The Archimedean limit of random sorting networks

Duncan Dauvergne

October 29, 2021

Abstract

A sorting network (also known as a reduced decomposition of the reverse permutation), is a shortest path from 12⋯n to n⋯21 in the Cayley graph of the symmetric group $S_n$ generated by adjacent transpositions. We prove that in a uniform random $n$-element sorting network $\sigma^n$, all particle trajectories are close to sine curves with high probability. We also find the weak limit of the time-$t$ permutation matrix measures of $\sigma^n$. As a corollary of these results, we show that if $S_n$ is embedded into $\mathbb{R}^n$ via the map $\tau \mapsto (\tau(1), \tau(2), \ldots, \tau(n))$, then with high probability, the path $\sigma^n$ is close to a great circle on a particular $(n - 2)$-dimensional sphere in $\mathbb{R}^n$. These results prove conjectures of Angel, Holroyd, Romik, and Virág.

Figure 1: The half-way permutation matrix evolution for a uniform 500-element sorting network. As a corollary of our main results, we prove that all of the paths in this evolution are localized, and that the distribution of these localized paths is the projected surface area measure of the 2-sphere onto the unit disk. This figure is from [AHRV07].

MSC classes – Primary: 60C05, Secondary: 05E15, 05A05, 68P10

1
1 Introduction

Consider \( n \) numbers \( 1, \ldots, n \) listed in increasing order. An \( n \)-element sorting network is a way of reversing this list from increasing to decreasing order by using a minimal number of adjacent swaps (see Figure 2 for an example with \( n = 4 \)). This minimal number is \( N = \binom{n}{2} \). For \( i \in \{1, \ldots, N\} \), let \( k_i \in \{1, \ldots, n-1\} \) be the location of the \( i \)th swap: that is, the particles at locations \( k_i \) and \( k_{i+1} \) get swapped at stage \( i \).

We can equivalently define sorting networks in the following way. Let \( \Gamma(S_n) \) be the Cayley graph of the symmetric group \( S_n \) with generators given by the adjacent transpositions \((i, i+1)\) for \( i = 1, \ldots, n-1 \). Then a sorting network is a shortest path from the identity permutation \( \text{id}_n = 12\cdots n \) to the reverse permutation \( \text{rev}_n = n\cdots 21 \) in \( \Gamma(S_n) \).
The name sorting network comes from computer science, where sorting networks are viewed as $N$-step algorithms for sorting a list of $n$ numbers. At step $i$, the sorting network algorithm sorts the numbers at positions $k_i$ and $k_i + 1$ into increasing order. This process sorts any list in $N$ steps.

In combinatorics, sorting networks are known as **reduced decompositions** or **reduced words** for the reverse permutation, as any sorting network can be represented as a minimal length decomposition of the reverse permutation as a product of adjacent transpositions: $\text{rev}_n = (k_N, k_N + 1)\cdots(k_1, k_1 + 1)$.

Stanley [Sta84] showed that the number of $n$-element sorting networks is equal to

$$\frac{n!}{(2n-3)!(2n-5)^2\cdots3n-21n-1},$$

and he observed that this is the same as the number of standard Young tableaux of staircase shape $(n-1, n-2, \ldots, 1)$. Stanley’s argument was based on properties of symmetric functions and did not yield a bijective proof of the connection with Young tableaux. A few years later, Edelman and Greene found an explicit bijection [EG87].

Since these seminal works, the combinatorics of sorting networks and reduced decompositions of other permutations and of elements of other Coxeter groups has been studied in great detail, revealing interesting connections with many other areas. We highlight a few of these here. For more connections and background, see Björner and Brenti [BB06] and Garsia [Gar02].

Reduced decompositions can be used to give a combinatorial interpretation of Schubert polynomials, see Billey, Jockusch, and Stanley [BJS93], Billey and Haiman [BH95] and Manivel [Man01]. They arise in the study of Bott-Samelson resolutions of Schubert varieties, see Magyar [Mag98], and are a useful tool in studying matrix factorizations, total positivity, and canonical bases, e.g. see Berenstein, Fomin, and Zelevinsky [BFZ96] and Leclerc and Zelevinsky [LZ98].

Equivalence classes of reduced decompositions of permutations are in bijection with rhombic tilings of certain polygons, see Elnitsky [Eln97] and Tenner [Ten06]. Permutations of Coxeter group elements that avoid certain patterns can also be characterized by properties of their reduced decompositions, see Lascoux and Schützenberger [LS85], Stembridge [Ste97], and Tenner [Ten06]. Reduced decompositions are also connected to the study of point configurations in the plane and pseudoline arrangements, e.g. see the work of Goodman and Pollack [GP80] and subsequent papers.

A natural direction of inquiry regarding sorting networks is to try to understand their asymptotic properties as the number of elements $n$ approaches infinity. By applying Stirling’s formula to Equation (1), it is fairly straightforward to find asymptotics for the number
of sorting networks. Beyond this, the next logical question to ask is what a typical (i.e. uniformly chosen) sorting network looks like as $n$ approaches infinity. This direction of inquiry has been incredibly fruitful for understanding other objects in algebraic combinatorics, e.g. Young tableaux and domino/lozenge tilings.

In [AHRV07] Angel, Holroyd, Romik, and Virág initiated the study of uniform random $n$-element sorting networks. They studied the global limiting behaviour of the space-time swap distribution, rescaled particle trajectories, time-$t$ permutation matrices, and the Cayley graph path itself.

They proved a law of large numbers for the space-time swap distribution, and based on strong numerical evidence, made conjectures about the limiting behaviour of the other three objects. One of the main difficulties they faced in proving these conjectures is that unlike with many probabilistic models where global limiting results have been proven (e.g. random walks, classical interacting particle systems, random tilings, random graphs), the combinatorics of sorting networks cannot be easily reduced to a series of local rules. In this paper, we overcome this difficulty and prove their conjectures.

1.1 Main limit theorems

We will think of the elements $\{1, \ldots, n\}$ as particles being sorted in time (see Figure 2). We use the notation $\sigma(x,t) = (k_1, k_1 + 1) \ldots (k_1, k_1 + 1)(x)$ for the position of particle $x$ at time $t \in [0, N]$ in a sorting network $\sigma$. We call $(k_1, \ldots, k_N)$ the swap sequence for $\sigma$.

I. Rescaled particle trajectories. For a sorting network $\sigma$, define the global trajectory

$$\sigma_G(x,t) = \frac{2\sigma(x,Nt)}{n} - 1.$$ 

The function $\sigma_G(x,\cdot) : [0,1] \rightarrow [-1,1]$ is the trajectory of particle $x$, with time rescaled so that the sorting process finishes at time 1, and space rescaled so that the trajectory stays
in the interval $[-1, 1]$. In [AHRV07], Angel et al. conjectured that with high probability, all particle trajectories in a uniform random sorting network are close to sine curves (see Figure 3). They proved that with high probability, all global trajectories in a random sorting network are close in uniform norm to some Hölder-$1/2$ curve with Hölder constant $\sqrt{8}$.

Our first theorem proves the sine curve conjecture from [AHRV07]. Here and throughout the paper we use the notation $\sigma^n$ for a uniform random $n$-element sorting network.

**Theorem 1** (Sine curve limit). For each $n$ there exist random variables $(A_n^i, \Theta_n^i) \in [0, 1] \times [0, 2\pi])_{i \in \{1, \ldots, n\}}$ such that for any $\epsilon > 0$, we have that

$$
P\left(\max_{i \in \{1, n\}} \max_{t \in [0, 1]} |\sigma_n^G(i, t) - A_n^i \sin(\pi t + \Theta_n^i)| > \epsilon\right) \to 0 \quad \text{as } n \to \infty.
$$

**II. Permutation Matrices.** For a uniform $n$-element sorting network $\sigma^n$, define the random measure

$$
\rho_t^n = \frac{1}{n} \sum_{i=1}^{n} \delta(\sigma_G^n(i, 0), \sigma_G^n(i, t)).
$$

Here $\delta(x, y)$ is a $\delta$-mass at the point $(x, y)$. The measure $\rho_t^n$ rescales the time-$t$ permutation matrix of $\sigma^n$, placing atoms of weight $1/n$ at the positions of the ones. Define the
Archimedean measure $\text{Arch}$ on the square $[-1,1]^2$ to be the measure with Lebesgue density

$$f(x,y) = \frac{1}{2\pi \sqrt{1-x^2-y^2}}$$

on the unit ball $B(0,1)$, and 0 elsewhere. The Archimedean measure is the unique rotationally symmetric measure on $[-1,1]^2$ whose linear projections are all uniform. It is obtained by projecting the surface area measure of the 2-sphere onto the unit disk. The fact that this measure has uniform linear projections follows from Archimedes' theorem that if two parallel planes slice through a 2-sphere, then the surface area cut out is simply proportional to the distance between the planes. Define the time-$t$ Archimedean measure by

$$(X, X\cos(\pi t) + Y\sin(\pi t)) \overset{d}{=} \text{Arch}_t, \quad \text{where} \quad (X,Y) \overset{d}{=} \text{Arch}.$$

Note that $\text{Arch} = \text{Arch}_{1/2}$. In [AHRV07], Angel et al. conjectured that $\rho^n_t$ converges weakly to $\text{Arch}_t$ for every $t$ (see Figure 4). They proved that for any $t$, the support of $\rho^n_t$ lies in a particular octagon with high probability. In [DV20], Dauvergne and Virág showed that the support of any limit of $\rho^n_t$ lies within the elliptical support of $\text{Arch}_t$.

Our second theorem proves the weak convergence of the random measures $\rho^n_t$. We also show that the support of $\rho^n_t$ and $\text{Arch}_t$ are close. To state this theorem, recall that the Hausdorff distance between two sets $A, B \subset \mathbb{R}^2$ is

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a,b), \sup_{b \in B} \inf_{a \in A} d(a,b) \right\}.$$ 

**Theorem 2** (Permutation matrix limit). For any $t \in [0,1]$, $\rho^n_t \to \text{Arch}_t$ weakly in probability as $n \to \infty$. That is, for any weakly open set $U$ in the space of probability measures on $[-1,1]^2$
containing Arch, we have that
\[ \mathbb{P}(\rho^n_t \in U) \to 1 \quad \text{as} \quad n \to \infty. \]

Moreover,
\[ d_H(\text{supp}(\rho^n_t), \text{supp}(\text{Arch}_t)) \to 0 \quad \text{in probability as} \quad n \to \infty. \]

Angel, Holroyd, Romik, and Virág [AHRV07] also considered a permutation matrix evolution for \( \sigma^n \). Let \( j \in \{1, \ldots, n\} \), and consider the random complex-valued function
\[ Z^n_j(t) = e^{\pi i t} [\sigma^n_G(j, t) + i \sigma^n_G(j, t + 1/2)], \quad t \in [0, 1/2]. \]
For a fixed \( t \), \( (Z^n_1(t), \ldots, Z^n_n(t)) \) is the set of points in the scaled permutation matrix for
\[ \sigma^n(\cdot, N(t + 1/2))(\sigma^n)^{-1}(\cdot, N) \]
after a counterclockwise rotation by \( \pi t \) (see Figure 1). Theorem 1 guarantees that each of the paths \( Z^n_j \) localize.

**Theorem 3 (Path Localization).** Let \( \sigma^n \) be a uniform random \( n \)-element sorting network. Then
\[ \max_{j \in [1, n]} \max_{s, t \in [0, 1]} |Z^n_j(t) - Z^n_j(s)| \to 0 \quad \text{in probability as} \quad n \to \infty. \]

### III. Great Circles.
We can embed the vertices of \( \Gamma(S_n) \) into \( \mathbb{R}^n \) by sending the permutation \( \tau \in S_n \) to the point \( \bar{\tau} = (\tau(1), \ldots, \tau(n)) \in \mathbb{R}^n \). For any \( \tau \in S_n \), the point \( \bar{\tau} \) lies on the \((n - 2)\)-sphere \( S^{n-2} = \mathbb{L}_n \cap \mathbb{K}_n \), where
\[
\mathbb{L}_n = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^n x_i = \frac{n(n+1)}{2} \right\}
\]
and
\[
\mathbb{K}_n = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 = \frac{n(n+1)(2n+1)}{6} \right\}.
\]

If we also embed the edges of \( \Gamma(S_n) \) as straight lines between the embedded vertices, we get an object called the **permutahedron** (see Figure 5).

In [AHRV07], Angel et al. conjectured that with high probability, a uniform random sorting network is close to a great circle in \( S^{n-2} \) under this embedding. They showed that this conjecture implies the other global limiting results for uniform random sorting networks. Our strongest theorem proves this conjecture.

For two functions \( f, g : [0, 1] \to \mathbb{R}^n \), define the distance function
\[ d_\infty(f, g) = \sup_{t \in [0, 1]} \|f(t) - g(t)\|_\infty. \]
This is the uniform norm on functions, where the pointwise distance is the $L^\infty$ distance. Also, for an $n$-element sorting network $\sigma$, we define the embedded and time-rescaled path $\bar{\sigma} : [0, 1] \to S^{n-2}$ by letting

$$\bar{\sigma}(t) = (\sigma(1, Nt), \sigma(2, Nt), \ldots, \sigma(n, Nt)).$$

**Theorem 4** (Great circles). Let $\bar{\sigma}^n$ be the embedding of $\sigma^n$ into $S^{n-2}$. For every $n$, there exists a random path $C_n : [0, 1] \to S^{n-2}$ such that $C_n$ is a constant-velocity parametrization of an arc of a great circle in $S^{n-2}$ starting at $(1, \ldots, n)$ and finishing at $(n, \ldots, 1)$, and such that

$$\frac{d_{\infty}(\bar{\sigma}^n, C_n)}{n} \to 0 \quad \text{in probability as } n \to \infty.$$

Note that it is easy to find sorting networks that aren’t close to a great circle in $S^{n-2}$. For example, the “bubble sort” sorting network given by the swap sequence

$$(1, 2, \ldots, n - 1, 1, 2, \ldots, n - 2, \ldots, 1, 2, 1)$$

is $d_{\infty}$-distance $n - 1 - o(1)$ from any great circle.

**1.2 Random trajectory limits**

To prove the main theorems of Section 1.1, we first analyze the limit of a random particle trajectory. This approach was first considered by Rahman, Virág, and Vizer [RVV19].
Let $D$ be the closure in the uniform norm $\| \cdot \|_u$ of the space of all possible sorting network trajectories $\sigma_G(x, \cdot) : [0, 1] \to [-1, 1]$. The space $(D, \| \cdot \|_u)$ is a complete separable metric space. The only functions in $D$ are continuous functions and the sorting network trajectories themselves.

Let $Y_n \in D$ be a uniform $n$-element sorting network trajectory. That is, if $\sigma_n$ is a uniform $n$-element sorting network, and $I_n$ is an independent uniform random variable on $\{1, \ldots, n\}$, then

$$Y_n = \sigma_n^G(I_n, \cdot).$$

We refer to $Y_n$ as the trajectory random variable of $\sigma^n$. In [DV20], Dauvergne and Virág proved that the sequence $\{Y_n\}_{n \in \mathbb{N}}$ is precompact in distribution, and that any subsequential limit is almost surely Lipschitz. In this paper, we show that $\{Y_n\}_{n \in \mathbb{N}}$ converges in distribution, and identify its limit as the Archimedean path, a random element of $D$ first described in [RVV19], see Conjecture 1 from that paper and surrounding discussion.

**Theorem 5** (The weak trajectory limit). Let $(X, Z) \sim \text{Arch}$, and define the Archimedean path $A \in D$ by $A(t) = X \cos(\pi t) + Z \sin(\pi t)$. Then

$$Y_n \overset{d}{\to} A \quad \text{as} \quad n \to \infty.$$

Theorem 5 will be used in the proof of all our main theorems from Section 1.1. Most of the paper is devoted to its proof. We note here that we can equivalently write

$$A(t) = \sqrt{1 - V^2} \sin(\pi t + 2\pi U),$$

where $V$ and $U$ are independent uniform random variables on $[0, 1]$.

**Random $m$-out-of-$n$ sorting networks.** We will also use Theorem 5 to identify the limit of random $m$-out-of-$n$ subnetworks. This answers a question of Angel and Holroyd [AH10]. This limit can also be by found by using the stronger great circle theorem (Theorem 4), as was noted in [AH10].

Let $\sigma$ be an $n$-element sorting network. For $A \subset \{1, \ldots, n\}$, let $\sigma|_A$ be the $|A|$-element sorting network given by restricting $\sigma$ to the set $A$. Specifically, for $i \in \{1, \ldots, N\}$, let $\sigma^*_A(\cdot, i)$ be the relative ordering of the particles in $A$ in the permutation $\sigma(\cdot, i)$. This gives a sequence of $N$ permutations $\{\sigma^*_A(\cdot, i) \in S_{|A|} : i \in \{1, \ldots, N\}\}$. Removing duplicates gives the permutation sequence for the sorting network $\sigma|_A$. For $m < n$, let the random $m$-out-of-$n$ subnetwork $\tau^m_m$ be the restriction of $\sigma^n$ to a uniform $m$-element subset of $\{1, \ldots, n\}$, chosen independently from $\sigma^n$.

Let $\{x_1, \ldots, x_n\}$ be a set of points in $\mathbb{R}^2$ in general position, and such that no two pairs of points determine parallel lines. Label the points in order of increasing $x$-coordinate. For all but finitely many angles $\theta$, listing the labels of the points $\{x_1, x_2, \ldots, x_n\}$ in increasing order
of their horizontal projections after rotation by $\pi \theta$ gives a permutation $\tau_\theta$. The **geometric sorting network** associated to $\{x_1, \ldots, x_n\}$ is simply the sequence of permutations $\{\tau_\theta, \theta \in [0, 1]\}$ listed in order of increasing $\theta$.

**Theorem 6 (The subnetwork limit).** Let $\{X_1, X_2, \ldots, X_m\}$ be random points in the unit ball $B(0,1)$ sampled from the Archimedean distribution $\mathcal{A}$, and let $\tau_m$ be the associated geometric sorting network. Then

$$\tau_m \xrightarrow{d} \tau_m \quad \text{as } n \to \infty.$$

Note that while random $m$-out-of-$n$ subnetworks are geometric with high probability for fixed $m$, the same is not true of entire sorting networks. Indeed, Angel, Gorin, and Holroyd [AGH12] showed that a uniform sorting network is, with high probability, not geometric.

### 1.3 The local speed distribution

As a by-product of the proof of Theorem 5, we find the distribution of speeds in the local limit of random sorting networks. To state this result, we first give an informal description of this limit (a precise description is given in Section 2). The existence of this limit was established independently by Angel, Dauvergne, Holroyd, and Virág [ADHV19], and by Gorin and Rahman [GR18]. Define

$$U_n(x, t) = \sigma_n([n/2] + x, nt) - [n/2].$$

Each path $U_n(x, \cdot)$ is a locally scaled particle trajectory. With an appropriate notion of convergence, we have that

$$U_n \xrightarrow{d} U,$$

where $U$ is a random function from $\mathbb{Z} \times [0, \infty) \to \mathbb{Z}$. The process $U$ is the local limit at the centre of the sorting network. We can also take a local limit centred at particle $[\alpha n]$ for any $\alpha \in (0, 1)$. The result is the process $U$ with time rescaled by a semicircle factor $2\sqrt{\alpha(1-\alpha)}$.

In [DV20], Dauvergne and Virág used a stationarity argument to conclude that particles in $U$ have asymptotic speeds. Specifically, they showed that for every $x \in \mathbb{Z}$, the limit

$$S(x) = \lim_{t \to \infty} \frac{U(x, t) - U(x, 0)}{t}$$

exists almost surely.

By spatial stationarity of $U$, the speed $S(x)$ has distribution $\mu$ independent of $x$. In this paper, we identify $\mu$.

**Theorem 7.** The measure $\mu$ is the arcsine distribution on $[-\pi, \pi]$ given by the Lebesgue density

$$f(x) = \frac{1}{\pi \sqrt{\pi^2 - x^2}}.$$
1.4 Related work and some random sorting variants

All of the prior work discussed so far on random sorting networks relies crucially on Edelman-Greene bijection. This bijection has also been used by Reiner [Rei05] and Tenner [Ten14] to study the frequency of particular substrings in the swap sequence of a random sorting network. Little [Lit03] found another bijection between the set of sorting networks and Young tableaux of staircase shape, and Hamaker and Young [HY14] proved that these bijections coincide.

Problems involving limits of sorting networks under different measures have been considered by Angel, Holroyd, and Romik [AHR09], and also by Young [You14]. Uniform “relaxed” random sorting networks have been analyzed by Kotowski and Virág [Kot16] (see also [RVV19]). Uniform random sorting networks that avoid intermediate permutations with a 132-pattern have been analyzed by Linusson, Potka, and Sulzgruber [LPS18].

A natural follow-up to studying random sorting networks would be to try to prove analogous theorems for random reduced decompositions of other permutations. When the target permutation is vexillary (i.e. it avoids the pattern 2143) then the Edelman-Greene bijection allows for efficient sampling of random reduced decompositions so the limit object can at least be guessed at, see Gross and Marsaglia [GM12] for simulations. Much of the combinatorics that is necessary for analyzing random reduced decompositions of the reverse permutation breaks down for general vexillary permutations, so the proofs from this paper and previous random sorting network papers cannot be adapted directly. Reduced decompositions of random 321-avoiding permutations are in correspondence with skew standard Young tableaux of a shape determined by the permutation, see Section 2 in [BJS93], so a general theory of random reduced decompositions would also subsume the theory of random Young tableaux.

There are also closely related sorting contexts which exhibit similar Archimedean limiting behaviour. For the Coxeter groups \( \{ B_n : n \in \mathbb{N} \} \), reduced decompositions of the long element can again be efficiently sampled by a variant of the Edelman-Greene bijection, see Haiman [Hai92]. It seems possible that the same broad strategy used to show the Archimedean limit in this paper could work for type \( B \) reduced decompositions.

Another variant of the Edelman-Greene bijection allows for sampling of random involution words for the reverse permutation in \( S_n \), see Marberg [Mar20] for relevant definitions and the description of the algorithm. In this case, while simulations still suggest an Archimedean limit, a proof along the lines of the current paper seems further away. Indeed, random involution words lack a basic time stationarity property used heavily in this paper and in previous work on random sorting networks.
1.5 A connection with fluid mechanics

It turns out that the Archimedean path appeared in the literature on fluid mechanics long before the first papers on random sorting networks. 1 We give a brief description of the context here. For more details and motivation, see Brenier [Bre08] and references therein.

Incompressible fluid flow in a compact connected subset $D \subset \mathbb{R}^d$ can be modelled by a system of PDEs: the Euler equations. A solution to these equations from time 0 to 1 is a function $g : [0, 1] \times D \to D$ with the property that $g(t, \cdot)$ is a volume and orientation preserving diffeomorphism on $D$ for all times $t$. When $g(0, a) = a$, we can think of the value $g(t, a)$ as representing the time-$t$ location of the parcel of fluid that started at location $a$. One way of finding solutions to the Euler equations is by looking for minimizers of the Dirichlet energy

$$E(g) = \int_0^1 \int_D \frac{1}{2} |\partial_t g(t, a)|^2 \, da \, dt$$

among functions with the same values at time 0 and 1. For a given admissible end state (i.e. a volume and orientation preserving diffeomorphism) $g_1 : D \to D$, there may not exist a minimizer of (2) with $g(0, a) = a$ and $g(1, \cdot) = g_1$, see [Shn]. Because of this difficulty, Brenier [Bre89] introduced a notion of generalized solutions of the Euler equations. These solutions allow particles to ‘split’ and ‘cross each other’. Formally, solutions are now random functions $G : [0, 1] \times D \to D$ that minimize expected Dirichlet energy $\mathbb{E}E(G)$ subject to the constraint that $G(t, \cdot)$ is uniformly distributed on $D$ for all $t \in [0, 1]$ (this is incompressibility), and subject to the initial and final conditions $G(0, a) = a$ and $G(1, a) = g_1(a)$ for all $a \in D$. The end state $g_1$ no longer needs to be either a diffeomorphism or orientation-preserving; it just needs to preserve volume and it can even be taken to be random. For a given $g_1$, a generalized solution of the Euler equations always exists and solves a set of ‘generalized’ Euler equations.

