GEOMETRY OF PERMUTATION LIMITS

MUSTAZEE RAHMAN, BÁLINT VIRÁG, AND MÁTÉ VIZER

ABSTRACT. This paper initiates a limit theory of permutation valued processes, building on the recent theory of permutons. We apply this to study the asymptotic behaviour of random sorting networks. We prove that the Archimedean path, the conjectured limit of random sorting networks, is the unique path from the identity to the reverse permuton having minimal energy in an appropriate metric. Together with a recent large deviations result (Kotowski, 2016), it implies the Archimedean limit for the model of relaxed random sorting networks.

1. INTRODUCTION

The objective of this paper is two-fold. First, to develop a limit theory of permutation valued stochastic processes that it is applicable to study their asymptotic properties. Second, to apply this theory to study random sorting networks.

1.1. **Permutations limits.** Recently, the language of permutons has been developed to study asymptotic properties of permutations. Examples within this theory include the study of finite forcibility, pattern avoidance, property testing, pseudorandomness, the Mallows model, etc. (see [7, 9, 11, 16] and references therein). We begin with a discussion of permutons as it is pertinent to Theorem 1, our first result.

Let $\mathfrak{S}_n$ be the symmetric group of n elements. The *empirical measure* of $\sigma \in \mathfrak{S}_n$ is

$$(1.1) \qquad \mu^\sigma = \frac{1}{n} \sum_i \delta_{\left(\frac{2i}{n}-1, \frac{2\sigma(i)}{n}-1\right)}.$$

This is a probability measure of $[-1,1]^2$. One defines a sequence of permutations $\{\sigma_n\}$ with size $|\sigma_n| \to \infty$ to converge if μ^{σ_n} converges weakly to a Borel measure μ.

A *permuton* μ is a Borel probability measure on $[-1,1]^2$ with uniform marginals. (In the literature, this definition is often w.r.t. the unit square $[0,1]^2$ but it is convenient for us to use $[-1,1]^2$.) It is proven in [9] that limits of permutations in the above sense are permutons and that every permuton may be realized as a limit of permutations.

In this paper we study not a sequence of single permutations but rather a sequence of sequences of permutations. For an integer n, let $[n] = \{1, \ldots, n\}$. Suppose that

$$(1.2) \qquad \sigma^n = \left(\sigma_t^n; \, t \in [t_n]\right)$$

is a $\mathfrak{S}_n$-valued sequence. We say that σ^n is a *permutation process* of $\mathfrak{S}_n$, or simply a permutation process when there is no ambiguity. We always set σ_0^n to be the identity permutation. Our goal is to find an appropriate representation of the limits of sequences of

permutation processes growing both in the size, n, of the permutations and the length, t_n, of the sequence.

An example of a permutation process is a *sorting network*. A sorting network of $\mathfrak{S}_n$ is a path of minimal length from the identity $\mathbf{id}_n = 1, \ldots, n$ to the reverse permutation $\mathbf{rev}_n = n, n-1, \ldots, 1$ in the Cayley graph of $\mathfrak{S}_n$ generated by the adjacent transpositions $(i, i+1)$, for $1 \leq i \leq n-1$. It is a permutation process of length $t_n = \binom{n}{2}$. The adjacent transpositions are also called swaps. An example of a sorting network is the bubble sort algorithm applied to $\mathbf{rev}_n$ and viewed in reverse time.

Another example of a permutation process is the interchange process on finite paths. In this setting the permutation process is random. As the path lengths tend to infinity, the interchange process converges in probability to a limit: stationary Brownian motion on an interval. This is explained in Section 2.1.

1.2. **Limits of permutation processes.** Given a permutation process σ^n of $\mathfrak{S}_n$ as in (1.2), the rescaled trajectory of particle $i \in [n]$ is the function

$$(1.3) \qquad T_i^n(t/t_n) = \frac{2\sigma_t^n(i)}{n} - 1 \ \text{ for } t \in [t_n].$$

After linearly interpolating between the discrete times t/t_n, we may consider T_i^n as a continuous function from $[0, 1]$ to $[-1, 1]$. The *trajectory process* of σ^n, denoted X^n, is the trajectory of a particle chosen uniformly at random:

$$(1.4) \qquad X^n = \frac{1}{n} \sum_{i=1}^n \delta_{T_i^n}.$$

Observe that for every $t \in [t_n]$ the distribution of $X^n(t/t_n)$ is uniform over the set $\{\frac{2i}{n} - 1; i \in [n]\}$ due to σ_t^n being a permutation. Moreover, σ^n can be reconstructed from X^n and t_n.

