisson measure was allocated or coupled to the Lebesgue measure, by some allocation map Ψ.

It is fairly natural to introduce a non-negative, increasing *cost function* c that gives the *allocation* or *transport* cost $c(|x-y|)$ between two points $x, y \in \mathbb{R}^d$. A standard and extensively studied question is the following: For a given cost function c, does there exist an *optimal* allocation rule Ψ. Optimality means that for any other allocation rule Ψ', we have

$$\mathbb{E}c(Y_\Psi) \leq \mathbb{E}c(Y_{\Psi'}),$$

where Y_Ψ and $Y_{\Psi'}$ are the extra head schemes, i.e. the distances between the origin and the point, where the origin is assigned to, under the allocation Ψ and Ψ', respectively. This was the main subject of [21], where this question was answered in a general context.

Instead of taking the Lebesgue measure, we might as well take another homogeneous Poisson process with the same unit intensity and try to couple the two Poisson measures. Informally, assume that there are blue and red points in $\mathbb{R}^d$ given by two independent Poisson processes with the same unit intensity. We aim to find a scheme that matches the blue points to the red points in a translation invariant way. Then the matching distance Y between a typical blue point and its matched red pair can be studied. A wealth of results concerning the optimal tail and moment properties of Y was established in [19], which contains some problems that are still open.

1.5 Optimal transport

Optimal transport has been a vast area of intensive mathematical research in the past three decades and apart from its probabilistic aspects it has a deep analytic connection to PDE theory and Riemannian geometry. We refer to Villani's book [33], as a solid account on the topic.

Since the main results in this thesis drew some inspiration from the optimal transport literature, we think that the classical problem deserves a description and it may help the reader to put these results into a richer context.

As already mentioned in the preceding subsection 1.4.1, it is natural to consider allocations or transport maps together with some cost. In the classical *Monge-Kantorovich* setting - instead of Poisson and Lebesgue measures - we take two *probability measures* μ, ν on Polish spaces $\mathcal{X}, \mathcal{Y}$. A set $\Theta(\mu, \nu)$ of transport maps consists of all measures on $\mathcal{X} \times \mathcal{Y}$ with $\mathcal{X}$-marginal μ and $\mathcal{Y}$-marginal ν. Given a non-negative *cost function* $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$, we define a transport cost functional by

$$C(\theta) = \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\theta(x, y),$$

that assigns to every transport map $\theta \in \Theta(\mu,\nu)$ a *transport cost*. The *Monge-Kantorovich problem* is to find

$$\underline{C} := \inf\{C(\theta) : \theta \in \Theta(\mu,\nu)\}$$

together with a transport map θ that minimizes $\underline{C}$.

When we look for tractable characteristics of the minimizers, it turns out that the right description is the *cyclically monotone* property their support sets. A Borel set $B \subseteq \mathcal{X} \times \mathcal{Y}$ is cyclically monotone if and only if

$$c(x_1,y_2) - c(x_1,y_1) + \cdots + c(x_{n-1},y_n) - c(x_{n-1},y_{n-1}) + c(x_n,y_1) - c(x_n,y_n) \geq 0,$$

for all $(x_1,x_1),(x_2,y_2),\ldots,(x_n,y_n) \in B$. A transport map θ is *cyclically monotone* if it is concentrated on some cyclically monotone set B. A fundamental result says that every optimal transport map is cyclically monotone, see e.g. [4] and the references therein. Intuitively this should be clear, since the cyclically monotone property means that no cost improvement can be achieved by cyclical rerouting the mass along a cycle. The converse property, whether a cyclically monotone transport map is always optimal, is a more subtle issue. Under mild condition on the transport map, this converse can indeed be proved. For a short and concise treatment of this topic, see [4], which establishes this result as a corollary of the ergodic theorem. This might be particularly interesting, since the main optimality results in this thesis also use an ergodic argument, see Theorem 12 and Theorem 13 in Chapter 2 and 3, respectively.

1.6 Embedding aspects

The fundamental problem also has a 'embedding aspect', as we pointed out at the beginning of this chapter.

Let us first recall the classical *Skorokhod embedding problem (SEP)*. Let X be a real valued random variable with distribution μ and $W = \{W(t) : t \in \mathbb{R}_+\}$, where $\mathbb{R}_+$ is the usual notation for the non-negative real line, be a standard Brownian motion. Our goal is to find a stopping time T with $\mathbb{E}[T] < \infty$, such that $W(T)$ has law μ. Since we assume that T is an integrable stopping time, it is a corollary of Wald's lemmas, see e.g. Theorem 2.44 and Theorem 2.48 in [26], that

$$\mathbb{E}[W(T)] = 0 \qquad \text{and} \quad \mathbb{E}[W^2(T)] = \mathbb{E}[T] < \infty.$$

This implies that if X is a centered random variable, distributed according to μ, with finite variance, i.e. $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$, then there exists a stopping time T satisfying $\mathbb{E}[T] < \infty$ and $W(T) \sim \mu$. For a proof of this standard result, see Theorem 5.15 in [26].

Now consider a *two-sided* Brownian motion $B = \{B(t) : t \in \mathbb{R}\}$ with $B(0) = 0$, then the fundamental problem translates to the following problem. Given a probability measure μ on the real line, how do we find a random time T, such that $B(T)$ has law μ and $\{B(T+t) - B(T) : t \in \mathbb{R}\}$ is again a two sided Brownian motion. The main difference to the classical SEP is that we need to guarantee the 'Brownian property' on both sides. Recall that if T is a stopping time, then the forward looking part is a Brownian motion by the strong Markov property. However the backward looking part is the main source of difficulty. This problem was first studied in [23].

In Chapter 3 we elaborate more formally on this topic and we prove a strong result related to the optimality of such times T.

1.6.1 Variations on embedding

Recently there have been a few interesting directions emerged, all of which share some key features of either the problems or the techniques we have introduced so far.

In [5], the authors establish a subtle connection between optimal transport and the classical SEP. Based on concepts and geometric ideas in optimal transport theory, they develop a new approach to study and characterize solutions to SEP.

The novelty lies in the starting point. Consider the set of stopped paths

$$S = \{(f, s) : f : [0, s] \to \mathbb{R} \quad \text{is continuous,} \quad f(0) = 0\},$$

and a functional $\varphi : S \to \mathbb{R}$. The optimal Skorokhod embedding problem is to construct a stopping time optimizing

$$P_\varphi(\mu) = \inf\{\mathbb{E}[\varphi((B_t)_{0\le t\le\tau}, \tau)] : \tau \quad \text{solves (SEP)}\},$$

where μ is the target distribution we would like to embed, i.e. $B(\tau)$ has distribution μ. In this approach it is usually assumed that this problem is well-posed that is $\mathbb{E}[\varphi((B_t)_{0\le t\le\tau}, \tau)]$ exists and takes values in $(-\infty, \infty]$ for all stopping times τ that solve SEP. A first example is given by the Root's solution to SEP, where the cost functional φ is given by $\varphi(f, s) = s^2$. For further exciting examples and a detailed treatment, see [5].

Recently, another interesting embedding related problem was presented in [7], where the authors introduce and study *forward Brownian motions*. These processes are defined on the whole real line and they appear to be Brownian motion when observed from random points in space-time in the forward time direction. Formally speaking, $\{X(t) : t \in \mathbb{R}\}$ is a forward Brownian motion (FBM), if there exists a sequence $\{S_n : n \le 0\}$ of random times satisfying $\lim_{n\to-\infty} S_n = -\infty$, a.s., and for all n, the process $\{X(S_n + t) - X(S_n) : t \ge 0\}$ is a standard Brownian motion on $[0, \infty)$.

The first example of an FBM is a two-sided Brownian motion. The authors of [7] give less trivial examples of FBM, which are substantially different from a two-sided Brownian motion. One of the intriguing questions they also address is the following. Given that we can observe a process only in one direction, what can we infer - from these observations - regarding the opposite direction. In particular, if we see a process that shows 'Brownian properties' in forward time, is it necessarily a 'Brownian like' process in backward time as well? It turns out that an example can be constructed which demonstrates that the answer to the previous question is a no. However this immediately leads to a natural question of homogeneity in both, space and time. In other words, are there processes that generate i.i.d. pieces of trajectories under shifts? Again, for the detailed exposition, we refer to the original paper [7].

As a last example we mention the recent work of Pitman and Tang, [29], where they embed Brownian bridge in Brownian motion by a spacetime shift. More precisely, given a standard Brownian motion $\{B(t) : t \ge 0\}$, the authors show that there exists a non-negative random time T, such that $\{B(T+u) - B(T) : 0 \leq u \leq 1\}$ has the same distribution as a standard Brownian bridge.

This problem led to the study of the *bridge-like process*, which appeared as a natural candidate for a solution to the bridge-embedding problem. The bridge-like process is defined by

$$\{B(F+u) - B(F) : 0 \leq u \leq 1\},$$

where $F := \inf\{t \geq 0 : B(t+1) - B(t) = 0\}$. It is still an open problem, whether the bridge-like process is a standard Brownian bridge. Another open problem, intimately related to this is that in case of a negative answer, is the distribution of a standard Brownian bridge absolutely continuous with respect to the distribution of the bridge-like process? The reason why this proves to be a hard problem is the rather unusual nature of the random time F. It is also unknown, whether the bridge-like process has the Markov property or not.

Chapter 2

Skorokhod embeddings for two-sided Markov chains

This chapter is a joint work with my supervisor, Peter Mörters and was published in Probability Theory and Related Fields, see [27].

2.1 Introduction and statement of main results

Let $\mathcal{S}$ be a finite or countable state space and $p = (p_{ij} \colon i, j \in \mathcal{S})$ an irreducible and recurrent transition matrix. Then there exists a stationary measure $(m_i \colon i \in \mathcal{S})$ with positive weights, which is finite in the positive recurrent case, and infinite otherwise. The two-sided stationary Markov chain $X = (X_n \colon n \in \mathbb{Z})$ with initial measure $(m_i \colon i \in \mathcal{S})$ and transition matrix p is characterized by

- $\mathbb{P}(X_n = i) = m_i$ for all $n \in \mathbb{Z}$, $i \in \mathcal{S}$;
- $\mathbb{P}(X_n = j \,|\, X_{n-1}, X_{n-2}, \ldots) = p_{X_{n-1} j}$ for all $n \in \mathbb{Z}$, $i, j \in \mathcal{S}$.

This chain always exists, if we allow $\mathbb{P}$ to be a σ-finite measure. For the simplest construction, let $(X_n \colon n \geq 0)$ be the chain with initial measure $(m_i \colon i \in \mathcal{S})$ and transition matrix p, and $(X_{-n} \colon n \geq 0)$ be the chain with given initial state X_0 and dual transition probabilities given by $p^*_{ij} = (m_j / m_i) p_{ji}$.

By conditioning the stationary chain X on the event $\{X_0 = i\}$, we define the two-sided Markov chain with transition matrix p with fixed initial state $X_0 = i$. Its law, denoted by $\mathbb{P}_i$, does not depend on the choice of $(m_i \colon i \in \mathcal{S})$ and is always a probability law. Note that we can equivalently define this chain, or indeed the two-sided Markov chain with transition matrix p and arbitrary initial distribution ν, by picking X_0 according to ν and letting the forward and backward chains $(X_n \colon n \geq 0)$, resp. $(X_{-n} \colon n \geq 0)$, evolve as in the case of the stationary chain.

A natural version of the *Skorokhod embedding problem* in this context asks, given the two-sided Markov chain $(X_n \colon n \in \mathbb{Z})$ with transition matrix p and initial state $X_0 = i$ and a probability measure ν on the state space $\mathcal{S}$, whether there exists a random time T such that $(X_{n+T}\colon n \in \mathbb{Z})$ is a two-sided Markov chain with transition matrix p such that X_T has law ν. If this is the case we say that T is an *embedding* of the *target distribution* ν. Our interest here is mainly in times T which are *non-randomized*, which means that T is a measurable function of the sample chain X. The random times T are often stopping times, but this is not a necessary requirement.

Finding embeddings of two-sided Markov chains is a subtle problem, because even for stopping times T the shifted process $T^{-1}X := (X_{n+T}\colon n \in \mathbb{Z})$ often will *not* be a two-sided Markov chain. For example, take a simple symmetric random walk on the integers, started in $X_0 = 0$, and let T be the first positive hitting time of the integer $a > 0$. Then T embeds the Dirac measure δ_a, but the increment $T^{-1}X_0 - T^{-1}X_{-1}$ always takes the value $+1$, hence $T^{-1}X$ is not a two-sided simple random walk. A similar argument shows that even shifting the simple random walk by a nonzero fixed time does not preserve the property of being a simple random walk with given distribution of the state at time zero.

The first main result of this paper gives a necessary and sufficient condition on the initial state, the target measure and the stationary distribution for the existence of a Skorokhod embedding for an arbitrary two-sided Markov chain.

Theorem 9. *Let X be a two-sided irreducible and recurrent Markov chain with transition matrix p and initial state $X_0 = i$. Take $\nu = (\nu_j\colon j \in \mathcal{S})$ to be any probability measure on $\mathcal{S}$. Then the following statements are equivalent.*

- (a) *There exist a non-randomized random time T such that $(X_{n+T}\colon n \in \mathbb{Z})$ is a Markov chain with transition matrix p and X_T has law ν.*
- (b) *The stationary measure $(m_j\colon j \in \mathcal{S})$ satisfies $\frac{m_i}{m_j}\,\nu_j \in \mathbb{Z}$ for all $j \in \mathcal{S}$.*

If the random time T in (a) exists it can always be taken to be a stopping time.

Example 2.1.1 (Embedding measures with mass in the initial state) Assume that the target measure ν charges the initial state $i \in \mathcal{S}$ of the Markov chain, i.e. $\nu_i > 0$. Choosing $i = j$ in (b) shows that a non-randomized random time T with the properties of (a) can exist only if $\nu = \delta_i$. In this case a natural family of embeddings can be constructed using the concept of point stationarity, see for example [32], as follows: Let $r \in \mathbb{N}$ and let T_r be the the time of the rth visit of state i after time zero. Then it is easy to check, and follows from [24, Theorem 6.3], that the process $T_r^{-1}X$ is a Markov chain with transition matrix p and $X_{T_r} = i$. ■

Example 2.1.2 (Extra head problem) Take a doubly-infinite sequence of tosses of a (possibly biased) coin, or more precisely let $X = (X_n\colon n \in \mathbb{Z})$ be i.i.d. random variables with distribution $\mathbb{P}(X_n = \texttt{head}) = p$, $\mathbb{P}(X_n = \texttt{tail}) = 1 - p$, for some $p \in (0, 1)$. Our aim is to find, without using any randomness generated in a way different from looking at coins in the sequence, a coin showing head in this sequence in such a way that the two semi-infinite sequences of coins to the left and to the right of this coin remain independent i.i.d. sequences of coins with the same bias. This is known as *extra head problem* and was investigated and fully answered by

Liggett [25] and Holroyd and Peres [20]. To relate this to our setup, we can assume that $X_0 = \mathtt{tail}$, as otherwise the coin at the origin is the extra head. Then the extra head problem becomes the Skorokhod embedding problem for X with initial state $X_0 = \mathtt{tail}$ and target measure $\nu = \delta_{\mathtt{head}}$. Theorem 9 shows (as proved by Holroyd and Peres before) that the extra head problem has a solution if and only if $(1-p)/p \in \mathbb{Z}$, i.e. if and only if p is the inverse of an integer. Moreover, Liggett [25] gives an explicit solution of the extra head problem which we generalize to our setup in Theorem 2 below. ■

Example 2.1.3 (Inverse extra head problem) If in the setup of Example 2.1.2 the state of the coin at the origin has been revealed, we ask whether it is possible to shift the sequence in such a way that this information is lost, i.e. the shifted sequence is an i.i.d. sequence of coins with the original bias. This means that we wish to embed the invariant distribution $\nu = m$ given by $m_{\mathtt{head}} = p, m_{\mathtt{tail}} = 1 - p$. Theorem 9 shows that this is impossible. ■

Example 2.1.4 (Extra head problem with a finite pattern) In the setup of Example 2.1.2 we now ask to find a particular *finite pattern* of successive outcomes, such that the coins to its left and right remain an i.i.d. sequence of coins with the same bias. Looking, for example, for the pattern `head/tail` we would first reveal the coin at the origin, and then if this shows `head` its right neighbour, and if this shows `tail` its left neighbour. The underlying Markov chain has the state space $\{\mathtt{tail/tail}, \mathtt{tail/head}, \mathtt{head/tail}, \mathtt{head/head}\}$, the transition matrix

$$\begin{pmatrix} 1-p & p & 0 & 0 \\ 0 & 0 & 1-p & p \\ 1-p & p & 0 & 0 \\ 0 & 0 & 1-p & p \end{pmatrix},$$

and invariant measure $((1-p)^2, p(1-p), p(1-p), p^2)$. Our theorem shows that, if we initially reveal `tail/tail` then we need $1/p$ to be an integer, and if we reveal `head/head` then we need $1/(1-p)$ to be an integer. Hence we can only embed `head/head` if $p = \frac{1}{2}$. More generally, the problem can be solved for patterns that are repetitions of the single symbol `head` if and only if $1/p$ is an integer, for patterns that are repetitions of the single symbol `tail` if and only if $1/(1-p)$ is an integer, and for patterns containing both symbols `tail` and `head` if and only if $p = \frac{1}{2}$. ■

Example 2.1.5 (Simple random walk) Let X be a two-sided simple symmetric random walk on the integers, with $X_0 = i$ for some $i \in \mathbb{Z}$. In this case the invariant measure is $m_i = 1$ for all $i \in \mathbb{Z}$, hence Theorem 9 shows that the target measures that can be embedded are precisely the Dirac measures δ_j, $j \in \mathcal{S}$. The same result holds for the simple symmetric random walk on the square lattice $\mathbb{Z}^2$. ■

The proof of Theorem 9 extends the ideas developed by Liggett [25] and Holroyd and Peres [20] for the extra head problem to the more general Markov chain setup. In particular, under the additional assumption that the target measure does not charge the initial state, we are able to generalize Liggett's construction of an elegant explicit solution, in analogy to the Brownian motion case studied in Last et al. [23]. Recall that the case when the target measure charges the initial state was already discussed in Example 2.1.1. To describe this solution we define the local time L^j spent by X at state $j \in \mathcal{S}$ to be the normalized counting measure given by

$$L^j(A) := \frac{1}{m_j}\, \#\{n \in A \colon X_n = j\} \qquad \text{for any } A \subset \mathbb{Z}.$$

Theorem 10. *Let X be a two-sided irreducible and recurrent Markov chain with $X_0 = i$ and further assume that the target measure ν satisfies $\nu_i = 0$ and the conditions in Theorem 9(b). Then*

$$T_* := \min\Big\{n \geq 0 \colon L^i([0,n]) \leq \sum_{j\in\mathcal{S}} \nu_j\, L^j([0,n])\Big\} \tag{2.1.1}$$

is a finite, non-randomized stopping time satisfying the conditions of Theorem 9(a).