The case when $D = [-1, 1]$ and $g_1(a) = (-a)$ is one of the only scenarios where a non-deterministic generalized solution can be computed. In this case, the unique generalized solution to the Euler equations is $G(\cdot, a) = A_a(\cdot)$, where $A_a$ is the Archimedean path conditioned to start at $a$ (see [Bre89], Proposition 6.3 or [Bre08], Figure 3,6). In other words, if $U$ is an independent uniform random variable, then for every $a, b$, the random function $G(\cdot, U) \overset{d}{=} A_a$.

The fact that the Archimedean path minimizes Dirichlet energy was later also observed by Rahman, Virág and Vizer [RVV19] and used to study random ‘relaxed’ sorting networks by Kotowski and Virág [Kot16].

In one sense, it is not all that surprising that the limit of random sorting networks is an energy minimizer; hydrodynamic limits of interacting particle systems often minimize

---

1And as far as I know, no one studying either sorting networks or fluid mechanics was aware of this connection until Laurent Miclo pointed this out at one of my talks after the first version of this paper appeared online.
energy. On the other hand, it is the limiting version of the surprising and beautiful fact that random sorting networks typically follow great circles on a sphere. Moreover, unlike in many interacting particle systems, there does not seem to be an obvious way to prove that random sorting networks tend to minimize energy. In fact, in this paper we take an orthogonal approach to proving the main limit theorems, though still one based around minimizing a particular functional.

It is interesting to note that for a general target permutation, the trajectory limit will not necessarily minimize Dirichlet energy. For example, the limit of random reduced decompositions of the permutation \((\lfloor n/2 \rfloor, \ldots, n, 1, \ldots, \lceil n/2 \rceil - 1)\) is not an energy minimizer (this can be deduced from Romik [Rom15], Theorem 3.24). However, it is possible that a time-changed version of the limiting process will minimize energy and therefore align with a generalized solution of the Euler equations.

1.6 Outline of the proofs and the paper

Most of the paper is devoted to proving Theorem 5. The key to proving this theorem is the notion of particle flux across a curve \(h\). Heuristically, particle flux measures the number of times that particles cross \(h\) in a typical large-\(n\) sorting network. It is defined in terms of the a priori unknown local speed distribution \(\mu\).

Particle flux is a useful quantity because any limit \(Y\) of the trajectory random variables \(Y_n\) must minimize this quantity among curves \(h\) with \(h(0) = -h(1)\). This is due to the fact that in any sorting network, every pair of particles swaps exactly once, whereas any particle must cross a line with \(h(0) = -h(1)\) at least once. We define the particle flux functional and establish that it does indeed measure what it intends to in Section 3.

The next three sections of the paper (Section 4, Section 5, and Section 6) are devoted to understanding properties of minimal flux paths and how they interact. The goal of these sections is to work towards a classification of minimal flux paths. First, by stationarity and symmetry arguments it will be enough to classify minimal flux paths with \(h(0) = h(1) = 0\) satisfying \(h \geq 0\). Our end goal will be to show that the only minimal flux paths in this class are the sine curves \(a\sin(\pi t)\) for \(a \in [0, 1]\). When going through these sections, we encourage the reader to keep in mind this goal.

To understand minimal flux paths, we use a combination of analytic reasoning about how minimizers must behave with intuition based on both the combinatorics of sorting and previously established results about sorting networks. Two of the most useful observations guiding the proofs are:

(i) the number of particles that get within distance \(r\) of the edge can be understood by using the semicircular law of large numbers for the swap distribution,

(ii) since sorting network trajectories can only cross each other once, if they start at the same location then their initial speeds and their maximum heights must be in the same relative order.
In Section 4, we show that for every height \( h \in [0, 1] \), there is exactly one minimal flux path \( h_a \) with \( h_a(0) = -h_a(1) = 0 \), \( h \geq 0 \), and maximum height \( a \). Moreover, paths in this family are ordered by their maximum height, and all these paths are unimodal and concave.

While these ordering, uniqueness, and regularity properties do not yet allow us to conclude that \( h_a(t) = a \sin(\pi t) \), they do give enough structure to allow us to identify certain features of any limit \( Y \) of \( Y_n \). First, they allow us show that the local speed distribution \( \mu \) matches the initial derivative distribution of any limit \( Y \). Intuitively, this means that trajectories in a uniform sorting network typically follow straight lines on all time scales \( t \) with \( n \ll t \ll n^2 \). Second, these properties allow us to explicitly identify the maximum height distribution of \( Y \). This also uses the observation (i) above. All this is done in Section 5.

Now, at this point we have an explicit formula for the probability that \( Y \) reaches a certain maximum height. Observation (ii) above relates this to the initial derivative distribution of \( Y \), and hence to the local speed distribution \( \mu \). Exploring this connection yields an integral transform formula involving \( \mu \), which we obtain in Section 6.

By proving injectivity of this integral transform, we are able to identify \( \mu \) as the arcsine distribution and then use this to prove that the only minimal flux paths are sine curves. This is done in Section 7 and completes the proof of Theorem 5.

In the last two sections of the paper, we use Theorem 5 to establish our stronger limit theorems. The basic idea behind the stronger limit theorems is that if most particles in a sorting network move along sine curves, then the fact that every pair of particles swaps exactly once forces all other particles to follow sine curves as well: the movements of the many control the movements of the few. In Section 8, we combine Theorem 5 with bounds from [AHRV07] to prove Theorems 1, 2, and 3. Finally, in Section 9, we combine Theorem 5 with Theorem 1 to prove Theorem 4.

2 Preliminaries

In this section, we introduce the necessary background about sorting networks. The most basic observation about uniform \( n \)-element sorting networks is that they exhibit a type of time stationarity. This was first observed in [AHRV07].

**Theorem 2.1.** Let \((K^n_1, K^n_2, \ldots, K^n_N)\) be the swap sequence of a random \( n \)-element sorting network \( \sigma^n \). Then

\[
(K^n_1, K^n_2, \ldots, K^n_N) \overset{d}{=} (n - K^n_N, K^n_1, K^n_2, \ldots, K^n_{N-1}).
\]

We will repeatedly use time stationarity of sorting networks to reduce proofs to statements about the beginning of a sorting network. Using the Edelman-Greene bijection between sorting networks and Young tableaux, Angel, Holroyd, Romik, and Virág [AHRV07] also found an explicit formula for the distribution of \( K^n_1 \).
Theorem 2.2. Let $K^n_1$ be the location of the first swap of $\sigma^n$. For any $i \in \{1, \ldots, n-1\}$, we have that
\[
P(K^n_1 = i) = \frac{1}{N} \left[ \frac{3}{2} \cdot 5 \cdot \cdots \cdot (2i - 1) \right] \left[ \frac{3}{2} \cdot 5 \cdot \cdots \cdot (2n - i - 1) \right] \leq \frac{3}{n}.
\]
Moreover, if $\{i_n\}_{n \in \mathbb{N}}$ is any sequence of integers such that $i_n/n \to (\alpha + 1)/2$ for some $\alpha \in (-1, 1)$, then
\[
nP(K^n_1 = i_n) \to \frac{4\sqrt{1 - \alpha^2}}{\pi} \quad \text{as } n \to \infty.
\]

The local limit. Define a swap function as a map $V : \mathbb{Z} \times [0, \infty) \to \mathbb{Z}$ with the following properties:

(i) For each $x$, $V(x, \cdot)$ is cadlag with nearest neighbour jumps.
(ii) For each $t$, $V(\cdot, t)$ is a bijection from $\mathbb{Z}$ to $\mathbb{Z}$ (i.e. a permutation) and $V(\cdot, 0)$ is the identity.
(iii) Define $V^{-1}(x, t)$ by $V(V^{-1}(x, t), t) = x$. Then for each $x$, $V^{-1}(x, \cdot)$ is a cadlag path.
(iv) For any time $t \in (0, \infty)$ and any $x \in \mathbb{Z},$
\[
\lim_{s \to t^-} V^{-1}(x, s) = V^{-1}(x + 1, t) \quad \text{if and only if} \quad \lim_{s \to t^-} V^{-1}(x + 1, s) = V^{-1}(x, t).
\]

We think of a swap function as a collection of particle trajectories $\{V(x, \cdot) : x \in \mathbb{Z}\}$. Condition (iv) guarantees that the only way that a particle at position $x$ can move up at time $t$ is if the particle at position $x + 1$ moves down. That is, particles move by swapping with their neighbours.

Let $\mathcal{A}$ be the space of swap functions endowed with the following topology. A sequence of swap functions $V_n \to V$ if each of the cadlag paths $V_n(x, \cdot) \to V(x, \cdot)$ and $V_n^{-1}(x, \cdot) \to V^{-1}(x, \cdot)$. Convergence of cadlag paths is convergence in the Skorokhod topology. We refer to a random swap function as a swap process.

For a swap function $V$ and a time $t \in (0, \infty)$, define
\[
V(\cdot, t, s) = V(V^{-1}(t, \cdot), t + s).
\]
The function $V(\cdot, t, s)$ is the increment of $V$ from time $t$ to time $t + s$.

Now for $i \in \{1, \ldots, n - 1\}$, define the semicircle scaling factor
\[
r_n(i) = \sqrt{1 - (2i/n - 1)^2},
\]
and consider the shifted, time-scaled swap process
\[
U_n^i(x, s) = \sigma_n \left( i + x, \frac{ns}{r_n(i)} \right) - i.
\]
Here recall that $\sigma_n(i,t)$ is the location of particle $i$ at time $t$ in a uniform $n$-element sorting network $\sigma_n$. To ensure that $U_n^i$ fits the definition of a swap process, we can extend it to a random function from $\mathbb{Z} \times [0,\infty) \to \mathbb{Z}$ by letting $U_n^i$ be constant after time $(n-1)r_n(i)/2$, and with the convention that $U_n^i(x,s) = x$ whenever $x \notin \{1-i, \ldots, n-i\}$. In the swap processes $U_n^i$, all particles are labelled by their initial positions. The following is shown in [ADHV19], and also essentially in [GR18].

**Theorem 2.3.** There exists a swap process $U$ such that the following holds. For any $\alpha \in (-1,1)$, and any sequence of integers $\{i_n\}_{n \in \mathbb{N}}$ such that $i_n/n \to (\alpha+1)/2$, we have that

$$U_n^{i_n} \xrightarrow{d} U \quad \text{as } n \to \infty.$$ 

The swap process $U$ has the following properties:

(i) $U$ is stationary and mixing of all orders with respect to the spatial shift $\tau U(x,t) = U(x+1,t) - 1$.

(ii) $U$ has stationary increments in time: for any $t \geq 0$, the process $U(\cdot, t,s)_{s \geq 0}$ has the same law as $U(\cdot, s)_{s \geq 0}$.

(iii) $U$ is symmetric: $U(\cdot, \cdot, \cdot) \overset{d}{=} -U(-\cdot, \cdot, \cdot)$.

(iv) For any $t \in [0,\infty)$, $\mathbb{P}(\text{There exists } x \in \mathbb{Z} \text{ such that } U(x,t) \neq \lim_{s \to t^-} U(x,t)) = 0$.

Moreover, for any sequence of times $\{t_n : n \in \mathbb{N}\}$ such that $(n-1)r_n(i)/2 - t_n \to \infty$ as $n \to \infty$,

$$U_n^{i_n}(\cdot, t_n, \cdot) \xrightarrow{d} U \quad \text{as } n \to \infty.$$ 

The main takeaway from Theorem 2.3 is that at every location in a large random sorting network, the local swap process converges to a universal limiting object. The only difference between the limits at different locations comes from a rescaling of time by the semicircle factor $\sqrt{1-\alpha^2}$ anticipated by Theorem 2.2.

Now, for a swap function $V$, let $W(V,t)$ be the number of times that the particles at locations 0 and 1 swap in the interval $[0,t]$. That is,

$$W(V,t) = \left| \left\{ s \in (0,t] : \lim_{r \to s^-} V^{-1}(0,r) = V^{-1}(1,s) \right\} \right|.$$ 

As a by-product of the proof of convergence of $U_n^{i_n}$ to $U$, Angel et al. [ADHV19] also found the expectation of $W(U,t)$.

**Theorem 2.4 (Proposition 7.10, [ADHV19]).** Let $\alpha \in (-1,1)$, and let $\{i_n\}_{n \in \mathbb{N}}$ be any sequence of integers converging to $(\alpha+1)/2$. Then for any $t \in [0,\infty)$, we have that

$$\mathbb{E}W(U_n^{i_n}, t) \to \mathbb{E}W(U,t) = \frac{4t}{\pi}.$$ 

16
Note that the expected value $4t/\pi$ in Theorem 2.4 is immediate from Theorem 2.2. Only the convergence in Theorem 2.4 is non-trivial. Now let $Q(V,t)$ be the number of swaps that particle 0 makes by time $t$ in a swap function $V$. That is,

$$Q(V,t) = \left| \left\{ s \in (0,t] : \lim_{r \to s^-} V(0,r) \neq V(0,s) \right\} \right|.$$

Dauvergne and Virág [DV20] used a straightforward stationarity argument to prove an analogous result to Theorem 2.4 for $Q(U,t)$.

**Lemma 2.5** (Lemma 3.2, [DV20]). For any $t \in [0,\infty)$, we have that

$$\mathbb{E}Q(U,t) = \frac{8t}{\pi}.$$

The fact that $U$ is stationary in both time and space implies that the point process of swaps of a given particle $x$ in $U$ is also stationary. This realization combined with the ergodic theorem allows us to conclude that all particles in $U$ have asymptotic speeds. This observation was used in [DV20] to prove results about the relationship between the local and global limit. We use their results as a starting point in our proofs.

**Theorem 2.6** (Theorem 1.7, [DV20]). For every $x \in \mathbb{Z}$, the limit

$$S(x) = \lim_{t \to \infty} \frac{U(x,t) - U(x,0)}{t}$$

exists almost surely. The random variables $S(x), x \in \mathbb{Z}$ satisfy $S(x) \overset{d}{=} -S(x) \overset{d}{=} S(0)$, and $S(0) \in [-\pi,\pi]$ almost surely.

Throughout the paper we let $\mu$ be the law of $S(0)$ and refer to $\mu$ as the **local speed distribution**. One of the main difficulties overcome in this paper is in finding the local speed distribution $\mu$. We will slowly learn more information about $\mu$ throughout the paper, culminating in the proof that $\mu$ is the arcsine distribution on $[-\pi,\pi]$.

Dauvergne and Virág [DV20] were able to express limiting swap rates in $U$ in terms of $\mu$. Define

$$D^*(\mu) = \int (y-c)^+ d\mu(y), \quad D^\mu(c) = \int (y-c)^- d\mu(y) \quad \text{and} \quad D_\mu(c) = \int |y-c| d\mu(y).$$

**Theorem 2.7** (Theorem 1.8, [DV20]). Let $S(0)$ be the asymptotic speed of particle 0 in $U$. For any $x \in \mathbb{Z}$,

$$\frac{Q(U,t)}{t} \to D_\mu(S(0)) \quad \text{almost surely and in } L^1.$$

In particular, the random variables $Q(U,t)/t$ are uniformly integrable and $\mathbb{E}D_\mu(S(0)) = 8/\pi$. 
We also need an analogous result regarding crossings of lines in the local limit. Let $L(t) = ct + d$ be a line of constant slope $c$. For a swap function $V$, define

$$C^+(V, L, t) = \left\{ x \in \mathbb{Z} : V(x, 0) \leq L(0), V(x, t) > L(t) \right\}.$$ 

The quantity $C^+(V, L, t)$ is the total number of particles that are below $L$ at time 0 and above $L$ at time $t$. We symmetrically define $C^-(V, L, t)$ as the total number of particles that are above $L$ at time 0 and below $L$ at time $t$, and let $C(V, L, t) = C^-(V, L, t) + C^+(V, L, t)$.

**Theorem 2.8** (Theorem 5.7, [DV20]). Let $L(t) = ct + d$. Then almost surely and in $L^1$, we have that

$$\lim_{t \to \infty} \frac{C^+(U, L, t)}{t} = D_\mu^+(c), \quad \lim_{t \to \infty} \frac{C^-(U, L, t)}{t} = D_\mu^-(c), \quad \lim_{t \to \infty} \frac{C(U, L, t)}{t} = D_\mu(c).$$

We also record here a few basic facts about the functions $D_\mu$ and $D_\mu^+$ that will be used throughout the paper. These properties can be proven using basic facts about integrals, and the fact that $\mu$ is symmetric and supported in $[-\pi, \pi]$.

**Lemma 2.9.** (i) Both $D_\mu$ and $D_\mu^+$ are convex, 1-Lipschitz functions.

(ii) For all $x$, we have that $D_\mu^+(x) \leq D_\mu(x) \leq |x| \vee \pi$.

(iii) Suppose that $L(t) = at + b$ is tangent to either $D_\mu$ or $D_\mu^+$. Then $b \in [0, \pi]$.

(iv) $D_\mu$ is a symmetric function, and hence minimized at 0.

(v) $D_\mu^+(\pi) = 0$, and $D_\mu(\pm \pi) = \pi$.

**Subsequential limits of $Y_n$.** Recall that $Y_n$ is the trajectory random variable of $\sigma^n$. We record here the main result of [DV20] regarding subsequential limits of $Y_n$. Here and throughout the paper, the phrase “subsequential limit of $Y_n$” always refers to a subsequential limit of $Y_n$ in distribution.

We say that a path $y \in \mathcal{D}$ is $g(y)$-Lipschitz if $y$ is absolutely continuous and if for almost every $t$, $|y'(t)| \leq g(y(t))$.

**Theorem 2.10** (Theorem 1.4, [DV20]). (i) The sequence $\{Y_n\}$ is precompact in distribution.

(ii) Suppose that $Y$ is a subsequential limit of $Y_n$ (in distribution). Then

$$\mathbb{P}\left(Y \text{ is } \pi\sqrt{1-y^2} \text{-Lipschitz}, Y(0) = -Y(1) \right) = 1.$$

Moreover, $Y(t)$ is uniformly distributed on $[-1, 1]$ for every $t$.

In addition, we observe here that any subsequential limit $Y$ of $Y_n$ inherits certain symmetries from $\sigma^n$. 

18
Proposition 2.11. Let $Y$ be any subsequential limit of $Y_n$.

(i) Define $Y_t \in D$ by

$$Y_t(s) = \begin{cases} Y(s + t), & s \leq t. \\ -Y(s + t - 1), & s > t. \end{cases}$$

For any $t \in [0, 1]$, we have that $Y_t \overset{d}{=} Y$.

(ii) $Y \overset{d}{=} -Y$.

(iii) Define $Z \in D$ by $Z(t) = Y(1 - t)$. Then $Z \overset{d}{=} Y$.

Proof. Property (i) follows from time stationarity of random sorting networks (Theorem 2.1). Properties (ii) and (iii) follow from the corresponding properties of the swap sequence $(K^n_1, \ldots, K^n_N)$ of $\sigma^n$:

$$(K^n_1, K^n_2, \ldots, K^n_N) \overset{d}{=} (n - K^n_1, n - K^n_2, \ldots, n - K^n_N) \overset{d}{=} (K^n_N, \ldots, K^n_2, K^n_1).$$

3 Particle flux for Lipschitz paths

In this section, we introduce particle flux and prove that it measures the amount of particles that cross a line in a typical sorting network. Let $\text{Lip}$ be the set of Lipschitz paths from $[0, 1] \to [-1, 1]$ (we will use this notation throughout the paper). Define the local speed of a function $h \in \text{Lip}$ at time $t$ by

$$s(t) = \frac{d(\arcsin(h))}{dt} = \frac{h'(t)}{\sqrt{1 - h^2(t)}}.$$  

The local speed is not well-defined by the above formula at points when $h(t) = \pm 1$. In this case we use the convention that $s(t) = 0$. The local speed $s(t)$ exists for almost every time $t$ for any $h \in \text{Lip}$. Recalling the definition of $D_\mu$ from the previous section, we then define the particle flux of $h$ over a set $A$ by

$$J(h; A) = \frac{1}{2} \int_A D_\mu(s(t)) \sqrt{1 - h^2(t)} dt. \quad (3)$$

We define $J(h) = J(h; [0, 1])$. Note that $J(h) < \infty$ for any Lipschitz function $h$. This follows from Lemma 2.9 (ii), which implies that

$$D_\mu(s(t)) \sqrt{1 - h^2(t)} \leq [\pi \vee |s(t)|] \sqrt{1 - h^2(t)} \leq \pi \vee |h'(t)|.$$  

We will also consider positive particle flux $J^+(h; A)$ and negative particle flux $J^-(h; A)$ defined by

$$J^+(h; A) = \frac{1}{2} \int_A D_\mu^+(s(t)) \sqrt{1 - h^2(t)} dt, \quad J^-(h; A) = \frac{1}{2} \int_A D_\mu^-(s(t)) \sqrt{1 - h^2(t)} dt.$$
Again, we let \( J^+(h) = J^+(h; [0,1]) \) and \( J^-(h) = J^-(h; [0,1]) \).

We now connect flux to random sorting networks. If a random sorting network resembles the local limit in a local window around the global space-time position \((t, h(t))\), then by Theorem 2.8, the number of distinct particles that cross \( h \) in this window should be proportional to

\[
D_{\mu} \left( \frac{h'(t)}{\sqrt{1 - h^2(t)}} \right) \sqrt{1 - h^2(t)}.
\]

The scaling factors of \( \sqrt{1 - h^2} \) come from the semicircle rescaling of time in the local limit away from the center.

Therefore in a typical large-\( n \) sorting network, where most local windows resemble the local limit, \( J^+(h) \) should be proportional to the number of particles that cross the line \( h \), counting multiple crossings for a given particle if and only if the crossings happen at globally distinguishable locations.

The factor of \( 1/2 \) is to account for the difference between the \( n^{-1} \) scaling in the global limit and the \( n^{-1} \) scaling in the local limit. Similarly, \( J^+(h) \) and \( J^-(h) \) should be proportional to the number of upcrossings and downcrossings of the line \( h \) in a large \( n \) sorting network.

Now let \( \text{Lip}^* \) be the set of paths \( h \in \text{Lip} \) with \( h_0 = -h(1) \). If \( h \in \text{Lip}_r \), then in any sorting network, every particle must cross \( h \) at least once unless the particle starts at \( h(0) \). Therefore \( J^+(h) \) should be bounded below for such paths. Since any two particles in a sorting network cross each other exactly once, \( J^+(h) \) should achieve this lower bound when \( h \) is a trajectory limit.

The next theorem makes rigorous this intuition behind the definition of particle flux. To state the theorem, for a random variable \( Y \in D \) and a path \( h \in \text{Lip} \), we define

\[
P^+_Y(h; [a, b]) = \mathbb{P} \left( \exists s < t \in [a, b] \text{ such that } Y(s) < h(s) \text{ and } Y(t) > h(t) \right).
\]

In other words, \( P^+_Y(h; [a, b]) \) is the probability that \( Y \) upcrosses \( h \) in the interval \([a, b]\). We similarly define the downcrossing probability

\[
P^-_Y(h; [a, b]) = \mathbb{P} \left( \exists s < t \in [a, b] \text{ such that } Y(s) > h(s) \text{ and } Y(t) < h(t) \right).
\]

**Theorem 3.1.** Let \( Y \) be any subsequential limit of \( Y_n \).

(i) Let \( h \in \text{Lip} \) and \([a, b] \subset [0,1] \). Then \( P^+_Y(h; [a, b]) \leq J^+(h; [a, b]) \) and \( P^-_Y(h; [a, b]) \leq J^-(h; [a, b]) \).

(ii) Let \( h \in \text{Lip}^* \). Then \( J^+(h) \geq 1 \).
To prove this theorem, we first need two key lemmas about convergence to the local limit. Recall that $C^+(V, L, t)$ is the number of upcrossings of a line $L(s) = cs + d$ in the interval $[0, t]$ in a swap function $V$. Recall also the definition of $U_n^i$ from Section 2.

**Lemma 3.2.** Let $\alpha \in (-1, 1)$, and suppose that $\{i_n\}_{n \in \mathbb{N}}$ is a sequence of integers such that $i_n/n \to (1 + \alpha)/2$. Let $X$ be a uniform random variable on $[0, 1]$ that is independent of $U_n^i, L_n^i$, let $\{c_n\}_{n \in \mathbb{N}}$ be a sequence of real numbers converging to $c \in \mathbb{R}$, and let $\{d_n\}_{n \in \mathbb{N}}$ be a sequence of real numbers in $[0, 1]$. Define

$$L_n(s) = c_n s + d_n + X, \quad \text{and} \quad L(s) = cs + X.$$  

Then for any time $t \in (0, \infty)$, we have that

(i) $C^+(U_n^i, L_n, t) \overset{d}{\to} C^+(U, L, t)$ as $n \to \infty$.