Let $\mathbf{C}$ denote the space of continuous functions from $[0, 1]$ to $[-1, 1]$ in the topology of uniform convergence. The trajectory process is then a Borel probability measure on $\mathbf{C}$. Given a sequence of permutation processes $\{\sigma^n\}$, its limit is defined to be the weak limit of its associated trajectory processes as Borel probability measures on $\mathbf{C}$. In other words, $\{\sigma^n\}$ converges if there is a stochastic process $X = (X(t), 0 \leq t \leq 1)$ with continuous sample paths such that for every uniformly continuous and bounded $F : \mathbf{C} \to \mathbb{R}$,

$$(1.5) \qquad \mathbb{E}\left[F(X^n)\right] = \frac{1}{n} \sum_{i=1}^n F(T_i^n) \ \overset{n \to \infty}{\Longrightarrow} \ \mathbb{E}\left[F(X)\right].$$

Our first result characterizes the limits of permutation processes.

Definition 1. A *permuton process* is a $[-1, 1]$-valued stochastic process $X = (X(t), 0 \leq t \leq 1)$ with continuous sample paths and such that $X(t) \sim \mathrm{Uniform}[-1, 1]$ for every t.

Theorem 1. *For each n, let $\sigma^n = (\sigma_t^n; t \in [t_n])$ be a permutation process of $\mathfrak{S}_n$ with $t_n \to \infty$ as $n \to \infty$. Suppose $\{\sigma^n\}$ converges to a limit X in the sense of (1.5). Then X is a permuton process. Conversely, given any permuton process X, there is a sequence of permutation processes that converges to X.*

Theorem 1 extends the limit theory of single permutations to permutation processes. Indeed, if a sequence of permutation processes $\{\sigma^n\}$ has a limit X then for every $s \in [0, 1]$ the limit of $\sigma^n_{\lfloor s \, t_n \rfloor}$ is the permuton with the distribution of $(X(0), X(s))$. Moreover, for any set of times $s_1, \ldots, s_k$, the empirical measure of the k-tuples $\left(\sigma^n_{\lfloor s_j \, t_n \rfloor}(i); 1 \le j \le k\right)$ as i ranges over $[n]$ converges weakly as measures rescaled onto $[-1, 1]^k$ to the distribution of $(X(s_1), \ldots, X(s_k))$.

1.3. **Random sorting networks.** Recall from the previous section that a sorting network of $\mathfrak{S}_n$ is a shortest path from $\mathbf{id}_n$ to $\mathbf{rev}_n$ in the Cayley graph generated by adjacent transpositions. The number of sorting networks of $\mathfrak{S}_n$ was enumerated by Stanley [15], and later a combinatorial bijection with staircase shaped Young tableaux was provided by Edelman and Greene [6].

The number of permutations in a sorting network of $\mathfrak{S}_n$ is always $N := \binom{n}{2}$. A *random sorting network* of $\mathfrak{S}_n$ is a sorting network of $\mathfrak{S}_n$ chosen uniformly at random. We denote this random permutation process as $\mathbf{RSN}^n = (\mathbf{RSN}^n_t; 0 \le t \le N)$ (thus, $\mathbf{RSN}^n_t$ is the t-th permutation in $\mathbf{RSN}^n$).

The asymptotic behaviour of $\mathbf{RSN}^n$ was first studied by Angel et. al. [3]. It is shown that, as $n \to \infty$, the spacetime process of swaps of $\mathbf{RSN}^n$ converges to the product of semicircle law and Lebesgue measure. It is also shown that, in the limit, the particle trajectories are Hölder-1/2 continuous, and the support of the permutation matrix lies within a certain octagon. Additional results about the asymptotic behaviour of $\mathbf{RSN}^n$ have since been proved; see for example [2] and the references therein. However, the main conjecture of [3], the Archimedean path conjecture, remains open. To state the conjecture and our results we first introduce the Archimedean measure.

The *Archimedean measure* is the unique probability measure on the plane with the property that all of its projections onto lines through the origin have the Uniform$[-1, 1]$ distribution. Its density, supported on the unit disk, is given by $\left(2\pi\sqrt{1 - x^2 - y^2}\right)^{-1} dx dy$. It is in fact the projection of the normalized surface area measure of the 2-sphere onto the unit disk. Let $(\mathbf{A}_x, \mathbf{A}_y)$ denote a random variable whose distribution is the Archimedean measure. The *Archimedean process* $\mathcal{A} = (\mathcal{A}(t); 0 \le t \le 1)$ is the permuton process defined by

(1.6) $$\mathcal{A}(t) = \cos(\pi t)\, \mathbf{A}_x + \sin(\pi t)\, \mathbf{A}_y.$$

The Archimedean path conjecture [3, Conjecture 2] states that for every t the random permutation $\mathbf{RSN}^n_{\lfloor tN \rfloor}$ converges to the deterministic permuton $(\mathcal{A}(0), \mathcal{A}(t))$. The *Archimedean path* is the permuton valued path $\mathbf{A} = (\mathbf{A}(t); 0 \le t \le 1)$ such that

(1.7) $$\mathbf{A}(t) \sim (\mathcal{A}(0), \mathcal{A}(t)) \quad \text{for every } 0 \le t \le 1.$$