Example 2.1.6 We take a stationary three state Markov chain with transition probabilities given by $p_{12} = p_{32} = 1$ and $p_{21} = 1 - p$ and $p_{23} = p$. If $1/p$ is an integer we can shift the chain so that it starts in the third state and the chain property is preserved, as follows: Uncover the state at the origin. If it is the third state we are done; if it is the second state we move along the chain until the number of visits to the third state is at least p times the number of visits to the second state; if it is the first state we move until the number of visits to the third state is at least $\frac{p}{1-p}$ times the number of visits to the first state. Note that if the state of the origin is the first state it is *not* a solution to wait one time step, whence you are in the second state, and then apply the strategy for start in the second state as this creates a bias in the backward chain. ■

Skorokhod embedding problems usually concern embedding times with finite expectation. However in the extra head problem it is not possible to achieve finite expectation of the random time T. In fact Liggett [25] shows that in this case always $E\sqrt{T} = \infty$, see Theorem 2 in Chapter 1 above or Holroyd and Liggett [18]. For the simple random walk on the integers we expect in analogy to the Brownian motion case studied by Last et al. [23] that always $E\sqrt[4]{T} = \infty$. Our aim here is to understand the general picture.

To this end we now recall the notion of asymptotic Green's function of the Markov chain. Given states $i, j \in \mathcal{S}$ we first define the normalized truncated Green's function by

$$a_{ij}(n) = \mathbb{E}_i L^j([0,n]) = \frac{1}{m_j}\, \mathbb{E}_i\Big[\sum_{k=0}^{n} \mathbb{1}\{X_k = j\}\Big],$$

that is $a_{ij}(n)$ gives the normalized expected number of visits to state j between time 0 and time n, by the Markov chain with initial state $X_0 = i$. By Orey's ergodic theorem, see, e.g., Chen [9], for any states $i, j, k, l \in \mathcal{S}$, the functions a_{ij} and a_{kl} are asymptotically equivalent in the sense that

$$\lim_{n\to\infty} \frac{a_{ij}(n)}{a_{kl}(n)} = 1.$$

We then define the *asymptotic Green's function* $a(n)$ as the equivalence class of the truncated

Green's functions under asymptotic equivalence. Observe that finiteness of moments is a class property, i.e. expressions of the form $E[a(Y)] < \infty$, where a is an equivalence class and Y an integer-valued random variable, are meaningful.

Theorem 11. *Let X be a two-sided irreducible and recurrent Markov chain with $X_0 = i$ and ν be any target measure different from the Dirac measure δ_i. If T_* is the stopping time defined in* (2.1.1)*, then*

(i) $\mathbb{E}_i\big[a(T_*)^{1/2}\big] = \infty$.

If additionally ν has finite support, then

(ii) $\mathbb{E}_i\big[a(T_*)^{\beta}\big] < \infty$ *for all* $0 \leq \beta < \frac{1}{2}$.

As $a(n)$ cannot grow faster than n, our solutions T_* always have 'bad' moment properties as even for the nicest Markov chain T_* can never have finite square root moments. However, our next theorem shows that no other solution of the embedding problem has better moment properties than T_*.

In fact, it turns out that T_* has a strong optimality property, as it simultaneously minimizes all concave moments of non-negative solutions of the embedding problem. This striking result is new even for the case of the extra head problem and therefore, in our opinion, constitutes the most interesting contribution in this paper.

Theorem 12. *Let X be a two-sided irreducible and recurrent Markov chain with $X_0 = i$ and ν be a target measure satisfying the conditions in Theorem 10. If T_* is the solution of the Skorokhod embedding problem constructed in* (2.1.1) *and T any other non-negative (possibly randomized) solution, then*

$$\mathbb{E}_i\big[\psi(T_*)\big] \leq \mathbb{E}_i^{\oplus}\big[\psi(T)\big],$$

for any non-negative concave function ψ defined on the non-negative integers, where the expectation on the right is with respect to the chain as well as any possible extra randomness used to define T.

Theorem 12 is inspired by exciting recent developments connecting the classical Skorokhod embeddings for Brownian motion with optimal transport problems. In a recent paper, Beiglböck, Cox and Huesmann [5] exploit this connection to characterize certain solutions to the Skorokhod embedding problem by a geometric property. In a similar vein, our solution T_* is characterized by a geometric property, the 'non-crossing' condition, which yields the optimality. See also our concluding remarks in Section 2.6 for possible extensions of this result.

Example 2.1.7 Suppose the underlying Markov chain is *positive recurrent*. Then the asymptotic Green's function satisfies $a(n) \sim n$. Therefore all non-negative solutions T of the Skorokhod embedding problem satisfy $\mathbb{E}_i[\sqrt{T}] = \infty$, while the solution constructed in Theorem 10 satisfy $\mathbb{E}_i[T_*^{\beta}] < \infty$ for all $0 \leq \beta < 1/2$. This applies in particular to Examples 2.1.2 and 2.1.4. ∎

Example 2.1.8 The situation is much more diverse for *null-recurrent* chains. Looking at Example 2.1.5, for a two-sided simple symmetric random walk on the integers we have $a(n) \sim \sqrt{n}$. Hence the solution T_* constructed in Theorem 10 satisfies $\mathbb{E}_i[T_*^\alpha] < \infty$ for all $0 \le \alpha < 1/4$, while any non-negative solution has infinite 1/4 moment. This is similar to the case of Brownian motion on the line, which is discussed in [23], although in that paper other than here the discussion is restricted to solutions which are non-randomized stopping times. In contrast to this, for simple symmetric random walk on the square lattice $\mathbb{Z}^2$ we have $a(n) \sim \log n$, and therefore $\mathbb{E}_i[\sqrt{\log T}]$ is infinite for any non-negative solution T, while the solution T_* constructed in Theorem 10 satisfies $\mathbb{E}_i[(\log T_*)^\alpha] < \infty$, for all $0 \le \alpha < 1/2$. ■

2.2 Relating embedding and allocation problems

In this section we relate our embedding problem to an equivalent allocation problem. The section specializes some results from Last and Thorisson [24] which are themselves based on ideas from [20]. We give complete proofs of the known facts in order to keep this paper self-contained. Generalizing from [23] we call a random time T an *unbiased shift* of the Markov chain X if the shifted process $T^{-1}X$ is a two-sided Markov chain with the same transition matrix as X. Note that this definition allows T to be randomized, i.e. it does not have to be a function of the sample chain X alone.

Let $\Omega = \{(\omega_i)_{i\in\mathbb{Z}} \colon \omega_i \in \mathcal{S}\}$ be the set of trajectories of X. A *transport rule* is a measurable function $\theta \colon \Omega \times \mathbb{Z} \times \mathbb{Z} \to [0,1]$ satisfying

$$\sum_{y\in\mathbb{Z}} \theta_\omega(x,y) = 1 \qquad \text{for all } x \in \mathbb{Z} \text{ and } \mathbb{P}\text{-almost every } \omega.$$

Note that we write the dependence on the trajectory ω by a subindex, which we drop from the notation whenever convenient. Transport rules are interpreted as distributing mass from x to $\mathbb{Z}$ in such a way that the site y gets a proportion $\theta(x,y)$ of the mass. For sets $A, B \subset \mathbb{Z}$ we define

$$\theta_\omega(A,B) := \sum_{x\in A, y\in B} \theta_\omega(x,y).$$

A transport rule θ is called *translation invariant* if

$$\theta_{z\omega}(x+z, y+z) = \theta_\omega(x,y),$$

for all $\omega \in \Omega$ and $x, y, z \in \mathbb{Z}$, where $z\omega$, defined by $z\omega_n = \omega_{n-z}$ for any $n \in \mathbb{Z}$, is the trajectory shifted by $-z$. A transport rule *balances* the random measures ξ and ζ on $\mathbb{Z}$ if

$$\sum_{z\in\mathbb{Z}} \theta_\omega(z,A)\xi(z) = \zeta(A), \tag{2.2.1}$$

for any $A \subset \mathbb{Z}$ and $\mathbb{P}$-almost all ω. Given a two-sided Markov chain X as before recall the definition of the local times L^i, and given a probability measure $\nu = (\nu_i \colon i \in \mathcal{S})$ we further define

$$L^\nu = \sum_{i\in\mathcal{S}} \nu_i \, L^i.$$

Proposition 2.2.1. *Assume that there is a measurable family of probability measures* $(\mathbb{Q}_\omega\colon \omega\in\Omega)$ *on some measurable space* Ω' *and* $T\colon \Omega\times\Omega'\to\mathbb{Z}$ *is measurable. The random time* T *and a translation invariant transport rule* θ *are associated if*

$$\mathbb{Q}_\omega\big(\omega'\in\Omega'\colon T(\omega,\omega')=t\big)=\theta_\omega(0,t) \qquad \text{for all } t\in\mathbb{Z} \text{ and } \mathbb{P}\text{-almost all } \omega\in\Omega. \tag{2.2.2}$$

For any probability measure $\mu=(\mu_i\colon i\in\mathcal{S})$ *we define the probability measure* $\mathbb{P}^\oplus_\mu$ *on* $\Omega\times\Omega'$ *by*

$$\mathbb{P}^\oplus_\mu(d\omega\, d\omega') = \sum_{i\in\mathcal{S}} \mu_i\, \mathbb{P}_i(d\omega)\, \mathbb{Q}_\omega(d\omega'). \tag{2.2.3}$$

Then, if μ,ν *is any pair of probability measures on* $\mathcal{S}$ *and the random time* T *and translation invariant transport rule* θ *are associated, the following statements are equivalent.*

- (a) *Under* $\mathbb{P}^\oplus_\mu$ *the random time* T *is an unbiased shift of* X *and* X_T *has law* ν*.*
- (b) *The transport rule* θ *balances* L^μ *and* L^ν $\mathbb{P}$*-almost everywhere.*

Note that in the last proposition unbiased shifts need not be non-randomized. The transport rules associated to non-randomized shifts are the *allocation rules*. These are given by a measurable map $\tau\colon\Omega\times\mathbb{Z}\to\mathbb{Z}$ such that $\theta_\omega(x,y)=1$ if $\tau_\omega(x)=y$ and zero otherwise.

Proposition 2.2.2. *If the random time* T *in Proposition 2.2.1 is non-randomized, then there is an associated transport rule* θ*, which is an allocation rule. Conversely if* θ *in Proposition 2.2.1 is an allocation rule, then there exists an associated non-randomized random time* T*.*

We give proofs of the propositions for completeness. For a transport rule θ we define

$$J_\mu(\omega) := \sum_{k\in\mathbb{Z}} \theta_\omega(k,0)\, L^\mu(k), \tag{2.2.4}$$

which is interpreted as the total mass received by the origin. We recall the following simple fact, see [20] for a more general version.

Lemma 2.2.3. *Let* $m\colon\mathbb{Z}\times\mathbb{Z}\to[0,\infty]$ *be such that* $m(x+z,y+z)=m(x,y)$ *for all* $x,y,z\in\mathbb{Z}$*. Then*

$$\sum_{y\in\mathbb{Z}} m(x,y) = \sum_{y\in\mathbb{Z}} m(y,x).$$

The following calculation is at the core of the proof.

Lemma 2.2.4. *Suppose that* T *and* θ *are related by* (2.2.2)*. Then, for any measurable function* $f\colon\Omega\to[0,\infty]$*, we have*

$$\mathbb{E}^\oplus_\mu\big[f(T^{-1}X)\big] = \mathbb{E}\big[J_\mu(X) f(X)\big],$$

where $\mathbb{E}^\oplus_\mu$ *is the expectation with respect to* $\mathbb{P}^\oplus_\mu$ *defined in* (2.2.3)*.*

Proof of Lemma 2.2.4. Writing $\mathbb{P}_\mu = \sum_{i\in\mathbb{Z}} \mu_i \mathbb{P}_i$ we get

$$\begin{aligned}\mathbb{E}^{\oplus}_{\mu}\big[f(T^{-1}X)\big] &= \int d\mathbb{P}_\mu(\omega) \int f\big(T(\omega,\omega')^{-1}X(\omega)\big)\,\mathbb{Q}_\omega(d\omega')\\ &= \int d\mathbb{P}_\mu(\omega) \sum_{t\in\mathbb{Z}} \mathbb{Q}_\omega(T=t) f(t^{-1}X(\omega)).\end{aligned}$$

Using relation (2.2.2) and the definition of $\mathbb{P}_\mu$ we continue with

$$\begin{aligned}&= \sum_{i\in\mathbb{Z}} \mu_i \int d\mathbb{P}_i(\omega) \sum_{t\in\mathbb{Z}} \theta_\omega(0,t) f(t^{-1}X(\omega))\\ &= \sum_{i\in\mathbb{Z}} \mu_i \int d\mathbb{P}(\omega) \sum_{t\in\mathbb{Z}} \theta_\omega(0,t) L^i(0) f(t^{-1}X(\omega)),\end{aligned}$$

as $L^i(0) = \frac{1}{m_i}$ and $L^j(0) = 0$ under $\mathbb{P}_i$ for $j \neq i$. Applying Lemma 2.2.3 gives

$$\begin{aligned}&= \sum_{i\in\mathbb{Z}} \mu_i \int d\mathbb{P}(\omega) \sum_{t\in\mathbb{Z}} \theta_\omega(t,0) L^i(t) f(X(\omega))\\ &= \int d\mathbb{P}(\omega) \sum_{t\in\mathbb{Z}} \theta_\omega(t,0) L^\mu(t) f(X(\omega))\\ &= \mathbb{E}\big[J_\mu(X) f(X)\big],\end{aligned}$$

using first the definition of L^μ and second the definition of $J_\mu(X)$. □

Proof of Proposition 2.2.1. First assume that θ is a translation invariant transport rule. Then, for any non-negative measurable f, by Lemma 2.2.4, we have

$$\mathbb{E}^{\oplus}_{\mu}\big[f(T^{-1}X)\big] = \mathbb{E}\big[J_\mu(X) f(X)\big] = \mathbb{E}\Big[\sum_{k\in\mathbb{Z}} \theta_\omega(k,0)\, L^\mu(k) f(X)\Big]. \tag{2.2.5}$$

If θ balances L^μ and L^ν this equals

$$\mathbb{E}\big[L^\nu(0) f(X)\big] = \sum_{j\in\mathbb{Z}} \nu_j\, \mathbb{E}\big[L^j(0) f(X)\big] = \sum_{j\in\mathbb{Z}} \nu_j\, \mathbb{E}_j\big[f(X)\big] = \mathbb{E}_\nu\big[f(X)\big].$$

Hence under $\mathbb{P}^{\oplus}_{\mu}$ the random variable $T^{-1}X$ has the law of X under $\mathbb{P}_\nu$. In other words T is an unbiased shift and X_T has distribution ν.