(ii) $\mathbb{E}C^+(U_n^i, L_n, t) \to \mathbb{E}C^+(U, L, t)$ as $n \to \infty$.

(iii) $\mathbb{E}C^+(U_n^i, L_n, t) \leq 3t + |c_n| + 2$ for all $n$.

**Proof.** First note that $d_n + X$ (mod 1) $\overset{d}{\to} X$ for all $n$. Therefore by the stationarity of $U$ with respect to integer-valued spatial shifts (Theorem 2.3 (i)), we have that

$$C^+(U_n^i, L_n, t) \overset{d}{\to} C^+(U, L, t) \quad \text{if and only if} \quad C^+(U_n^i, c_n s + X, t) \overset{d}{\to} C^+(U, L, t). \quad (4)$$

Now if $V_n \in A$ is a sequence of swap functions converging to a swap function $V$ (in the swap function topology introduced in Section 2), then

$$C^+(V_n, c_n s + X, t) \to C^+(V, L, t)$$

unless $V$ either has a swap at time $t$, or either $X$ or $ct + X \in \mathbb{Z}$. By Theorem 2.3 (iv), for any time $t$, the event where $U$ has a swap at time $t$ has probability 0. Moreover, the probability that either $X$ or $ct + X \in \mathbb{Z}$ is also 0. Therefore

$$C^+(U_n^i, c_n s + X, t) \overset{d}{\to} C^+(U, L, t),$$

and hence (i) follows by statement (4). Now recall that $W(V, t)$ is the number of swaps at location 0 in a swap function $V$ in the interval $[0, t]$. For any swap function $V$ and any line $H(s) = as + b$ with $b \in [0, 1]$, we have that

$$C^+(V, H, t) \leq W(V, t) + |a|t + 2. \quad (5)$$
To see why this is true, observe that every particle \( x \) with \( x \leq 0 \) and \( V(x, t) > 1 \) must move from position 0 to position 1 at least once in the interval \([0, t]\), therefore contributing to \( W(V, t) \). Every particle \( x \) that upcrosses \( H \) in the interval \([0, t]\) fits this description, unless \( at \leq V(x, t) \leq 1 \).

There are at most \(|at| + 2\) such values of \( x \), proving (5).

Now again since \( U \) has no swap at time \( t \) almost surely (Theorem 2.3 (iv)), \( W(U_n, t) \rightarrow W(U, t) \). Also, by Theorem 2.4, we have that

\[
\mathbb{E}W(U_n, t) \rightarrow \mathbb{E}W(U, t).
\]

Hence, the random variables \( W(U_n, t) \) are uniformly integrable (see Proposition 3.12, [Kal06]). Therefore by (5) applied to the swap processes \( U_n \) and the lines \( L_n \), the random variables \( C^+(U_n, L_n, t) \) are also uniformly integrable, and hence converge in expectation since they converge in distribution (again by Proposition 3.12, [Kal06]).

Finally, the bound on the probability distribution for the first swap location in a random sorting network (Theorem 2.2) and time stationarity (Theorem 2.1) allows us to conclude that for any \( n, i_n \) and \( t \), that

\[
\mathbb{E}W(U_n, t) \leq 3t,
\]

which in turn proves the bound on the expectation of \( C^+(U_n, L_n, t) \) via (5).

This next lemma is similar to Lemma 3.2, but deals with particle speeds rather than upcrossing rates. For a swap function \( V \) we define

\[
S(V, t) = \frac{V(0, t)}{t}.
\]

This is the average speed of particle 0 in the interval \([0, t]\).

**Lemma 3.3.** Let \( \{U_i^n : i \in \{1, \ldots, n\}, n \in \mathbb{N}\} \) be the array of locally scaled random sorting networks defined in Section 2, and let \( U \) be the local limit. Let \( I \) be a uniform random variable on \((-1, 1)\), independent of everything else. For each \( n \), define \( I_n = \lfloor n(I + 1)/2 \rfloor \), let

\[
R_n = \sqrt{1 - \left[ \frac{2I_n}{n} - 1 \right]^2}, \quad \text{and} \quad R = \sqrt{1 - I^2}.
\]

Then the following statements hold.

(i) For any \( t \in (0, \infty) \), we have that \( R_n D_\mu(S(U_n, R_n t)) \overset{d}{\longrightarrow} RD_\mu(S(U, R t)) \).

(ii) For any \( t \in (0, \infty) \), we have that \( \mathbb{E}R_n D_\mu(S(U_n, R_n t)) \overset{n \to \infty}{\longrightarrow} \mathbb{E}RD_\mu(S(U, R t)) \).
(iii) \( \mathbb{E}RD_\mu (S(U, Rt)) \to 2 \) as \( t \to \infty \).

Note that the processes \( \{U_n^\circ: n \in \mathbb{N}\} \) were not defined in Section 2. These processes plays no real role in the above lemma as \( P(I_n = n) \to 0 \) as \( n \to \infty \), so we simply set \( U_n^\circ = U_{n-1} \).

Proof. Fix \( t \in (0, \infty) \). If \( V_n \) is a sequence of swap functions converging to a swap function \( V \) with no swap at time \( t \), then \( D_\mu (S(V_n, t)) \to D_\mu (S(V, t)) \). Now condition on \( I \). This fixes \( R_n, I_n, \) and \( R \). For any fixed time \( t \in (0, \infty) \), the process \( U \) has no swap at time \( t \) almost surely (Theorem 2.3 (iv)). Therefore for any bounded continuous test function \( f \), we have that

\[
\mathbb{E}\left[f\left(R_n D_\mu (S(U_n^{I_n}, R_n t))\right) \mid I\right] \to \mathbb{E}\left[f\left(RD_\mu (S(U, Rt))\right) \mid I\right].
\]

Taking expectations proves the distributional convergence in (i). Now, by Lemma 2.9, we have that \( D_\mu (s) \leq |s| + \pi \) for any \( s \in \mathbb{R} \). Recalling that \( Q(V, t) \) is the number of swaps made by particle 0 in the interval \([0, t]\) in a swap process \( V \), this implies that

\[
R_n \left(D_\mu (S_{I_n}(U_n^{I_n}, R_n t))\right) \leq R_n (|S_{I_n}(U_n^{I_n}, R_n t)| + \pi) \leq \frac{Q(U_n^{I_n}, R_n t)}{t} + \pi. \tag{6}
\]

Now, we similarly have that \( Q(U_n^{I_n}, R_n t) \overset{d}{\to} Q(U, Rt) \), again since \( U \) has no swap at time \( t \) almost surely. Moreover,

\[
\mathbb{E}Q(U_n^{I_n}, R_n t) = \frac{2\lfloor nt \rfloor}{n},
\]

since this expectation is simply the expected number of swaps made by a uniformly random particle in a sorting network after \( 2\lfloor nt \rfloor \) steps. By Lemma 2.5, we also have that

\[
\mathbb{E}Q(U, Rt) = \frac{4t}{\pi} \int_{-1}^{1} \sqrt{1-x^2} \, dx = 2t,
\]

so \( \mathbb{E}Q(U_n^{I_n}, R_n t) \to \mathbb{E}Q(U, Rt) \). Again, by Proposition 3.12 from [Kal06], this implies that the random variables \( \{Q(U_n^{I_n}, R_n t), n \in \mathbb{N}\} \) are uniformly integrable, and hence so are the random variables \( \{R_n D_\mu (S(U_n^{I_n}, R_n t)), n \in \mathbb{N}\} \). Since these random variables converge in distribution to \( RD_\mu (S(U, Rt)) \), they must also converge in expectation, proving (ii).

Now we prove (iii). First, Theorem 2.6 and the Lipschitz property of \( D_\mu \) imply that

\[
RD_\mu (S(U, Rt)) \xrightarrow{t \to \infty} RD_\mu (S(0)) \quad \text{almost surely}, \tag{7}
\]

where \( S(0) \) is the speed of particle 0 in \( U \). Analogously to (6), we also have that

\[
RD_\mu (S(U, Rt)) \leq \frac{Q(U, t)}{t} + \pi.
\]
By Theorem 2.7, the right hand side above is uniformly integrable over \( t \in [1, \infty) \), and hence so is the left hand side. Therefore since \( RD_\mu(S(U,Rt)) \) has an almost sure limit by (7), it also converges in expectation. Finally, Theorem 2.7 implies that

\[
\mathbb{E}RD_\mu(S(0)) = \mathbb{E}RD_\mu(S(0)) = 2. \quad \square
\]

To prove Theorem 3.1, we also need two auxiliary results. The first is a simple lemma about uniform convergence of functions. This will be used in the proof of Theorem 3.1 (i).

**Lemma 3.4.** Let \( f_n : [0, 1] \to [-1, 1] \) be a sequence of functions converging uniformly to a continuous function \( f \), and let \( h : [0, 1] \to [-1, 1] \) be any continuous function. Let

\[
\{\Pi_n = \{t_{n,0} = 0 < t_{n,1} < \cdots < t_{n,m(n)} = 1\}\}_{n \in \mathbb{N}}
\]

be a sequence of partitions of \([0, 1]\) such that

\[
\text{mesh}(\Pi_n) = \max_{i \in \{0, 1, \ldots, m(n)-1\}} |t_{n,i+1} - t_{n,i}| \to 0 \quad \text{as } n \to \infty.
\]

Let \([a, b] \subset [0, 1]\). If there exist times \( s < t \in [a, b] \) such that \( f(s) < h(s) \) and \( f(t) > h(t) \), then for all large enough \( n \), there exists a time \( t_{n,i} \in [a, b] \) such that \( f_n(t_{n,i}) \leq h(t_{n,i}) \) and \( f_n(t_{n,i+1}) > h(t_{n,i+1}) \).

**Proof.** By the continuity of \( f \) and \( h \), there exists an \( \epsilon > 0 \) and disjoint intervals \([s, s_*] \) and \([t_-, t] \) such that \( f(r) < h(r) - \epsilon \) for all \( r \in [s, s_*] \) and \( f(r) > h(r) + \epsilon \) for all \( r \in [t_-, t] \). Therefore by uniform convergence, for all large enough \( n \) we have that \( f_n(r) < h(r) - \epsilon/2 \) for all \( r \in [s, s_*] \) and \( f(r) > h(r) + \epsilon/2 \) for all \( r \in [t_-, t] \). Now, for large enough \( n \) we also have that

\[
\text{mesh}(\Pi_n) < \min(s_* - s, t - t_-).
\]

Therefore for such \( n \), there must exist \( i_1 < i_2 \in \Pi_n \cap [a, b] \) such that \( f_n(t_{n,i_1}) < h(t_{n,i_1}) \) and \( f_n(t_{n,i_2}) > h(t_{n,i_2}) \). Thus for some \( i \in \{i_1, \ldots, i_2 - 1\} \), we must have that \( f_n(t_{n,i}) \leq h(t_{n,i}) \) and \( f_n(t_{n,i+1}) > h(t_{n,i+1}) \). \quad \square

To prove part (iii), we also need that \( J(\cdot) \) is lower semicontinuous.

**Proposition 3.5.** Let \( h_n \in \text{Lip} \) be a sequence converging uniformly to \( h \). Then for any set \( A \subset [0, 1] \), we have that

\[
J(h; A) \leq \liminf_{n \to \infty} J(h_n; A).
\]

**Proof.** Note that \( J(h; A) \) is of the form \( \int_A G(h(t), h'(t))dt \), where \( G \) is a continuous, positive function, such that for any fixed value of \( a \), \( G(a, \cdot) \) is convex. This follows from the convexity of \( D_\mu \) (Lemma 2.9). Functionals of this form are lower semicontinuous in the uniform norm by Theorem 1.6 from [Str96]. \quad \square

24
Proof of Theorem 3.1. Proof of (i): Note first that by the symmetry of $Y$ (Proposition 2.11), that
\[
P_Y^+(h;[a,b]) = P_Y^+(-h;[a,b]) = P_Y^+(-h;[a,b])
\]
for any path $h \in \text{Lip}$ and any interval $[a,b]$. Also, $J^+(h) = J^-(h)$ by the symmetry of $\mu$ (Theorem 2.6). Therefore the assertion that $P_Y^+(h;[a,b]) \leq J^-(h)$ is equivalent to the assertion that $P_Y^+(h;[a,b]) \leq J^+(h)$, and so to prove Theorem 3.1 (i), it suffices to prove that $P_Y^+(h;[a,b]) \leq J^+(h)$ for every path $h \in \text{Lip}$.

We first prove this for $h \in \text{Lip}$ with range in the open interval $(-1,1)$, since there is a technical difficulty around the definition of local speed when $|h| = 1$. We first give a somewhat informal explanation of the proof in this case. Let $t \in \mathbb{Z} \cap (0,\infty)$, and for $s \in [0,1]$, define
\[
s_{n,t} = \frac{2t}{n-1} \left\lfloor \frac{(n-1)s}{2t} \right\rfloor, \quad \text{and} \quad s^*_{n,t} = \min\left(s_{n,t} + \frac{2t}{n-1},1\right).
\]
Note that we will also use this notation with $s = a$ and $s = b$. Let $X$ be a uniform random variable on $[0,1]$, independent of the sequence $\{Y_n\}$, and let
\[
A^+_{n,t,s} = \left\{ Y_n(s_{n,t}) < h(s_{n,t}) + \frac{2X}{n} \quad \text{and} \quad Y_n(s^*_{n,t}) \geq h(s^*_{n,t}) + \frac{2X}{n} \right\}.
\]
For $t$ fixed and $n$ large, the event where $Y_n$ upcrosses $h$ in the interval $[a,b]$ is approximately equal to the union $\bigcup_{s \in [a,b]} A^+_{n,t,s}$, in a sense that can be made rigorous using Lemma 3.4. Therefore by taking $n \to \infty$, we can relate $P_Y^+(h;[a,b])$ to $\lim_{n \to \infty} P A^+_{n,t,s}$. This limiting probability is connected to counting upcrossings in the local limit. Taking $t \to \infty$ will relate $\lim_{n \to \infty} P A^+_{n,t,s}$ to particle flux. The technical tool needed to take both the limit in $n$ and in $t$ is Lemma 3.2.

The details are as follows. We first find a better way to calculate $P A^+_{n,t,s}$. Observe that when $s < b_{n,t}$, then $s^*_{n,t} \leq b_{n,t}$, and so $s^*_{n,t} = s_{n,t} + 2t/(n-1)$. Therefore for such $s$, time stationarity of random sorting networks (Theorem 2.1) implies that $A^+_{n,t,s}$ has the same probability as the event
\[
\left\{ Y_n(0) < h(s_{n,t}) + \frac{2X}{n} \quad \text{and} \quad Y_n\left(\frac{2t}{n-1}\right) \geq h(s^*_{n,t}) + \frac{2X}{n} \right\}.
\]
Here have used that $t \in \mathbb{Z}$ to apply time stationarity. We can then express the upcrossing probability $P A^+_{n,t,s}$ in terms of the expected number of upcrossings in the local scaling. For $u \in [-1,1]$, define
\[
[u]_n = \left\lfloor \frac{n(u+1)}{2} \right\rfloor, \quad \{u\}_n = \frac{n(u+1)}{2} - \left\lfloor \frac{n(u+1)}{2} \right\rfloor, \quad g_n(u) = \sqrt{1 - \left(\frac{2[u]_n}{n} - 1\right)^2}.
\]
Then for \( s < b_{n,t} \), we have that

\[
  n \mathbb{P}(A_{n,t,s}^+) = EC^+ \left( U_n^{[h(s_{n,t})]}, L_{n,s}, g_n(h(s_{n,t}))t \right),
\]

where

\[
  L_{n,s}(r) = \frac{n(h(s_{n,t}^+) - h(s_{n,t}))}{2tg_n(h(s_{n,t}))}r + \{h(s_{n,t})\}_n + X.
\]

Now, let

\[
  L_s(r) = \frac{h'(s)}{\sqrt{1 - h^2(s)}}r + X.
\]

Since \( h \) is Lipschitz and hence differentiable almost everywhere, Lemma 3.2 (ii) implies that for almost every \( s \in [0, b) \),

\[
  \lim_{n \to \infty} n \mathbb{P}(A_{n,t,s}^+) = \lim_{n \to \infty} EC^+ \left( U_n^{[h(s_{n,t})]}, L_{n,s}, g_n(h(s_{n,t}))t \right) = EC^+(U, L_s, \sqrt{1 - h^2(s)} t). \quad (8)
\]

Here we have used that the range of \( h \) is in \((-1, 1)\) to ensure convergence. Moreover, there exists a constant \( c \) such that

\[
  n \mathbb{P}(A_{n,t,s}^+) \leq ct \quad (9)
\]

for all \( n \in \mathbb{N}, s < b_{n,t} \). This follows from Lemma 3.2 (iii) and the fact that \( h \) is Lipschitz.

Now let \( Z_{n,t} \) be the number of times \( r \) of the form \( s_{n,t} \) in the interval \([a_{n,t}, b_{n,t})\) such that the upcrossing event \( A_{n,t,r}^+ \) occurs. We have that

\[
  Z_{n,t} = \sum_{r \in [a_{n,t}, b_{n,t}), r = s_{n,t}} 1(A_{n,t,r}^+) = \int_{a_{n,t}}^{b_{n,t}} \frac{(n-1)1(A_{n,t,s}^+)}{2t} \, ds.
\]

Therefore by the bounded convergence theorem, we have that

\[
  \mathbb{E}Z_{n,t} = \int_{a_{n,t}}^{b_{n,t}} \frac{(n-1)\mathbb{P}(A_{n,t,s}^+)}{2t} \, ds \xrightarrow{n \to \infty} \frac{1}{2} \int_{a}^{b} \frac{EC^+(U, L_s, \sqrt{1 - h^2(s)} t)}{t} \, ds. \quad (10)
\]

Next, let \( Y \) be any distributional subsequential limit of \( Y_n \), and consider a subsequence \( n_i \) and a coupling of \( Y_{n_i}, X, Y \) where \( Y_{n_i} - 2X/n_i \to Y \) almost surely. On the event where this almost sure convergence holds, Lemma 3.4 implies that

\[
  \{ \exists s < t \in [a, b] \text{ such that } Y(s) < h(s) \text{ and } Y(t) > h(t) \} \subset \limsup_{i \to \infty} \{ Z_{n_i,t} \geq 1 \}.
\]

Therefore

\[
  P_Y^+(h; [a, b]) \leq \liminf_{n \to \infty} \mathbb{P}(Z_{n_i,t} \geq 1) \leq \lim_{n \to \infty} \mathbb{E}Z_{n,t}.
\]
The integrand on the right hand side of (10) is bounded above by \( c \) for all \( t \in \mathbb{Z} \cap (0, \infty) \) by (8) and (9). Therefore by Theorem 2.8 and the bounded convergence theorem, the right hand side of (10) converges to

\[
\frac{1}{2} \int_a^b D_\mu^* \left( \frac{h'(s)}{\sqrt{1 - h^2(s)}} \right) \sqrt{1 - h^2(s)} ds = J^+(h; [a, b]) \quad \text{as } t \to \infty.
\]

This proves Theorem 3.1 (i) for \( h \in \text{Lip} \) with range in \((-1, 1)\). Now we extend this to all \( h \in \text{Lip} \). Define \( h_m = h \vee (-1 + 1/m) \land (1 - 1/m) \), and let

\[ A_m = \{ t \in [0, 1] : h(t) \geq 1 - 1/m \text{ or } h(t) \leq -1 + 1/m \}. \]

Letting \( \mathcal{L} \) be Lebesgue measure on \([0, 1]\), we have that

\[
J^+(h_m; A_m \cap [a, b]) = D_\mu^+(0) \sqrt{1 - (1 - 1/m)2} \mathcal{L}(A_m \cap [a, b]), \quad \text{and}
\]

\[
J^+(h_m; A_m^c \cap [a, b]) = J(h; A_m^c \cap [a, b]).
\]

The flux \( J^+(h_m; A_m) \to 0 \) as \( m \to \infty \), so

\[
\limsup_{m \to \infty} J^+(h_m; [a, b]) \leq J^+(h; [a, b]).
\]

Moreover, \( h_m \) converges uniformly to \( h \), so if \( Y \) upcrosses \( h \), then \( Y \) will eventually upcross \( h_m \). Therefore

\[
\liminf_{m \to \infty} P_Y^+(h_m; [a, b]) \geq P_Y^+(h; [a, b]).
\]

Putting these two inequalities together completes the proof.

**Proof of (ii):** Let \( h \in \text{Lip}^* \), and let \( Y \) be a subsequential limit of \( Y_n \). Therefore by Theorem 3.1 (i),

\[
\mathbb{P} \left( \exists t \in [0, 1] \text{ such that either } Y(r) < h(r), \ Y(t) > h(t), \text{ or } Y(r) > h(r), \ Y(t) < h(t) \right) \leq J^+(h) + J^-(h) = J(h).
\]

The event above holds unless \( Y(0) = h(0) \) since \( Y(0) = -Y(1) \) almost surely by Theorem 2.10. Since \( Y(0) \) is uniformly distributed (Theorem 2.10), \( Y(0) \neq h(0) \) with probability one, so the left hand side above is equal to 1.

**Proof of (iii):** Let \( Y \) be any subsequential limit of \( Y_n \). Fix \( t \in (0, \infty) \cap \mathbb{Z} \), and define

\[
t_{n,j} = \frac{2jt}{n-1} \quad \text{for } j \in \{0, \ldots, \lfloor (n-1)/(2t) \rfloor \} \quad \text{and let } t_n \left\lfloor \frac{n+1}{2t} \right\rfloor + 1 = 1.
\]
Define \( Y_{n,t} \) so that \( Y_{n,t}(t_{n,j}) = Y_n(t_{n,j}) \) for \( j \in \{0,\ldots,[(n-1)/(2t)] + 1\} \), and so that \( Y_{n,t} \) is linear at times in between. By time stationarity of random sorting networks (Theorem 2.1), we can write

\[
\mathbb{E}J(Y_{n,t}) = \sum_{j=0}^{[n-1]/(2t)} \mathbb{E}J(Y_{n,t}; [t_{n,j}, t_{n,j+1}]) \leq \left( \frac{n-1}{2t} \right) \mathbb{E}J(Y_{n,t}; [0, t_{n,1}]).
\]  

(11)

Here we have used that \( t \in \mathbb{Z} \) to apply time stationarity of sorting networks. The final term in the sum above may be slightly smaller than the previous terms since the length of interval may be less than \( 2t/(n-1) \); this gives rise to the inequality.