Thus, the Archimedean path conjecture is that the empirical measures of permutations in $\mathbf{RSN}^n$ converges to the Archimedean path; see Figure 1. Observe that the Archimedean process is a random sine curve. The sine curve conjecture [3, Conjecture 1] asserts that the trajectories of particles are close to random sine curves with high probability; see Figure 2.
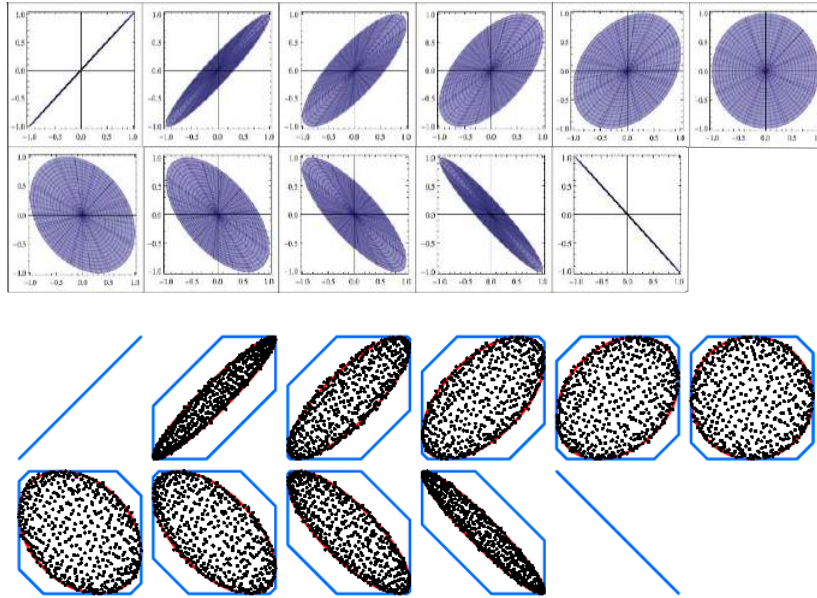
Figure 1. Support of the Archimedean path (top) and $\mathbf{RSN}^{500}$ (bottom). Bottom figure is from [3, Figure 5].
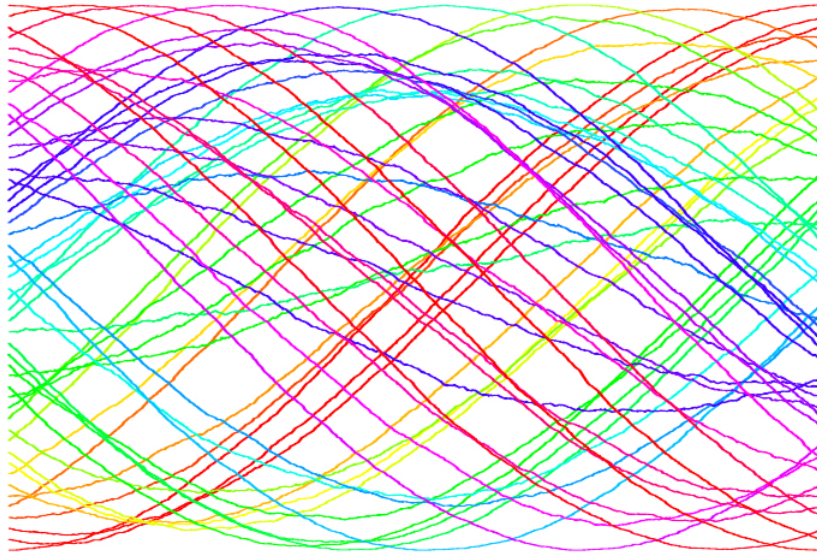
Figure 2. Some scaled particle trajectories from $\mathbf{RSN}^{2000}$ from [3, Figure 1].

The Archimedean path conjecture and the sine curve conjecture may be viewed under a common framework as permuton processes as follows. It is a stochastic process version of the Archimedean path conjecture (and implies it). It also implies that typical trajectories of $\mathbf{RSN}^n$ are close to sine curves.

Conjecture 1. $\mathbf{RSN}^n$ converges in probability as a permutation process to the Archimedean process (1.6).

We emphasize that this conjecture states the random trajectory process of $\mathbf{RSN}^n$ concentrates around a deterministic limit, which is the Archimedean process.

1.4. **Variational characterization of Archimedean process.** The (Dirichlet) energy of a stochastic process $X = (X(t); 0 \leq t \leq 1)$ is

$$\mathcal{E}[X] = \sup_{\Pi} \sum_{i=1}^{k} \frac{\mathbb{E}\left[|X(t_i) - X(t_{i-1})|^2\right]}{t_i - t_{i-1}},$$

where the supremum is over all finite partitions $\Pi = \{0 = t_0 < t_1 < \cdots < t_k = 1\}$ of $[0,1]$. If X has continuously differentiable sample paths then $\mathcal{E}[X] = \int_0^1 \mathbb{E}\left[X'(t)^2\right] dt$. Thus, for example, a simple calculation shows that $\mathcal{E}[\mathcal{A}] = \pi^2/3$.