Conversely, assume that T is an unbiased shift and X_T has distribution ν. Hence $\mathbb{E}^{\oplus}_{\mu}[f(T^{-1}X)] = \mathbb{E}_\nu[f(X)] = \mathbb{E}[L^\nu(0) f(X)]$. Plugging this into (2.2.5) gives

$$\mathbb{E}\Big[\sum_{k\in\mathbb{Z}} \theta_\omega(k,0)\, L^\mu(k) f(X)\Big] = \mathbb{E}\big[L^\nu(0) f(X)\big].$$

As f was arbitrary we get $\sum_{k\in\mathbb{Z}} \theta_\omega(k,0)\, L^\mu_\omega(k) = L^\nu_\omega(0)$ for $\mathbb{P}$-almost all ω, where we emphasise the dependence of the measures L^μ and L^ν on the trajectories by a subscript. As θ is translation invariant we get, substituting $m := k - \ell$,

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \theta_\omega(k, A)\, L^\mu_\omega(k) &= \sum_{k\in\mathbb{Z}} \sum_{\ell\in A} \theta_\omega(k, \ell)\, L^\mu_\omega(k) = \sum_{\ell\in A} \sum_{m\in\mathbb{Z}} \theta_{-\ell\omega}(m, 0)\, L^\mu_{-\ell\omega}(m) \\ &= \sum_{\ell\in A} L^\nu_{-\ell\omega}(0) = \sum_{\ell\in A} L^\nu_\omega(\ell) = L^\nu_\omega(A),\end{aligned}$$

for every $A \subset \mathbb{Z}$ and $\mathbb{P}$-almost every ω. □

Proof of Proposition 2.2.2. Suppose $T = T(\omega)$ is non-randomized. Define $\tau_\omega\colon \mathbb{Z} \to \mathbb{Z}$ by $\tau_\omega(k) = T(-k\omega)+k$ and let $\theta_\omega(x, y) = 1$ if $\tau_\omega(x) = y$ and zero otherwise. Then θ is a translation invariant allocation rule. Moreover, $\mathbb{Q}_\omega(T = t) = \mathbb{1}\{t = T(\omega)\} = \mathbb{1}\{t = \tau_\omega(0)\} = \theta_\omega(0, t)$, hence T and θ are associated. Conversely, if θ is a translation invariant allocation rule given by $\tau\colon \Omega \times \mathbb{Z} \to \mathbb{Z}$ define a non-randomized time T by $T = \tau_\omega(0)$. As before, $\mathbb{Q}_\omega(T = t) = \mathbb{1}\{t = T(\omega)\} = \mathbb{1}\{t = \tau_\omega(0)\} = \theta_\omega(0, t)$, and hence T and θ are associated. □

2.3 Existence of allocation rules: Proof of Theorems 9 and 10

In the light of the previous section our Theorems 9 and 10 can be formulated and proved as equivalent statements about allocation rules. We start with the result on non-existence of non-randomized unbiased shifts, which is implicit in Theorem 9.

Suppose that statement (a) in Theorem 9 holds and for the Markov chain X with $X_0 = i$ there exists a non-randomized unbiased shift T such that X_T has law ν. Then by Proposition 2.2.2 there exists a translation-invariant allocation rule τ associated with T and by Proposition 2.2.1 this rule balances the measures L^i and L^ν. Recall that L^i is the measure on $\mathbb{Z}$ which has masses of fixed size $1/m_i$ at the times when the stationary chain X visits state i. By the balancing property (2.2.1) for allocation rules, all masses of L^ν must have sizes which are integer multiples of $1/m_i$. As these masses are ν_j/m_j we get that $\frac{m_i}{m_j}\nu_j$ must be integers for all $j \in \mathcal{S}$, which is statement (b).

The remainder of this section is devoted to the proof of *existence* of non-randomized unbiased shifts of the Markov chain X with $X_0 = i$, embedding ν under the assumption of Theorem 9 (b). By Example 2.1.1 we may additionally assume that for the initial state i of the Markov chain we have $\nu_i = 0$. Our claim is that the stopping time T_* defined in Theorem 10 is an unbiased shift with the required properties. The next proposition shows that an associated allocation rule balances the measures L^i and L^ν which, once accomplished, implies Theorem 10 and completes the proof of Theorem 9.

Proposition 2.3.1. *Under the assumptions set out above, the following holds.*

(a) *The mapping $\tau\colon \Omega \times \mathbb{Z} \to \mathbb{Z}$ defined by*

$$\tau_\omega(k) = \min\left\{n \ge k\colon L^i_\omega([k, n]) \le L^\nu_\omega([k, n])\right\}$$

is a translation-invariant allocation rule associated with the T_* defined in (2.1.1).

(b) *For $\mathbb{P}$-almost every ω and all $A \subset \mathbb{Z}$ we have*

$$\sum_{k\in\mathbb{Z}} \mathbb{1}\{\tau_\omega(k) \in A\}\, L^i_\omega(k) = L^\nu_\omega(A), \tag{2.3.1}$$

in other words the allocation rule balances L^i and L^ν.

The proof of the proposition is similar to that of [23, Theorem 5.1] in the diffuse case. We prepare it with two lemmas. The first lemma is a pathwise statement which holds for every fixed trajectory ω satisfying the integer assumption stated in Theorem 9 and also the assumption $\nu_i = 0$.

Lemma 2.3.2. *Suppose $b \in \mathbb{Z}$ is such that $X_b = j$ for some $j \in \mathcal{S}$ with $\nu_j > 0$, and $a \in \mathbb{Z}$ is given by*

$$a := \max\big\{k < b\colon L^i([k,b]) \ge L^\nu([k,b])\big\}.$$

Then

$$\sum_{k\in[a,b]} \mathbb{1}\{\tau(k) \in A\}\, L^i(k) = L^\nu(A), \tag{2.3.2}$$

holds for any $A \subset [a,b]$.

Proof. We define the function $\Delta f\colon \mathbb{Z} \to [0,\infty)$ by

$$\Delta f(k) := L^i(k) - L^\nu(k) = \begin{cases} \frac{1}{m_i} & \text{if} \quad X_k = i, \\ -\frac{\nu_j}{m_j} & \text{if} \quad X_k = j \neq i. \end{cases}$$

Recall that by our assumption $\frac{\nu_j}{m_j}$ is an integer multiple of $\frac{1}{m_i}$. Hence, denoting

$$f_u^v := \sum_{n=u}^{v} \Delta f(n) \quad \text{ for all } u, v \in \mathbb{Z} \text{ and } u \le v,$$

we have $a = \max\big\{k < b\colon f_k^b = 0\big\}$ and hence $f_a^b = 0$.
By the additivity of both sides of (2.3.2) it suffices to prove

$$\sum_{k\in[a,b]} \mathbb{1}\{\tau(k) = z\}\, L^i(k) = L^\nu(z) \quad \text{for all sites } z \in [a,b]. \tag{2.3.3}$$

Fix $z \in [a,b]$ and let $j = X_z$. Observe that $\tau(k) = z$ if and only if $f_k^z \le 0$ but $f_k^\ell > 0$ for all $k \le \ell < z$. Hence we may assume $\nu_j > 0$ as otherwise both sides of (2.3.3) are zero. We also have that $f_a^z > 0$ if $z < b$. Indeed, suppose that $f_a^z \le 0$. Then $f_{z+1}^b = f_a^b - f_a^z \ge 0$ contradicting the choice of a.
As $f_a^z \ge 0, f_z^z = -\frac{\nu_j}{m_j} < 0$ and ν_j/m_j is an integer multiple of $1/m_i$ we find a $k_1 \ge a$ with $f_{k_1}^z = 0$ and $f_j^z < 0$ for all $k_1 < j \le z$. Similarly, we find $k_1 < k_2 < \cdots < k_N$ where $N := (\frac{m_i}{m_j})\,\nu_j$ such

that

$$f^z_{k_n} = \tfrac{1-n}{m_i} \text{ and } f^z_j < \tfrac{1-n}{m_i} \text{ for all } k_n < j \le z.$$

As $\tau(k) = \min\{n \ge k \colon f^n_k \le 0\}$ we infer that $\tau(k_n) = z$ for all $n \in \{1, \dots, N\}$ and there are no other values k with $\tau(k) = z$. Each of these values contributes a summand $\frac{1}{m_i}$ to the left hand side in (2.3.3). Therefore this side equals $\frac{N}{m_i} = \frac{\nu_j}{m_j}$, as does the right hand side. This completes the proof. □

The second lemma is probabilistic and ensures in particular that the mapping τ described in Proposition 2.3.1 (a) is well defined.

Lemma 2.3.3. *For $\mathbb{P}$-almost every ω the following two events hold*

(E1) *for all k with $X_k = i$ we have $\tau(k) < \infty$;*

(E2) *for all b such that $X_b = j$ for some $j \in \mathcal{S}$ with $\nu_j > 0$ there exists $a < b$ such that $X_a = i$ and $L^i([a,b]) = L^\nu([a,b])$.*

Proof. To show this we use an argument from [20], see Theorem 17 and the following remark. We formulate the negation of the two events. The complement of (*E1*) is the event that there exists k such that $X_k = i$ and $L^i([k,\ell]) > L^\nu([k,\ell])$, for all $\ell > k$. The complement of (*E2*) is that there exists b such that $X_b = j$ for some $j \in \mathcal{S}$ with $\nu_j > 0$ and $L^i([a,b]) < L^\nu([a,b])$, for all $a < b$ with $X_a = i$. We first show that, for $\mathbb{P}$-almost every ω, both complements cannot occur simultaneously.

Indeed, for a fixed ω, it is clear that there cannot be k and b as above such that $k < b$. Assume for contradiction that the set of trajectories ω for which there exists $k > b$ as above has positive probability. On this event the minimum over all $\ell > k$ with $X_\ell = i$ and $\tau(\ell) = \infty$ is finite, we denote it by K. By translation invariance $\mathbb{P}(K = 0) > 0$ from which we infer by conditioning on the event $\{X_0 = i\}$ that $\mathbb{P}_i(K = 0) > 0$. If $(T_n \colon n \in \mathbb{N})$ is the collection of return times to state i, by the invariance described in Example 1.1 we have $\mathbb{P}_i(K = T_n) = \mathbb{P}_i(K = 0) > 0$ for all $n \in \mathbb{N}$ contradicting the finiteness of $\mathbb{P}_i$. Therefore we have shown that, for $\mathbb{P}$-almost every ω, either (*E1*) or (*E2*) occurs.

As the last step we show that event (*E1*) cannot occur without event (*E2*). To this end define $m(x,y) = \mathbb{E}[\mathbb{1}\{\tau(x) = y, X_x = i\}]$ and apply Lemma 2.2.3 to get

$$\mathbb{E}\Big[\sum_{k\in\mathbb{Z}} \mathbb{1}\{\tau(k) = 0, X_k = i\}\Big] = \mathbb{E}\Big[\sum_{k\in\mathbb{Z}} \mathbb{1}\{\tau(0) = k, X_0 = i\}\Big].$$

The left-hand side in this equation equals m_i if and only if (*E2*) occurs $\mathbb{P}$-almost every ω, and the right-hand side equals m_i if and only if (*E1*) occurs $\mathbb{P}$-almost every ω. As these two events cannot fail at the same time, both events (*E1*) and (*E2*) occur for $\mathbb{P}$-almost every ω. □

Proof of Proposition 2.3.1. Recall that τ is well-defined and note that translation-invariance of the allocation rule defined in terms of τ follows easily from the fact that $\tau_\omega(k) = \tau_{k\omega}(0) + k$. As $T_*(\omega) = \tau_\omega(0)$ by definition, the allocation rule is associated with T_*. This proves (a).

To prove (b) we note that it suffices to fix $z \in \mathbb{Z}$ and show that for $\mathbb{P}$-almost every ω equation (2.3.1) holds for $A = \{z\}$. We let $b = \tau(z)$. By Lemma 2.3.3 for $\mathbb{P}$-almost every ω there exists $a < b$ such that $X_a = i$ and $L^i([a,b]) = L^\nu([a,b])$. Then the interval $[a,b]$ contains z and all k with $\tau(k) = z$. Hence the results follows by application of Lemma 2.3.2. □

2.4 Moment properties of T_*: Proof of Theorem 3

The critical exponent $\frac{1}{2}$ occurring in Theorem 11 originates from the behaviour of the first passage time below zero by a mean zero random walk. We summarize the results required for such random walks in the following lemma.

Lemma 2.4.1. *Let $\xi, \xi_1, \xi_2, \dots$ be independent identically distributed random variables with $E\xi = 0$ taking values in the integers. Define the associated random walk by $S_n = \sum_{i=1}^n \xi_i$ and its first passage time below zero as $N = \min\{n \in \mathbb{N} \colon S_n \leq 0\}$.*

(a) *If the walk is skip-free to the right, i.e. $P(\xi > 1) = 0$, then $E\big[N^{1/2}\big] = \infty$.*

(b) *If the walk has finite variance, then there exists $C > 0$ such that $P\big(N > n\big) \sim C \,\dfrac{1}{\sqrt{n}}$.*

Proof. (a) Denote by $N^{(j)}$ the first passage time for the walk given by $S_n^{(j)} = \sum_{i=1}^n \xi_{i+j-1}$. Then

$$E[N \wedge n] = \sum_{j=1}^n P(N \geq j) = \sum_{j=1}^n P(N^{(j)} \geq n-j+1) = E\Big[\sum_{j=1}^n \mathbb{1}\{N^{(j)} \geq n-j+1\}\Big],$$

If $\underline{S}_n$ denotes the minimum of $\{S_0, S_1, \dots, S_n\}$ we have, using that the walk is skip-free to the right,

$$\sum_{j=1}^n \mathbb{1}\{N^{(j)} \geq n-j+1\} = S_n - \underline{S}_n.$$

This implies $E[N \wedge n] \geq E[(S_n)_+]$. By a concentration inequality for arbitrary sums of independent random variables, see [28, Theorem 2.22], there exists a constant $C > 0$ such that, for all $\varepsilon > 0$ and $n \in \mathbb{N}$, we have $P(S_n \in [-\varepsilon\sqrt{n}, \varepsilon\sqrt{n}]) \leq C\,\varepsilon$. Hence, by Markov's inequality, for any $\varepsilon > 0$,

$$E\big[(S_n)_+\big] = \frac{1}{2} E|S_n| \geq \frac{1}{2}\,\varepsilon\sqrt{n}\,P\big(|S_n| > \varepsilon\sqrt{n}\big) \geq \frac{1}{2}\,\varepsilon(1 - C\varepsilon)\,\sqrt{n}.$$

We infer that $\liminf \frac{1}{\sqrt{n}}\,E[N \wedge n] > 0$. But if we we had $E[N^{1/2}] < \infty$ dominated convergence would imply that this limit is zero, which is a contradiction.

(b) This is a classical result of Spitzer [30]. A good proof can be found in [16, Theorem 1a in Section XII.7], see also [16, Section XVIII.5] for a proof that random walks with finite variance satisfy Spitzer's condition. □

2.4.1 Proof of Theorem 11 (i).

We start by proving a variant of the upper half in the Barlow-Yor inequality [2] for Markov chains. This result, usually given in the context of continuous martingales, estimates the moments of the local time at a stopping time, by moments of the stopping time itself.

Lemma 2.4.2. *For any $0 < p < 1$, there exists a constant C_p such that, for any state $i \in \mathcal{S}$ and any stopping time T,*

$$\mathbb{E}_i[L^i([0,T])^p] \leq C_p\, \mathbb{E}_i[a_{ii}(T)^p]. \tag{2.4.1}$$

The lemma relies on the following classical inequality, we refer to [3, (6.9)] for a proof.