Now we have that

\[
J(Y_{n,t}; [0, t_{n,1}]) = \frac{1}{2} \int_0^{t_{n,1}} \int |Y'_{n,t}(0) - \sqrt{1 - Y_{n,t}^2(r)} x| d\mu(x) dr
\]

\[
= \frac{1}{2} \int_0^{t_{n,1}} \int |Y'_{n,t}(0) - \sqrt{1 - Y_{n,t}^2(0)} x| d\mu(x) dr + \epsilon_n
\]  

(12)

for some error term \( \epsilon_n \). In the first equality we have used that \( Y'_{n,t}(r) = Y_{n,t}(0) \) for all \( r \in [0, t_{n,1}] \) by piecewise linearity of \( Y'_{n,t} \). Now if \( |Y'_{n,t}(0)| \geq \pi \), then since \( \mu \) is symmetric and supported in \([-\pi, \pi]\) (Theorem 2.6), the two inner integrals are the same. Therefore \( \epsilon_n = 0 \) in this case. Also, when \( |Y'_{n,t}(0)| < \pi \) and \( x \in [-\pi, \pi] \), the difference of the integrands is bounded above by

\[
\left| \sqrt{1 - Y_{n,t}^2(0)} x - \sqrt{1 - Y_{n,t}^2(r)} x \right| \leq |x| \sqrt{1 - \left( 1 - \frac{2\pi t}{n-1} \right)^2} \leq \pi \sqrt{\frac{4\pi t}{n-1}}.
\]

Hence \( |\epsilon_n| \leq k(t/n)^{3/2} \) for some constant \( k \). Now, letting \( I_n = n(Y_{n,t}(0) + 1)/2 \), and using the notation of Lemma 3.3, (12) is equal to

\[
\frac{t}{n-1} R_n D_\mu \left( S(U_{n,t}^I, R_n t) \right) + \epsilon_n.
\]

Therefore by Lemma 3.3 (ii) and the bound on \( \epsilon_n \), (11) implies that

\[
\mathbb{E}J(Y_{n,t}) \leq \left( \frac{n-1}{2t} \right) \left[ \frac{t}{n-1} \mathbb{E}R_n D_\mu \left( S(U_{n,t}^I, R_n t) \right) + \mathbb{E}\epsilon_n \right] \longrightarrow_{n \to \infty} \frac{1}{2} \mathbb{E}R D_\mu \left( S(U, Rt) \right).
\]

Letting \( t \to \infty \), Lemma 3.3 (iii) then implies that

\[
\lim_{t \to \infty} \lim_{n \to \infty} \mathbb{E}J(Y_{n,t}) \leq 1.
\]

(13)

Since subsequential limits of \( Y_n \) are Lipschitz by Theorem 2.10, \( Y \) is a subsequential limit of \( Y_n \) if and only if \( Y \) is a subsequential limit of \( Y_{n,t} \). Therefore, by Fatou’s lemma and the lower semicontinuity of \( J(\cdot) \) (Proposition 3.5),

\[
\lim_{n \to \infty} \mathbb{E}J(Y_{n,t}) \geq \mathbb{E}J(Y)
\]

for all \( t \), so \( \mathbb{E}J(Y) \leq 1 \) by (13). Moreover, \( J(Y) \geq 1 \) almost surely by Theorem 3.1 (ii) and \( Y(0) = -Y(1) \) almost surely by Theorem 2.10 (ii), so \( J(Y) = 1 \) almost surely. \( \square \)
4 Characterization of minimal flux paths

In this section, we show that any subsequential limit $Y$ of $Y_n$ is uniquely determined by its initial position, maximum absolute value, and whether it is initially increasing or decreasing. By Theorem 2.10 (ii) and Theorem 3.1 (iii), it is enough to characterize paths $h \in \text{Lip}^*$ with minimal flux $J(h) = 1$.

Let $\text{Lip}_0$ be the set of path $h \in \text{Lip}$ with $h(0) = h(1) = 0$. We can first recognize that to characterize minimal flux paths $h \in \text{Lip}^*$, it is enough to characterize minimal flux paths $h \in \text{Lip}_0$. This is a simple consequence of the following simple fact.

Lemma 4.1. Let $h \in \text{Lip}^*$ and $t_0 = \inf \{t : h(t) = 0\}$. Define the path

$$\tilde{h}(t) = \begin{cases} h(t + t_0), & t \leq 1 - t_0 \\ -h(t + t_0 - 1), & t > 1 - t_0. \end{cases}$$

Then $J(h) = J(\tilde{h})$. In particular, every path $h \in \text{Lip}^*$ with flux $J(h) = 1$ can be shifted to a path $\tilde{h} \in \text{Lip}_0$ with $J(\tilde{h}) = 1$.

We build up to the following characterization of minimal flux paths in $h \in \text{Lip}_0$. Define the maximum height $m(h)$ of a continuous path $h : [0, 1] \to [-1, 1]$ by

$$m(h) = \max \{|h(t)| : t \in [0, 1]\}.$$

Recall also the definition of $g(y)$-Lipschitz from Section 2.

Theorem 4.2. For every $m \in [0, 1]$, there exists exactly one $\pi\sqrt{1 - y^2}$-Lipschitz path $h_m \in \text{Lip}_0$ such that $J(h_m) = 1$, $m(h_m) = m$, and $h_m \geq 0$. The paths $h_m$ satisfy the following properties.

(i) $h_m(1/2) = m$, and $h_m(t) = h_m(1 - t)$ for $t \in [0, 1/2]$.
(ii) $h_m$ is strictly increasing on the interval $[0, 1/2]$ for $m > 0$.
(iii) For any $\ell \in [0, m]$, we have that $h_\ell(s) \leq h_m(s)$ for all $s \in [0, 1]$.

Now define $h_{-m} = -h_m$. If $h \in \text{Lip}_0$ is a $\pi\sqrt{1 - \tilde{h}^2}$-Lipschitz path with $J_m(h) = 1$ and $m(h) = m$, then either $h = h_m$ or $h = h_{-m}$.

Theorem 4.2 will be proven as Proposition 4.9, Proposition 4.10, Corollary 4.11, Lemma 4.14, and Proposition 4.17. When thinking about the proofs in this section, it may be useful to keep in mind that our end goal is to show that $h_m(t) = m \sin(\pi t)$ for all $m \in [-1, 1]$.

We also state an analogue of Theorem 4.2 for minimal flux paths $h \in \text{Lip}^*$. First, for a path $h \in \text{Lip}$, define

$$S(h) = \sup_{t \in [0, 1]} h(t) \quad \text{and} \quad I(h) = \inf_{t \in [0, 1]} h(t).$$
Theorem 4.3. Fix \(a \in [-1,1] \) and \(k \in ([a],1] \). Then the following hold:

(i) There is exactly one \(\pi \sqrt{1 - y^2}\)-Lipschitz path \(h_{a,k} \in \text{Lip}^*\) such that \(h_{a,k}(0) = -h_{a,k}(1) = a\) and \(m(h) = S(h) = k\).

(ii) There is exactly one \(\pi \sqrt{1 - y^2}\)-Lipschitz path \(h_{a,-k} \in \text{Lip}^*\) such that \(h_{a,-k}(0) = -h_{a,-k}(1) = a\) and \(-m(h) = J(h) = -k\). We have that \(h_{a,k}(t) = -h_{a,-k}(1-t)\) for all \(t\).

(iii) If \(a \neq 0\), there is a unique time \(t \in (0,1)\) such that \(h_{a,k}(t) = 0\). Moreover, the path

\[
    g(s) = \begin{cases} 
        h(t + s), & s \leq 1 - t, \\
        -h(t + s - 1), & s > 1 - t
    \end{cases}
\]

is equal to \(h_k\) if \(a < 0\) and \(h_{-k}\) if \(a > 0\).

(iv) For \(k_1 \leq k_2 \in [-1,-|a|] \cup [|a],1\], we have that \(h_{a,k_1}(t) \leq h_{a,k_2}(t)\) for all \(t \in [0,1]\).

(v) \(h_{a,a} = h_{a,-a}\).

All parts of this theorem follow from applying Lemma 4.1 to Theorem 4.2, except part (iv). This will be proven in Lemma 4.14.

4.1 Basic bounds on \(D_\mu\)

In order to work with the functional \(J(\cdot)\), we first prove a few basic bounds on \(D_\mu\).

Lemma 4.4. Let \(Y\) be a subsequential limit of \(Y_n\). For all \(a \in [0,1]\), we have that

\[
    \mathbb{P}(m(Y) > a) \leq \frac{D_\mu(0) \sqrt{1 - a^2}}{2}.
\]

Proof. Since \(Y(0) = -Y(1)\) and \(Y(0)\) is uniformly distributed (Theorem 2.10), the left hand side of this inequality can be written as

\[
    \mathbb{P}\left(Y(0) < a \text{ and } Y(t) > a \text{ for some } t > 0, \text{ or } Y(0) > -a \text{ and } Y(t) < -a \text{ for some } t > 0\right).
\]

By Theorem 3.1, this is bounded above by \(J^+(g_a) + J^-(g_{-a})\), where \(g_a\) is the path of constant height \(a\). Finally,

\[
    J^+(g_a) + J^-(g_{-a}) = \frac{D_\mu(0) \sqrt{1 - a^2}}{2}. \tag*{\Box}
\]

Lemma 4.5. \(D_\mu(0) = 2\). That is, if \(X\) is a random variable with distribution \(\mu\), then \(E|X| = 2\).

Proof. Let \(Y\) be a subsequential limit of \(Y_n\), and let \(a = D_\mu(0)\). By Lemma 4.4, we have that

\[
    1 = \mathbb{P}(m(Y) > 0) \leq \frac{a}{2},
\]

30
so \( \alpha \geq 2 \). The equality above follows since \( m(Y) \geq |Y(0)| \) and \( Y(0) \) is uniformly distributed on \([-1, 1]\) (Theorem 2.10). Now, for every height \( a \) such that

\[
a > \sqrt{1 - \left(\frac{2}{\alpha}\right)^2},
\]

Lemma 4.4 guarantees that \( m(Y) \leq a \) with positive probability. Therefore by Theorem 3.1 (iii), for any \( \epsilon > 0 \) there is a \( \pi \sqrt{1 - y^2} \)-Lipschitz path \( h_\epsilon \in \text{Lip}^* \) with \( J(h_\epsilon) = 1 \) such that

\[
m(h_\epsilon) \leq m_\epsilon := \sqrt{1 - \left(\frac{2}{\alpha} - \epsilon\right)^2}.
\]

(14)

Using that \( D_\mu \) is minimized at 0 (Lemma 2.9), we have the bound

\[
1 = J(h_\epsilon) \geq \frac{\alpha}{2} \int_0^1 \sqrt{1 - h_\epsilon^2(t)} dt.
\]

(15)

The above integrand is always bounded below by \( 2/\alpha - \epsilon \). Also, since \( h_\epsilon(0) = -h_\epsilon(1) \) and \( h_\epsilon \) is \( \pi \)-Lipschitz, the amount of time that \( h_\epsilon \) spends in the interval \([-m_\epsilon/2, m_\epsilon/2]\) is at least \( m_\epsilon/(2\pi) \). Therefore

\[
\frac{2}{\alpha} \geq \int_0^1 \sqrt{1 - h_\epsilon^2(t)} dt \geq \left(\frac{2}{\alpha} - \epsilon\right) \frac{m_\epsilon}{2\pi} + \frac{m_\epsilon}{2\pi} \sqrt{1 - \frac{m_\epsilon^2}{4}}.
\]

Letting \( \epsilon \to 0 \), we get an inequality in \( \alpha \) which implies that \( \alpha = 2 \).

Combining Lemmas 4.4 and 4.5 shows that \( \mathbb{P}(m(Y) > a) \leq \sqrt{1 - a^2} \). This is optimal, as can be seen by a quick calculation involving the Archimedean path. Later on in the paper we will prove the opposite inequality, which is much more involved.

**Lemma 4.6.** There is a sequence \( u_n \to 0 \) such that

\[
D_\mu(u_n) \leq \frac{2}{\cos(u_n/2)}.
\]

**Proof.** For any subsequential limit \( Y \) of \( Y_n \) and any \( \epsilon > 0 \), Lemmas 4.4 and 4.5 imply that

\[
\mathbb{P}(m(Y) \leq \epsilon) \geq 1 - \sqrt{1 - \epsilon^2} > 0.
\]

Also, \( Y(0) \) is uniformly distributed by Theorem 2.10, so \( \mathbb{P}(m(Y) = 0) = 0 \). Therefore by Theorem 3.1 (iii), we can find a sequence of positive numbers \( \alpha_n \to 0 \) and a sequence of paths \( h_n \in \text{Lip}^* \) with \( m(h_n) = \alpha_n \) and \( J(h_n) = 1 \).

Let \( s_n \) be the local speed of \( h_n \). The total variation of each of the paths \( h_n \) is at least \( 2\alpha_n \), and hence the average absolute local speed of each \( h_n \) is at least \( 2 \arcsin(\alpha_n) \). The convexity and symmetry of \( D_\mu \) (Lemma 2.9) then gives the following bound.

\[
J(h_n) \geq \sqrt{1 - \alpha_n^2} \int_0^1 D_\mu(s_n(t)) dt \geq \sqrt{1 - \alpha_n^2} D_\mu(2 \arcsin(\alpha_n)).
\]

Letting \( u_n = 2 \arcsin(\alpha_n) \) and rearranging completes the proof. \qed
4.2 Monotonicity and uniqueness for minimal flux paths

In this subsection, we prove that minimal flux paths $h \in \text{Lip}_0$ with a particular maximum height $m(h)$ are unique up to sign, and that they satisfy a particular monotonicity relation.

We start with two simple lemmas. The first lemma shows that minimal flux paths minimize flux on every subinterval of $[0, 1]$.

**Lemma 4.7.** Let $h \in \text{Lip}^*$ be a path with $J(h) = 1$. For any interval $[a, b] \subset [0, 1]$ and any path $g \in \text{Lip}$ with $g(a) = h(a)$ and $g(b) = h(b)$, we have that

$$J(h; [a, b]) \leq J(g; [a, b]).$$

Moreover, if $f \in \text{Lip}^*$ is another path with $J(f) = 1$, $f(a) = h(a)$, and $f(b) = h(b)$, then we can form a new path $p \in \text{Lip}^*$ with $J(p) = 1$ by letting

$$p(t) = \begin{cases} f(t) & t \leq t_1, t \geq t_2 \\ h(t) & t \in [t_1, t_2]. \end{cases}$$

**Proof.** If there is some $g$ with $J(h; [a, b]) > J(g; [a, b])$, then we can make a new path $p \in \text{Lip}^*$ which is equal to $g$ on $[a, b]$ and equal to $h$ on $[a, b]^c$. This path $p$ will have $J(p) < 1$, contradicting Theorem 3.1 (ii). The second part of the lemma is a consequence of the fact that

$$J(p) = J(p; [a, b]) + J(p; [a, b]^c).$$

The next lemma uses the bounds established in Section 4.1 to eliminate plateaus in minimal flux paths.

**Lemma 4.8.** For any height $\alpha \in (0, 1)$ and any interval $[t_1, t_2] \subset [0, 1]$, we have that

$$\inf_{h \in \text{Lip}} \{ J(h; [t_1, t_2]) : h(t_1) = \alpha, h(t_2) = \alpha \} < (t_2 - t_1) \sqrt{1 - \alpha^2}.$$

**Proof.** Without loss of generality, we may assume that $t_1 = 0$. Let $\{u_n\}_{n \in \mathbb{N}}$ be as in Lemma 4.6, and define a sequence of paths $h_n \in \text{Lip}$ by letting

$$h_n(t) = \begin{cases} \sin(\arcsin(\alpha) + u_n t), & t \leq t_2/2 \\ h_n(t_2 - t), & t \in [t_2/2, t_2] \\ \alpha, & t > t_2. \end{cases}$$

Lemma 4.6 gives the following bound on the flux of $h_n$ on the interval $[0, t_2]$.

$$J(h_n; [0, t_2]) \leq \frac{2}{\cos(u_n/2)} \int_0^{t_2/2} \cos(\arcsin(\alpha) + u_n t) dt = t_2 \sqrt{1 - \alpha^2} - \frac{t_2^2 \alpha u_n}{4} + O(u_n^2).$$

For small enough $u_n$, this calculation shows that $J(h_n) < t_2 \sqrt{1 - \alpha^2}$. 

\[32\]
We can now prove that minimal flux paths \( h \in \text{Lip}_0 \) are unimodal.

**Proposition 4.9.** Let \( h \in \text{Lip}_0 \) be such that \( J(h) = 1 \), and \( m(h) \in (0, 1) \). Then \( \|h(1/2)\| = m(h) \). Moreover, if \( h(1/2) = m(h) \), then \( h \) is strictly increasing on \([0, 1/2]\) and strictly decreasing on \([1/2, 1]\). If \( -h(1/2) = m(h) \), then \( h \) is strictly decreasing on \([0, 1/2]\) and strictly increasing on \([1/2, 1]\).

**Proof.** First consider the case where \( h \geq 0 \). Let \( t_m \) be any time when \( h(t_m) = m(h) \). Suppose that there exist times \( s_1 < s_2 \in [0, t_m] \) such that \( h(s_1) = h(s_2) \), and \( h(t) \leq h(s_1) \) on the interval \([s_1, s_2]\). Define a new path

\[
  r(t) = \begin{cases} \h(t), & t \leq s_1, t \geq t_m \\ h(t + (s_2 - s_1)), & t \in (s_1, t_m - (s_2 - s_1)) \\ m(h), & t \in [t_m - (s_2 - s_1), t_m] \end{cases}
\]

The path \( r \) replaces the segment of \( h \) in the interval \([s_1, s_2]\) with a plateau at height \( t_m \) at a later time. This operation cannot increase flux since \( D_{\mu} \) is minimized at 0 (Lemma 2.9), so \( r \) must be a minimal flux path in \( \text{Lip}^* \). However

\[
  J(r; [t_m - (s_2 - s_1), t_m]) = (s_2 - s_1)\sqrt{1 - m^2(h)}
\]

which contradicts Lemmas 4.7 and 4.8. Therefore there is no interval \([s_1, s_2] \subset [0, t_m] \) where \( h(s_1) = h(s_2) \) and \( h(t) \leq h(s_1) \) for \( t \in [s_1, s_2] \), so \( h \) must be strictly increasing on \([0, t_m]\). Similarly, \( h \) is strictly decreasing on \([t_m, 1]\). The point \( t_m \) is the unique time when \( h \) achieves its maximum.

We now show that \( t_m = 1/2 \). Without loss of generality, assume that \( t_m \leq 1/2 \). Define a new path \( g(t) = h(1 - t) \). The path \( g \in \text{Lip}_0 \) also satisfies \( J(g) = 1 \) and \( m(g) = m \). We have that \( g(1/2) = h(1/2) \), and so by Lemma 4.7 we can create a path \( p \in \text{Lip}^* \) with \( J(p) = 1 \) and \( m(p) = m \) by letting

\[
  p(t) = \begin{cases} \h(t), & t \leq 1/2 \\ g(t), & t \geq 1/2 \end{cases}
\]

The path \( p(t) \) achieves its maximum height at both \( t_m \) and \( 1 - t_m \), so we must have that \( t_m = 1 - t_m \), and hence \( t_m = 1/2 \).

Now if \( h \) is not a strictly non-negative path, then we can create a non-negative path \( p(t) = |h(t)| \). By the symmetry of \( \mu \) (Theorem 2.6), the path \( p \) must again have minimal flux, so since \( p \geq 0 \), by the above argument \( p \) is strictly increasing on \([0, 1/2]\) and strictly decreasing on \([1/2, 1]\). Therefore either \( h = p \) or \( h = -p \), completing the proof.

We can now use Proposition 4.9 to prove uniqueness of minimal flux paths with a given maximum height.
Proposition 4.10. For every \( m \in [0,1] \), there is at most one \( \pi \sqrt{1 - y^2} \)-Lipschitz path \( h \in \text{Lip}_0 \) with \( J(h) = 1 \), \( m(h) = m \), and \( h \geq 0 \). If such a path \( h \) exists, then the only other \( \pi \sqrt{1 - y^2} \)-Lipschitz path \( g \in \text{Lip}_0 \) with \( J(g) = 1 \) and \( m(g) = m \) is \( g = -h \).

Proof. First observe that the only \( \pi \sqrt{1 - y^2} \)-Lipschitz paths in \( \text{Lip}_0 \) with \( m(h) = 1 \) are 

\[ h = \pm \sin(\pi t) \] 

Similarly, the only path \( h \in \text{Lip} \) with \( m(h) = 0 \) is \( h = 0 \). This proves the proposition for \( m \in \{0,1\} \). Now we assume \( m \in (0,1) \).

Suppose that \( h, g \in \text{Lip}_0 \) are two distinct non-negative paths with \( J(h) = J(g) = 1 \) and \( m(h) = m(g) = m \). By Proposition 4.9, \( h(1/2) = g(1/2) = m \). Without loss of generality, there exists a time \( t_2 \in [0,1/2) \) such that \( g(t_2) < h(t_2) \). Since \( g(t_2) \geq 0 \), \( h(0) = 0 \) and \( h \) is continuous we can find a time \( t_1 < t_2 \) such that \( h(t_1) = g(t_2) \). Define the shifted path

\[ g^*(t) = \begin{cases} g(t + (t_2 - t_1)), & t \leq 1 - (t_2 - t_1) \\ -g(t + (t_2 - t_1) - 1), & t \geq 1 - (t_2 - t_1). \end{cases} \]

By Proposition 4.9, we have that

\[ m = g^*(1/2 - (t_2 - t_1)) > h(1/2 - (t_2 - t_1)) \quad \text{and} \quad g^*(1/2) < h(1/2) = m. \]

Therefore there is some time \( t_3 \in (1/2 - (t_2 - t_1), 1/2) \) such that \( g^*(t_3) = h(t_3) \). Moreover, \( g^* \in \text{Lip}^* \), \( J(g^*) = 1 \), and \( g^*(t_1) = h(t_1) \). Therefore by Lemma 4.7 we can create a path \( p \in \text{Lip}_0 \) with \( J(p) = 1 \) by letting

\[ p(t) = \begin{cases} h(t), & t \leq t_1 \text{ or } t \geq t_3 \\ g^*(t), & t \in (t_1, t_3). \end{cases} \]

This new path \( p \) is a non-negative minimal flux path in \( \text{Lip}_0 \) which does not uniquely achieve its maximum value at \( 1/2 \), contradicting Proposition 4.9.

Now, if \( g \in \text{Lip}_0 \) is another minimal flux path with \( J(g) = 1 \) and \( m(g) = m \), then \( |g| \) is a non-negative minimal flux path in \( \text{Lip}_0 \), so \( |g| = h \). Since \( h(t) > 0 \) for \( t \in (0,1) \) by Proposition 4.9, either \( g = h \) or \( g = -h \).

Proposition 4.10 gives us the following easy corollary.

Corollary 4.11. Any \( \pi \sqrt{1 - h^2} \)-Lipschitz path \( h \in \text{Lip}_0 \) with \( J(h) = 1 \) has \( h(t) = h(1-t) \) for all \( t \in [0,1] \).

Proof. Without loss of generality, assume that \( h \geq 0 \). Both \( h(t) \) and \( h(1-t) \) are non-negative minimal flux paths satisfying all the conditions of Proposition 4.10 and with the same maximum height, so by that proposition, \( h(t) = h(1-t) \) for all \( t \).

Remark 4.12. Note that the uniqueness proofs in this section automatically give the uniqueness of the paths \( h_{a,k} \) in Theorem 4.3. The fact that uniqueness immediately carries over to shifted paths follows from the strict monotonicity in Proposition 4.9.
4.3 Existence of minimal flux paths

We now show that for every \( m \in [0, 1] \), there exists a minimal flux path in \( \text{Lip}_0 \) with maximum height \( m \). This is a consequence of the following proposition, which shows that the inequality in Theorem 3.1 is an equality for minimal flux paths. Recall the definition of the upcrossing probability \( P^*_Y(h;[a,b]) \) and the downcrossing probability \( P^-_Y(h;[a,b]) \) from Section 3.

**Proposition 4.13.** Let \( h \in \text{Lip}^* \) be a path with \( J(h) = 1 \). Let \( Y \) be a subsequential limit of \( Y_n \). Then for any interval \([a,b]\), we have that

\[
P^*_Y(h;[a,b]) = J^*(h;[a,b]) \quad \text{and} \quad P^-_Y(h;[a,b]) = J^-(h;[a,b]).
\]

**Proof.** We will only prove the first equality, as the second one follows by the symmetry of \( Y \) and \( \mu \) (Theorem 2.6, Proposition 2.11). We have that

\[
J(h) = J^*(h) + J^-(h) = 1 \leq P^*_Y(h;[0,1]) + P^-_Y(h;[0,1]).
\]

Here the inequality uses that \( Y(0) = -Y(1) \) almost surely and \( Y(0) \) is uniformly distributed (Theorem 2.10) and a union bound. Therefore by Theorem 3.1 (i), we have that \( P^*_Y(h;[0,1]) = J^*(h) \). Now,

\[
P^*_Y(h;[0,1]) \leq P^*_Y(h;[0,a]) + P^*_Y(h;[a,b]) + P^*_Y(h;[b,1])
\leq J^*(h;[a,b]) + P^*_Y(h;[a,b])
\leq J^*(h;[a,b]) + J^*(h;[a,b]) = J^*(h).
\]

To see the first inequality above, note that the union of the three events on the right hand side contains the event on the left hand side minus the set \( \{Y(a) = h(a) \text{ or } Y(b) = h(b)\} \). This set has probability 0 since \( Y(t) \) is uniformly distributed for all \( t \) (Theorem 2.10). The second and third inequalities follow from Theorem 3.1 (i).

Since \( P^*_Y(h;[0,1]) = J^*(h) \), all the inequalities in (16) must in fact be equalities, proving the proposition.