Theorem 2. *Among all permuton processes X with the property that $X(1) = -X(0)$, the Archimedean process $\mathcal{A}$ uniquely minimizes the energy.*

The theorem allows for a characterization of stationary random permutation processes which converge to the Archimedean process in terms of the second moment of their speed. Random sorting networks are invariant under ϵ-shifts which take a trajectory $X(t)$ to $X(\epsilon + t)$. More precisely, the ϵ-shift makes the trajectory periodic with reversing boundary conditions, so the ϵ-shift is defined as

$$\left((-1)^{\lfloor \epsilon + t \rfloor} X(\epsilon + t \bmod 1), \ 0 \leq t \leq 1\right).$$

We call a random trajectory process ϵ-stationary if its distribution is invariant under ϵ-shift of all individual particle trajectories.

Corollary 2.1. *Let X^n be a tight sequence of random trajectory processes that are ϵ_n-stationary with $\epsilon_n \to 0$. Assume that*

(1.8)
$$\limsup_{t \to 0} \limsup_{n \to \infty} \frac{\mathbb{E}\left[(X^n(t) - X^n(0))^2\right]}{t^2} \leq \frac{\pi^2}{3}.$$

Then the sequence converges in probability to the deterministic limit given by the Archimedean process.

The trajectory process of $\mathbf{RSN}^n$ is N^{-1}–stationary because of stationarity of the swaps of $\mathbf{RSN}^n$, [3, Theorem 1]. Tightness of the random trajectory process of $\mathbf{RSN}^n$ follows from [3, Theorem 3], which states that for any $\delta > 0$, with probability tending to 1, all individual trajectories T^n of the particles in $\mathbf{RSN}^n$ satisfy

$$|T^n(t) - T^n(s)| \leq \sqrt{8}\,|s - t|^{1/2} + \delta \ \text{ for every } s, t.$$

Corollary 2.1 is proved in Section 4.

Random sorting networks can also be studied in the setting of large deviation theory of the interchange process on paths. Consider the discrete time interchange process on the n-path, which is the path graph with n vertices. A random sorting network is the interchange process on the n-path conditioned to be at $\mathbf{rev}_n$ in the shortest possible time N. Instead, one can consider *relaxed random sorting network*, which is the interchange

process conditioned to be close to $\mathbf{rev}_n$ in time $n^{2+\alpha}$ for some $\alpha \in (0,1)$. One can study the relaxed network using large deviation theory in the following sense.

Suppose we fix a permuton process X that satisfies $X(0) = -X(1)$. One can ask what is the probability that the trajectory of a relaxed random sorting network on the n-path is close to X. This is the problem addressed in [12]. It is shown that this probability satisfies a large deviation principle whose rate function is the energy of X. Then Theorem 2 implies relaxed random sorting networks have to be close to the Archimedean process with high probability. This proves the Archimedean path conjecture for the relaxed networks.

1.5. **Variational characterization of Archimedean path.** The 2-Wasserstein distance (henceforth, Wasserstein distance) between two Borel probability measures μ, ν on a metric space K is defined by

$$(1.9) \qquad W(\mu, \nu)^2 = \inf_{\text{couplings } (V,W) \text{ s.th. } V \sim \mu, W \sim \nu} \mathbb{E}\left[d(V,W)^2\right].$$

We study permutons in the Wasserstein metric, whereby $K = [-1,1]^2$ in the Euclidean metric. Let **id** denote the identity permuton (X,X) and **rev** denote the reverse permuton $(X, -X)$, where $X \sim \mathrm{Uniform}[-1,1]$.

Theorem 3. *Let $\mu = (\mu(t); 0 \le t \le 1)$ be a permuton valued path from $\mu(0) = \mathbf{id}$ to $\mu(1) = \mathbf{rev}$. Then the energy of μ in the Wasserstein metric satisfies $\mathcal{E}[\mu] \ge \mathcal{E}[\mathbf{A}] = \pi^2/6$, where $\mathbf{A}$ is the Archimedean path (1.7). If there is equality then $\mu(t) = \mathbf{A}(t)$ for every t.*

The main tool used in proving Theorem 3 should be of independent interest. We show that for a permuton valued path μ, there exists a $[-1,1]^2$–valued stochastic process X such that the fixed time distributions of X is given by μ and the energy of X in the L^2-metric equals the energy of μ in the Wasserstein metric. One may think of X as being an optimal coupling of the measures along μ. We prove such a 'realization theorem' for measure valued paths in a fairly general setting as stated in Theorem 5.

A motivation for Theorem 3 is that the Wasserstein distance is a natural metric on permutons. It is also related to sorting networks in the following way. A two-sided random sorting network is a shortest sequence of permutations from $\mathbf{id}_n$ to $\mathbf{rev}_n$ so that in each step the permutation is multiplied by an adjacent transposition either on left or on the right. This means that in each step, two adjacent columns or two adjacent rows of the permutation matrix are exchanged. Thus the 1s in the permutation matrix can be thought of as particles moving horizontally or vertically.