Lemma 2.4.3 (Good λ inequality). *For every $0 < p < 1$ there is a constant $C_p > 0$ such that, for any pair of non-negative random variables (X, Y) satisfying*

$$P(X > 3\lambda, Y < \delta\lambda) \leq \delta\, P(X > \lambda) \quad \textit{for all } 0 < \delta < 3^{-p-1} \textit{ and } \lambda > 0, \tag{2.4.2}$$

we have

$$E\left[X^p\right] \leq C_p\, E\left[Y^p\right].$$

Proof of Lemma 2.4.2. If we show that (2.4.2) holds with random variables $X = m_i L^i([0,T])$ and $Y = m_i a_{ii}(T)$ under $\mathbb{P}_i$, the result follows immediately from Lemma 2.4.3. If $\lambda \leq 1$ the left hand side of (2.4.2) is zero and there is nothing to show. We may therefore assume that $\lambda > 1$. Define $m_i a_{ii}^{-1}(x) := \max\{n\colon m_i a_{ii}(n) < x\}$. Let $T_0 = 0$ and T_k be the time of the kth visit of state i after time zero. Finally assume, without loss of generality, that $\mathbb{P}_i(X > \lambda) > 0$. Then,

$$\begin{aligned}\mathbb{P}_i\big(X > 3\lambda, Y < \delta\lambda \,\big|\, X > \lambda\big) &= \mathbb{P}_i\big(T_{\lfloor 3\lambda\rfloor+1} \leq T, m_i a_{ii}(T) < \delta\lambda \,\big|\, T_{\lfloor\lambda\rfloor+1} \leq T\big)\\ &\leq \mathbb{P}_i\big(T_{\lfloor 3\lambda\rfloor - \lfloor\lambda\rfloor} \leq m_i a_{ii}^{-1}(\delta\lambda)\big) \leq \mathbb{P}_i\big(L^i([0, m_i a_{ii}^{-1}(\delta\lambda)]) \geq \lfloor 2\lambda\rfloor\big).\end{aligned}$$

By Markov's inequality the last expression above can be bounded by

$$\lfloor 2\lambda\rfloor^{-1}\, \mathbb{E}_i\left[L^i([0, m_i a_{ii}^{-1}(\delta\lambda)])\right] = \lfloor 2\lambda\rfloor^{-1}\, m_i a_{ii}\left(m_i a_{ii}^{-1}(\delta\lambda)\right) \leq \delta\frac{\lambda}{\lfloor 2\lambda\rfloor},$$

which is smaller than δ, as required. □

We define $T_0 = 0$ and $T_k = \min\{n > T_{k-1}\colon X_n = i\}$, for $k \geq 1$. Recall that $\mathbb{E}_i L^j([T_{k-1}, T_k)) = 1/m_i$ and hence, by the strong Markov property, the random variables $\xi_k := 1 - m_i\, L^\nu([T_{k-1}, T_k))$ are independent and identically distributed with mean zero. By Lemma 2.4.1 (a) the first passage time of zero for this walk satisfies $\mathbb{E}_i[N^{1/2}] = \infty$. As $m_i L^i([0, T_*]) \geq N - 1$ the result follows.

2.4.2 Proof of Theorem 11 (ii).

We first prove the result in the simple case that the state space $\mathcal{S}$ is finite. In this case the chain is positive recurrent and we have $a(n) \sim n$.

Lemma 2.4.4. *Suppose $\mathcal{S}$ is finite. Then, for any $i \in \mathcal{S}$, we have $\mathbb{E}_i\big[T_*^{\beta}\big] < \infty$, for all $0 \le \beta < \frac{1}{2}$.*

Proof. Let $T_0 = 0$ and, for $k \in \mathbb{N}$, define $T_k = \min\{n > T_{k-1} \colon X_n = i\}$. Denote by h_{ij} the probability that the chain started in i hits j before returning to i, and observe that irreducibility implies that $h_{ij} > 0$. By the strong Markov property we have $m_j L^j[0,T_1] = YZ$ where Y is a Bernoulli variable with mean h_{ij} and Z is an independent geometric with success parameter h_{ji}. Hence $\mathbb{E}_i L^j[0,T_1) = h_{ij}/m_j h_{ji}$, which also equals $1/m_i$. Recalling that $\mathbb{E}[Z^2] \le 2/h_{ji}^2$ we get $\mathbb{E}_i[L^j[0,T_1)^2] \le 2h_{ij}/m_j^2 h_{ji}^2$, and hence $L^{\nu}[0,T_1)$ has finite variance. Define $\xi_k := 1 - m_i\, L^{\nu}([T_{k-1},T_k))$, and observe that $\xi_1, \xi_2, \ldots$ are independent and identically distributed variables with mean zero and finite variance. Let $N := \min\{n \colon \sum_{k=1}^n \xi_k \le 0\}$, and observe that $T_* \le T_N$. Fix $\varepsilon > 0$ and note that

$$\mathbb{P}_i\big(T_* > n\big) \le \mathbb{P}_i\big(N > \varepsilon n\big) + \mathbb{P}_i\Big(\sum_{k=1}^{\lceil \varepsilon n\rceil}(T_k - T_{k-1}) > n\Big).$$

By Lemma 2.4.1 (b) the first term on the right-hand side is bounded by a constant multiple of $(\varepsilon n)^{-1/2}$. For the second term we note that the random variables $T_1 - T_0, T_2 - T_1, \ldots$ are independent and identically distributed with finite variance. By Chebyshev's inequality we infer that, for sufficiently small $\varepsilon > 0$, the term is bounded by a multiple of $1/n$. Altogether we get that $\mathbb{P}_i(T_* > n)$ is bounded by a constant multiple of $n^{-1/2}$, from which the result follows immediately. $\square$

We return to the general case. The next result, which is an auxiliary step in the proof of Theorem 11 (ii), may be of independent interest. The short proof given here, which does not make any regularity assumptions on the chain, is due to Vitali Wachtel.

Lemma 2.4.5. *Fix a state $i \in \mathcal{S}$ and let $T = \min\{n > 0 \colon X_n = i\}$ be the first return time to this state. Then*

$$\mathbb{E}_i\big[a_{ii}(T)^{\alpha}\big] < \infty, \qquad \textit{for all } 0 \le \alpha < 1.$$

Proof. By Lemma 1 in Erickson [14], we have for $m(n) := \int_0^n \mathbb{P}_i(T > x)\,dx$ that

$$\frac{n}{m(n)} \le m_i a_{ii}(n) \le 2\,\frac{n}{m(n)} \qquad \text{for all positive integers } n.$$

As $m(n) \ge n\mathbb{P}_i(T > n-1)$ we infer that $m_i a_{ii}(n) \le 2/\mathbb{P}_i(T > n-1)$ and therefore

$$\mathbb{E}_i\big[a_{ii}(T)^{\alpha}\big] \le \big(\tfrac{2}{m_i}\big)^{\alpha} \sum_{n=1}^{\infty} \big(\mathbb{P}_i(T > n-1)\big)^{-\alpha}\,\mathbb{P}_i(T = n) = \big(\tfrac{2}{m_i}\big)^{\alpha} \sum_{n=1}^{\infty} (1 - s_{n-1})^{-\alpha}\big(s_n - s_{n-1}\big),$$

where $s_n := \mathbb{P}_i(T \le n)$. Letting $s(t) := s_{n-1} + (t-(n-1))(s_n - s_{n-1})$, for $n-1 \le t < n$, we can bound the sum by $\int_0^{\infty}(1-s(t))^{-\alpha}\,ds(t)$, which is finite for all $0 \le \alpha < 1$, as required. $\square$

We now look at the reduction of our Markov chain to the finite state space $\mathcal{S}' = \{0\} \cup \{j \in \mathcal{S} \colon \nu_j > 0\}$. More explicitly, let $t_0 = 0$ and $t_k = \min\{n > t_{k-1} \colon X_n \in \mathcal{S}'\}$ for $k \in \mathbb{N}$, and

$t_k = \max\{n < t_{k+1} \colon X_n \in \mathcal{S}'\}$ for $k \in -\mathbb{N}$. Then $Y_n = X_{t_n}$ defines an irreducible Markov chain $Y = (Y_n \colon n \in \mathbb{Z})$ with finite state space $\mathcal{S}'$, and its invariant measure is $(m_i \colon i \in \mathcal{S}')$. If N is the stopping time constructed in Theorem 2 for the reduced chain Y, then the solution T_* for the original problem is $T_* = t_N$.

Given two states $i, j \in \mathcal{S}'$ we denote by S_{ij} a random variable whose law is given by $\mathsf{P}(S_{ij} = s) = \mathbb{P}_i(t_1 = s \,|\, Y_1 = j)$ for all $s \in \mathbb{N}$, if $\mathbb{P}_i(Y_1 = j) > 0$, and $S_{ij} = 0$ otherwise. We construct a probability space on which there are independent families $(S_{ij}, S_{ij}^{(k)} \colon k \in \mathbb{N})$ of independent random variables with this law, together with an independent copy of Y and hence N. We denote probability and expectation on this space by P, resp. E. Observe that on this space we can also define a copy of the process $(t_k \colon k \in \mathbb{N})$ by $t_0 = 0$ and

$$t_k = t_{k-1} + \sum_{i,j \in \mathcal{S}'} S_{ij}^{(k)} \mathbb{1}\{Y_{k-1} = i, Y_k = j\} \qquad \text{for } k \in \mathbb{N}.$$

For any non-decreasing, subadditive representative a of the class of the asymptotic Green's function,

$$\mathbb{E}_i\big[a(T_*)^\beta\big] = \mathsf{E}\Big[a\Big(\sum_{k=1}^{N} t_k - t_{k-1}\Big)^\beta\Big] \le \mathsf{E}\Big[a\Big(\sum_{k=1}^{N} \sum_{i,j \in \mathcal{S}'} S_{ij}^{(k)}\Big)^\beta\Big] \le \sum_{i,j \in \mathcal{S}'} \mathsf{E}\Big[a\Big(\sum_{k=1}^{N} S_{ij}^{(k)}\Big)^\beta\Big].$$

It therefore suffices to show that

$$\mathsf{E}\Big[a_{ii}\Big(\sum_{k=1}^{N} S_{ij}^{(k)}\Big)^\beta\Big] < \infty.$$

Let $n \in \mathbb{N}$ and use first subadditivity of a_{ii} and then Jensen's inequality to get, for $2\beta < \alpha < 1$, that

$$\mathsf{E}\Big[a_{ii}\Big(\sum_{k=1}^{n} S_{ij}^{(k)}\Big)^\beta\Big] \le \mathsf{E}\Big[\Big(\sum_{k=1}^{n} a_{ii}^\alpha\big(S_{ij}^{(k)}\big)\Big)^{\beta/\alpha}\Big] \le \Big(\sum_{k=1}^{n} \mathsf{E}\big[a_{ii}^\alpha\big(S_{ij}^{(k)}\big)\big]\Big)^{\beta/\alpha} = n^{\beta/\alpha} \mathsf{E}\big[a_{ii}^\alpha\big(S_{ij}\big)\big]^{\beta/\alpha}.$$

We now note that, if T_{ij} denotes the first hitting time of state j for X under $\mathbb{P}_i$, we have $\mathsf{P}(S_{ij} > x) \le C_0\, \mathbb{P}_i(T_{ij} > x)$ for all $x > 0$, where C_0 is the maximum of the inverse of all nonzero transition probabilities from i to all other states, by the chain Y. Hence

$$\mathsf{E}\big[a_{ii}^\alpha\big(S_{ij}\big)\big] \le C_0\, \mathbb{E}_i\big[a_{ii}^\alpha\big(T_{ij}\big)\big].$$

In the case $i = j$ the right hand side is finite by Lemma 2.4.5 and, as a_{ii} grows no faster than linearly, the right hand side is finite for all choices of $i, j \in \mathcal{S}'$ by application of Theorem 1.1 in Aurzada et al. [1]. Summarising, we have found a constant $C > 0$ such that

$$\mathsf{E}\Big[a_{ii}\Big(\sum_{k=1}^{n} S_{ij}^{(k)}\Big)^\beta\Big] \le C n^{\beta/\alpha}.$$

Using the independence of N and $(S_{ij}^{(k)} : k \in \mathbb{N})$ and Lemma 2.4.4 we get

$$\mathsf{E}\Big[a_{ii}\Big(\sum_{k=1}^{N} S_{ij}^{(k)}\Big)^{\beta}\Big] \le C\mathbb{E}_i\big[N^{\beta/\alpha}\big] < \infty,$$

as required.

2.5 Optimality of T_*: Proof of Theorem 4

In this section we prove Theorem 12. In what follows we only consider forward looking transport rules, all the results here contain this assumption implicitly. We start by introducing an intuitive and convenient way to talk about allocation rules. A path of the Markov chain X can be viewed as leaving white and coloured balls on the integers, in the following way: At each site $k \in \mathbb{Z}$ we place one white ball if $X_k = i$, and $\frac{m_i}{m_j}\nu_j$ balls of colour j if $X_k = j$. By our assumption there is always an integer number of balls at each site. We call a bijection from the set of white balls to the set of coloured balls a *matching*. Given a matching we define an allocation rule $\tau : \Omega \times \mathbb{Z} \to \mathbb{Z}$ by letting

- $\tau(k) = k$ if there is no white ball at site k,
- $\tau(k) = \ell$ if the white ball at site k is matched to a coloured ball at site ℓ.

Every allocation rule thus constructed balances L^μ and L^ν, for $\mu = \delta_i$. Conversely, every balancing allocation rule agrees L^μ-almost everywhere with an allocation rule constructed from a matching. We denote by $\tau_* : \Omega \times \mathbb{Z} \to \mathbb{Z}$ the allocation rule associated with T_* constructed in Proposition 2.3.1.

The allocation rule τ_* is associated with the following one-sided *stable matching* or *greedy algorithm*, which is a variant of the famous Gale–Shapley stable marriage algorithm [17].

(1) If the next occupied site to the right of a white ball carries one or more coloured balls, map the white ball to one of those coloured balls.

(2) Remove all white and coloured balls used in step (1) and repeat.

By Lemma 2.3.3 the algorithm matches every ball after a finite number of steps, and it is easy to see that this leads to the allocation rule τ_*.

Now recall from Section 2 that non-negative, possibly randomized, times T are associated to transport rules $\theta : \Omega \times \mathbb{Z} \times \mathbb{Z} \to [0, 1]$ balancing L^μ and L^ν with the property that $\theta_\omega(x, y) = 0$ whenever $x > y$. Without loss of generality we may assume that $\theta_\omega(x, x) = 1$ if the site x does not carry a white ball. This implies that, for $x < y$, we can have $\theta_\omega(x, y) > 0$ only if the site x carries a white ball, and the site y carries a coloured ball. Moreover, if y carries a ball of colour j, we have

$$\sum_{x<y} \theta_\omega(x, y) = \frac{m_i}{m_j}\nu_j.$$

Suppose that $u, v \in \mathbb{Z}$ with $u < v$. We say that the pair (u, v) is *crossed* by θ if there exist sites $x < u < v < y$ such that $\theta(x, v) > 0$ and $\theta(u, y) > 0$. In this case (x, u, v, y) is called a *crossing.*

For a transport rule θ we repair the crossing (x, u, v, y) by letting

- $\theta'(x, y) = \theta(x, y) + (\theta(x, v) \wedge \theta(u, y))$,
- $\theta'(u, v) = \theta(u, v) + (\theta(x, v) \wedge \theta(u, y))$,
- $\theta'(x, v) = \theta(x, v) - (\theta(x, v) \wedge \theta(u, y))$,
- $\theta'(u, y) = \theta(u, y) - (\theta(x, v) \wedge \theta(u, y))$,

and setting $\theta'(w, z) = \theta(w, z)$ if $w \notin \{x, u\}$ or $z \notin \{y, v\}$, see Figure 2.1. Note that θ' is still a transport rule, the crossing has been repaired, i.e. (x, u, v, y) is not a crossing by θ', and if θ balances L^μ and L^ν then so does θ'.

We now explain how to repair a pair (u, v) crossed by θ by sequentially repairing its crossings and taking limits, so that (u, v) is not crossed by the limiting transport rule. For this purpose we define that a sequence of transport rules θ_n *converges uniformly* to a transport rule θ if

$$\lim_{n\to\infty} \sum_{x,y\in\mathbb{Z}} \big|\theta_n(x, y) - \theta(x, y)\big| = 0.$$

Denote by $y_1, y_2, \ldots$ the sequence of sites $v < y_1 < y_2 < \cdots$ such that $\theta(u, y_n) > 0$, and by $x_1, x_2, \ldots$ the sequence of sites $u > x_1 > x_2 > \cdots$ such that $\theta(x_n, v) > 0$. Note that both sequences could be finite or infinite. First we successively repair the crossings $x_1 < u < v < y_n$, for $n = 1, 2, \ldots$. The total mass moved in the nth repair is bounded by $4\theta(u, y_n)$ and because $\sum_n \theta(u, y_n) \le 1$ we can infer that the sequence of repaired transport rules converges uniformly to a transport rule θ_1. Of course, here and below if a sequence is finite we take the last element of the sequence as limit. We continue by repairing the crossings $x_2 < u < v < y_n$ of θ_1, for $n = 1, 2, \ldots$, obtaining θ_2, and so on. We obtain a sequence $\theta_1, \theta_2, \ldots$ of transport rules. The amount of mass moved when going from θ_{n-1} to θ_n is bounded by $4\theta(x_n, v)$. As $\sum_n \theta(x_n, v) < \infty$, we infer that the sequence $(\theta_n)_n$ converges uniformly to a limiting transport rule. We observe that this transport rule balances L^μ and L^ν and that (u, v) is not crossed by it.