To apply the above proposition in order to prove the existence of minimal flux paths at every height, we need three lemmas. The first lemma shows that any two minimal flux paths in \( \text{Lip}^* \) cannot cross each other more than once. This lemma also proves part (iii) of Theorem 4.2 and part (iv) of Theorem 4.3.

**Lemma 4.14.** Suppose that \( h, g \in \text{Lip}^* \) are non-negative \( \pi \sqrt{1 - y^2} \)-Lipschitz paths with \( J(h) = J(g) = 1 \). Then there cannot exist times \( t_0 < t_1 < t_2 \in [0, 1] \) such that \( h(t_0) < g(t_0) \), \( h(t_1) > g(t_1) \), and \( h(t_2) = g(t_2) \). In particular, if \( g(0) = h(0) \), then either \( h \leq g \) or \( g \leq h \).
Proof. Suppose there exist times $t_0 < t_1 < t_2 \in [0,1]$ such that $h(t_0) < g(t_0)$, $h(t_1) > g(t_1)$, and $h(t_2) = g(t_2)$. First, by possibly shifting the paths as in Lemma 4.1, we may assume that $t_2 = 1$. Hence $g(1) = h(1) = -g(0) = -h(0)$. Now let $s \in (t_0, t_1)$ be such that $h(s) = g(s)$. Letting $a = g(0)$, Remark 4.12 implies that $h = h_{a,k_1}$ and $g = h_{a,k_2}$ for some $k_1, k_2 \in [-1,-a] \cup [a,1]$. Here the paths $h_{a,k}$ are as in Theorem 4.3. Without loss of generality, assume that $k_2 > k_1$ and that $k_2 \geq a$; the other cases follow by symmetric arguments.

Define $p = \max(h,g)$. By Lemma 4.7, $J(p) = 1$. Moreover, $p(0) = -p(1) = a$, and since $k_2 > k_1$ and $k_2 \geq |a|$, we have that $m(p) = S(p) = k_2$. This contradicts the uniqueness of the path $h_{a,k_2}$ established in Remark 4.12. □

The next lemma establishes a strong concavity property for minimal flux paths.

Lemma 4.15. For every $k \in [0,1]$, if there exists a minimal flux, $\sqrt{1-y^2}$-Lipschitz path $h_k \in \text{Lip}_0$ with $h_k \geq 0$ and $m(h_k) = k$, then arcsin($h_k$) is concave. In other words, the local speed $s_k$ of $h_k$ is a non-increasing function of $t$.

Proof. By the symmetry of $h_k$ (Corollary 4.11), it suffices to prove that arcsin($h_k$) is concave on $[0,1/2]$. Suppose that there exist times $t_1 < t_2 < t_3 \leq 1/2$ such that $h(t_2) < p(t_2)$, where $p$ is the path of constant local speed $\tilde{s} \in [0,\pi]$ with $p(t_1) = h(t_1)$ and $p(t_2) = h(t_2)$. Let

$$t_4 = \sup\{t \in [t_1,t_2] : p(t) \leq h(t)\} \quad \text{and} \quad t_5 = \inf\{t \in [t_2,t_3] : p(t) \leq h(t)\}.$$ 

Then $h(t_4) = p(t_4)$ and $h(t_5) = p(t_5)$, and for every $t \in (t_4,t_5)$, we have that $h(t) < p(t)$. By Lemma 2.9, we can find a line $L(s) = as + b$ with $b \geq 0$ such that $L(s) \leq D_{\mu}(s)$ for all $s$, and such that $L(s) = D_{\mu}(s)$. Therefore

$$J(h; [t_4,t_5]) \geq \int_{t_4}^{t_5} (as(t) + b)\sqrt{1-h^2(t)}dt \geq \int_{t_4}^{t_5} ah'(t)dt + \int_{t_4}^{t_5} b\sqrt{1-p^2(t)}dt.$$ 

The inequality in the second line follows since $p(t) > h(t) \geq 0$ for all $t \in (t_4,t_5)$. By the fundamental theorem of calculus, since $p(t_4) = h(t_4)$ and $p(t_5) = h(t_5)$, we have

$$\int_{t_4}^{t_5} ah'(t)dt + \int_{t_4}^{t_5} b\sqrt{1-p^2(t)}dt = \int_{t_4}^{t_5} ap'(t)dt + \int_{t_4}^{t_5} b\sqrt{1-p^2(t)}dt = J(p; [t_4,t_5]).$$

We can then create a non-negative $\pi \sqrt{1-y^2}$-Lipschitz path $g \in \text{Lip}_0$ with $m(g) = k$ by letting $g(t) = p(t)$ for $t \in [t_4,t_5]$ and $g(t) = h(t)$ otherwise. We have $J(g) \leq 1$, contradicting the uniqueness of $h_k$. □
The next lemma is a consequence of the symmetry and concavity of $Y$.

**Lemma 4.16.** Let $Y$ be a subsequential limit of $Y_n$, let $t \in [0,1)$, and let $[a,b] \subset [0,1]$. Then $Y'(t)$ exists almost surely and

(i) $\mathbb{P}(Y'(t) = 0) = 0$.

(ii) $\mathbb{P}(Y'(t) > 0, Y(t) \in [a,b]) = \mathbb{P}(Y'(t) < 0, Y(t) \in [a,b]) = (b-a)/4$.

**Proof.** By time stationarity of $Y$ (Proposition 2.11 (i)) it suffices to consider $t = 0$. This also implies that

$$Y'(0) \overset{d}{=} Y'(U),$$

for an independent uniform random variable $U$ on $[0,1]$. Since $Y$ is Lipschitz and hence almost everywhere differentiable, this proves that $Y'(0)$ exists almost surely. By the concavity and strict monotonicity of minimal flux paths (Lemma 4.15 and Proposition 4.9), we have that

$$\mathbb{P}(Y'(U) = 0) = \mathbb{P}(m(Y) = |Y(U)|) = 0. \tag{17}$$

Define $Z(t) = -Y(1-t)$. Since $Y(0) = -Y(1)$ (Theorem 2.10), we have that $Z(0) = Y(0)$ and $Z'(0) = -Y'(0)$. By Proposition 2.11, $Y \overset{d}{=} Z$, so

$$\mathbb{P}(Y'(0) > 0, Y(0) \in [a,b]) = \mathbb{P}(Y'(0) < 0, Y(0) \in [a,b]).$$

Putting this together with (17) and the fact that $\mathbb{P}(Y(0) \in [a,b]) = (b-a)/2$ completes the proof. \hfill \square

We are now ready to prove the existence of minimal flux paths at every height.

**Proposition 4.17.** For every $m \in [0,1]$, there exists a $\pi\sqrt{1-y^2}$-Lipschitz path $h_m \in \text{Lip}_0$ such that $m(h_m) = m$, $J(h_m) = 1$, and $h \geq 0$.

**Proof.** First observe that we can set $h_0(t) = 0$ and $h_1(t) = \sin(\pi t)$. Both of these paths satisfy all of the above properties. For this, we use that $D_\mu(\pm \pi) = \pi$ and $D_\mu(0) = 2$ (Lemma 2.9 and Lemma 4.5).

Let $A \subset [0,1]$ be the set of all $m$ such there is a $\pi\sqrt{1-y^2}$-Lipschitz path $h_m \in \text{Lip}_0$ with $h_m \geq 0$, $J(h_m) = 1$ and $m(h_m) = m$. Let $m \in A$, and let $m_k \in A$ be a sequence converging to $m$.

By the Arzelà-Ascoli Theorem, there exists a $\pi\sqrt{1-h^2}$-Lipschitz subsequential limit $h_m \in \text{Lip}_0$ of $h_{m_k}$ with $m(h_m) = m$ and $h_m \geq 0$. By the lower semicontinuity of $J$ (Proposition 3.5), $J(h_m) = 1$. Therefore $A$ is closed.
Now suppose that $A \neq [0,1]$. Then there exists an interval $(\ell, m) \in [0,1]$ such that $\ell, m \in A$ and $(\ell, m) \cap A = \emptyset$. Let $k = (\ell + m)/2$, and choose $t_1 < t_2 \in [0,1/2]$ so that $h_m(t_1) = \ell$ and $h_m(t_2) = k$. We will show that

$$P_Y^+(h_m;[t_1,t_2]) + P_Y^-(h_m;[t_1,t_2]) = P_Y^+(k;[t_1,t_2]) + P_Y^-(k;[t_1,t_2]).$$

(18)

Here we use $k$ as shorthand for the line $h(t) = k$.

Let $g \in \text{Lip}^*$ be any minimal flux $\pi \sqrt{1 - y^2}$-Lipschitz path. We will analyze how $g$ crosses the paths $k$ and $h_m$ in the interval $[t_1, t_2]$. By the unimodality established in Proposition 4.9, $g$ upcrosses $k$ at most once in the interval $[0,1]$, and downcrosses $k$ at most once. Moreover, by Lemma 4.14, $g$ crosses $h_m$ at most once.

If $g(t_1) \notin [\ell, k]$, and $g(t_2) \notin k$, then the total number of crossings of the lines $h_m$ and $k$ is even. Hence, either $g$ doesn’t cross either path in the interval $[t_1, t_2]$, or else $g$ first crosses $k$ going down, and then crosses $k$ going up, all while staying above $h_m$. This would force $g$ to have a positive local minimum in the interval $[t_1, t_2]$. This is forbidden by the monotonicity properties of minimal flux paths (Proposition 4.9).

If $g(t_1) \in (\ell, k)$ and $g'(t_1) < 0$, then again by the monotonicity properties of $g$, $g$ downcrosses $h_m$, does not upcross $h_m$, and neither upcrosses nor downcrosses $k$.

If $g(t_1) \in (\ell, k)$, $g'(t_1) > 0$, and $g(t_2) \notin k$, then $m(g) \geq m$ since $(\ell, m) \cap A = \emptyset$. By the monotonicity of $g$ (Proposition 4.9), $g(t^*) = m(g)$ for some $t^* > t_1$. If $g(t_2) < h(t_2)$, then since both $g$ and $h_m$ are minimal flux paths, the restrictions on crossings imposed by Lemma 4.14 imply that $t^* < t_2$. In this case, $g$ both upcrosses and downcrosses $k$ in the interval $[t_1, t_2]$, and downcrosses but does not upcross $h_m$.

On the other hand, if $g(t_2) > h(t_2)$, then $g$ upcrosses $k$ in the interval $[t_1, t_2]$ but does not downcross $k$, and does not cross $h$ at all in this interval.

Now let $Y$ be any subsequential limit of $Y_n$. By Theorem 2.10 and Lemma 4.16 (i), we have that

$$\mathbb{P}(Y(t_1) \notin \{\ell, k\}, Y(t_2) \notin k, Y'(t_1) \text{ exists and is not equal to } 0) = 1.$$

Therefore by the above analysis of minimal flux $\pi \sqrt{1 - y^2}$-Lipschitz paths $g \in \text{Lip}^*$, we have that

$$P_Y^+(h_m;[t_1,t_2]) - P_Y^+(k;[t_1,t_2]) = -\mathbb{P}(Y'(t_1) > 0, Y(t_1) \in [\ell, k]),$$

and

$$P_Y^-(h_m;[t_1,t_2]) - P_Y^-(k;[t_1,t_2]) = \mathbb{P}(Y'(t_1) < 0, Y(t_1) \in [\ell, k]).$$

Equation (18) then follows from the symmetry established in Lemma 4.16 (ii). Now by Proposition 4.13, the left hand side of (18) is equal to $J(h_m;[t_1,t_2])$, and by Theorem 3.1
(i), the right hand side of (18) is bounded above by \( J(k;[t_1,t_2]) \). However, since \( D_\mu \) is minimized at 0 by Lemma 2.9, we can also easily calculate that

\[
J(h_m;[t_1,t_2]) \geq \frac{D_\mu(0)}{2} \int_{t_1}^{t_2} \sqrt{1-h_m^2(t)} \, dt \\
> \frac{D_\mu(0)}{2} (t_2 - t_1) \sqrt{1-k^2} = J(k;[t_1,t_2]).
\]

This is a contradiction, so \( A \) must be the whole interval \([0,1]\).

5 The derivative distribution

Let \( \mathbb{P}_a(Y \in \cdot) \) be any regular conditional distribution of \( Y \) given that \( Y(0) = a \). In this section, we use the structure of minimal flux paths to find \( \mathbb{P}_a(Y'(0) \in \cdot) \).

**Proposition 5.1.** Let \( Y \) be any subsequential limit of \( Y_n \). With probability one, \( Y'(0) \) exists. Moreover,

\[
\mathbb{P}_a\left( \frac{Y'(0)}{\sqrt{1-a^2}} \in \cdot \right) = \mu,
\]

for Lebesgue-a.e. \( a \in (-1,1) \).

As an immediate corollary of Proposition 5.1, we will be able to find the distribution of the maximum height function \( m(Y) \) for any subsequential limit \( Y \) of \( Y_n \). This will allow us to show that the sequence \( \{Y_n\} \) has a unique limit point.

Moreover, Proposition 5.1 will be used later to prove an integral transform formula for the local speed distribution \( \mu \), which will allow us to determine \( \mu \), and then \( Y \). This is done in Section 6.

To prove Proposition 5.1, we first show that minimal flux paths fill space. For this lemma, we use the notation of Theorem 4.3.

**Lemma 5.2.** Let \( a \in [0,1] \), and let

\[
M_a = \{ h_{a,k} : k \in [-1,-a] \cup [a,1] \}.
\]

Then for any \( t \in [0,1] \), we have that

\[
\{ h(t) : h \in M_a \} = [\sin(-\pi t + \arcsin(a)), \sin(\pi t + \arcsin(a))].
\]

**Proof.** Let \( K_{a,t} = \{ h(t) : h \in M_a \} \). The set \( K_{a,t} \) contains the endpoints of the above interval, since \( M_a \) contains the two extremal functions \( h_{a,-1}(t) = \sin(-\pi t + \arcsin(a)) \) and \( h_{a,1}(t) = \sin(\pi t + \arcsin(a)) \). Moreover, the ordering on minimal flux paths (Theorem 4.3 (iv)) implies that

\[
K_{a,t} \subset [\sin(-\pi t + \arcsin(a)), \sin(\pi t + \arcsin(a))].
\]
Now suppose that \( c \in (\sin(-\pi t + \arcsin(a)), \sin(\pi t + \arcsin(a))) \). Let

\[
A = \{ h_{a,k} \in M_a : h_{a,k}(t) \leq c \} \quad \text{and} \quad B = \{ h_{a,k} \in M_a : h_{a,k}(t) > c \}.
\]

Both of these sets are non-empty. By Theorem 4.3 (iv), \( M_a \) is a linearly ordered subset of the set of \( \pi \sqrt{1 - y^2} \)-Lipschitz paths in \( \text{Lip} \) with the usual partial order on functions. Therefore the sets \( A \) and \( B \) have a supremum and infimum in \( \text{Lip} \) by the Arzelà-Ascoli Theorem. By the lower semicontinuity of flux (Proposition 3.5), \( \sup A \) and \( \inf B \) lie in the set \( M_a \). This, combined with the monotonicity from Theorem 4.3 (iv), implies that

\[
\begin{align*}
\sup A &= h_{a,k_A}, & k_A &= \sup \{ k : h_{a,k} \in A \}, & \text{and} & \\
\inf A &= h_{a,k_B}, & k_B &= \inf \{ k : h_{a,k} \in B \}.
\end{align*}
\]

Since \( A \cup B = M_a \), we either have that \( k_A = k_B \), or else \( k_A = -a \) and \( k_B = a \). Since \( h_{a,a} = h_{-a,a} \) by Theorem 4.3 (v), this implies that \( h_{a,k_A}(t) = h_{a,k_B}(t) = c \), so \( c \in K_{a,t} \).

**Proof of Proposition 5.1.** First, \( Y'(0) \) exists almost surely by Lemma 4.16. We now show that for every \( q \in (-\pi, \pi) \) and \( b < c \in (-1, 1) \), that

\[
\int_b^c \mathbb{E}_a(Y'(0) - \sqrt{1 - a^2})^+ da = D^*_\mu(q) \int_b^c \sqrt{1 - a^2} da. \tag{20}
\]

Here \( \mathbb{E}_a \) is the expectation taken with respect to \( \mathbb{P}_a \). Lemma 5.2 guarantees that there exists a time \( t_0 > 0 \) such that for any \( r \in [0, t_0] \) and \( a \in [b, c] \), there is a minimal flux \( \pi \sqrt{1 - y^2} \)-Lipschitz path \( g \in \text{Lip}^+ \) with \( g(0) = a \) and \( g(r) = \sin(\arcsin(a) + qr) \). Call this path \( g_{a,r} \). Letting

\[
s_{a,r}(t) = \frac{g'_{a,r}(t)}{\sqrt{1 - g^2_{a,r}(t)}}
\]

be the local speed of \( g_{a,r} \), we have that

\[
\begin{align*}
J^+(g_{a,r}; [0,r]) &\geq \frac{1}{2} \min_{t \in [0,r]} \int_0^r D^*_\mu(s_{a,r}(t)) dt \\
&\geq \frac{1}{2} \min_{t \in [0,r]} \sqrt{1 - g^2_{a,r}(t)} r D^*_\mu(q). \tag{21}
\end{align*}
\]

Here the second inequality follows by Jensen’s inequality, using that \( D^*_\mu \) is convex (Lemma 2.9) and that the average local speed of \( g_{a,r} \) is \( q \). Now letting \( g_a(t) = \sin(\arcsin(a) + qt) \), Lemma 4.7 implies that

\[
J^+(g_{a,r}; [0,r]) \leq J^+(g_a; [0,r]) \leq \frac{1}{2} \max_{t \in [0,r]} \sqrt{1 - g^2_a(t)} r D^*_\mu(q). \tag{22}
\]
Combining this inequality with (21) implies that
\[
\lim_{r \to 0^+} \frac{J^+(g_{a,r}; [0, r])}{r} = \frac{D_\mu^+(q)}{2} \sqrt{1 - a^2}.
\]
Moreover, (22) implies that \( J^+(g_{a,r}; [0, r]) \leq r D_\mu^+(q) \) for all \( a \in [b, c], r \in [0, t_0] \). Therefore by the bounded convergence theorem, we have that
\[
\lim_{r \to 0^+} \int_b^c \frac{J^+(g_{a,r}; [0, r])}{r} \, da = \frac{D_\mu^+(q)}{2} \int_b^c \sqrt{1 - a^2} \, da.
\]
(23)
Now recall that \( P_Y^+(g_{a,r}; [0, r]) \) is the probability that \( Y \) upcrosses \( g_{a,r} \) in the interval \([0, r]\). Since minimal flux paths cross at most once (Lemma 4.14),
\[
\mathbb{P}(Y(0) < a, Y(r) > g_{a,r}(r)) = P_Y^+(g_{a,r}; [0, r]).
\]
Therefore by Proposition 4.13, we can write
\[
\int_b^c \frac{J^+(g_{a,r}; [0, r])}{r} \, da = \int_b^c \frac{\mathbb{P}(Y(0) < a, Y(r) > g_{a,r}(r))}{r} \, da.
\]
Then defining \( P_x(a, r) = \mathbb{P}(Y(r) > g_{a,x}(r))r \), the right hand side above is equal to
\[
\int_b^c \frac{1}{2} \int_{-1}^a P_x(a, r) \, dx \, da = \int_{-1}^c \int_{xb}^c P_x(a, r) \, dx \, da.
\]
(24)
Now define \( S_{x,y} = \sin(\arcsin(x) + (\pi - q)y) \). Since \( Y \) is \( \pi \sqrt{1 - h^2} \)-Lipschitz almost surely (Theorem 2.10), for almost every \( x \in [-1, 1] \), we have that
\[
\mathbb{P}_x(Y(t) > g_{a,r}(r)) = 0
\]
whenever \( a > S_{x,r} \). Therefore we can rewrite the right hand side of (24) as
\[
\frac{1}{2} \left[ \int_b^c \int_{x,c} S_{x,r} P_x(a, r) \, dx \, da - \int_{S_{x,r}}^c \int_c S_{x,r} P_x(a, r) \, dx \, da + \int_b^c \int_{S_{b,r}} S_{x,r} P_x(a, r) \, dx \, da \right].
\]
(25)
In the second and third terms above, the integrand is of size \( O(r^{-1}) \), whereas the region of integration is of size \( O(r^2) \). Therefore these terms go to zero as \( r \to 0 \). Now letting
\[
f(a, x, s, r) = \frac{\sin(\arcsin(a) + s r) - x}{r},
\]
and making the substitution \( a \to u = f(a, x, q, r) \), we can write the first term of (25) as
\[
\int_b^c \int_{f(x,q,r)} f(x,q,r) \frac{\mathbb{P}_x(Y(r) > Y(0))}{r} \frac{\cos(\arcsin(r u + x) - qr)}{\sqrt{1 - (r u + x)^2}} \, dudx.
\]
(26)
Moreover, for every $x \in [-1, 1]$ such that $Y'(0)$ exists $\mathbb{P}_x$-almost surely, and for $u \in [q\sqrt{1-x^2}, \pi\sqrt{1-x^2}]$ for which $u$ is not an atom of the distribution $\mathbb{P}_x(Y'(0) \in \cdot)$, we have that

$$\lim_{r \to 0} \mathbb{P}_x \left( \frac{Y(r) - Y(0)}{r} > u \right) = \mathbb{P}_x(Y'(0) > u).$$

Therefore by the dominated convergence theorem, (26) converges to

$$\int_b^c \int_{q\sqrt{1-x^2}}^{\pi\sqrt{1-x^2}} \mathbb{P}_x(Y'(0) > u)du \, dx = \int_b^c \mathbb{E}_x(Y'(0) - \sqrt{1-x^2}q)^+ \, dx \quad \text{as } r \to 0.$$  

The equality above again uses that $|Y'(0)| \leq \pi\sqrt{1-Y^2(0)}$ almost surely, since $Y$ is $\pi\sqrt{1-y^2}$-Lipschitz. Combining this with (23), (24), and (25) proves (20). Now define

$$d(Y) = \frac{Y'(0)}{\sqrt{1-Y^2(0)}}.$$  

By (20), for almost every $a \in (-1, 1)$, for every $q \in \mathbb{Q} \cap (\pi, \pi)$, we have that

$$\sqrt{1-a^2} \mathbb{E}_a (d(Y) - q)^+ = \mathbb{E}_a \left( Y'(0) - \sqrt{1-a^2}q \right)^+ = \sqrt{1-a^2} D^+_a(q).$$

Therefore by continuity in $s$ of the functions $\mathbb{E}_a (d(Y) - s)^+$ and $D^+_a(s)$, we have that

$$\mathbb{E}_a (d(Y) - s)^+ = D^+_a(s) \quad \text{for all } s \in (-\pi, \pi). \quad (27)$$

Now, $D^+_a(s)$ is Lipschitz (Lemma 2.9) and hence differentiable almost everywhere, so we can differentiate both sides of the above equation to get that for almost every $s \in (-\pi, \pi)$, that

$$\mathbb{P}_a(d(Y) > s) = \mu(s, \infty).$$

Finally, for almost every $a \in (-1, 1)$, we have that

$$\mathbb{P}_a(d(Y) \in [-\pi, \pi]) = \mu[\pi, \pi] = 1$$

since $Y$ is almost surely $\pi\sqrt{1-y^2}$-Lipschitz (Theorem 2.10). Therefore $\mathbb{P}_a(d(Y) \in \cdot) = \mu$. \hfill \Box

### 5.1 The max-height distribution and the uniqueness of $Y$

As an application of Proposition 5.1, we can find $\mathbb{P}(m(Y) > a)$ for all $a \in [0, 1]$. This allows us to deduce the uniqueness of $Y$.

**Theorem 5.3.** Let $Y$ be any subsequential limit of $Y_n$. For all $a \in [0, 1]$, we have that

$$\mathbb{P}(m(Y) > a) = \sqrt{1-a^2}. \quad \text{That is, } m(Y) \stackrel{d}{=} \sqrt{1-U^2} \text{ for a uniform random variable } U \text{ on } [0, 1].$$
Proof. Let $V(Y)$ be the total variation of $Y$ on the interval $[0,1]$. We have that

$$
\mathbb{E}V(Y) = \mathbb{E}\int_0^1 |Y'(t)|dt = \mathbb{E}\left(\frac{|Y'(0)|}{\sqrt{1-Y^2(0)}}\right)\sqrt{1-Y^2(0)} = 2\mathbb{E}\sqrt{1-U^2},
$$

where $U$ is a uniform random variable on $[0,1]$. Here the second equality follows from time stationarity of $Y$ (Proposition 2.11 (i)). The third equality follows by Proposition 5.1, the fact that the first moment of $\mu$ is 2 (Lemma 4.5), and the fact that $Y(0)$ is uniformly distributed (Theorem 2.10).