After scaling, we may consider the $[-1,1]^2$-valued trajectories for the n particles in a uniformly chosen two-sided sorting network. It can be shown that Conjecture 1 would imply that the trajectory process of two-sided random sorting networks converges to an optimal coupling of the Archimedean path $\mathbf{A}$.

1.6. **Permuton geometry.** Let $\mathcal{P}$ denote the space of all permutons. Motivated by the great circle conjecture about random sorting networks it is natural to study $\mathcal{P}$ in the Wasserstein metric since it is an infinite dimensional analogue of the permutohedron embedded

into the Euclidean sphere. (The permutohedron of order n embeds naturally into an $(n-2)$-dimensional sphere in $\mathbb{R}^n$.) In analogy with the sphere, one may ask whether the sum of distances squared, $W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2$, is uniquely minimized by the Archimedean measure over all $\mathbf{P} \in \mathcal{P}$?

Theorem 4. *The function* $\mathbf{P} \mapsto W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2$ *is minimized by a permuton* $\mathbf{P} \sim (X, Y)$ *if and only if the pair* $\left(\frac{X-Y}{\sqrt{2}}, \frac{X+Y}{\sqrt{2}}\right)$ *is also a permuton. In particular, the Archimedean measure is not the unique minimizer.*

In proving this theorem we will find a nice formula for the Wasserstein distance from any permuton to the identity, as stated in Theorem 6.

One can also ask whether there is a unique minimal energy path from $\mathbf{id}$ to a given permuton $\mathbf{P}$? A motivation for this question is to understand minimal length paths from the identity to arbitrary permutations of the permutohedron. These are called reduced decompositions. Counting reduced compositions is a deep and difficult combinatorial problem. We may get insights by studying related questions in the space of permutons. For instance, are there analogues of the Archimedean path conjecture for reduced decompositions of permutations approximating a target permuton $\mathbf{P}$? Can large deviation theory provide an asymptotic count for the number of relaxed reduced decompositions of $\mathbf{P}$, à la sorting networks? We pose the following two open problems.

Question 1.1 (Uniqueness of minimal energy paths). Under what condition does there exist an unique minimal energy path in $\mathcal{P}$ from $\mathbf{id}$ to a given permuton $\mathbf{P}$? What about for the Lebesgue permuton?

Question 1.2 (Diameter of permuton space). Suppose $\mathbf{P}$ is a permuton. Does the minimal energy path(s) from $\mathbf{id}$ to $\mathbf{P}$ have energy at least that of the Archimedean path?

Remark. Proof of the conjectures from [3] have been announced in [4]. The proof uses the framework of permuton processes and the main step involves proving Conjecture 1. The proof also relies on the local structure of random sorting networks and additional local-to-global properties from [1, 5, 8].

Outline of the paper. We prove Theorem 1 in Section 2. In Section 3 we define path energy in metric spaces and discuss some of its basic properties. In Section 3.1 we prove Theorem 5 about realizing measure valued paths as stochastic processes. In Section 4 we prove Theorem 2 and Corollary 2.1. In Section 5 we prove Theorem 3. Finally, in Section 6 we prove Theorem 4.

2. Limits of permutation processes

In proving Theorem 1 we state a lemma about approximating continuous processes by their piecewise linear parts. The proof is in the Appendix.

Lemma 2.1. *Let* $Y = (Y(t); 0 \le t \le 1)$ *be a continuous* $[-1, 1]$*-valued process. Consider its modulus of continuity* $m^\delta(Y) = \sup_{s,t:|s-t|\le\delta} |Y(s) - Y(t)|$. *Then* $\mathbb{E}\left[m^\delta(Y)\right] \to 0$ *as* $\delta \to 0$.

Moreover, if Y and $\hat{Y}$ are continuous processes then $|m^\delta(Y) - m^\delta(\hat{Y})| \leq 2||Y - \hat{Y}||_\infty$. Finally, let $\mathrm{Lin}(n, Y)$ be the process obtained from Y such that it agrees with Y at times $t = i/n$ for $0 \leq i \leq n$ and is linear in between. Then $\mathbb{E}\left[||\mathrm{Lin}(n, Y) - Y||_\infty\right] \to 0$ as $n \to \infty$.