Lemma 2.5.1. *Suppose that θ is a transport rule balancing L^μ and L^ν and $A \subset \mathbb{Z}$ a finite interval. Then, by repairing pairs crossed by θ in a given order, we obtain a transport rule θ_* balancing L^μ and L^ν, such that if $u, v \in A$ then (u, v) is not crossed by θ_*.*

Proof. Without loss of generality the left endpoint of A carries a white ball, and its right endpoint carries a coloured ball. Let $v_1, \ldots, v_n$ be the sites in A carrying coloured balls, ordered from left to right. We go through these sites in order, starting with v_1. Take u_1 to be the rightmost site to the left of v_1 carrying a white ball. Repair the pair (u_1, v_1) as above, and observe that the resulting transport rule transports a unit mass from u_1 to v_1. We declare the white ball at site u_1 and one of the coloured balls at v_1 *cancelled.* If v_1 carries an uncancelled

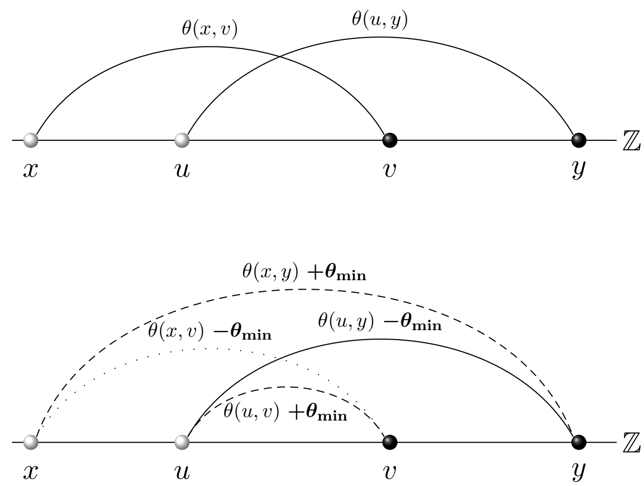

Figure 2.1: The picture above shows a crossing. Its weight $\theta_{\min} := \theta(x,v) \wedge \theta(u,y)$ is assumed to be $\theta(x,v)$, so that in the picture below we see that after the repair the dotted edge has weight zero, and the crossing is therefore removed.

ball and there are uncancelled white balls on sites of A to the left of v_1, we choose the rightmost of those, say u_2, repair the pair (u_2, v_1), and cancel two balls as above. We continue until we run out of uncancelled balls. The resulting transport rule has the property that none of the pairs (u, v_1), with $u \in A$, is crossed, and from all sites carrying cancelled white balls a unit mass is transported to site v_1.

We now move to the next coloured ball v_2 and repair all pairs (u, v_2), where u goes from right to left through all sites in $A \cap (-\infty, v_2)$ carrying uncancelled white balls. We do this until we run out of uncancelled white balls to the left of, or coloured balls on the site v_2. Observe that at the end of this step none of the pairs (u, v_1) or (u, v_2), with $u \in A$, is crossed by the resulting transport rule. We continue, moving to the next coloured ball until all coloured balls in A are exhausted. At the end of this finite procedure we obtain a transport rule θ_* balancing L^μ and L^ν, such that if $u, v \in A$ then (u, v) is not crossed by θ_*. □

We call a set A an *excursion* if it is an interval $[m, n]$ such that that there is the same number of white and coloured balls on the sites of A, but the number of white balls exceeds the number of coloured balls on every subinterval $[m, k]$, for $m \le k < n$. Observe that if A is an excursion, then it is an interval of the form $[m, \tau_*(m)]$ where m carries a white ball, but not all such intervals are excursions. Moreover, for every $x \in A$, we have both $\tau_*(x) \in A$ and $\tau_*^{-1}(x) \subset A$.

Lemma 2.5.2. *Let A be an excursion and θ_* a transport rule balancing L^μ and L^ν, such that any pair (u, v) with $u, v \in A$ is not crossed by θ_*. Then θ_* agrees in A with the allocation rule τ_*, in the sense that $\theta_*(x, y) = \mathbb{1}\{\tau_*(x) = y\}$ and $\theta_*(y, x) = \mathbb{1}\{\tau_*(y) = x\}$, for all $x \in A$ and $y \in \mathbb{Z}$.*

Proof. We start by fixing a site $x \in A$ carrying a white ball, and note that, by definition of an excursion, we also have $\tau_*(x) \in A$. We show by contradiction that θ_* transports no mass from x to a point other than $\tau^*(x)$.

First, suppose that there exist $x < v < \tau_*(x)$ with $\theta_*(x, v) > 0$. As there are more white than coloured balls on the sites in $[x, v]$, which is a consequence of the definition of τ_*, and as every site carries at most one white ball, we find $x' \in (x, v)$ such that the sites of $[x', v]$ carry the same number of white and coloured balls. As $\theta_*(x, v) > 0$ not all white balls in $[x', v]$ are matched within that interval, and there must also exist $u \in [x', v)$ and $y > v$ such that $\theta_*(u, y) > 0$. So we have found a pair (u, v) with $u, v \in A$, which is crossed by θ_*, and hence a contradiction.

Second, suppose that there exist $v > \tau_*(x)$ with $\theta_*(x, v) > 0$. As there are at least as many coloured balls as white balls in $[x, \tau_*(x)]$ not all coloured balls are matched within that interval, and hence there exists a $y \in (x, \tau_*(x)]$ and a site $u < x$ with $\theta_*(u, y) > 0$. So we have found a pair (x, y) with $x, y \in A$, which is crossed by θ_*, and hence a contradiction. We conclude that $\theta_*(x, y) = \mathbf{1}\{\tau_*(x) = y\}$ for all $x \in A$.

Now fix a site $x \in A$ carrying balls of colour j. Then $\tau_*^{-1}(x)$ is a set of $(m_i/m_j)\nu_j$ points in A. Hence, by the first part, $\theta_*(y, x) = \mathbf{1}\{\tau_*(y) = x\}$ for all $y \in \tau_*^{-1}(x)$. Moreover,

$$\sum_{y \in \tau_*^{-1}(x)} \theta_*(y, x) = (m_i/m_j)\nu_j = \sum_{y \in \mathbb{Z}} \theta_*(y, x).$$

Hence $\theta_*(y, x) = 0 = \mathbf{1}\{\tau_*(y) = x\}$ also for all $y \notin \tau_*^{-1}(x)$. □

We now let ψ be a non-negative, concave function on the non-negative integers $\mathbb{N}_0$. Note that this implies that $\psi\colon \mathbb{N}_0 \to [0, \infty)$ is non-decreasing. We further assume that $\psi(0) = 0$, an assumption which causes no loss of generality in Theorem 12. We write $\psi(n) = 0$ for $n \leq 0$ to simplify the notation.

Lemma 2.5.3. *Let A be an excursion and suppose θ is a transport rule balancing L^μ and L^ν. Then*

$$\sum_{x \in A} \psi\big(\tau_*(x) - x\big) + \sum_{x \in \tau_*^{-1}(A)} \psi\big(\tau_*(x) - x\big) \leq \sum_{\substack{x \in A \\ y \in \mathbb{Z}}} \theta(x, y)\, \psi(y - x) + \sum_{\substack{x \in \mathbb{Z} \\ y \in A}} \theta(x, y)\, \psi(y - x).$$

Proof. Observe that, by concavity, for all $a, b, c \in \mathbb{N}_0$, we have

$$\psi(a + b) + \psi(b + c) \geq \psi(a + b + c) + \psi(b). \tag{2.5.1}$$

Fix a crossing $x < u < v < y$ with $u, v \in A$, and let θ' be the result of repairing the crossing. We show that repairing the crossing does not increase

$$\sum_{\substack{x \in A \\ y \in \mathbb{Z}}} \theta(x, y)\, \psi(y - x) + \sum_{\substack{x \in \mathbb{Z} \\ y \in A}} \theta(x, y)\, \psi(y - x)$$

by looking at the difference of the repaired and original state of the sum. If $x, y \notin A$ we get

$$\begin{aligned}\theta'(u,y)\psi(y-u) + 2\theta'(u,v)\psi(v-u) + \theta'(x,v)\psi(v-x) \\ - \big(\theta(u,y)\psi(y-u) + 2\theta(u,v)\psi(v-u) + \theta(x,v)\psi(v-x)\big) \\ = \big(\theta(x,v) \wedge \theta(u,y)\big)\big(2\psi(v-u) - \psi(v-x) - \psi(y-u)\big) \le 0,\end{aligned}$$

as ψ is non-decreasing. If $x \in A, y \notin A$ we get

$$\begin{aligned}\theta'(u,y)\psi(y-u) + 2\theta'(u,v)\psi(v-u) + 2\theta'(x,v)\psi(v-x) + \theta'(x,y)\psi(y-x) \\ - \big(\theta(u,y)\psi(y-u) + 2\theta(u,v)\psi(v-u) + 2\theta(x,v)\psi(v-x) + \theta(x,y)\psi(y-x)\big) \\ = \big(\theta(x,v) \wedge \theta(u,y)\big)\big(2\psi(v-u) + \psi(y-x) - 2\psi(v-x) - \psi(y-u)\big) \\ \le \big(\theta(x,v) \wedge \theta(u,y)\big)\big(\psi(v-u) + \psi(y-x) - \psi(v-x) - \psi(y-u)\big) \le 0,\end{aligned}$$

using first that ψ is non-decreasing and then (2.5.1). The case $x \notin A, y \in A$ is analogous. If $x, y \in A$ the difference is twice

$$\begin{aligned}\theta'(x,v)\psi(v-x) + \theta'(u,y)\psi(y-u) + \theta'(u,v)\psi(v-u) + \theta'(x,y)\psi(y-x) \\ - \big(\theta(x,v)\psi(v-x) + \theta(u,y)\psi(y-u) + \theta(u,v)\psi(v-u) + \theta(x,y)\psi(y-x)\big) \\ = \big(\theta(x,v) \wedge \theta(u,y)\big)\big(\psi(y-x) + \psi(v-u) - \psi(v-x) - \psi(y-u)\big) \le 0,\end{aligned}$$

by application of (2.5.1), which shows that in all cases the sum above is not increased by the repair.

Repairing crossings successively as described in Lemma 2.5.1, we get

$$\sum_{\substack{x\in A\\ y\in\mathbb{Z}}} \theta_*(x,y)\,\psi(y-x) + \sum_{\substack{x\in\mathbb{Z}\\ y\in A}} \theta_*(x,y)\,\psi(y-x) \le \sum_{\substack{x\in A\\ y\in\mathbb{Z}}} \theta(x,y)\,\psi(y-x) + \sum_{\substack{x\in\mathbb{Z}\\ y\in A}} \theta(x,y)\,\psi(y-x).$$

By Lemma 2.5.2 we have $\theta_*(x,y) = \mathbf{1}\{\tau_*(x) = y\}$ if $x \in A$ or $y \in A$, and this allows us to rewrite the left hand side as stated. □

Lemma 2.5.4. *Let $T \ge 0$ be a (possibly randomized) unbiased shift and $\theta\colon \Omega \times \mathbb{Z} \times \mathbb{Z} \to [0,1]$ be the associated transport rule. Denote by $(T_n : n \in \mathbb{Z})$ the times in which X visits the state i, in order so that $T_0 = 0$. Let $\psi\colon \mathbb{Z} \to [0,\infty)$ be concave. Then, $\mathbb{P}_i$-almost surely,*

$$\lim_{n\to\infty} \frac{1}{n}\Big\{ \sum_{k=0}^{T_n-1} \sum_{\ell=k+1}^{\infty} \theta(k,\ell)\psi(\ell-k) + \sum_{\ell=-\infty}^{k-1} \theta(\ell,k)\psi(k-\ell)\Big\} = 2\mathbb{E}_i^{\oplus}\psi(T),$$

and

$$\lim_{m\to\infty} \frac{1}{m}\Big\{ \sum_{k=T_{-m}}^{-1} \sum_{\ell=k+1}^{\infty} \theta(k,\ell)\psi(\ell-k) + \sum_{\ell=-\infty}^{k-1} \theta(\ell,k)\psi(k-\ell)\Big\} = 2\mathbb{E}_i^{\oplus}\psi(T),$$

where the expectations on the right hand sides are taken with respect to the chain as well as any possible extra randomness used to define T, formally $\mathbb{E}_i \sum_{\ell=1}^{\infty} \psi(\ell)\theta_\omega(0,\ell) = \mathbb{E}_i^{\oplus}\psi(T)$.

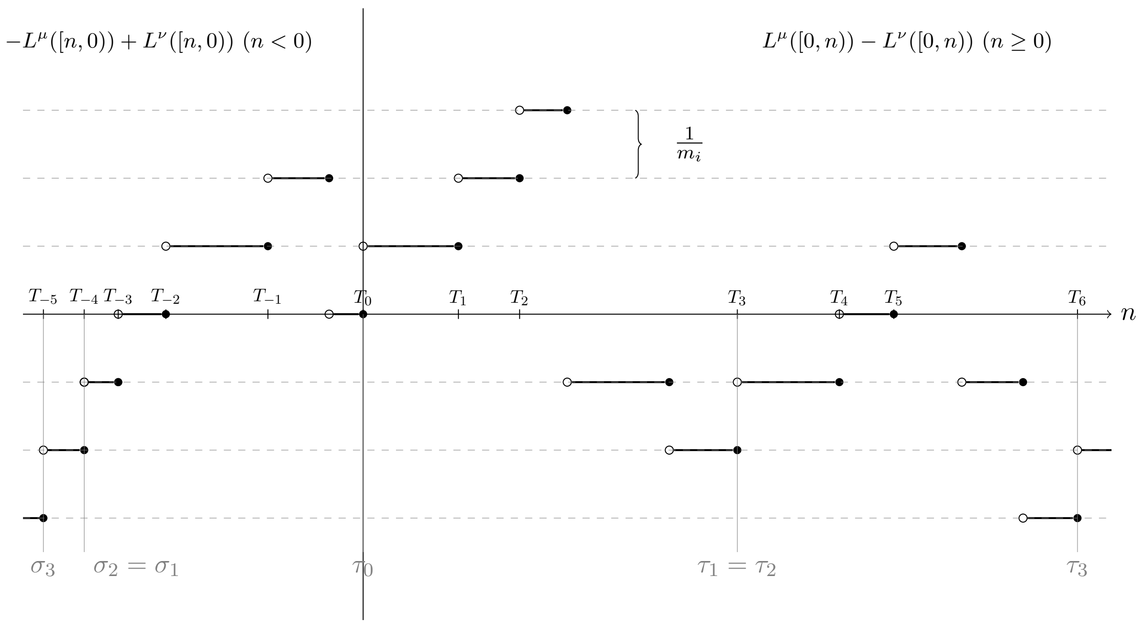

Figure 2.2: A possible profile of local time differences over the excursion $[\sigma_3, \tau_3 - 1]$. Upward jumps are of size $1/m_i$, downward jumps are a positive integer multiple of $1/m_i$, the actual value depending on the colour of the ball at the location of the jump.

Proof. We observe, from the strong Markov property, that $\xi_n = (X_{T_{n-1}+1}, \dots, X_{T_n})$, $n \in \mathbb{Z}$, are independent and identically distributed random vectors. Hence their shift is stationary and ergodic, see for example [12, 8.4.5]. By the ergodic theorem, see e.g. [12, 8.4.1], $\mathbb{P}_i$-almost surely,

$$\lim_{n\to\infty} \frac{1}{n} \Big\{ \sum_{k=0}^{T_n - 1} \sum_{\ell=k+1}^{\infty} \theta(k,\ell)\psi(\ell - k) \Big\} = \mathbb{E}_i \sum_{\ell=1}^{\infty} \theta_\omega(0,\ell)\psi(\ell) = \mathbb{E}_i^\oplus \psi(T).$$

Similarly,

$$\lim_{n\to\infty} \frac{1}{n} \Big\{ \sum_{k=0}^{T_n - 1} \sum_{\ell=-\infty}^{k-1} \theta(\ell,k)\psi(k-\ell) \Big\} = \mathbb{E}_i \sum_{k=0}^{T_1 - 1} \sum_{\ell=-\infty}^{k-1} \theta_\omega(\ell,k)\psi(k-\ell).$$

The expectation equals

$$\sum_{j\in\mathcal{S}} \frac{m_j}{m_i} \mathbb{E}_j \sum_{\ell=-\infty}^{-1} \theta_\omega(\ell,0)\psi(-\ell) = \frac{1}{m_i} \mathbb{E} \sum_{\ell=-\infty}^{-1} \theta_\omega(\ell,0)\psi(-\ell) = \frac{1}{m_i} \mathbb{E} \sum_{\ell=1}^{\infty} \theta_\omega(0,\ell)\psi(\ell) = \mathbb{E}_i^\oplus \psi(T),$$

using translation invariance of θ. Also recall that if $y \neq 0$ and $X_0 \neq i$, then $\theta_\omega(0,y) = 0$. The second statement follows in the same manner. $\square$

Proof of Theorem 12. We now look at the sequence

$$\tau_n = \min\big\{T_k \geq 0\colon L^\mu([0,T_k)) - L^\nu([0,T_k)) \leq \tfrac{-n}{m_i}\big\}.$$

Let $d_n = L^\mu([0,\tau_n)) - L^\nu([0,\tau_n))$ and define

$$\sigma_n = \max\big\{k \leq 0\colon -L^\mu([k,0)) + L^\nu([k,0)) = d_n\big\}.$$

(τ_n) and (σ_n) are well-defined subsequences of $(T_n\colon n \in \mathbb{Z})$, $\mathbb{P}_i$-almost surely, by Lemma 2.3.3. Moreover, $\tau_n \uparrow \infty$, $\sigma_n \downarrow -\infty$ and by construction $[\sigma_n, \tau_n - 1]$ is an excursion, see Figure 2.2. By Lemma 2.5.3

$$\sum_{k=\sigma_n}^{\tau_n-1} \Big\{\psi\big(\tau_*(k)-k\big) + \sum_{\ell\in\tau_*^{-1}(k)} \psi\big(\tau_*(\ell)-\ell\big)\Big\} \leq \sum_{\substack{\sigma_n\leq k\leq\tau_n-1\\ \ell\in\mathbb{Z}}} \theta(k,\ell)\psi(k-\ell) + \sum_{\substack{\sigma_n\leq \ell\leq\tau_n-1\\ k\in\mathbb{Z}}} \theta(k,\ell)\psi(k-\ell).$$

Lemma 2.5.4 shows that the left hand side is asymptotically equivalent to $2m_i\, L^i([\sigma_n,\tau_n])\, \mathbb{E}_i\psi(T_*)$ and the right hand side to $2m_i\, L^i([\sigma_n,\tau_n])\, \mathbb{E}_i^{\oplus}\psi(T)$, from which we conclude that $\mathbb{E}_i\psi(T_*) \leq \mathbb{E}_i^{\oplus}\psi(T)$. □

2.6 Concluding remarks and open problems

Non-Markovian setting. Theorem 9 and Theorem 10 remain valid in a more general non-Markovian setting. We require that under the σ-finite measure $\mathbb{P}$ the stochastic process X, taking values in the countable state space $\mathcal{S}$, is stationary with a strictly positive stationary σ-finite measure $(m_i : i \in \mathcal{S})$. The probability measure $\mathbb{P}_i$ is then defined by conditioning X on the event $\{X_0 = i\}$. We further require that, for every $i, j \in \mathcal{S}$, the random sets $\{n \in \mathbb{N}\colon X_n = j\}$ and $\{n \in \mathbb{N}\colon X_{-n} = j\}$ are infinite $\mathbb{P}_i$-almost surely. Then both theorems carry over to this conditioned process. Further technical conditions are required to generalize Lemma 2.5.4 and hence extend Theorem 4 to the non-Markovian setting. Theorem 3 however fully exploits the Markov structure and cannot be generalized easily.