Moreover, the characterization of minimal flux paths (Theorem 4.2) implies that $V(Y) = 2\mathbb{E}m(Y)$, so $\mathbb{E}m(Y) = \mathbb{E}\sqrt{1-U^2}$. Finally, the bound in Lemma 4.4 implies that

$$
\mathbb{P}(m(Y) > a) \leq \frac{D\mu(0)}{2} \sqrt{1-a^2} = \sqrt{1-a^2} = \mathbb{P}(\sqrt{1-U^2} > a),
$$

so the random variable $m(Y)$ is stochastically dominated by $\sqrt{1-U^2}$. Hence $m(Y) \stackrel{d}{=} \sqrt{1-U^2}$.

Now we can prove the existence of a unique limit $Y$ of $Y_n$. First extend the paths $h_m$ defined in Theorem 4.2 to paths $h_m : [0, \infty) \to [-1,1]$ by letting $h_m(t) = -h_m(t-1)$ for all $t > 1$.

**Theorem 5.4.** The sequence $Y_n$ has a distributional limit $Y$ given by

$$
Y(t) = h_{\sqrt{1-U^2}}(V + t).
$$

Here $U$ is a uniform random variable on $[0,1]$ and $V$ is uniform on $[0,2]$, independent of $U$.

**Proof.** By Theorem 4.3 and Theorem 3.1 (iii), any subsequential limit $Y$ of $Y_n$ is supported on the set of paths

$$
\{h_m(\cdot + u) : m \in [0,1], u \in [0,2]\},
$$

so we can write $Y(\cdot) = h_M(\cdot + V)$, for a pair of random variables $(M,V)$. By Theorem 5.3, $M = \sqrt{1-U^2}$, for a uniform random variable $U$ on $[0,1]$. By the time stationarity of $Y$ (Proposition 2.11 (i)), for any $t \in [0,2]$, we have that

$$
h_M(\cdot + t + V) \stackrel{d}{=} h_M(\cdot + V),
$$

which in turn implies that $(M,V) \stackrel{d}{=} (M,V+t \pmod{2})$ for any $t \in [0,2]$. This implies that $V$ has uniform distribution, and that $M$ is independent of $V$. $\square$

43
6 The integral transform formula

Let \( \mu_+ \) be the pushforward of the local speed distribution \( \mu \) under the map \( x \mapsto |x| \). In this section we prove an integral transform formula for \( \mu_+ \). This transform formula allows us to identify \( \mu \) as the arcsine distribution on \([-\pi, \pi]\). Once we know \( \mu \), we can find the set of minimal flux paths \( h_m \) introduced in Theorem 4.2, and then in turn identify \( Y \).

To find the integral transform formula, we will calculate \( \Pr(m(Y) > k) \) in two different ways. We first give a heuristic explanation of how to do this when the local speed distribution \( \mu \) has no atoms. By Theorem 5.3, we have that

\[
\Pr(m(Y) > k) = \sqrt{1 - k^2}.
\]

We can also calculate \( \Pr(m(Y) > k) \) by integrating the marginal probabilities \( \Pr(a)(m(Y) > k) \). This gives that

\[
\sqrt{1 - k^2} = \frac{1}{2} \int_{-1}^{1} \Pr(a)(m(Y) > k) \, da = 1 - k + \frac{1}{2} \int_{-k}^{k} \Pr(a)(m(Y) > k) \, da.
\]

Now we want to find an expression for \( \Pr(a)(m(Y) > k) \) when \( |a| < k \). By the ordering on minimal flux paths, if \( Y(0) = a \), then \( m(Y) > k \) if and only if \( |Y'(0)| \) is greater than some threshold value. Therefore for some \( s_{a,k} \in \mathbb{R} \), we have that

\[
\Pr(a)(m(Y) > k) = \Pr(|Y'(0)| > \sqrt{1 - a^2 s_{a,k}}) = \mu_+(s_{a,k}, \infty).
\]

The final equality here follows from Proposition 5.1. To find \( s_{a,k} \), we calculate \( \Pr(m(Y) > k \mid m(Y) > a) \). Set

\[
T_a = \inf\{t \in [0,1] : |Y(t)| = a\}.
\]

Note that the time \( T_a \) may not exist, in which case we set \( T_a = \infty \). We should have that

\[
\Pr(m(Y) > k \mid m(Y) > a) = \Pr(|Y'(T_a)| > \sqrt{1 - a^2 s_{a,k}} \mid T_a < \infty).
\]

Now, the “amount of time” that \( Y \) spends with \( |Y| = a \) is inversely proportional to its speed at that location. Therefore the distribution of \( |Y'(T_a)| \) should be a size-biased version of the distribution of \( Y'(0) \) given that \( |Y(0)| = a \). Hence,

\[
\Pr(|Y'(T_a)| > \sqrt{1 - a^2 s_{a,k}}) = \hat{\mu}_+(s_{a,k}, \infty),
\]

where \( \hat{\mu}_+ \) is the size-biased distribution of \( \mu_+ \) (we define this formally in the next paragraph). Finally, we can also calculate \( \Pr(m(Y) > k \mid m(Y) > a) \) using Theorem 5.3. This gives that

\[
\hat{\mu}_+(s_{a,k}, \infty) = \Pr(m(Y) > k \mid m(Y) > a) = \frac{\Pr(m(Y) > k)}{\Pr(m(Y) > a)} = \frac{\sqrt{1 - k^2}}{\sqrt{1 - a^2}}.
\]
We can combine this with (29) and (28) to get an integral transform formula involving the function \( r_\mu_+(x) = S(\hat{S}^{-1}(x)) \), where \( S \) and \( \hat{S} \) are the survival functions of \( \mu_+ \) and \( \hat{\mu}_+ \), respectively.

We now precisely define everything that is needed to state the integral transform formula. For a probability measure \( \nu \) on \((0, \infty)\) with finite mean, define the size-biased distribution \( \hat{\nu} \) on \((0, \infty)\) by the Radon-Nikodym derivative formula

\[
\frac{d\hat{\nu}}{d\nu}(x) = \frac{x}{\int_0^\infty x d\nu(x)}.
\]

In order to define the integral transform when \( \mu \) has atoms, we define the extended survival function \( S : \mathbb{R} \times [0, 1] \to [0, 1] \) of a probability measure \( \nu \) on \((0, \infty)\) with finite mean by

\[
S(x, q) = \nu(x, \infty) + (1 - q) \nu(x).
\]

\( S \) is a non-increasing, continuous function in the lexicographic ordering on \( \mathbb{R} \times [0, 1] \). The function \( S \) can be thought of as the survival function in the lexicographic ordering for \( \nu \times L \), where \( L \) is uniform measure on \([0, 1]\). Now define the size-bias ratio function \( r_\nu(x) : [0, 1] \to [0, 1] \) of a probability measure \( \nu \) on \((0, \infty)\) with finite mean by

\[
r_\nu(x) = S(\hat{S}^{-1}(x)),
\]

where \( \hat{S} \) is the extended survival function of \( \hat{\nu} \) and \( S \) is the extended survival function of \( \nu \). Here the inverse function is given by

\[
\hat{S}^{-1}(x) = \sup\{(y, q) \in \mathbb{R} \times [0, 1] : \hat{S}(y, q) = x\},
\]

where the supremum is taken with respect to the lexicographic ordering. When \( \nu \) has no atoms, we can define \( r_\nu \) in terms of the usual survival functions.

**Proposition 6.1** (The integral transform). Let \( \mu_+ \) be the pushforward of the measure \( \mu \) under the map \( f(x) = |x| \). For every \( k \in (0, 1) \), we have that

\[
1 - k + \int_0^k r_\mu_+ \left( \frac{\sqrt{1 - k^2}}{\sqrt{1 - x^2}} \right) dx = \sqrt{1 - k^2}.
\]

(30)

To prove Proposition 6.1, we first establish that \( \mu(0) = 0 \).

**Lemma 6.2.** \( \mu(0) = 0 \).

**Proof.** By Lemma 4.16, \( \mathbb{P}\left( \frac{Y''(0)}{\sqrt{1 - Y'(0)^2}} = 0 \right) = 0 \). By Proposition 5.1, this is equal to \( \mu(0) \). \( \Box \)
Now let \( m < 1 \), and let \( h_m \) be as in Theorem 4.2. For any \( a \in [0, m] \), define

\[
s_{a,m} = \frac{1}{\sqrt{1 - a^2}} \liminf_{r \to 0} \frac{|h_m(t^* + r) - h_m(t^*)|}{r},
\]

where \( t^* \) is any point where \( h_m(t^*) = a \). Note that \( s_{a,m} \) is independent of \( t^* \) by the symmetry of minimal flux paths (Theorems 4.2 (i)). If \( h_m \) is differentiable at \( t^* \), then \( s_{a,m} = |h_m'(t^*)|/\sqrt{1 - a^2} \). We now prove an integral formula relating the speeds \( s_{a,m} \) to the local speed distribution.

**Lemma 6.3.** Let \( \nu \) be the law of \( m(Y) \). For almost every \( a \in [0, 1] \), for every \( k \in (a, 1) \) there exists a constant \( q_{a,k} \in [0, 1] \) such that

\[
\mathbb{P}_a(m(Y) > k) = \frac{1}{\sqrt{1 - a^2}} \int \mathbf{1}(s_{a,m} > s_{a,k} \text{ or } s_{a,m} = s_{a,k}, m > k) \frac{2}{s_{a,m}} d\nu(m) = \mu_+(s_{a,k}, \infty) + q_{a,k} \mu_+(s_{a,k}).
\]

Moreover, \( \mathbb{P}_a(m(Y) > k) \) is a continuous function of \( k \in (a, 1) \) for almost every \( a \in [0, 1] \).

**Proof.** Let \( h_m \) be as in Theorem 4.2, and let \( a < b \in [0, 1] \). We first compute the amount of time that \( h_m \) spends in the interval \([a, b] \). Define \( h_m^{-1} : [0, m] \to [0, 1/2] \) by

\[
h_m^{-1}(x) = \inf \{t : h_m(t) = x\}.
\]

By the strict monotonicity and symmetry of \( h_m \) (Theorem 4.2 (i), (ii)), we can write

\[
\mathcal{L}\{t : |h_m(t)| \in [a, b]\} = 2[h_m^{-1}(b) - h_m^{-1}(a)],
\]

where \( \mathcal{L} \) is Lebesgue measure on \([0, 1]\). Now by the concavity of \( h_m \) (Lemma 4.15), the inverse \( h_m^{-1} \) is absolutely continuous with derivative \( 2/[s_{x,m} \sqrt{1 - x^2}] \) for almost every \( x \). Therefore the left hand side of (32) is equal to

\[
\int_a^b \frac{2}{s_{x,m} \sqrt{1 - x^2}} dx.
\]

Now let \( U \) be a uniform random variable on \([0, 1]\) that is independent of \( Y \), for any \( a < b < k \in [0, 1] \), we have that

\[
\mathbb{P}(m(Y) > k \text{ and } Y(0) \in [a, b]) = \frac{1}{2} \mathbb{P}(m(Y) > k \text{ and } |Y(U)| \in [a, b])
\]

\[
= \frac{1}{2} \int_k^1 \mathcal{L}\{t : |h_m(t)| \in [a, b]\} d\nu(m)
\]

\[
= \int_a^b \frac{1}{\sqrt{1 - x^2}} \int_k^1 \frac{1}{s_{x,m}} d\nu(m) dx.
\]
The first equality above follows by the time stationarity and symmetry of $Y$ (Proposition 2.11 (i) and (ii)). The second equality follows since $Y$ is supported on shifts of the minimal flux paths $h_m$ (Theorem 4.3). This implies that for almost every pair $a < k \in [0,1)$, we have

$$\mathbb{P}_a(m(Y) > k) = \frac{1}{\sqrt{1-a^2}} \int_0^1 \frac{2}{s_{a,m}} dv(m).$$  \hspace{1cm} (33)

Now by the ordering on minimal flux paths (Theorem 4.3(iv)), we have that

$$\text{if } s_{a,m(Y)} < s_{a,k}, \text{ then } m(Y) < k.$$ \hspace{1cm} (34)

This allows us to rewrite (33) to get that

$$\mathbb{P}_a(m(Y) > k) = \frac{1}{\sqrt{1-a^2}} \int_{s_{a,k}}^{s_{a,m}} 1(s_{a,m} > s_{a,k} \text{ or } s_{a,m} = s_{a,k}, m > k) \frac{2}{s_{a,m}} dv(m)$$ \hspace{1cm} (35)

for almost every pair $a < k \in [0,1)$. Moreover, for almost every $a \in [0,1)$, the derivative $Y'(0)$ exists almost surely (Proposition 5.1). In particular,

$$|Y'(0)| = \sqrt{1-a^2} s_{a,m(Y)} \mathbb{P}_a-\text{almost surely.}$$

This uses the fact that $Y$ is supported on shifts of $h_{m(Y)}$. Also, $Y'(0)/\sqrt{1-a^2}$ has distribution $\mu$ for almost every $a$ (Proposition 5.1). Therefore (34) implies that for almost every $a \in [0,1)$, for every $k \in (a,1)$ there exists a constant $q_{a,k} \in [0,1]$ such that

$$\mathbb{P}_a(m(Y) > k) = \mu_+(s_{a,k}, \infty) + q_{a,k} \mu_+(s_{a,k}).$$ \hspace{1cm} (36)

Now let $a$ be such that (36) holds for every $k \in (a,1)$ and (33) and (35) hold for almost every $k \in (a,1)$. By concavity of minimal flux paths (Lemma 4.15), $s_{a,m} > 0$ whenever $m > a$. Therefore since $\nu$ has a Lebesgue density by Theorem 5.3, the right hand side of (33) is continuous and non-increasing. Since both sides of (35) are also non-increasing, and are equal to the right hand side of (33) for almost every $k \in (a,1)$, they must be equal for every $k \in (a,1)$. Combining this with (36) proves the lemma. \hfill \Box

**Lemma 6.4.** Let $s_a$ be the law of $\mathbb{P}(s_{a,m(Y)} \in \cdot | m(Y) > a)$, and define the measure $d\bar{s}_a$ by the Radon-Nikodym formula

$$\frac{2}{s} ds_a(s) = d\bar{s}_a(s).$$

Then $\bar{s}_a = \mu_+$ for almost every $a \in [0,1)$. In particular, for such $a$, for all $k \in (a,1)$ we have

$$q_{a,k} \mu_+(s_{a,k}) = \frac{2}{s_{a,k} \sqrt{1-a^2}} \int_{s_{a,k}}^{s_{a,m}} 1(s_{a,m} = s_{a,k}, m > k) dv(m).$$ \hspace{1cm} (37)

Implicit in the statement of Lemma 6.4 is the claim that $s_a > 0$ almost surely for all $a$. This follows since $s_{a,k} > 0$ whenever $a < k$, which is a consequence of the concavity of minimal flux paths (Lemma 4.15).
Proof. Let \(a\) be such that (31) holds for every \(k \in (a, 1)\) and such that \(P_a(m(Y) > k)\) is continuous. Further assume that

\[
P_a(Y'(0) \text{ exists and is non-zero }, m(Y) < 1, J(Y) = 1) = 1.
\]

These conditions hold for almost every \(a \in (0, 1)\) (Proposition 5.1, Lemma 6.2). Noting that \(\sqrt{1 - a^2} = P(m(Y) > a)\) by Theorem 5.3, equation (31) implies that for every \(k \in (a, 1)\), there exists a \(p_{a,k} \in [0, 1]\) such that

\[
P_a(m(Y) > k) = \bar{\nu}_a(s_{a,k}, \infty) + p_{a,k}\bar{\nu}_a(s_{a,k}) = \mu_+(s_{a,k}, \infty) + q_{a,k}\mu_+(s_{a,k}). \tag{38}
\]

Now, since \(Y'(0) \neq 0\), \(P_a\)-almost surely, the concavity of minimal flux paths (Lemma 4.15) implies that \(P_a(m(Y) = a) = 0\). Moreover, since \(m(Y) < 1\), \(P_a\)-almost surely, we have that \(P_a(m(Y) \in (a, 1)) = 1\), and so

\[
\lim_{k \to 1} P_a(m(Y) > k) = 0 \quad \text{and} \quad \lim_{k \to a} P_a(m(Y) > k) = 1. \tag{39}
\]

Therefore, (38) and the continuity of \(P_a(m(Y) > k)\) implies that

\[
\mu_+(s_{a,k} : k \in (a, 1)) = \bar{\nu}_a(s_{a,k} : k \in (a, 1)) = 1. \tag{40}
\]

Now fix \(k \in (a, 1)\). Let

\[
k^* = \inf\{\ell \in [k, 1] : s_{a,k} < s_{a,m} \quad \text{for all} \quad m > \ell\}.
\]

Equation (38) and the continuity of \(P_a(m(Y) > k)\) then implies that

\[
P_a(m(Y) > k^*) = \bar{\nu}_a(s_{a,k}, \infty) = \mu_+(s_{a,k}, \infty).
\]

Combining this with (40) proves that \(\mu_+ = \bar{\nu}_a\). Equation (37) follows by using that \(\mu_+ = \bar{\nu}_a\) to simplify equation (31). \hfill \Box

Proof of Proposition 6.1. Fix \(a\) so that the conclusion of Lemma 6.4 holds, and let \(k > a\). By Theorem 5.3 and the ordering on minimal flux paths, we can write

\[
P(m(Y) > k \mid m(Y) > a) = \frac{1}{\sqrt{1 - a^2}} \left[ \int 1(s_{a,m} > s_{a,k})d\nu(m) + \int 1(s_{a,m} = s_{a,k}, m > k)d\nu(m) \right].
\]

By Lemma 6.4, we can rewrite the first integral above in terms of \(\bar{\nu}_a\), and then in terms of \(d\mu_+\). This gives that

\[
\frac{1}{\sqrt{1 - a^2}} \int 1(s_{a,m} > s_{a,k})d\nu(m) = \int 1(s > s_{a,k})d\bar{\nu}_a(s) = \int 1(s > s_{a,k})\frac{s}{2}d\mu_+(s).
\]

We can also rewrite the second integral using (37). This implies that

\[
P(m(Y) > k \mid m(Y) > a) = \int 1(s > s_{a,k})\frac{s}{2}d\mu_+(s) + \frac{s_{a,k}}{2}q_{a,k}\mu_+(s_{a,k}).
\]
By Lemma 4.5, we can recognize $\frac{1}{2}$ as the Radon-Nikodym derivative of the size-biased random variable $\hat{\mu}_+$ with respect to $\mu_+$, proving that

$$\frac{\sqrt{1-k^2}}{\sqrt{1-a^2}} = \mathbb{P}(m(Y) > k | m(Y) > a) = \hat{\mu}_+(s_{a,k}, \infty) + q_{a,k}\hat{\mu}_+(s_{a,k})$$

for almost every $a < k \in [0,1)$. Now, Lemma 6.3 allows us to conclude that

$$\mathbb{P}_a(m(Y) > k) = S(s_{a,k}, 1 - q_{a,k}) = S\left(\frac{\sqrt{1-k^2}}{\sqrt{1-a^2}}\right)$$

for almost every $a < k \in [0,1)$. Combining this with the symmetry of $Y$ (Proposition 2.11) and Theorem 5.3 implies that

$$\sqrt{1-k^2} = \mathbb{P}(m(Y) > k) = \int_0^1 \mathbb{P}_a(m(Y) > k) da = 1 - k + \int_0^k r_{\mu_+}\left(\frac{\sqrt{1-k^2}}{\sqrt{1-a^2}}\right) da. \quad \square$$

7 The weak trajectory limit

In this section we show that the integral transform in Proposition 6.1 determines the local speed distribution. This will allow us to immediately conclude that the weak limit of the trajectory random variables $Y_n$ is the Archimedean path. We begin with two basic lemmas about size-bias ratio functions.

**Lemma 7.1.** For any probability distribution $\nu$ on $(0, \infty)$ with finite first moment $m$, the size-bias ratio function $r_\nu$ is locally Lipschitz on $[0,1)$ and continuous on $[0,1]$.

**Proof.** Let $\pi_1(x, y) = x$. By calculus, when $x \in (0, 1)$ we have that

$$\partial_+ r_\nu(x) = \lim_{h \to 0^+} \frac{m}{\pi_1(\hat{S}^{-1}(x + h))} \quad \text{and} \quad \partial_- r_\nu(x) = \lim_{h \to 0^-} \frac{m}{\pi_1(\hat{S}^{-1}(x - h))}.$$ 

The first equation also holds at $x = 0$. As $\pi_1(\hat{S}^{-1}(y))$ is a decreasing function of $y$, and strictly positive for all $y \in [0,1)$, this shows that $r_\nu$ is locally Lipschitz on $[0,1)$. It is straightforward to check that $r_\nu$ is continuous at 1. \qed

**Lemma 7.2.** Suppose that $\nu_1$ and $\nu_2$ are probability measures on $(0, \infty)$ with the same first moment, such that $r_{\nu_1} = r_{\nu_2}$. Then $\nu_1 = \nu_2$.

**Proof.** We prove the contrapositive. Let $\nu_1$ and $\nu_2$ be measures with the same first moment, and suppose that $\nu_1 \neq \nu_2$. Since $\nu_1$ and $\nu_2$ have the same first moment, $\hat{\nu}_1 \neq \hat{\nu}_2$, so for some value of $(y, q) \in (0, \infty) \times [0,1]$, we have that $\hat{S}_{\nu_1}(y, q) \neq \hat{S}_{\nu_2}(y, q)$. Since the functions
\( \hat{S}_{\nu_i}(y, \cdot) \) and \( \hat{S}_{\nu_2}(y, \cdot) \) are linear for any fixed value of \( y \), this implies that \( \hat{S}_{\nu_1}(y, r) \neq \hat{S}_{\nu_2}(y, r) \) for some \( r \in \{0, 1\} \). Moreover, since
\[
\hat{S}_{\nu_i}(y, 1) = \sup \{ \hat{S}_{\nu_i}(z, 0) : z > y \}, \quad i = 1, 2,
\]
we can conclude that \( \hat{S}_{\nu_1}(z, 0) \neq \hat{S}_{\nu_2}(z, 0) \) for some \( z \in (0, \infty) \). Without loss of generality, assume that \( a_1 := \hat{S}_{\nu_1}(z, 0) > a_2 := \hat{S}_{\nu_2}(z, 0) \). Therefore using the notation of the previous lemma, letting \( b = (a_1 + a_2)/2 \), we have that
\[
\lim_{h \to 0^+} \pi_1(\hat{S}_{\nu_1}(b + h)) \geq z, \quad \text{whereas} \quad \lim_{h \to 0^+} \pi_1(\hat{S}_{\nu_2}(b + h)) < z.
\]
Since \( b \in (0, 1) \), the derivative computation in the previous lemma combined with the fact that \( \nu_1 \) and \( \nu_2 \) have the same first moment implies that \( r_{\nu_1} \neq r_{\nu_2} \).

Now let \( \mathcal{X} \) be the space of continuous functions from \([0, 1] \to \mathbb{R}\) that are locally Lipschitz on \([0, 1]\). Define an integral transform \( H \) on \( \mathcal{X} \) by
\[
H(r)(k) = \int_0^k r \left( \frac{\sqrt{1-k^2}}{\sqrt{1-x^2}} \right) dx,
\]
for \( k \in [0, 1] \). By Lemma 7.1, any size-bias ratio function is in \( \mathcal{X} \), so if the integral transform \( H \) is injective on \( \mathcal{X} \), then \( H(r_{\mu_+}) \) determines \( r_{\mu_+} \).