Proof that limit of permutation processes is a permuton processes. Suppose that the trajectory processes X^n of a sequence of permutation processes $(\sigma_t^n; t \in [t_n])$ converges to a continuous process X. By Skorokhod's representation Theorem, we may assume that the X^n and X are realized on a common probability space and $||X^n - X||_\infty \to 0$ almost surely. For a fixed t, we may choose $s_n \in \{i/t_n; 0 \leq i \leq t_n\}$ such that $|t - s_n| \leq 1/t_n$. Then by triangle inequality and Lemma 2.1,

$$|X(t) - X^n(s_n)| \leq ||X - X^n||_\infty + m^{1/t_n}(X^n) \leq 3||X - X^n||_\infty + m^{1/t_n}(X).$$

The term $m^{1/t_n}(X) \to 0$ almost surely in the sample outcomes of X due to continuity, and $||X^n - X||_\infty \to 0$ almost surely as well. Thus $|X(t) - X^n(s_n)| \to 0$ almost surely. The distribution of $X^n(s_n)$ is uniform on the set $\{\frac{2i}{n} - 1; i \in [n]\}$ as remarked earlier. Therefore, $X^n(s_n)$ converges weakly to $\mathrm{Uniform}[-1, 1]$ and it follows that $X(t) \sim \mathrm{Uniform}[-1, 1]$.

Proof that a permuton process is a limit of permutation processes. Let $X = (X(t); 0 \leq t \leq 1)$ be the permuton process that is to be approximated by permutation processes. We will construct a sequence of random permutation processes and show that it converges almost surely to X. For $n \geq 1$ set $\Pi_n = \{i/n; 0 \leq i \leq n\}$. The following defines a random permutation process $(\sigma_t^n; t \in [n])$ of $\mathfrak{S}_n$.

Let $X_1, X_2, \ldots$ be i.i.d. copies of X. We may assume that for every t in the countable set $\cup_n \Pi_n$ the values $X_1(t), X_2(t), \ldots$ are all distinct. For each n, consider the order statistics of $X_1(0), \ldots, X_n(0)$: $X_{(1)}(0) < X_{(2)}(0) < \cdots < X_{(n)}(0)$. Let $\pi(i)$ be the index such that $X_{(i)}(0) = X_{\pi(i)}(0)$. For $t \in [n]$, define the permutation σ_t^n by

$$\sigma_t^n(i) = \text{ rank of } X_{\pi(i)}(t/n) \text{ among } X_1(t/n), \ldots, X_n(t/n).$$

For fixed $t \in [0, 1]$, set $\Delta_{i,j} = \mathbf{1}_{\{X_j(t) \leq X_i(t)\}} - (X_i(t) + 1)/2$. As the rank of x_i among $x_1, \ldots, x_n$ is $\sum_j \mathbf{1}_{\{x_j \leq x_i\}}$, we have that for $t \in \Pi_n$,

$$(2.1) \qquad \frac{2\sigma_{nt}^n(\pi^{-1}(i))}{n} - 1 - X_i(t) = \frac{2}{n} \sum_{j=1}^n \Delta_{i,j}.$$

Observe that $|\Delta_{i,j}| \leq 1$. Also, for $j \neq i$, $\mathbb{E}[\Delta_{i,j} \mid X_i(t)] = 0$. This is where we use the fact that $X(t) \sim \mathrm{Uniform}[-1, 1]$ for every t. Moreover, for all j such that $j \neq i$, the $\Delta_{i,j}$s are mutually independent conditional on $X_i(t)$. Therefore by Bernstein's concentration inequality we infer that for $\epsilon \geq 0$ and every i,

$$(2.2) \qquad \mathbb{P}\left[\frac{1}{n-1} \left| \sum_{j:j \neq i} \Delta_{i,j} \right| > \epsilon \,\bigg|\, X_i(t)\right] \leq 2e^{-\frac{\epsilon^2(n-1)}{4}}.$$

The right hand side (r.h.s.) of (2.1) is bounded in absolute value by $\frac{2}{n} + \frac{2|\sum_{j \neq i} \Delta_{i,j}|}{n-1}$. Therefore, taking an union bound over all $t \in \Pi_n$, setting $\epsilon = n^{-1/4}$ in (2.2), then taking expectation over $X_i(t)$ and another union bound over all particles $i = 1, \ldots, n$, we infer

that for all large n,

$$(2.3) \qquad \mathbb{P}\left[\sup_{1\le i\le n,\, t\in\Pi_n}\left|\frac{2\sigma^n_{nt}(i)}{n}-1-X_{\pi(i)}(t)\right| > 2n^{-1/4}+2n^{-1}\right] \le 2n^2\, e^{-\frac{n^{1/2}}{8}}.$$

Recall that the trajectory of particle i is $T^n_i(t) = (2/n)\sigma^n_{nt}(i) - 1$ for $t\in\Pi_n$, and T^n_i is linearly interpolated in between the times in Π_n. Let

$$A_n = \sup_{1\le i\le n,\, t\in\Pi_n} \left|T^n_i(t) - X_{\pi(i)}(t)\right|.$$

Let $\mathrm{Lin}(n, X_i)$ be the piecewise linear function that agrees with X_i at times $t\in\Pi_n$. Observe that $||T^n_i - \mathrm{Lin}(n, X_{\pi(i)})||_\infty = \sup_{t\in\Pi_n}|T^n_i(t) - X_{\pi(i)}(t)|$ because both functions are piecewise linear between the times $t\in\Pi_n$. Let

$$B_n = \frac{1}{n}\sum_i ||X_i - \mathrm{Lin}(n, X_i)||_\infty.$$

From another application of Bernstein's inequality we deduce that

$$\mathbb{P}\left[B_n > \mathbb{E}\left[||X - \mathrm{Lin}(n, X)||_\infty\right] + n^{-1/4}\right] \le e^{-\frac{n^{1/2}}{16}}.$$