General initial distribution. Although our main focus is on the case where the initial distribution is the Dirac measure δ_i for some $i \in \mathcal{S}$, the statements of Proposition 2.2.1 and 2.2.2 allow general initial distributions μ. By conditioning on the initial state one can see that a *sufficient* condition for existence of the solution is that the target measure ν admits a decomposition $\nu = \sum_{i\in\mathcal{S}} \nu^{(i)}\mu_i$, where $\nu^{(i)}$ are probability measures on $\mathcal{S}$, such that $m_i\nu_j^{(i)}/m_j$ are integers for all $i, j \in \mathcal{S}$. We do not believe that this is also a *necessary* condition.

Randomized shifts. If the target measure ν fails to satisfy the integer condition in Theorem 9 (b), extra randomization is needed to solve the embedding problem. With extra randomness any target measure ν may be embedded in a way similar to the extra head schemes in [20]: Take a random variable $U \sim \mathrm{Uniform}(0,1)$ and define

$$T_{\texttt{rand}} := \min\Big\{n \geq 0\colon L^i([0,n]) - \sum_{j\in\mathcal{S}} \nu_j\, L^j([0,n]) \leq \tfrac{U}{m_i}\Big\}. \tag{2.6.1}$$

Then $T_{\mathtt{rand}}$ is an unbiased shift embedding ν. We see that if the integer condition holds, the sample value of U becomes irrelevant and we recover the non-randomized solution T_* defined in Theorem 10.

Brownian motion and optimal shifts. Last et al. [23] discuss the Skorokhod embedding problem for a two-sided Brownian motion $(B_t)_{t\in\mathbb{R}}$. In this context a random time T solves the embedding problem if $(B_{T+t} - B_T)_{t\in\mathbb{R}}$ is a standard two-sided Brownian motion independent of B_T and the law of B_T is ν. They show that for *any* target distribution ν not charging the origin the stopping time $T_* = \inf\{t > 0 \colon L^0_t = L^\nu_t\}$, where $(L^x_t \colon t > 0)$ is the process of local times at level x and $L^\nu_t := \int L^x_t \, \nu(dx)$, solves the embedding problem. They further show that every solution T that is a stopping time satisfies $\mathbb{E}[T^{\frac{1}{4}}] = \infty$ while under a mild condition on ν the constructed solution T_* satisfies $\mathbb{E}[T_*^\beta] < \infty$ for all $\beta < \frac{1}{4}$. Some techniques of the present chapter can be adapted to improve the results of [23] by showing that $\mathbb{E}[T^{\frac{1}{4}}] = \infty$ even for non-negative solutions which are not necessarily stopping times, and also to show a strong optimality result similar to Theorem 12, i.e. that $\mathbb{E}_0\psi(T_*) \le \mathbb{E}_0\psi(T)$ simultaneously for all non-negative concave functions ψ. Chapter 3 is entirely devoted to discuss the proof of this result.

Signed shifts. The optimality result of Theorem 12 cannot be extended easily to random times T that can take both positive and negative values. Indeed, starting from such a solution T and associating an allocation rule τ to it, we may still make local improvements by repairing crossings, but now there is more than one way to repair a crossing and the optimal way to do this appears to involve nonlocal choices. To get a feeling for the difficulties, we look at a two-sided stable matching strategy that at a first glance looks like a good candidate for an optimal solution. In the language of Section 2.5 we match a coloured ball to a white ball if *both* the coloured ball is the nearest coloured ball to the white ball, and the white ball is the nearest white ball to the coloured ball (resolving possible ties in some deterministic way). Locally, the resulting allocation rule may be better or worse than the one coming from our one-sided stable matching. Consider, for example, configuration of balls in the order white–coloured–white–coloured placed at distances a, b, c such that $b < a, c$. The two-sided algorithm matches the middle balls and, if other balls are sufficiently far away, the outer balls, which gives a contribution of $\psi(b) + \psi(a + b + c)$. One-sided stable matching matches the first pair and the second pair and gives $\psi(a) + \psi(c)$, and each contribution could be smaller or larger depending on the relative size of a, b, c. Even finding the optimal moment properties of signed shifts is an open problem.

Random fields. A vast open area of possible further research are embedding problems for multiparameter processes and random fields. In higher dimensions stable allocation procedures no longer have optimal moment properties, see for example Holroyd, Peres and Schramm [19], so other methods need to be considered. It would be particularly interesting to investigate embedding problems for spin systems such as the infinite volume Gibbs measure of the Ising model at high temperature.

Acknowledgments: We would like to thank Vitali Wachtel for valuable discussions on Wiener-Hopf decompositions and for suggesting the proof of Lemma 2.4.5, which allowed us to remove a regularity condition on the asymptotic Green's function of the Markov chain. We would like to thank Mathias Beiglböck and Martin Huesmann for enlightening discussions during visits to the Hausdorff Institute and Eurandom. We would particularly like to thank Martin Huesmann for conjecturing the result of Theorem 12. Last but not least, we would like to thank Günter Last and an anonymous referee for several insightful comments.

Chapter 3

Optimal embeddings by unbiased shifts of Brownian motion

This chapter is a joint work with my supervisor, Peter Mörters and has been submitted for publication.

3.1 Introduction

A random time T is called an *unbiased shift* of the two sided Brownian motion $(B_t \colon t \in \mathbb{R})$ if the shifted process $(B_{T+t} \colon t \in \mathbb{R})$ is again a two-sided Brownian motion. This notion was introduced and studied by Last et al. in [23], where the additional requirement of measurability of T with respect to the process $(B_t \colon t \in \mathbb{R})$ is made, which we drop here for greater generality. The concept of unbiased shifts goes back to a similar idea for coin tosses, known as the extra head problem in [18, 25, 20], and, more generally, the concept of shift-couplings, see the work of Thorisson [32] and Chapter 1.

In [23] the authors solve the *embedding problem* for unbiased shifts. Namely, given two orthogonal probability measure μ, ν on the real line such that $B_0 \sim \mu$ they construct a nonnegative unbiased shift T_* such that $B_{T_*} \sim \nu$. The solution of [23] can easily be described explicitly. Let $(L_t^x \colon x \in \mathbb{R}, t \geq 0)$ be a continuous version of the local time for $(B_t \colon t \geq 0)$. We use this to build two continuous additive functionals by letting

$$L_t^\nu := \int L_t^x \, \nu(dx), \text{ and } L_t^\mu := \int L_t^x \, \mu(dx),$$

and obtain the solution as

$$T_* := \inf\{t > 0 \colon L_t^\nu = L_t^\mu\}. \tag{3.1.1}$$

Note that T_* occurred in the context of one-sided (Skorokhod) embedding problems in the work of Bertoin and Le Jan [6], is reminiscent of extra head schemes in [25, 20] or [27], and turns out to be closely related to allocation and transport problems, see [19, 5].

The present paper is concerned with the problem whether this natural and explicit solution of the embedding problem is optimal in the sense that it minimises certain moments. To see which moments should be looked at, Last et al. [23] have investigated finiteness of moments for unbiased shifts under the assumption $B_0 = 0$, and have shown that, for any unbiased shift T embedding a measure without atom at zero, the random variable $|T|$ has infinite square-root moment. Under the additional assumption that T is a nonnegative stopping time they obtain

$$\mathbb{E}\big[T^{1/4}\big] = \infty.$$

It is conjectured in [23] that this holds without the assumption that T is a stopping time, and this conjecture will be confirmed in this paper, see Remark 1(c) below. The results of [23] show that in order to understand optimality we need to focus on moments of fractional order smaller than $\frac{1}{2}$, or more generally moments taken with respect to concave gauge functions.

We denote by $\mathbb{P}_\mu, \mathbb{E}_\mu$ probability and expectation on a probability space supporting a two-sided Brownian motion $(B_t \colon t \in \mathbb{R})$ with $B_0 \sim \mu$. We can now state our main result.

Theorem 13. *Assume that μ and ν are two orthogonal probability measures on $\mathbb{R}$. For any non-negative unbiased shift T embedding ν and all $\psi \colon [0,\infty) \to [0,\infty)$ concave,*

$$\mathbb{E}_\mu \psi(T_*) \leq \mathbb{E}_\mu \psi(T),$$

where T_* is the unbiased shift constructed in (3.1.1)*.*

Remark 4

(a) This result is closely related to a similar optimality result in [27], or in Chapter 2 for the case of discrete-time Markov chains. Although our proof relies on a discrete approximation, we have been unable to derive Theorem 13 directly from the results of the previous Chapter. Instead, we use a concavity inequality, stated as Lemma 3.3.2(*b*) below, that may be of independent interest.

(b) We make no assumptions on measurability of T with respect to the Brownian motion, i.e., T is allowed to use additional randomisation beyond that taken from the Brownian motion.

(c) As T_* is a stopping time, we have $\mathbb{E}_\mu T_*^{1/4} = \infty$ by Theorem 8.1 in [23], and hence

$$\mathbb{E}_\mu T^{1/4} = \infty$$

for *all* non-negative solutions T of the embedding problem.

(d) Under mild conditions on ν we have $\mathbb{E}_\mu T_*^{\alpha} < \infty$ for $\alpha < \frac{1}{4}$ by Theorem 8.2 in [23].

(e) The result is strongly reminiscent of optimality results for Skorokhod embeddings, as given, for example, in the classical papers [10, 11]. Optimality of Skorokhod embeddings and optimal transport is also the topic of a lot of current research, of which [5] is a major highlight.

3.2 Preliminaries and outline of the proof

We recall the framework of [23], which will be used as a main reference throughout this paper. Let $(\Omega, \mathcal{A}, \mathbb{P}_0)$ be such that Ω is the space of continuous functions $\omega\colon \mathbb{R} \to \mathbb{R}$ - with the natural topology that is uniform convergence on compact sets - equipped with the Borel σ-algebra $\mathcal{A}$ and the distribution $\mathbb{P}_0$ of two-sided Brownian motion with $\omega_0 = 0$. For all $x \in \mathbb{R}$ we define $\mathbb{P}_x$ to be the distribution of $\omega + x$ and introduce the σ-finite measure $\mathbb{P} := \int \mathbb{P}_x \, dx$, which is invariant under the shifts $\omega \mapsto s\omega$ defined by $(s\omega)(\cdot) = \omega(\cdot - s)$, for all $s \in \mathbb{R}$. As usual, expectations with respect to $\mathbb{P}, \mathbb{P}_x$ will be denoted by $\mathbb{E}, \mathbb{E}_x$, respectively.

Let $(\ell^x\colon x \in \mathbb{R})$ be a continuous, in time and space, version of the family of local times of ω at level x, and for any locally finite measure ν on $\mathbb{R}$ set $\ell^\nu := \int \ell^x \, \nu(dx)$. We note that the local times ℓ^x and ℓ^ν are random measures and depend on $\omega \in \Omega$. When necessary, this dependence will be indicated as $\ell^{\cdot}_\omega$. We may and shall assume that $\ell^x_{z+\omega} = \ell^{x-z}_\omega$, for all $x, z \in \mathbb{R}$, for $\mathbb{P}$-almost every ω.

A *transport rule* is a Markov kernel $\theta\colon \Omega \times \mathbb{R} \times \mathcal{B} \to [0,1]$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. We use the notation $\theta_\omega(ds, dt) = \theta_\omega(s, dt)\ell^\mu(ds)$, and $\theta_\omega(A, B) = \int_A \theta_\omega(s, B)\ell^\mu(ds)$ for any $A, B \in \mathcal{B}$. An interpretation of the Markov kernels $\theta(s, \cdot)$ is that each site $s \in \mathbb{R}$ sends out unit mass to the real line. A transport rule θ is called *translation invariant* if $\theta_\omega(s, A) = \theta_{t\omega}(s - t, A - t)$, for any $(s, A) \in \mathbb{R} \times \mathcal{B}$, $t \in \mathbb{R}$, for $\mathbb{P}$-almost every ω.

We now briefly summarise results of [23], which are derived from the abstract results of [24]. Given a translation invariant transport rule θ we obtain an unbiased shift T by letting

$$\mathbb{P}_\mu(\omega \in A, T \in B) = \iint_A \theta_\omega(0, B)\mathbb{P}_x(d\omega)\, \mu(dx), \text{ for } A \in \mathcal{A}, B \in \mathcal{B}. \tag{3.2.1}$$

Conversely, given an unbiased shift T we can construct a translation invariant transport rule θ by letting $\theta_\omega(s, B) = \mathbb{P}_\mu(T \in B - s \,|\, s\omega)$, for $s \in \mathbb{R}, B \in \mathcal{B}, \omega \in \Omega$, where we use a suitably regularised version of conditional probabilities so that (3.2.1) holds.

The transport rules associated in this way with nonnegative unbiased shifts are *forward looking* in the sense that $\theta_\omega(s, (-\infty, s)) = 0$ for all $s \in \mathbb{R}$. In the chapter we only consider forward looking transport rules and all results contain this assumption implicitly. An unbiased shift T solves the embedding problem for a pair of orthogonal probability measures μ, ν if and only if the associated transport rule satisfies the *balancing property*

$$\int \theta_\omega(s, dt)\, \ell^\mu_\omega(ds) = \ell^\nu_\omega(dt) \qquad \text{for } \mathbb{P}\text{-almost all } \omega.$$

If the unbiased shift is not randomised, i.e. a measurable function of ω, then the associated θ is an *allocation rule*, i.e. it is of the form $\theta_\omega(s, A) = 1_A(\tau_\omega(s))$ for some $\tau_\omega\colon \mathbb{R} \to \mathbb{R}$. In this case each site $s \in \mathbb{R}$ is assigned to a new site $\tau_\omega(s) \in \mathbb{R}$. The allocation rule associated to the random times T_* is given by

$$\tau_*(s) := \inf\{t > s\colon \ell^\mu[s, t] = \ell^\nu[s, t]\}, \qquad \text{for all } s \in \mathbb{R}. \tag{3.2.2}$$

Let us now outline the proof of Theorem 13. The first part looks at what happens *pathwise*.

With every transport rule we associate a local cost. Given a fixed ω, we show that on carefully chosen intervals called excursions, the best possible cost is offered by the allocation rule τ_*. We do this by proving an analogous discrete result and then taking a suitable limit. The second part of the proof uses *ergodic theory* to translate the local cost optimality of τ_* into a result on the moments of T_*.

3.3 Pathwise level

In what follows, we fix a path ω. An *excursion* $\mathcal{E}$ is any interval of the form $[s, \tau_*(s)]$, for $s \in \mathbb{R}$. Observe that τ_* maps $\mathcal{E}$ onto itself. Our main goal is to show that the allocation rule τ_* offers the best possible cost inside an arbitrary excursion $\mathcal{E}$. More precisely we will devote this section to proving the following.