**Lemma 7.3.** The integral transform \( H \) is injective on \( \mathcal{X} \).

**Proof.** Let \( r \in \mathcal{X}, r \neq 0 \). We will show that \( H(r) \neq 0 \). Without loss of generality, we may assume that
\[
\max_{x \in [0, 1]} r(x) = \delta > 0.
\]
Letting \( u = \frac{\sqrt{1-k^2}}{\sqrt{1-x^2}} \) and \( y = \sqrt{1-k^2} \), we have that
\[
H(r)(\sqrt{1-y^2}) = \int_0^1 r(u) \frac{y^2}{u^2 \sqrt{u^2 - y^2}} du.
\]
For \( y \in (0, 1] \), we have that
\[
\overline{H}(r)(y) := \frac{H(r)(\sqrt{1-y^2})}{y^2} = \int_{y}^1 \frac{r(u)}{u^2 \sqrt{u^2 - y^2}} du. \tag{41}
\]
It suffices to show that \( \overline{H}(r)(y) \neq 0 \) for some \( y \in (0, 1] \). Observe that if \( r(1) > 0 \), then there would exist a \( \gamma > 0 \) such that \( r(x) > 0 \) for all \( x \geq \gamma \), so \( \overline{H}(r)(\gamma) > 0 \). Also, if \( r(0) > 0 \), then \( \overline{H}(r)(y) \to \infty \) as \( y \to 0 \). Therefore there must exist \( y \in (0, 1) \) be such that \( r(y) = \delta \).
Since $r$ is locally Lipschitz on $[0, 1)$, we can find $k, \epsilon_0 > 0$ such that $r(y + x) \geq \delta - kx$ for all $x \in [0, \epsilon_0]$. Therefore for $\epsilon < \epsilon_0$ we have

$$
\int_y^{y+\epsilon} \frac{r(u)}{u^2\sqrt{u^2 - y^2}} du \geq \int_y^{y+\epsilon} \frac{(\delta - k\epsilon)\sqrt{\epsilon^2 + 2\epsilon y}}{y^2(y + \epsilon)} du = \frac{(\delta - k\epsilon)\sqrt{\epsilon^2 + 2\epsilon y}}{y^2(y + \epsilon)}.
$$

(42)

Now consider the difference between $\overline{\Pi}(r)(y)$ and $\overline{\Pi}(r)(y + \epsilon)$. We have

$$
\begin{align*}
\overline{\Pi}(r)(y) - \overline{\Pi}(r)(y + \epsilon) & = \int_y^{y+\epsilon} \frac{r(u)}{u^2\sqrt{u^2 - y^2}} du + \int_1^{y+\epsilon} \frac{r(u)}{u^2\sqrt{u^2 - y^2}} du - \int_{y+\epsilon}^{y+1} \frac{r(u)}{u^2\sqrt{u^2 - (y + \epsilon)^2}} du \\
& \geq \int_y^{y+\epsilon} \frac{r(u)}{u^2\sqrt{u^2 - y^2}} du + \delta \int_1^{y+1} \frac{1}{u^2\sqrt{u^2 - y^2}} - \frac{1}{u^2\sqrt{u^2 - (y + \epsilon)^2}} du \\
& \geq \frac{(\delta - k\epsilon)\sqrt{\epsilon^2 + 2\epsilon y}}{y^2(y + \epsilon)} + \delta \left[ \frac{(y + \epsilon)^2\sqrt{1 - y^2} - y^2\sqrt{1 - (y + \epsilon)^2} - (y + \epsilon)\sqrt{\epsilon^2 + 2\epsilon y}}{(y + \epsilon)^2 y^2} \right].
\end{align*}
$$

(43)

In above calculation the first inequality comes from the fact that $r(x) \leq \delta$ for all $x$, and the observation that

$$
\frac{1}{u^2\sqrt{u^2 - y^2}} < \frac{1}{u^2\sqrt{u^2 - (y + \epsilon)^2}}
$$

for all $u \in (y + \epsilon, 1)$. The second equality follows by integration and plugging in the bound in (42). Now expanding in $\epsilon$ about $\epsilon = 0$, we get that (43) is equal to

$$
\frac{(2 - y^2)\delta \epsilon}{y^3\sqrt{1 - y^2}} + O(\epsilon^{3/2}).
$$

This is strictly greater than 0 for small enough $\epsilon$, so $\overline{\Pi}(r)(y) - \overline{\Pi}(r)(y + \epsilon) \neq 0$ for such $\epsilon$. Hence $\overline{\Pi}(r)(x) \neq 0$ for some $x \in [0, 1)$.

**Proposition 7.4.** Let $\text{arcc}$ be the arcsine distribution on $[-\pi, \pi]$, and let $\text{arcc}_x$ be the push-forward of $\text{arcc}$ under the map $x \mapsto |x|$. Then for every $k \in [0, 1]$, we have that

$$
1 - k + \int_0^k r_{\text{arcc}_x} \left( \frac{\sqrt{1 - k^2}}{\sqrt{1 - x^2}} \right) dx = \sqrt{1 - k^2}.
$$

(44)

**Proof.** The distribution $\text{arcc}$ has density $(\pi\sqrt{\pi^2 - x^2})^{-1}$ on $[-\pi, \pi]$. From this, we can calculate that

$$
r_{\text{arcc}_x}(y) = 1 - \frac{2}{\pi}\text{arcsin}(\sqrt{1 - y^2}).
$$
We now use the connection between the arcsine distribution and the Archimedean measure $\Arch$ to help evaluate the integral in (44). Let $(X_1, X_2) \sim \Arch$ and let

$$A(t) = X_1 \cos(\pi t) + X_2 \sin(\pi t)$$

be the Archimedean path. Writing $\mathbb{P}(m(A) > k) = \mathbb{P}(X_1^2 + X_2^2 > k^2)$ in both polar and Cartesian coordinates, we get that

$$\int_0^1 \frac{rdr}{\sqrt{1-r^2}} = 4 \int_0^1 \int_0^{\sqrt{1-y^2}} \frac{dxdy}{2\pi \sqrt{1-x^2 - y^2}}.$$

The left hand side can easily be evaluated as $\sqrt{1-k^2}$. The right hand side is equal to

$$(1-k) + \int_0^k \frac{1}{\pi} \left[ \frac{\pi}{2} - \arcsin \left( \frac{\sqrt{k^2 - y^2}}{\sqrt{1 - y^2}} \right) \right] dy = (1-k) + \int_0^k r_{arc} \left( \frac{\sqrt{1-k^2}}{\sqrt{1-y^2}} \right) dy. \Box$$

We can now prove Theorem 7, and in turn use that to prove Theorem 5.

**Proof of Theorem 7.** By Proposition 6.1, Lemma 7.3 and Proposition 7.4, we have that $r_{\mu_+} = r_{\text{arc}_+}$. Moreover, the first moment of $\text{arc}_+$ is 2. By Lemma 4.5, this matches the first moment of $\mu_+$. Therefore by Lemma 7.2, $\mu_+ = \text{arc}_+$. Finally, symmetry of both measures implies that $\mu = \text{arc}$. \Box

**Proof of Theorem 5.** Fix $m > 0$. Let $g_m(t) = m \sin(\pi t)$, and let $g_m^{-1}$ be the inverse of $g_m$ on the interval $[0, 1/2]$. Let

$$s_m(t) = \frac{g_m'(t)}{\sqrt{1 - g_m^2(t)}}$$

be the local speed of $g_m$. We can calculate that

$$J(g_m) = 2J(g_m; [0, 1/2]) = \int_0^{1/2} D_{\mu}(s_m(t)) \sqrt{1 - g_m^2(t)} dt$$

$$= \int_0^m \frac{D_{\mu}(s_m(g_m^{-1}(x)))}{s_m(g_m^{-1}(x))} dx. \quad (45)$$

Here we have made the substitution $x = g_m(t)$ to go from the first to the second line. Now using Theorem 7, we can calculate

$$D_{\mu}(c) = D_{\text{arc}}(c) = \frac{2}{\pi} \left[ c \arcsin \left( \frac{c}{\pi} \right) + \sqrt{\pi^2 - c^2} \right].$$

From here we can use that $s_m(g_m^{-1}(x)) = \frac{\pi \sqrt{m^2 - x^2}}{\sqrt{1-x^2}}$ to compute that

$$J(g_m) = \int_0^m \frac{2}{\pi} \left[ \frac{\sqrt{1-m^2}}{\sqrt{m^2 - x^2}} + \arcsin \left( \frac{\sqrt{m^2 - x^2}}{\sqrt{1-x^2}} \right) \right] dx$$

52
The first part of this integral can be easily evaluated, and the second part can be evaluated by comparing with the integral in the final line of the proof of Proposition 7.4. Putting this all together yields that \( J(g_m) = 1 \), as desired. Therefore by Theorem 5.4, we can write \( Y(t) = \sqrt{1 - V^2} \sin(\pi t + 2\pi U) \), where \( U \) and \( V \) are independent uniform random variables on \([0, 1]\). This is the Archimedean path! 

7.1 The empirical distribution of trajectories

By a compactness argument, we can immediately prove a slightly stronger version of Theorem 5. This will allow us to conclude Theorem 6. This theorem will also be necessary for establishing the stronger limits in Sections 8 and 9.

For a Polish space \( S \), let \( \mathcal{M}(S) \) be the space of probability measures on \( S \) with the topology of weak convergence. Note that \( \mathcal{M}(S) \) is itself a Polish space. Recall the notation \( \sigma_G \) introduced in Section 1. For a fixed sorting network \( \sigma \), let \( \nu_\sigma \in \mathcal{M}(D) \) be uniform measure on the set \( \{ \sigma_G(i, \cdot) \}_{i\in\{1,\ldots,n\}} \). Now let \( \Omega_n \) be the space of all \( n \)-element sorting networks, and define

\[
\nu_n = \frac{1}{|\Omega_n|} \sum_{\sigma \in \Omega_n} \delta(\nu_\sigma).
\]

For each \( n \), \( \nu_n \in \mathcal{M}(\mathcal{M}(D)) \). We now extend Theorem 5 to give a limit theorem for the sequence \( \nu_n \).

**Proof.** By Remark 2.4 from [DV20], the sequence \( \nu_n \) is precompact. For any subsequential limit \( \nu \) of \( \nu_n \), Theorem 5 implies that \( \nu \)-almost every \( \rho \) is supported on curves of the form \( a \sin(\pi t) + b \cos(\pi t) \). Hence if \( Z \) is a random path with law \( \rho \), then

\[
Z(t) = X_1 \cos(\pi t) + X_2 \sin(\pi t)
\]

for some random variables \((X_1, X_2)\). Moreover, \( Z(t) \) is uniform for every \( t \) since each measure \( \nu_n \) is supported on the set \( \{ \nu_\sigma \}_{\sigma \in \Omega_n} \). Therefore \((X_1, X_2) \sim \text{Arch} \), so \( Z \) is the Archimedean path, and hence \( \nu = \delta_{\text{Arch}} \).

**Proof of Theorem 6.** We use the notation introduced in the paragraphs preceding Theorem 6. Fix \( m \in \mathbb{N} \). First, by Theorem ??, we have

\[
\frac{1}{|\Omega_n|} \sum_{\sigma \in \Omega_n} \nu_\sigma \rightarrow P_\text{Arch}^m \quad \text{as} \quad n \rightarrow \infty.
\]

(46)

Here \( \nu^m \) refers to the \( m \)-fold product measure \( \nu \times \nu \times \ldots \times \nu \) and the convergence in (46) is simply weak convergence in \( \mathcal{M}(D) \). Next, let \( \tilde{\nu}_\sigma^m \) denote the product measure \( \nu_\sigma^m \) conditioned to lie in the set

\[
T := \{(x_1, \ldots, x_m) \in D^m : x_i \neq x_j \text{ for all } i, j \}
\]
Then for every \( n, m \) and every \( \sigma \in S_n \), we have \( \nu_\sigma^m(T) \geq 1 - m^2/n \), and so (46) also holds with \( \tilde{\nu}_\sigma^m \) in place of \( \nu_\sigma^m \). Now, \( \tau_m^m \) is \( n \) is equal in law to the pushforward of \( \frac{1}{\mu(n)} \sum_{\sigma \in \Omega} \tilde{\nu}_\sigma^m \) under a certain ‘subnetwork map’ \( F \), which we describe formally in the next paragraph. Therefore the limit of \( \tau_m^m \) should be the pushforward of \( \mathbb{P}_A^m \) under \( F \), which we can check to be \( \tau_m \). The details are as follows.

Let \( R_m \) be the set of \( k \)-tuples \((x_1, \ldots, x_m)\) \( D^m \) where

- For every pair \( i \neq j \), there exists a unique time \( t_{i,j} \in [0,1] \) for which either \( x_i(t) > x_j(t) \) for \( t < t_{i,j} \) and \( x_i(t) > x_j(t) \) for \( t > t_{i,j} \) or else \( x_i(t) > x_j(t) \) for \( t > t_{i,j} \) and \( x_i(t) > x_j(t) \) for \( t < t_{i,j} \). Moreover, \( t_{i,j} \neq 0 \).
- For all \((i,j) \neq (i',j')\) we have \( t_{i,j} = t'_{i',j'} \).

The set \( R_m \) should be thought of as a generalization of the space of all \( m \)-element sorting networks. The subnetwork map \( F \) takes an element \( x \in R_m \) and maps it to an \( m \)-element sorting network, as follows: for \( t \in [0,1] \), let \( \sigma_t \) be the relative order of \((x_1(t), \ldots, x_m(t))\) to \((x_1(0), \ldots, x_m(0))\). This is well-defined outside of the points \( t_{i,j} \); we define the map \( t \mapsto \sigma_t \) to be right continuous at these points. Letting \((t_0, t_1, \ldots, t_{(n-1)/2})\) be the order statistics of the points \( 0, t_{i,j}, i \neq j \in \{1, \ldots, m\} \), setting \( F(x) = (\sigma_{t_0}, \sigma_{t_1}, \ldots, \sigma_{t_{(n-1)/2}}) \) gives an \( m \)-element sorting network.

The function \( F \) is continuous on the subset of \( R_m^* \) of \( R_m \) consisting of continuous paths. Moreover, \( \mathbb{P}_A^m(R_m^*) = 1 \), so by the continuous mapping theorem, \( \tau_m^m \) converges to the pushforward of \( \mathbb{P}_A^m \) under \( F \). Since \( \mathbb{P}_A^m \) is supported on sine curves of period 2, this is simply the geometric sorting network \( \tau_m \).

\[ \square \]

## 8 The strong sine curve limit

Theorem 5 shows that for any \( \epsilon > 0 \), with high probability \((1 - \epsilon)n\) particle trajectories in a random sorting network are close to sine curves. In this section, we extend this result to all particle trajectories, thus proving Theorem 1. By combining Theorem 1 and Theorem 5, we also prove Theorems 2 and 3.

The idea behind the proof is as follows. By Theorem 5, we know that most trajectories in a typical large-\( n \) sorting network are close to sine curves. Since any two trajectories must cross exactly once, this restricts the type of behaviour that the remaining trajectories can have. Specifically, this forces all other trajectories to be either sine curves themselves, or to spend a lot of time at the edge of the sorting network. We can eliminate this second case by using the octagon bound from [AHRV07]. To state this bound, let \( A_{n,\gamma} \) be the event where

\[
|\sigma_{C}^n(i, t) - \sigma_{C}^n(i, 0)| < 2\sqrt{2t - t^2 + \gamma} \quad \text{and} \quad |\sigma_{C}^n(i, t) - \sigma_{C}^n(i, 1)| < 2\sqrt{1 - t^2 + \gamma}
\]

for all \( t \in [0,1], i \in \{1, \ldots, n\} \).
Theorem 8.1 (Octagon bound, [AHRV07]). For any $\gamma > 0$, we have that
$$
\lim_{n \to \infty} P(A_n, \gamma) = 1.
$$

Now define
$$
L_i^n(\delta) = \mathcal{L}\{t : |\sigma^n_G(i, t)| \geq 1 - \delta\},
$$
where $\mathcal{L}$ is Lebesgue measure on $[0, 1]$. This is the amount of time that particle $i$ spends within $\delta$ of the edge in the random sorting network $\sigma^n$.

Lemma 8.2. For every $\epsilon > 0$, we have that
$$
\lim_{n \to \infty} P\left(\max_{i \in [1, n]} L_i^n(\epsilon^2/16) > \epsilon\right) = 0.
$$

Proof. Fix $\epsilon > 0$, $n \in \mathbb{N}$, and let $i$ be such that $|\sigma^n_G(i, 0)| \geq 1 - \epsilon^2/16$. On the event $A_{n, \epsilon^4/64}$, a simple calculation shows that
$$
-1 + \epsilon^2/16 < \sigma^n_G(i, t) < 1 - \epsilon^2/16 \quad \text{for all } t \in [\epsilon/2, 1 - \epsilon/2].
$$
Therefore by Theorem 8.1, we have that
$$
\lim_{n \to \infty} P\left(\max_{i \in [1, n]} L_i^n(\epsilon^2/16) > \epsilon\right) = 0. \tag{47}
$$

By time stationarity of random sorting networks (Theorem 2.1), for each $n$ the above probability is greater than or equal to
$$
\epsilon P\left(\max_{i \in [1, n]} L_i^n(\epsilon^2/16) > \epsilon\right).
$$
Therefore (47) implies the lemma. \hfill \Box

Recall that $\nu_\sigma$ is uniform measure on the trajectories of a sorting network $\sigma$, and that $P_A$ is the law of the Archimedean path. To prove Theorem 1 from here, we must show that if $\nu_\sigma$ is close to $P_A$ in the weak topology, then for any particle $i$, either $\sigma_G(i, \cdot)$ spends a lot of time at the edge of the sorting network, or else $\sigma_G(i, \cdot)$ is close to a sine curve.

Recall that $\mathcal{D}$ is the closure of all sorting network trajectories in the uniform norm. To metrize weak convergence on the space of probability measures on $\mathcal{D}$, we use the Lévy-Prokhorov metric $d_{LP}$. For a set $A \subset \mathcal{D}$, define
$$
A' = \{f \in \mathcal{D} : \|f - g\|_u < \epsilon \text{ for some } g \in A\}.
$$
Here and throughout the remaining proofs, $\|\cdot\|_u$ is the uniform norm. For two probability measures $\nu_1, \nu_2$ on $\mathcal{D}$, define
$$
d_{LP}(\nu_1, \nu_2) = \inf \{\epsilon > 0 : \nu_1(A) \leq \nu_2(A') + \epsilon, \nu_2(A) \leq \nu_1(A') + \epsilon \text{ for all Borel sets } A \subset \mathcal{D}\}. 
$$
We now prove two lemmas characterizing particle behaviour when \( \nu_\sigma \) is close to \( \mathbb{P}_A \). The first gives conditions under which a particle spends time close to the edge of a sorting network. Let \( a_{n,i} = 2i/n - 1 \), and let

\[
h_{a_{n,i},1}(t) = \sin(\arcsin(a_{n,i}) + \pi t) \quad \text{and} \quad h_{a_{n,i},-1}(t) = \sin(\arcsin(a_{n,i}) - \pi t).
\]

This agrees with the notation \( h_{a,k} \) from Theorem 4.3. Recalling the definition of the Archimedean measure \( \mathcal{A}_{\text{Arch}} \) from Section 1, for \( \epsilon \in (0,1] \) define

\[
L(\epsilon) = \mathcal{A}_{\text{Arch}}(r \geq 1 - \epsilon, \theta \in [2\pi x, 2\pi(x + \epsilon)]) = \epsilon \sqrt{2\pi - \epsilon^2},
\]

where \((r, \theta)\) are polar coordinates. Note that \( L(\epsilon) \) is independent of \( x \) by the rotational symmetry of \( \mathcal{A}_{\text{Arch}} \).

**Lemma 8.3.** Let \( \gamma \in (0, 2] \) and let \( \epsilon \in (0, \gamma^2/100) \). Suppose that for a fixed \( n \)-element sorting network \( \sigma \) and a particle \( i \in \{1, \ldots, n\} \), that \( d_{LP}(\nu_\sigma, \mathbb{P}_A) < L(\epsilon)/2 \), and either

\[
\max_{t \in [0,1]} [\sigma_G(i,t) - h_{a_{n,i},1}(t)] \geq \gamma \quad \text{or} \quad \max_{t \in [0,1]} [h_{a_{n,i},-1}(t) - \sigma_G(i,t)] \geq \gamma.
\]

Then

\[
\mathcal{L}\{t \in [0,1] : |\sigma_G(i,t)| \geq 1 - 3\epsilon\} \geq \gamma/6.
\]

**Proof.** Without loss of generality, we can assume that there exists an \( s \in [0, \arccos(a_{n,i})/\pi] \) such that

\[
\sigma_G(i,s) - h_{a_{n,i},1}(s) \geq \gamma. \tag{48}
\]

This is case where particle \( i \) passes \( h_{a_{n,i},1} \) while \( h_{a_{n,i},1} \) is increasing. The other cases (where particle \( i \) passes \( h_{a_{n,i},1} \) while \( h_{a_{n,i},1} \) is decreasing, and the two cases where particle \( i \) passes \( h_{a_{n,i},-1} \)) follow by symmetric arguments. We can only be in this case if \( a_{n,i} \leq 1 - \gamma \), so \( \arccos(a_{n,i}) > \gamma \).

Since \( d_{LP}(\nu_\sigma, \mathbb{P}_A) < L(\epsilon)/2 \), for every \( x \in [0,1] \), there exists a particle \( j(x) \) such that

\[
||\sigma_G(j(x),t) - r \sin(\pi t + 2\pi(x + \beta))||_u < L(\epsilon)/2
\]

for some \( r \geq 1 - \epsilon \) and \( \beta \in [0, \epsilon] \). Now since

\[
||r \sin(\pi t + 2\pi(x + \beta)) - \sin(\pi t + 2\pi x)||_u \leq 2\epsilon,
\]

this implies that

\[
||\sigma_G(j(x),t) - \sin(\pi t + 2\pi x)||_u < L(\epsilon)/2 + 2\epsilon \tag{49}
\]

for all \( x \in [0,1] \). Define \( \delta = \gamma - \frac{L(\epsilon)}{2} - 2\epsilon \). Note that \( \delta \) is positive. For all \( \alpha \in [0, \delta] \), the inequality (48) implies that

\[
\sigma_G(i,s) - \sin(\arcsin(a_{n,i}) + \alpha + \pi s) > \gamma - \delta = L(\epsilon)/2 + 2\epsilon. \tag{50}
\]

56
Now define \( j_\alpha = j((\arcsin(a_{n,i}) + \alpha)/(2\pi)) \). Combining (50) with (49), we have that
\[ \sigma_G(i, s) > \sigma_G(j_\alpha, s) \quad \text{for } \alpha \in [0, \delta]. \]
Moreover, if \( 1 - \cos(\alpha) \geq L(\epsilon)/2 + 2\epsilon \) and \( \alpha \leq \arccos(a_{n,i}) \), then
\[ \sigma_G(i, 0) = a_{n,i} \leq \arcsin(a_{n,i}) + \alpha - (L(\epsilon)/2 + 2\epsilon), \]
and so
\[ \sigma_G(i, 0) < \sigma_G(j_\alpha, 0). \]
Now, noting that \( \delta < \gamma < \arccos(a_{n,i}) \), for every
\[ \alpha \in [\arccos(1 - L(\epsilon)/2 - 2\epsilon), \arccos(a_{n,i})] \cap [0, \delta] = [\arccos(1 - L(\epsilon)/2 - 2\epsilon), \delta], \]
the particles \( i \) and \( j_\alpha \) must cross during the interval \([0, s]\). Therefore
\[ \sigma_G(i, t) > \sigma_G(j_\alpha, t) \quad \text{for all } t > s, \ \alpha \in [\arccos(1 - L(\epsilon)/2 - 2\epsilon), \delta]. \quad (51) \]
We now show that this forces the particle \( i \) to spend a large amount of time close to the edge of the sorting network. For all \( \alpha \in [0, \delta] \), there must be some time \( t_\alpha \in [0, 1] \) such that
\[ \sin(\pi t_\alpha + \arcsin(a_{n,i}) + \alpha) = 1. \]
The time \( t_\alpha \in [0, s] \) since for every \( \alpha \in [0, \delta] \), we have that
\[ \| \sin(\pi t + \arcsin(a_{n,i}) + \alpha) - h_{a_{n,i},1}(t) \|_u < \gamma, \quad \text{and} \quad h_{a_{n,i},1}(t) \leq h_{a_{n,i},1}(s) \leq 1 - \gamma \]
for all \( t \in [0, s] \). We have used (48) to get the second inequality on the right side above. Therefore for all \( \alpha \in [\arccos(1 - L(\epsilon)/2 - 2\epsilon), \delta] \), (51) implies that
\[ \sigma_G(i, t_\alpha) > \sigma(j_\alpha, t_\alpha) \geq 1 - L(\epsilon)/2 - 2\epsilon \geq 1 - 3\epsilon. \]
Using the fact that \( \arccos(1 - x) \leq 2\sqrt{x} \) for \( x \in [0, \pi/2] \), we have that
\[ L\{t_\alpha : \alpha \in [\arccos(1 - L(\epsilon)/2 - 2\epsilon), \delta]\} \geq \delta - 2\sqrt{L(\epsilon)/2 + 2\epsilon} \geq \frac{\gamma - 3\epsilon - 2\sqrt{3\epsilon}}{\pi} \geq \frac{\gamma}{6}. \]
Here the final bound follows from the fact that \( \epsilon < \gamma^2/100. \)

The second lemma shows that if a curve \( \sigma_G(i, \cdot) \) stays close to the region between the curves \( h_{a_{n,i},-1} \) and \( h_{a_{n,i},1} \), then it must be close to a sine curve.