The r.h.s. above is summable over n and Lemma 2.1 implies that $\mathbb{E}\left[||X - \mathrm{Lin}(n, X)||_\infty\right] \to 0$ as $n\to\infty$. Furthermore, $\mathbb{P}\left[A_n \ge 4n^{-1/4}\right]$ is summable over n due to (2.3). Therefore the Borel-Cantelli lemma implies that there is a subset Ω of outcomes of the X_i's having probability 1 such that both $A_n, B_n \to 0$ if $\omega\in\Omega$.

Let $I \sim \mathrm{Uniform}([n])$ and let $\mathbb{E}_I$ denote expectation w.r.t. I, that is, the outcomes of the X_ks are kept fixed. Then,

$$\mathbb{E}_I[||T^n_I - X_{\pi(I)}||_\infty] \le \mathbb{E}_I[||T^n_I - \mathrm{Lin}(n, X_{\pi(I)})||_\infty] + \mathbb{E}_I[||\mathrm{Lin}(n, X_{\pi(I)}) - X_{\pi(I)}||_\infty]$$
$$\le A_n + B_n.$$

Consequently, if $\omega\in\Omega$ then for every uniformly continuous and bounded $F : \mathbf{C} \to \mathbb{R}$ we have that $\mathbb{E}_I[F(T^n_I) - F(X_{\pi(I)})] \to 0$.

The distribution of $X_{\pi(I)}$ over the random I is the empirical measure $(1/n)\sum_i \delta_{X_i}$ on $\mathbf{C}$. As $\mathbf{C}$ is a Polish space, the strong law of large numbers for empirical measures on Polish spaces [10] implies that there is a subset Ω' of outcomes of the X_ks having probability 1 such that $X_{\pi(I)}$ converges weakly to X if $\omega\in\Omega'$.

If $\omega\in\Omega\cap\Omega'$ then for every function F as above we have that $\mathbb{E}_I[F(T^n_I)] \to \mathbb{E}[F(X)]$. We thus conclude that almost surely in the outcomes of the X_ks, the function T^n_I converges weakly to X. This means that the sequence $\{\sigma^n\}$ converges almost surely to the process X. Selecting any such good outcome of the X_ks gives a deterministic sequence of permutation processes converging to X.

2.1. **Convergence of random permutation processes.** The limit notion for permutation processes naturally defines the limit notion for random permutation processes. More precisely, if σ^n is a sequence of random permutation processes of $\mathfrak{S}_n$ then it converges if its trajectory processes converge weakly as random measures on $\mathbf{C}$. The limit in this case is

a random permuton process, that is, a measure on permuton processes. We illustrate this with two contrasting examples.

First, consider the interchange process, Int^n, on the n-path. It is a random permutation process of $\mathfrak{S}_n$ generated by first sampling i.i.d. uniform random adjacent transpositions $\tau_1, \tau_2, \ldots$, and then setting $\mathrm{Int}^n_t = \mathrm{Int}^n_{t-1} \circ \tau_t$ for $t \geq 1$ with $\mathrm{Int}^n_0 = \mathbf{id}_n$. The stationary distribution of this process is the uniform measure of $\mathfrak{S}_n$ and its relaxation time is of order n^3. Thus, one expects a limit of this process to exist if it is run until time n^3.

It is shown in [13] that if Int^n is run until time n^3 then it converges to a deterministic permuton process: stationary Brownian motion on $[-1, 1]$. (Actually, [13] considers the continuous time interchange process but the conclusion also holds for the discrete time process.) Observe that the trajectory of each particle of Int^n is a simple random walk on the n-path. So by Donsker's Theorem [10], after appropriate rescaling, each trajectory converges to Brownian motion on $[-1, 1]$. However, this alone does not imply that the trajectory process converges to stationary Brownian motion.

The convergence to the deterministic limit occurs because the trajectories become "asymptotically independent", which means the following. Let X^n_ω be the trajectory process for a sample outcome ω of Int^n. Let $T^n_{I_1}$ and $T^n_{I_2}$ be two samples from X^n_ω, that is, the trajectory of two particles I_1 and I_2 chosen independently and uniformly at random from X^n_ω. Then, for every continuous and bounded function $F : \mathbf{C} \to \mathbb{R}$,

$$\mathbb{E}_{\omega, I_1, I_2} \left[F(T^n_{I_1}) F(T^n_{I_2}) \right] - \mathbb{E}_{\omega, I_1} \left[F(T^n_{I_1}) \right]^2 \overset{n \to \infty}{\longrightarrow} 0.$$

Asymptotic independence suffices to ensure that random permutation processes have deterministic limits; see [13] for the details.