Proposition 3.3.1. *Given an excursion $\mathcal{E}$, for any transport rule θ balancing ℓ^μ and ℓ^ν,*

$$\iint_{\mathcal{E}\times\mathbb{R}} \psi(t-s)\,\theta(ds,dt) + \iint_{\mathbb{R}\times\mathcal{E}} \psi(t-s)\,\theta(ds,dt) \geq 2\int_{\mathcal{E}} \psi(\tau_*(s)-s)\ell^\mu(ds), \tag{3.3.1}$$

for all $\psi\colon [0,\infty) \to [0,\infty)$ concave.

The left hand side in (3.3.1) is called the *cost* of the transport rule θ over the interval $\mathcal{E}$. Note that the right hand side is then the cost of the allocation rule given by τ_* over the same interval. By subtracting a constant from both sides of the equation we may henceforth assume that $\psi(0) = 0$. As a remark on the double integral notation, we note that first component $\mathcal{E}$ in the product range $\mathcal{E} \times \mathbb{R}$ refers to the first argument of $\theta(ds, \cdot)$ and the second component $\mathbb{R}$ refers to the second argument of $\theta(\cdot, dt)$.

We start by establishing a similar result in a discrete setting. The inequality we establish below is of a general nature and may be of independent interest. A map $\tau\colon A \to B$ between two discrete and disjoint sets $A, B \subset \mathbb{R}$ is the *stable allocation map* if

$$\tau(a) = \min\big\{b \in B\colon b > a, |B \cap [a,b]| = |A \cap [a,b]|\big\}.$$

Lemma 3.3.2. *Let $a_1 > a_2 > a_3 > \ldots$ and $b_1 < b_2 < b_3 < \ldots$ be disjoint real sequences, such that we have $a_n \searrow -\infty$ and $b_n \nearrow \infty$.*

(a) The stable allocation map $\tau\colon \{\ldots, a_3, a_2, a_1\} \to \{b_1, b_2, b_3, \ldots\}$ is well-defined and there exists $N \in \mathbb{N}$ such that $\tau(a_i) = b_i$ for all $i \geq N$.

(b) For every concave function $\psi\colon [0,\infty) \to [0,\infty)$ and nonnegative matrix $\pi = (\pi_{i,j}\colon i,j \in \mathbb{N})$ with the properties that

- $\pi_{i,j} = 0$ *if* $a_i > b_j$,
- $\sum_{j=1}^{\infty} \pi_{i,j} = 1$ *for all* $i \in \{1, \ldots, N\}$,

– $\sum_{i=1}^{\infty} \pi_{i,j} = 1$ *for all* $j \in \{1, \ldots, N\}$,

we have

$$\sum_{i=1}^{N}\sum_{j=1}^{\infty} + \sum_{i=1}^{\infty}\sum_{j=1}^{N} \pi_{i,j}\psi(b_j - a_i) \geq 2\sum_{i=1}^{N} \psi(\tau(a_i) - a_i). \tag{3.3.2}$$

In the special case that $a_1 < b_1$ we have $\tau(a_i) = b_i$ for all i, and hence $N \in \mathbb{N}$ can be chosen arbitrarily. Then our result becomes the following general result, which to the best of our knowledge is new, too.

Corollary 3.3.3. *For every double-sided increasing sequence* $(a_n : n \in \mathbb{Z})$ *that is unbounded from above and below, every stochastic matrix* $\pi = (\pi_{i,j} : i, j \in \mathbb{N})$, *and every concave function* $\psi : [0, \infty) \to [0, \infty)$ *we have*

$$\sum_{i=1}^{n}\sum_{j=1}^{\infty} + \sum_{i=1}^{\infty}\sum_{j=1}^{n} \pi_{i,j}\psi(a_i - a_{-j}) \geq 2\sum_{i=1}^{n} \psi(a_i - a_{-i}).$$

Proof of Lemma 3.3.2(a). Given $a \in A$ the set $A \cap [a, \infty)$ is finite, but the set $B \cap [a, \infty)$ is infinite. Hence, on the one hand, there exists $b \in B$ such that $|B \cap [a, b]| \geq |A \cap [a, \infty)| \geq |A \cap [a, b]|$, while on the other hand $|B \cap [a, a]| = 0 < 1 = |A \cap [a, a]|$. This implies that there exists $b' \in [a, b]$ with $|B \cap [a, b']| = |A \cap [a, b']|$, and hence τ is well-defined.

We now show that there exists $N \in \mathbb{N}$ such that, for all $n \geq N$, we have $\tau(a_n) = b_n$. If $a_1 < b_1$ one can choose $N = 1$. Otherwise define an integer-valued function $f : [b_1, \infty) \to \mathbb{R}$ by

$$f(x) = \big|A \cap [b_1, x]\big| - \big|B \cap [b_1, x]\big|.$$

Let $M \in \mathbb{Z}$ be the minimum of f on $[b_1, a_1]$. Note that $f(a_1) > M$ and on $[a_1, \infty)$ the function f is decreasing to $-\infty$ by downward jumps of size one. Hence there exists $n > 1$ with $f(b_n) = M - 1$. Observe that $f(b_1) = -1$ and $|B \cap [b_1, b_n]| = |A \cap [b_1, b_n]| - (M - 1)$. Clearly, for all $m \geq n$ we have $|A \cap [a_m, b_m]| = m = |B \cap [a_m, b_m]|$ while, for $j < m$, we have $|A \cap [a_m, b_j]| > j = |B \cap [a_m, b_j]|$. Hence, $\tau(a_m) = b_m$, as required. □

Proof of Lemma 3.3.2(b). This is a variant of Lemma 5.3 in [27]. We say that π crosses the pair (a_i, b_j) if there exist indices $k < i$ and $l > j$ such that $\pi_{kj} > 0$ and $\pi_{il} > 0$. Such a crossing can be repaired by replacing the matrix π by a new matrix π' given by

$$\pi'_{kj} = \pi_{kj} - (\pi_{kj} \wedge \pi_{il}), \quad \pi'_{il} = \pi_{il} - (\pi_{kj} \wedge \pi_{il}),$$
$$\pi'_{ij} = \pi_{ij} + (\pi_{kj} \wedge \pi_{il}), \quad \pi'_{kl} = \pi_{kl} + (\pi_{kj} \wedge \pi_{il}),$$

leaving all other entries untouched. If π satisfies the conditions of (b), then so does π'. By concavity of the function ψ we get

$$\psi(b_j - a_k) + \psi(b_l - a_i) \geq \psi(b_j - a_i) + \psi(b_l - a_k).$$

Hence the left hand side of (3.3.2) decreases when we replace π by π'. If we systematically repair all the crossings of pairs (a_i, b_j) with $1 \leq i, j \leq n$, using the repair algorithm described in Chapter 2, see Lemma 2.5.1 and Lemma 2.5.2, we end up with a matrix π^* with entries $\pi^*_{ij} = \mathbb{1}\{\tau(a_i) = b_j\}$, which satisfies

$$\sum_{i=1}^{N}\sum_{j=1}^{\infty} + \sum_{i=1}^{\infty}\sum_{j=1}^{N} \pi_{i,j}\psi(b_j - a_i) \geq \sum_{i=1}^{N}\sum_{j=1}^{\infty} + \sum_{i=1}^{\infty}\sum_{j=1}^{N} \pi^*_{i,j}\psi(b_j - a_i).$$

Plugging in the value of π^* the right hand side gives the form of (3.3.2). □

We now use Lemma 3.3.2 to get a continuous inequality. Let $\mathcal{E} = (b_0, a_0] \subset \mathbb{R}$ be an excursion and $M = \ell^\mu(\mathcal{E})$. Given $n \in \mathbb{N}$ we pick $a_1 > a_2 > \ldots > a_n$ such that $\ell^\mu(a_i, a_{i-1}) = \frac{M}{n}$ and $b_1 < b_2 < \ldots < b_n$ such that $\ell^\nu(b_{j-1}, b_j) = \frac{M}{n}$, for $1 \leq i, j \leq n$. Note that $a_n = b_0$ and $b_n = a_0$. Additionally, a_i, b_i, for $i > n$, are chosen in such a way that $a_i \searrow -\infty$, $b_i \nearrow \infty$, and

$$\sup_{i=n}^{\infty} \big(b_{i+1} - b_i\big), \quad \sup_{j=n}^{\infty} \big(a_j - a_{j+1}\big) \longrightarrow 0.$$

Then, if $\tau_n \colon \{\ldots, a_3, a_2, a_1\} \to \{b_1, b_2, b_3, \ldots\}$ is the (discrete) stable allocation map, in Lemma 3.3.2 we may choose $N = n$. Now suppose a balancing transport rule θ is given. We define

$$\pi_{i,j} = \frac{n}{M}\,\theta((a_i, a_{i-1}], (b_{j-1}, b_j])$$

and note that $\pi = (\pi_{i,j} \colon i, j \in \mathbb{N})$ satisfies the conditions of Lemma 3.3.2 (b). Hence we have

$$\sum_{i=1}^{n}\sum_{j=1}^{\infty} + \sum_{i=1}^{\infty}\sum_{j=1}^{n} \pi_{i,j}\psi(b_j - a_i) \geq 2\sum_{i=1}^{n} \psi(\tau_n(a_i) - a_i). \tag{3.3.3}$$

Multiply both sides by M/n and let n go to infinity. We will argue that the right and left hand side of (3.3.3) converge to those of (3.3.1).

First we show convergence of the left hand side of (3.3.3). Given $\{a_i\}_{i\in\mathbb{N}}$ and $\{b_j\}_{j\in\mathbb{N}}$ as constructed above, let $g_n(a) = a_i$, if $a \in (a_i, a_{i-1}]$ and $h_n(b) = b_j$, if $b \in (b_{j-1}, b_j]$.

Lemma 3.3.4. *For ℓ^μ-almost every a we have $g_n(a) \to a$, and for ℓ^ν-almost every b we have $h_n(b) \to b$.*

Proof. It suffices to prove the first claim. The result is trivial if $a \leq b_0$. Given $\varepsilon > 0$, for ℓ^μ-almost every a with $b_0 + \varepsilon < a \leq a_0$ there exists $\eta > 0$ such that $\ell^\mu(a - \varepsilon, a) \geq \eta$. Hence, for all $n > M/\eta$, there exists $a_i \in (a - \varepsilon, a)$ which implies $0 < a - g_n(a) < \varepsilon$. □

Note that

$$\frac{M}{n}\sum_{i=1}^{n}\sum_{j=1}^{\infty} \pi_{i,j}\psi(b_j - a_i) = \iint_{\mathcal{E}\times\mathbb{R}} \psi\big(h_n(b) - g_n(a)\big)\theta(da, db).$$

Now we compare the integrand with $\psi\big(b - a\big)$. Adding and subtracting $\psi(b - g_n(a))$ and then using the triangle inequality we get

$$\begin{aligned}|\psi(h_n(b)-g_n(a))-\psi(b-a)| &\le |\psi(h_n(b)-g_n(a))-\psi(b-g_n(a))|+|\psi(b-a)-\psi(b-g_n(a))|\\ &\le \psi(h_n(b)-b)+\psi(a-g_n(a)),\end{aligned}$$

where in the second inequality we used the sub-additivity of ψ as follows,

$$\begin{aligned}|\psi(h_n(b)-g_n(a))-\psi(b-g_n(a))| &= \psi(h_n(b)-g_n(a))-\psi(b-g_n(a))\\ &= \psi(h_n(b)-b+b-g_n(a))-\psi(b-g_n(a)) \le \psi(h_n(b)-b).\end{aligned}$$

Using this estimate on the integrand, we get

$$\iint_{\mathcal{E}\times\mathbb{R}} \psi\big(h_n(b)-g_n(a)\big)-\psi(b-a)\,\theta(da,db) \le \iint_{\mathcal{E}\times\mathbb{R}} \psi(h_n(b)-b)+\psi(a-g_n(a))\,\theta(da,db).$$

The integrand on the right is bounded and converges to zero, θ-almost everywhere, by Lemma 3.3.4. Hence the left hand side of (3.3.3), multiplied by $\frac{M}{n}$, converges to the required limit,

$$\iint_{\mathcal{E}\times\mathbb{R}} + \iint_{\mathbb{R}\times\mathcal{E}} \psi(b-a)\,\theta(da,db).$$

Second we show convergence of the right hand side of (3.3.3). The key to this is the following lemma.

Lemma 3.3.5. *For ℓ^μ-a.e. $a\in\mathcal{E}$, we have* $\lim_{n\to\infty}\tau_n(h_n(a))=\tau_*(a)$.

Proof. We define

$$f\colon [a,\tau_*(a)]\to[0,\infty), f(x)=\ell^\mu[a,x]-\ell^\nu[a,x],$$

and

$$f_n\colon [a,\tau_*(a)]\to\mathbb{R}, f_n(x)=\tfrac{M}{n}\big(|\{i\colon a_i\in[g_n(a),x]\}|-|\{j\colon b_j\in[g_n(a),x]\}|\big).$$

The proof is organised into five steps.

Step 1: $|f(x)-f_n(x)|\le\frac{4M}{n}$.

Proof. Denote $k_1=|\{i\colon a_i\in[g_n(a),x]\}|$ and $k_2=|\{j\colon b_j\in[g_n(a),x]\}|$, then $f_n(x)=\frac{M}{n}(k_1-k_2)$ and

$$\ell^\mu[a,x]=(k_1-1)\frac{M}{n}+\ell^\mu[g_n(x),x]-\ell^\mu[g_n(a),a],$$

and hence $|\ell^\mu[a,x]-k_1\frac{M}{n}|\le\frac{2M}{n}$. Similarly, $|\ell^\nu[a,x]-k_2\frac{M}{n}|\le\frac{2M}{n}$, which implies the statement. □

Recall that $\tau_*(a)=\inf\{x>a\colon f(x)=0\}$ and

$$\tau_n(g_n(a))=\inf\{x>g_n(a)\colon f_n(x)=0\}.$$

Step 2: For ℓ^μ-almost every $a \in \mathcal{E}$, there exists $\varepsilon > 0$ such that,

$$f(a+x) \geq \tfrac{5}{6}\ell^\mu[a, a+x] \text{ for all } x \in (0,\varepsilon), \text{ and}$$

$$f(\tau_*(a)-x) \geq \tfrac{5}{6}\ell^\nu[\tau_*(a)-x, \tau_*(a)] \text{ for all } x \in (0,\varepsilon).$$

Proof. For ℓ^μ-almost every $a \in \mathcal{E}$, there exists $\varepsilon > 0$ such that,

$$\ell^\nu[a, a+x] \leq \tfrac{1}{6}\ell^\mu[a, a+x] \text{ for all } x \in (0,\varepsilon).$$

see, e.g., Section 1.6, Theorem 3 in [15]. This implies the first property, and the second is analogous. □

Step 3: Let $G_n := \{a_i : i \in \{1,\dots,n\}, \nexists b_j \in [a_i, a_{i-7}]\}$. Then, for ℓ^μ-almost every a, there exists arbitrarily large n with $g_n(a) \in G_n$.

Proof. Fix $\eta > 0$ small for the moment. For ℓ^μ-almost every a, there exists $\delta_0 > 0$ such that

$$\ell^\nu[a-\delta, a+\delta] \leq \eta \ell^\mu[a-\delta, a+\delta], \qquad \text{for all } 0 < \delta < \delta_0.$$

Supposing that, for some $k \in \mathbb{N}$,

$$\tfrac{(k+2)M}{n} \leq \ell^\mu[a-\delta, a+\delta] \leq \tfrac{(k+3)M}{n}$$

we have

$$\big|\big\{a_i : i \in \{1,\dots,n\}, a_i \in [a-\delta, a+\delta]\big\}\big| \geq k$$

and

$$\big|\big\{b_j : j \in \{1,\dots,n\}, b_j \in [a-\delta, a+\delta]\big\}\big| \leq 1 + \eta(k+3).$$

By the pigeonhole principle therefore

$$\big|\big\{a_i \in G_n : a_i \in [a-\delta, a+\delta]\big\}\big| \geq k - 7(1+\eta(k+3)).$$

Hence, given $\varepsilon > 0$, we can find η, and hence $\delta_0 > 0$, such that

$$\ell^\mu\{x \in [a-\delta, a+\delta] : g_n(x) \notin G_n\} < \tfrac{\varepsilon}{4}\,\ell^\mu[a-\delta, a+\delta], \qquad \text{for all } 0 < \delta < \delta_0 \text{ and } n \geq n_0(\delta).$$

We can thus find a global δ, and a collection of intervals $[a-\delta, a+\delta]$ as above covering the set $[a, \tau_*(a)]$ at most twice, and at least up to a set E with $\ell^\mu(E) < \frac{\varepsilon}{2}$. The result follows as $\varepsilon > 0$ was arbitrary. □

Step 4: In this step $\varepsilon > 0$ is chosen according to the previous steps above. Using Step 3, for ℓ^μ-almost every a, we can choose n such that $g_n(a) \in G_n$ and

$$\tfrac{5M}{n} \leq \min\{f(x) : x \in [a+\varepsilon, \tau_*(a)-\varepsilon]\}.$$

Recall that $g_n(a) = a_i$, if $a \in (a_i, a_{i-1}]$. In this case we also write $g_n^{(k)}(a) = a_{i-k}$.