**Lemma 8.4.** Let \( \sigma \) be an \( n \)-element sorting network, \( i \in \{1, \ldots, n\} \), and \( \gamma \in (0, 1) \). Suppose that
\[ \max_{t \in [0, 1]} \| \sigma_G(i, t) - h_{a_{n,i},1}(t) \|_u \leq \gamma \quad \text{and} \quad \max_{t \in [0, 1]} \| h_{a_{n,i},-1}(t) - \sigma_G(i, t) \|_u \leq \gamma, \]
and that \( d_{LP}(\nu_\sigma, \mathbb{P}_\mathcal{A}) < \frac{\gamma^4}{128} \). Then there exist constants \( a \in [0, 1] \) and \( \theta \in [0, 2\pi] \) such that
\[ \| \sigma_G(i, t) - a \sin(\pi t + \theta) \|_u \leq 2\gamma + 2/n. \]
Proof. Suppose first that for some $s \in [0, 1]$, that
\[\sigma_G(i, s) \in [h_{a_{n,i},-1}(s) + \gamma, h_{a_{n,i},1}(s) - \gamma].\] (52)

Observe that $s \in [\arcsin(\gamma)/\pi, 1 - \arcsin(\gamma)/\pi]$, since otherwise the above interval is necessarily empty. By (52), we can find a point $(x_0, y_0) \in B(0, 1 - \gamma)$ such that
\[x_0 = \sigma_G(i, 0) \quad \text{and} \quad x_0 \cos(\pi s) + y_0 \sin(\pi s) = \sigma_G(i, s).\]

Since $s \leq 1 - \arcsin(\gamma)/\pi$, we can find another point $(x_1, y_1) \in B((x_0, y_0), \gamma) \subset B(0, 1)$ such that
\[x_1 > x_0 + \gamma^2/3 \quad \text{and} \quad x_1 \cos(\pi s) + y_1 \sin(\pi s) > x_0 \cos(\pi s) + y_0 \sin(\pi s) + \gamma^2/3.\] (53)

Now observe that
\[\gamma^4/64 \leq \inf \{\text{Arch}(B(x, \gamma^2/6)) : x \in B(0, 1)\}.\]

Therefore since $d_{LP}(\nu_\sigma, \mathbb{P}_{\mathbb{F}_x}) < \gamma^4/128$, there must be a point $(x_2, y_2) \in B(0, 1) \cap B((x_1, y_1), \gamma^2/6)$ and a particle $j \in \{1, \ldots, n\}$ such that
\[\|\sigma_G(j, t) - (x_1 \cos(\pi t) + y_1 \sin(\pi t))\|_u \leq \|\sigma_G(j, t) - (x_2 \cos(\pi t) + y_2 \sin(\pi t))\|_u + \gamma^2/6 < \gamma^2/3.\] (54)

By (53), this implies that
\[\sigma_G(j, 0) > \sigma_G(i, 0) \quad \text{and} \quad \sigma_G(j, s) > \sigma_G(i, s),\]

and hence $\sigma_G(j, t) > \sigma_G(i, t)$ for all $t \in [0, s]$. Combining this with (54) and the fact that $(x_1, y_1) \in B((x_0, y_0), \gamma)$ implies that
\[x_0 \sin(\pi t) + y_0 \cos(\pi t) - \sigma_G(i, t) \leq \gamma^2/3 + \gamma < 2\gamma \quad \text{for all} \quad t \in [0, s].\]

Symmetric arguments give the same upper bound on this difference over the interval $[s, 1]$, and on the difference $\sigma_G(i, t) - x_0 \sin(\pi t) + y_0 \cos(\pi t)$ over the interval $[0, 1]$. This implies that
\[\|\sigma_G(i, t) - [x_0 \sin(\pi t) + y_0 \cos(\pi t)]\|_u \leq 2\gamma.\]

Now suppose that there does not exist an $s \in [0, 1]$ such that $\sigma_G(i, s) \in [h_{a_{n,i},-1}(s) + \gamma, h_{a_{n,i},1}(s) - \gamma]$. Then either
\[\|\sigma_G(i, \cdot) - h_{a_{n,i},1}(\cdot)\|_u < 2\gamma + \frac{2}{n} \quad \text{or} \quad \|\sigma_G(i, \cdot) - h_{a_{n,i},-1}(\cdot)\|_u < 2\gamma + \frac{2}{n}.\]

This can be seen by observing that the difference $h_{a_{n,i},1} - h_{a_{n,i},-1}$ is unimodal and non-negative, so the path $\sigma_G(i, \cdot)$ must stay close of one of those paths for its entire trajectory. The additive factor of $2/n$ comes from the fact that the path $\sigma_G(i, \cdot)$ can make jumps of that size.

As all the functions $h_{a_{n,i},1}$ and $x_0 \sin(\pi t) + y_0 \cos(\pi t)$ are of the form $a \sin(\pi t + \theta)$, for some $(a, \theta) \in [0, 1] \times [0, 2\pi]$, this proves the lemma.
Proof of Theorem 1. Fix $\gamma \in (0, 1)$, and recall from Section 7.1 that $\nu_n$ is uniform measure on the set $\{\nu_\sigma\}_{\sigma \in \Theta_n}$. For small enough $\epsilon > 0$, Lemma 8.2 and Theorem ?? imply that
\[
\Pr\left(\max_{i \in [1, n]} \mathcal{L}\{t : |\sigma^n_G(i, t)| \geq 1 - 3\epsilon\} < \frac{\gamma}{6}, \ d_LP(\nu_n, \Pr_A) < \frac{\gamma^4}{128} \wedge \frac{L(\epsilon)}{2}\right) \to 1 \quad \text{as } n \to \infty.
\]
Combining Lemmas 8.3 and 8.4, this implies that there exist random variables $A_{n,i,\gamma} \in [0, 1]$ and $\Theta_{n,i,\gamma} \in [0, 2\pi]$ such that
\[
P_{n,\gamma} := \Pr\left(\max_{i \in [1, n]} \|\sigma^n_G(i, t) - A_{n,i,\gamma} \sin(\pi t + \Theta_{n,i,\gamma})\|_u < 2\gamma + \frac{2}{n}\right) \to 1 \quad \text{as } n \to \infty.
\]
As constructed, the random variables $A_{n,i,\gamma}$ and $\Theta_{n,i,\gamma}$ depend on $\gamma$. To remove this dependence, let $\gamma_n \to 0$ be a sequence such that $P_{n,\gamma_n} \to 1$ as $n \to \infty$. Let $A_{n,i} = A_{n,i,\gamma_n}$ and $\Theta_{n,i} = \Theta_{n,i,\gamma_n}$, and define
\[
B_{n,\gamma} = \left\{ \max_{i \in [1, n]} \|\sigma^n_G(i, t) - A_{n,i} \sin(\pi t + \Theta_{n,i})\|_u < \gamma \right\}, \quad (55)
\]
Then for any $\gamma > 0$, we have that $\Pr(B_{n,\gamma}) \to 1$ as $n \to \infty$. \hfill \square

We can also prove Theorem 2 and Theorem 3. For these we use the notation $B_{n,\gamma}$ from (55).

Proof of Theorem 2. Fix $t \in [0, 1]$. Since $\nu_n \to \delta_{\mathcal{A}\chi_t}$ (Theorem ??), the random measure $\rho^n_t$ converges in probability to the law of $(X, X \cos(\pi t) + Y \sin(\pi t))$, where $(X, Y) \sim \mathcal{A}\chi_t$. This law is simply $\mathcal{A}\chi_t$.

Now for a set $A \subset [-1, 1]^2$, let
\[
A^\gamma = \{ x \in [-1, 1]^2 : d(x, A) < \gamma \}.
\]
On the event $B_{n,\gamma}$, the support of the measure $\rho^n_t$ is contained in the set $\supp(\mathcal{A}\chi_t)^\gamma$. Moreover, since $\mathcal{A}\chi_t$ has a Lebesgue density that is bounded below, the weak convergence of $\rho^n_t$ to $\rho$ implies that with high probability, $\supp(\mathcal{A}\chi_t) \subset \supp(\rho^n_t)^\gamma$ for any $\gamma > 0$. Therefore since $\Pr(B_{n,\gamma}) \to 1$ as $n \to \infty$ by Theorem 1, we have that
\[
\lim_{n \to \infty} d_H(\supp(\rho^n_t), \supp(\mathcal{A}\chi_t)) = 0 \quad \text{in probability}. \hfill \square
\]

Proof of Theorem 3. First observe that for any $a \in [0, 1]$, $t \in [0, 1/2]$, and $\theta \in [0, 2\pi]$, that
\[
e^{\pi it} \left[ a \sin(\pi t + \theta) + ia \sin(\pi t + \pi/2 + \theta) \right] = a \sin(\theta).
\]
Therefore on the event $B_{n,\gamma}$, we have that
\[
\max_{j \in [1, n]} \max_{s,t \in [0, 1]} |Z^n_j(t) - Z^n_j(s)| \leq 2\gamma.
\]
Since $\Pr(B_{n,\gamma}) \to 1$ as $n \to \infty$ for any $\gamma > 0$ by Theorem 1, this proves the theorem. \hfill \square

59
9 The geometric limit

In this section, we use Theorem 1 and Theorem 5 to prove Theorem 4. Recall from Section 1 that \( d_\infty(f,g) \) is the uniform norm between two \( \mathbb{R}^n \)-valued functions \( f \) and \( g \), where the pointwise distance is the \( L^\infty \)-distance. Recall also that \( \tilde{\sigma} \) is the embedding of a sorting network \( \sigma \) into the \((n - 2)\)-dimensional sphere \( S^{n-2} \subset \mathbb{R}^n \).

By Theorem 1 and a change of variables, there exist random vectors \( X^n = (X^n_1, \ldots, X^n_n) \) and \( V^n = (V^n_1, \ldots, V^n_n) \) such that \( d_\infty(F_n, \tilde{\sigma}^n)/n \to 0 \) in probability, where

\[
F_n(t) = X^n \cos(\pi t) + V^n \sin(\pi t) + c,
\]

where \( c = \left( \frac{n+1}{2}, \ldots, \frac{n+1}{2} \right) \).

Moreover, we can assume that \( X^n_i = i-(n+1)/2 \) and that \( V^n_i = \sigma^n(i, N/2) - (n+1)/2 \), as these changes only shift the curve \( X^n \cos(\pi t) + V^n \sin(\pi t) \) by \( d_\infty \)-distance \( o(n) \) in probability.

The point \( c \) is the center of \( S^{n-2} \). It remains to show that we can shift the curve \( F_n \) by \( d_\infty \)-distance \( o(n) \) to obtain a great circle in \( S^{n-2} \). For this we need the following lemma.

**Lemma 9.1.** Let the vectors \( X^n \) and \( V^n \) be as above. Then

\[
\lim_{n \to \infty} \frac{\langle X^n, V^n \rangle}{n^3} = 0 \quad \text{in probability.}
\]

**Proof.** Let \( I_n \) be a uniform random variable on \( \{1, \ldots, n\} \), independent of \( X^n \) and \( V^n \), and define

\[
(\tilde{X}^n, \tilde{V}^n) = \left( \frac{2X^n_{I_n}}{n}, \frac{2V^n_{I_n}}{n} \right).
\]

We have that

\[
(\tilde{X}^n, \tilde{V}^n) \overset{d}{=} (Y_n(0) - 1/n, Y_n(1/2) - 1/n),
\]

where \( Y_n \) is the trajectory random variable of \( \sigma^n \). Therefore by Theorem 5,

\[
(\tilde{X}^n, \tilde{V}^n) \overset{d}{=} (X, V) \sim \text{Arch}.
\]

By the bounded convergence theorem, this implies that \( \mathbb{E} \tilde{X}^n \tilde{V}^n \to \mathbb{E}XV \) in probability. Observing that \( n^3 \mathbb{E} \tilde{X}^n \tilde{V}^n = 4 \langle X^n, V^n \rangle \) and that \( \mathbb{E}XV = 0 \) completes the proof. \( \qed \)

**Proof of Theorem 4.** Fix \( n \in \mathbb{N} \). For ease of bounding errors, we assume that \( n \geq 27 \). Let \( \sigma \) be a fixed \( n \)-element sorting network. Our goal is to perturb the vector \( V^n \) to a new vector \( W^n \) so that the path

\[
C_n(t) := X^n \cos(\pi t) + V^n \sin(\pi t) + c \quad (56)
\]

follows a great circle on \( S^{n-2} \), and so that \( d_\infty(C_n, F_n) \) is small. In order to do this, we just need to find \( W^n \) such that

\[
\sum_{i=1}^{n} W^n_i = 0, \quad \langle X^n, W^n \rangle = 0, \quad \text{and} \quad \|W^n\|_2 = (n^3 - n)/12.
\]

60
The first of these conditions guarantees that $C_n$ lies in the correct hyperplane $L_n \subset \mathbb{R}^n$, and the second and third conditions guarantee that $C_n$ traces out a great circle on the sphere $S^{n-2}$ within that hyperplane. We first perturb $V^n$ to a vector $Z^n$ satisfying the first two properties. Define the random variable

$$A_n = \frac{|(X^n, V^n)|}{n^3}.$$ 

Without loss of generality, assume that $(X^n, V^n) > 0$.

Choose any $i \geq 3(n+1)/4$. Let $k > 0$, and decrease the value of $V^n_i$ by $kn$ and increase the value of $V^n_{n+1-i}$ by $kn$. Call this new vector $V^n_k$. Note that $X^n_i \geq n/4$, and that $X^n_i = -X^n_{n+1-i}$. Therefore $\langle X^n, V^n \rangle \leq \langle X^n, V^n \rangle - kn^2/2$.

By iterating the above procedure to repeatedly lower the dot product $\langle X^n, V^n \rangle$, we can obtain a vector $K^n = (K_1^n, \ldots, K_n^n)$ with the following properties (here is where we require that $n \geq 27$).

(i) For all $i$, $|K_i| \leq 9A_n$, $-K_i = K_{n+1-i}$, and $K_i = 0$ if $i \in \left(\frac{n+1}{4}, \frac{3(n+1)}{4}\right)$.

(ii) $\langle X^n, V^n + K^n \rangle = 0$.

Let $Z^n = V^n + K^n$. Observe that $\sum_{i=1}^{n} Z^n_i = 0$ since both $\sum_{i=1}^{n} K_i = 0$ and $\sum_{i=1}^{n} V^n_i = 0$. The fact that $\sum_{i=1}^{n} V^n_i = 0$ follows since $V^n + c$ is a permutation of the vector $(1, 2, \ldots, n)$. For any $t$, we have that

$$\|F_n(t) - (X^n \cos(\pi t) + Z^n \sin(\pi t))\|_{\infty} = \max_{i \in [1, n]} |V^n_i \sin(\pi t) - Z^n_i \sin(\pi t)| \leq 9A_n n.$$ 

We can now define $W^n = M_n Z^n$, where $M_n$ is a random constant chosen so that $\|W^n\|_2 = (n^3 - n)/12$. Using the definition of $C_n$ in (56), for every $t \in [0, 1]$, we have that

$$\|F_n(t) - C_n(t)\|_{\infty} \leq 9A_n n + \max_{i \in [1, n]} |W^n_i \sin(\pi t) - Z^n_i \sin(\pi t)| < 9A_n n + |M_n - 1| \left[ \frac{n + 1}{2} + 9A_n n \right].$$

In the last inequality, we have used that $|Z^n_i - V^n_i| \leq 9A_n n$ to get that $|Z_{n,i}| \leq (n+1)/2 + 9A_n n$. The inequality (57) implies that

$$\frac{d_\infty(C_n, \Sigma^n)}{n} \leq 9A_n + |M_n - 1| \left[ \frac{n + 1}{2n} + 9A_n \right] + \frac{d_\infty(F_n, \Sigma^n)}{n}.$$ (58)

Now again using that $|Z^n_i - V^n_i| \leq 9A_n n$, we have that

$$\|Z^n\|_2^2 - \|V^n\|_2^2 \leq \sum_{i=1}^{n} |Z^n_i - V^n_i||Z^n_i + V^n_i| \leq 9A_n n^2 (n + 1 + 9A_n n).$$
Therefore since $\|V^n\|_2^2 = (n^3 - n)/12$, we have that

$$M_n \in \left[ \sqrt{1 - 217A_n}, \sqrt{1 + 217A_n} \right]$$

whenever $A_n < 1/217$. Lemma 9.1 implies that $A_n \to 0$ as $n \to \infty$ in probability, and hence $M_n \to 1$ in probability. Therefore since $d_\infty(F_n, \sigma^n)/n \to 0$ in probability as $n \to \infty$, (58) implies that $d_\infty(C_n, \sigma^n)/n$ does as well.

**Acknowledgements**

Thank you to Bálint Virág for many fruitful discussions about the problem, and for many constructive comments about previous drafts. Thank you to Laurent Miclo, Maxim Arnold, and Boris Khesin for pointing out the connections between random sorting networks and fluid dynamics. Thank you to Zachary Hamaker, Svante Linusson and Robin Sulzgruber for telling me about related families of sorting processes that also appear to converge to the Archimedean limit.

**References**

[ADHV19] Omer Angel, Duncan Dauvergne, Alexander Holroyd, and Bálint Virág. The local limit of random sorting networks. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 55(1):412–440, 2019.

[AGH12] Omer Angel, Vadim Gorin, and Alexander Holroyd. A pattern theorem for random sorting networks. *Electron. J. Probab.*, 17(99):1–16, 2012.

[AH10] Omer Angel and Alexander Holroyd. Random subnetworks of random sorting networks. *Elec. J. Combinatorics*, 17, 2010.

[AHR09] Omer Angel, Alexander Holroyd, and Dan Romik. The oriented swap process. *The Annals of Probability*, 37(5):1970–1998, 2009.

[AHRV07] Omer Angel, Alexander Holroyd, Dan Romik, and Bálint Virág. Random sorting networks. *Advances in Mathematics*, 215(2):839–868, 2007.

[BB06] Anders Björner and Francesco Brenti. *Combinatorics of Coxeter groups*, volume 231. Springer Science & Business Media, 2006.

[BFZ96] Arkady Berenstein, Sergey Fomin, and Andrei Zelevinsky. Parametrizations of canonical bases and totally positive matrices. *Advances in Mathematics*, 122(1):49–149, 1996.
[BH95] Sara Billey and Mark Haiman. Schubert polynomials for the classical groups. *Journal of the American Mathematical Society*, 8(2):443–482, 1995.

[BJS93] Sara C Billey, William Jockusch, and Richard P Stanley. Some combinatorial properties of Schubert polynomials. *Journal of Algebraic Combinatorics*, 2(4):345–374, 1993.

[Bre89] Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.

[Bre08] Yann Brenier. Generalized solutions and hydrostatic approximation of the Euler equations. *Physica D: Nonlinear Phenomena*, 237(14-17):1982–1988, 2008.

[DV20] Duncan Dauvergne and Bálint Virág. Circular support in random sorting networks. *Transactions of the American Mathematical Society*, 373(3):1529–1553, 2020.

[EG87] Paul Edelman and Curtis Greene. Balanced tableaux. *Advances in Mathematics*, 63(1):42–99, 1987.

[Eln97] Serge Elnitsky. Rhombic tilings of polygons and classes of reduced words in Coxeter groups. *Journal of Combinatorial Theory, Series A*, 77(2):193–221, 1997.

[Gar02] Adriano Mario Garsia. *The saga of reduced factorizations of elements of the symmetric group*. Université du Québec [Laboratoire de combinatoire et d’informatique mathématique (LACIM)], 2002.

[GM12] Max Gross and Nicole Marsaglia. Vexillary permutations. 2012.

[GP80] Jacob E Goodman and Richard Pollack. On the combinatorial classification of nondegenerate configurations in the plane. *Journal of Combinatorial Theory, Series A*, 29(2):220–235, 1980.

[GR18] Vadim Gorin and Mustazee Rahman. Random sorting networks: local statistics via random matrix laws. *Probability Theory and Related Fields*, pages 1–52, 2018.

[Hai92] Mark Haiman. Dual equivalence with applications, including a conjecture of Proctor. *Discrete Mathematics*, 99(1-3):79–113, 1992.

[HY14] Zachary Hamaker and Benjamin Young. Relating Edelman–Greene insertion to the Little map. *Journal of Algebraic Combinatorics*, 40(3):693–710, 2014.
[Kal06] Olav Kallenberg. *Foundations of Modern Probability*. Springer Science & Business Media, 2006.

[Kot16] Michal Kotowski. *Return probabilities on groups and large deviations for permutation processes*. PhD thesis, University of Toronto (Canada), 2016.

[Lit03] David Little. Combinatorial aspects of the Lascoux–Schützenberger tree. *Advances in Mathematics*, 174(2):236–253, 2003.

[LPS18] Svante Linusson, Samu Potka, and Robin Sulzgruber. On random shifted standard Young tableaux and 132-avoiding sorting networks. *arXiv preprint arXiv:1804.01795*, 2018.

[LS85] Alain Lascoux and Marcel-Paul Schützenberger. Schubert polynomials and the Littlewood-Richardson rule. *Letters in Mathematical Physics*, 10(2-3):111–124, 1985.

[LZ98] Bernard Leclerc and Andrei Zelevinsky. Quasicommuting families of quantum Plücker coordinates. In *AMS translation series*, volume 181, pages 85–108, 1998.

[Mag98] Peter Magyar. Schubert polynomials and Bott-Samelson varieties. *Commentarii Mathematici Helvetici*, 73(4):603–636, 1998.

[Man01] Laurent Manivel. *Symmetric functions, Schubert polynomials, and degeneracy loci*. Number 3. American Mathematical Soc., 2001.

[Mar20] Eric Marberg. A symplectic refinement of shifted hecke insertion. *Journal of Combinatorial Theory, Series A*, 173:105216, 2020.

[Rei05] Victor Reiner. Note on the expected number of Yang–Baxter moves applicable to reduced decompositions. *European Journal of Combinatorics*, 26(6):1019–1021, 2005.

[Rom15] Dan Romik. *The surprising mathematics of longest increasing subsequences*. Cambridge University Press, 2015.

[RVV19] Mustazee Rahman, Balint Virag, and Mate Vizer. Geometry of permutation limits. *Combinatorica*, 39(4):933–960, 2019.

[Shn] A Shnirelman. On the geometry of the group of diffeomorphisms and the dynamics of an ideal incompressible fluid. *Math Sbornik USSR 56 (1987) 79*, 105.

[Sta84] Richard Stanley. On the number of reduced decompositions of elements of Coxeter groups. *European Journal of Combinatorics*, 5(4):359–372, 1984.
[Ste97] John Stembridge. Some combinatorial aspects of reduced words in finite Coxeter groups. *Transactions of the American Mathematical Society*, 349(4):1285–1332, 1997.

[Str96] Michael Struwe. Variational methods, vol. 34 of *Results in Mathematics and Related Areas (3)*, 1996.

[Ten06] Bridget Eileen Tenner. Reduced decompositions and permutation patterns. *Journal of Algebraic Combinatorics*, 24(3):263–284, 2006.

[Ten14] Bridget Eileen Tenner. On the expected number of commutations in reduced words. *arXiv preprint arXiv:1407.5636*, 2014.

[You14] Benjamin Young. A Markov growth process for Macdonald’s distribution on reduced words. *arXiv preprint arXiv:1409.7714*, 2014.