By Step 1 and choice of n we have

$$f_n(x) \geq f(x) - \tfrac{4M}{n} \geq \tfrac{M}{n} \text{ for all } x \in \big[a+\varepsilon, \tau_*(a)-\varepsilon\big].$$

By Step 2, using again Step 1,

$$f_n(x) \geq f(x) - \tfrac{4M}{n} \geq \tfrac{5}{6}\ell^\mu[a, g_n^{(7)}(a)] - \tfrac{4M}{n} \geq \tfrac{M}{n} \text{ for all } x \in \big[g_n^{(7)}(a), a+\varepsilon\big],$$

and the fact that $g_n(a) \in G_n$ implies that

$$f_n(x) \geq \tfrac{M}{n} \text{ for all } x \in \big[g_n(a), g_n^{(7)}(a)\big].$$

We thus obtain that

$$\tau_n(g_n(a)) = \inf\{x > g_n(a) \colon f_n(x) = 0\} \geq \tau_*(a) - \varepsilon.$$

Step 5: For ℓ^μ-almost every a, we have, for sufficiently large n,

$$\tau_n(g_n(a)) \leq \tau_*(a) + \varepsilon.$$

Proof. Recall that $h_n(b) = b_j$, if $b \in (b_{j-1}, b_j]$. In this case we also write $h_n^{(k)}(b) = b_{j+k}$. We note from Step 1 that $f_n(\tau_*(a)) \leq \frac{4M}{n}$. Let $G_n' := \{b_j \colon i \in \{1,\ldots,n\}, \nexists a_i \in [b_j, b_{j+5}]\}$. Then, as in Step 3, for ℓ^μ-almost every a, there exists arbitrarily large n with $\tau_*(a) \in G_n'$. We infer that $\tau_n(g_n(a)) \leq h_n^{(5)}(\tau^*(a)) \leq \tau_*(a) + \varepsilon$, if n is large enough. □

The result follows by combining Steps 4 and 5, as $\varepsilon > 0$ was arbitrarily small. □

To conclude we note that, using Lemmas 3.3.4 and 3.3.5,

$$\begin{aligned}\lim_{n\to\infty} \frac{M}{n} \sum_{i=1}^{n} \psi(\tau_n(a_i) - a_i) &= \lim_{n\to\infty} \int_{\mathcal{E}} \psi(\tau_n(g_n(a)) - g_n(a))\, \ell^\mu(da) \\ &= \int_{\mathcal{E}} \lim_{n\to\infty} \psi(\tau_n(g_n(a)) - g_n(a))\, \ell^\mu(da) = \int \psi(\tau_*(a) - a)\, \ell^\mu(da),\end{aligned}$$

by bounded convergence.

As a concluding remark we note that immediately right after $\tau_*(a)$ the function f increases only if $\tau_*(a)$ is a local minimum. However there are only countably many local minima, hence $\tau_*(a)$ is almost surely none of them.

3.4 Ergodicity

Denote by $\mathbb{P}^{(\mu)} = \int \mathbb{P}_x\, \mu(dx)$ the law of two-sided Brownian motion with $\omega_0 \sim \mu$, and by S^r the generalized inverse of the local time ℓ^μ, that is

$$S^r := \begin{cases} \sup\{t \geq 0 : \ell^\mu[0,t] = r\}, & \text{if } \ r \geq 0, \\ \sup\{t < 0 : \ell^\mu[t,0] = -r\}, & \text{if } \ r < 0. \end{cases} \tag{3.4.1}$$

Then $\mathbb{P}^{(\mu)}$ is invariant under the shifts $S^r\colon (\omega_s) \mapsto (\omega_{S^r+s})$, for every $r \in \mathbb{R}$, by Theorem 3.4 in [23]. In order to show that the family of shifts (S^r) is ergodic, we need to show that any invariant set A is trivial, i.e. $\mathbb{P}^{(\mu)}(A) \in \{0,1\}$. We follow a classical approximation approach. By [13, Appendix A.3] there exist sets A_r with $A_r \in \sigma(B_s : s \in [S^{-r}, S^r])$ for all $r > 0$, such that $\mathbb{P}^{(\mu)}(A\Delta A_r) \to 0$ as $r \to \infty$, where Δ denotes the symmetric difference of the two sets. As A is invariant under the shift S^{2r} we get

$$\mathbb{P}^{(\mu)}(A\Delta S^{2r}A_r) = \mathbb{P}^{(\mu)}(S^{2r}A\Delta S^{2r}A_r) = \mathbb{P}^{(\mu)}(S^{2r}(A\Delta A_r)) = \mathbb{P}^{(\mu)}(A\Delta A_r) \to 0.$$

Hence there exists $r_n \nearrow \infty$ such that

$$\sum_{n=1}^{\infty} \mathbb{P}^{(\mu)}(A\Delta S^{2r_n}A_{r_n}) < \infty,$$

and $A \setminus \bigcap_{s>0}\bigcup_{r_n>s} S^{2r_n}A_{r_n}$ and $\bigcap_{s>0}\bigcup_{r_n>s} S^{2r_n}A_{r_n} \setminus A$ are $\mathbb{P}^{(\mu)}$-nullsets. This implies

$$\mathbb{P}^{(\mu)}(A) = \mathbb{P}^{(\mu)}\Big(\bigcap_{s>0}\bigcup_{r_n>s} S^{2r_n}A_{r_n}\Big) \in \{0,1\},$$

using that the latter event is a tail event and hence trivial, see e.g. [26, Theorem 2.9].

Lemma 3.4.1. *Let $T \geq 0$ be an unbiased shift embedding ν and θ be the associated transport rule. Furthermore let $\psi\colon [0,\infty) \to [0,\infty)$ be concave with $\psi(0) = 0$, with the extension $\psi(z) = 0$ for $z \leq 0$. Let $S^r_{\mu+\nu}$ be the inverse of the local time $\ell^{\mu+\nu}$. Then, $\mathbb{P}^{(\mu+\nu)}$-almost surely,*

$$\lim_{r\to\infty} \frac{1}{r} \iint_0^{S^r_{\mu+\nu}} \psi(t-s)\,\theta(ds,dt) = \lim_{r\to\infty} \frac{1}{r} \iint_{S^{-r}_{\mu+\nu}}^0 \psi(t-s)\,\theta(ds,dt) = \tfrac{1}{2}\,\mathbb{E}_\mu \psi(T),$$

and,

$$\lim_{r\to\infty} \frac{1}{r} \int_0^{S^r_{\mu+\nu}} \int \psi(t-s)\,\theta(ds,dt) = \lim_{r\to\infty} \frac{1}{r} \int_{S^{-r}_{\mu+\nu}}^0 \int \psi(t-s)\,\theta(ds,dt) = \tfrac{1}{2}\,\mathbb{E}_\mu \psi(T),$$

where the first (second) part of the double integrals should be understood as integration with respect to the first (second) argument of $\theta(\cdot,\cdot)$.

Proof. Recall that by the unbiased shift T embeds ν. By the ergodic theorem, $\mathbb{P}^{(\mu+\nu)}$-almost surely,

$$\begin{aligned}
\lim_{r\to\infty} \frac{1}{r} \iint_0^{S^r_{\mu+\nu}} \psi(t-s)\,\theta(ds,dt) &= \lim_{r\to\infty} \frac{1}{r} \int_0^{S^r_{\mu+\nu}} \psi(t-s) \int \theta(s,dt)\,\ell^{\mu+\nu}(ds) \\
&= \lim_{r\to\infty} \frac{1}{r} \int_0^r \psi(t - S^s_{\mu+\nu}) \int \theta(S^s_{\mu+\nu}, dt)\,ds \\
&= \tfrac{1}{2}\,\mathbb{E}^{(\mu+\nu)} \int \psi(t)\,\theta(0,dt) = \tfrac{1}{2}\,\mathbb{E}_\mu \psi(T).
\end{aligned}$$

Note that the measure $\mu+\nu$ has total mass of 2, hence a normalizing $\frac{1}{2}$ factor above. The last equality stems from the assumption that the initial measure is μ, i.e. $B_0 \sim \mu$, and μ and ν are orthogonal measures. Similarly, we obtain $\mathbb{P}^{(\mu+\nu)}$-almost surely,

$$\begin{aligned}\lim_{r\to\infty} \frac{1}{r} \int_0^{S^r_{\mu+\nu}} \int \psi(t-s)\, \theta(ds, dt) &= \tfrac{1}{2}\, \mathbb{E}^{(\mu+\nu)} \int \psi(-s) \theta(ds, 0) \\ &= \tfrac{1}{2}\, \mathbb{E}^{(\nu)} \int \psi(-s)\, \theta(ds, 0).\end{aligned}$$

Using the generalized Campbell formula, see [23, (2.4)], we get

$$\begin{aligned}\mathbb{E}^{(\nu)} \int \psi(-s)\, \theta(ds, 0) &= \mathbb{E}^{(\nu)} \int \Big(\int \psi(-s) \mathbf{1}\{s+r \in [0,1]\}\, \theta_\omega(ds, 0) \Big)\, dr \\ &= \mathbb{E} \int \Big(\int \psi(-s) \mathbf{1}\{s+r \in [0,1]\}\, \theta_{r\omega}(ds, 0) \Big)\, \ell^\nu(dr) \\ &= \mathbb{E} \iint \mathbf{1}\{t \in [0,1]\} \psi(r-t)\, \theta_\omega(dt, r) \ell^\nu(dr),\end{aligned}$$

where in the last equation we used the shift-invariance of θ. Using the balancing property first and then the generalized Campbell formula again this equals

$$\mathbb{E} \iint \mathbf{1}\{t \in [0,1]\} \psi(r-t)\, \theta_\omega(t, dr) \ell^\mu(dt) = \mathbb{E}^{(\mu)} \int \psi(t)\, \theta(0, dt) = \mathbb{E}_\mu \psi(T).$$

The claims about backward time follow in the same manner. □

Proof of Theorem 13. Define the following stopping times

$$\begin{aligned}\rho(u) &= \inf\{t \geq 0 : \ell^\mu([0,t]) - \ell^\nu([0,t]) = -u\}, \\ \sigma(u) &= \sup\{t \leq 0 : -\ell^\mu([t,0]) + \ell^\nu([t,0]) = -u\},\end{aligned}$$

for all $u \in \mathbb{R}_+$. We have $\rho(u) \nearrow \infty$ and $\sigma(u) \searrow -\infty$, as $u \to \infty$, and, for all $u \in \mathbb{R}$, the interval $[\sigma(u), \rho(u)]$ forms an excursion. Hence, by Proposition 3.3.1,

$$\int_{\sigma(u)}^{\rho(u)} \int \psi(t-s)\, \theta(ds, dt) + \int \int_{\sigma(u)}^{\rho(u)} \psi(t-s)\, \theta(ds, dt) \geq 2 \int_{\sigma(u)}^{\rho(u)} \psi(\tau_*(s) - s)\, \ell^\mu(ds).$$

Applying Lemma 3.4.1 the left hand side is asymptotically equal to $\frac{1}{2}\, \ell^{\mu+\nu}([\sigma(u), \rho(u)]) \mathbb{E}_\mu \psi(T)$, and the right hand side to $\frac{1}{2}\, \ell^{\mu+\nu}([\sigma(u), \rho(u)]) \mathbb{E}_\mu \psi(T_*)$, which concludes the proof.

As a remark, we note that the asymptotic equality comes from the following consideration. The random set $\{S^r_\nu : r > 0\}$ can be written as the set of points of left increase of ℓ^ν, that is $\{S^r_\nu : r > 0\} = \{s > 0 : \ell^\nu(s-\epsilon, s] > 0\}$. It is clear that $\{\rho(u) : u > 0\} \subset \{S^r_\nu : r > 0\}$, for any $\epsilon > 0$. This implies that for any sequence $\{u_n\}_{n\in\mathbb{N}}$ such that $u_n \to \infty$, there exists a sequence $\{r_n\}_{n\in\mathbb{N}}$ with $r_n \to \infty$. The ergodic theorem guarantees almost sure convergence along all such sequences. Analogously, this line of argument extends to $\{S^r_\nu : r < 0\}$ and also for sets involving $S^{\cdot}_{\mu}$. Furthermore, it is clear that $\{S^r_{\mu+\nu} : r > 0\} = \{S^r_\mu : r > 0\} \cup \{S^r_\nu : r > 0\}$. $\square$

References

[1] F. Aurzada, H. Döring, M. Ortgiese, and M. Scheutzow. Moments of recurrence times for Markov chains. *Electron. Commun. Probab.*, 16:296–303, 2011.

[2] M. T. Barlow and M. Yor. (Semi-) martingale inequalities and local times. *Z. Wahrsch. Verw. Gebiete*, 55:237–254, 1981.

[3] R. F. Bass. *Probabilistic techniques in analysis*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[4] M. Beiglböck. Cyclical monotonicity and the ergodic theorem. *Ergodic Theory Dynam. Systems*, 35(3):710–713, 2015.

[5] M. Beiglböck, A. M. G. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *arXiv:1307.3656*, 2013.

[6] J. Bertoin and Y. Le Jan. Representation of measures by balayage from a regular recurrent point. *Ann. Probab.*, 20:538–548, 1992.

[7] K. Burdzy and M. Scheutzow. Forward Brownian motion. *Probab. Theory Related Fields*, 160(1-2):95–126, 2014.

[8] S. Chatterjee, R. Peled, Y. Peres, and D. Romik. Gravitational allocation to Poisson points. *Ann. of Math. (2)*, 172(1):617–671, 2010.

[9] X. Chen. How often does a Harris recurrent Markov chain recur? *Ann. Probab.*, 27:1324–1346, 1999.

[10] A. M. G. Cox and D. G. Hobson. An optimal Skorokhod embedding for diffusions. *Stochastic Process. Appl.*, 111:17–39, 2004.

[11] A. M. G. Cox and D. G. Hobson. A unifying class of Skorokhod embeddings: connecting the Azéma-Yor and Vallois embeddings. *Bernoulli*, 13:114–130, 2007.

[12] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002.

[13] R. Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[14] K. B. Erickson. The strong law of large numbers when the mean is undefined. *Trans. Amer. Math. Soc.*, 185:371–381, 1973.

[15] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[16] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[17] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *Amer. Math. Monthly*, 69:9–15, 1962.

[18] A. E. Holroyd and T. M. Liggett. How to find an extra head: optimal random shifts of Bernoulli and Poisson random fields. *Ann. Probab.*, 29:1405–1425, 2001.

[19] A. E. Holroyd, R. Pemantle, Y. Peres, and O. Schramm. Poisson matching. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45:266–287, 2009.

[20] A. E. Holroyd and Y. Peres. Extra heads and invariant allocations. *Ann. Probab.*, 33:31–52, 2005.

[21] M. Huesmann and K.-T. Sturm. Optimal transport from Lebesgue to Poisson. *Ann. Probab.*, 41(4):2426–2478, 2013.

[22] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[23] G. Last, P. Mörters, and H. Thorisson. Unbiased shifts of Brownian motion. *Ann. Probab.*, 42:431–463, 2014.

[24] G. Last and H. Thorisson. Invariant transports of stationary random measures and mass-stationarity. *Ann. Probab.*, 37:790–813, 2009.

[25] T. M. Liggett. Tagged particle distributions or how to choose a head at random. In *In and out of equilibrium (Mambucaba, 2000)*, volume 51 of *Progr. Probab.*, pages 133–162. Birkhäuser, Boston, MA, 2002.

[26] P. Mörters and Y. Peres. *Brownian motion.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[27] P. Mörters and I. Redl. Skorokhod embeddings for two-sided Markov chains. *Probability Theory and Related Fields*, 165(1):483–508, 2016. 26 pages.

[28] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability.* Oxford University Press, New York, 1995.

[29] J. Pitman and W. Tang. The Slepian zero set, and Brownian bridge embedded in Brownian motion by a spacetime shift. *Electron. J. Probab.*, 20:no. 61, 28, 2015.

[30] F. Spitzer. A Tauberian theorem and its probability interpretation. *Trans. Amer. Math. Soc.*, 94:150–169, 1960.

[31] H. Thorisson. Transforming random elements and shifting random fields. *Ann. Probab.*, 24(4):2057–2064, 1996.

[32] H. Thorisson. *Coupling, Stationarity, and Regeneration.* Probability and its Applications (New York). Springer-Verlag, New York, 2000.

[33] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 2009. Old and new.