The second example illustrates a random limit of permutation process. Consider n particles placed on the vertices of the n-cycle. At each time step, rotate the cycle one unit clockwise or counter clockwise (by $2\pi/n$ radians) independently and uniformly at random. This gives a random permutation process σ^n whereby each particle performs a simple random walk on the n-cycle. However, note that the distances between particles remain fixed. When run until time n^3 this process has the following limit. Periodic Brownian motion on $[-1, 1]$, denoted B^{periodic}, is Brownian motion started from a uniform random point of $[-1, 1]$ and run in a period by identifying the endpoints ± 1. Let $U \sim \mathrm{Uniform}[-1, 1]$ be independent of B^{periodic}. The limit of $\{\sigma^n\}$ is the random permuton process $\omega \to X_\omega$ such that for a sample outcome U_ω of U,

$$X_\omega \overset{law}{=} U_\omega + B^{\mathrm{periodic}} \ \left(\mathrm{mod} \ [-1, 1] \right).$$

In other words, $[-1, 1]$ is first rotated by U_ω (by identifying ± 1) and then rotated independently according a periodic Brownian motion. Two samples from X_ω provide two periodic Brownian motions that start from a common point ω–almost surely.

3. Metric and energy for permutons

Let (K, d) be a metric space. A path $\gamma = (\gamma(t); 0 \leq t \leq 1)$ is a continuous function from the interval $[0, 1]$ into K. A *finite partition* of the interval $[a, b]$ is a set of ordered points

$\Pi = \{a = t_0 < t_1 < \cdots < t_n = b\}$. Let Part$[a, b]$ denote the set of all finite partitions of the interval $[a, b]$. The *mesh size* of a partition Π is $\Delta(\Pi) = \max_{1 \leq i \leq n}\{|t_i - t_{i-1}|\}$.

The energy of a path γ with respect to a partition $\Pi \in$ Part$[0, 1]$ is

$$\mathcal{E}[\gamma, \Pi] = \sum_{i=1}^{n} \frac{d^2(\gamma(t_i), \gamma(t_{i-1}))}{t_i - t_{i-1}}.$$

The energy of γ, denoted $\mathcal{E}[\gamma]$, is

(3.1)
$$\mathcal{E}[\gamma] = \sup_{\Pi \in \text{Part}[0,1]} \left\{\mathcal{E}[\gamma, \Pi]\right\}.$$

The energy of γ restricted to the interval $[a, b]$ is

$$\mathcal{E}[\gamma, [a, b]] = \sup_{\Pi \in \text{Part}[a,b]} \left\{\mathcal{E}[\gamma, \Pi]\right\}.$$

Notice that if $a \leq b \leq c$ then we have

(3.2)
$$\mathcal{E}[\gamma, [a, c]] \geq \mathcal{E}[\gamma, [a, b]] + \mathcal{E}[\gamma, [b, c]].$$

In particular, if $t_0 \leq t_1 \leq \ldots \leq t_n$ then $\mathcal{E}[\gamma, [t_0, t_n]] \geq \sum_{i=1}^{n} \mathcal{E}[\gamma, [t_{i-1}, t_i]]$.

For partitions Π, Π' of $[0, 1]$, we write $\Pi \subset \Pi'$ (Π' is a refinement of Π) if Π' contains all points of Π. The energy of a path is non-decreasing under refinements, as Lemma 3.1 below shows. The proof is in the Appendix. We will use this lemma throughout our arguments.

Lemma 3.1. *Suppose $\Pi \subset \Pi'$ are two finite partitions of $[0, 1]$. Then $\mathcal{E}[\gamma, \Pi] \leq \mathcal{E}[\gamma, \Pi']$ for any path γ in K.*

Using Lemma 3.1 we observe that for a path γ there is a sequence of nested finite partitions $\Pi_0 = \{0, 1\} \subset \Pi_1 \subset \Pi_2 \ldots$ such that $\mathcal{E}[\gamma, \Pi_n] \nearrow \mathcal{E}[\gamma]$. We may also assume that $\Delta(\Pi_n) \to 0$. Thus $\cup_n \Pi_n$ is a dense set of points in $[0, 1]$.

We consider paths in two types of metric spaces. First, given a probability space (Ω, Σ, P) we take $K = L^2(P, [-1, 1])$, the Hilbert space of all square integrable random variables $Z : \Omega \to [-1, 1]$. A path in K is then a stochastic process $X = (X(t); 0 \leq t \leq 1))$. Permuton processes fall within this setup. Second, we take K to be the space of permutons $\mathcal{P}$ in the Wasserstein metric, which is the setting of Theorem 3. The Wasserstein metric induces the topology of weak convergence on $\mathcal{P}$ (see Lemma 7.1), and $\mathcal{P}$ is compact in the weak topology by Phokhorov's Theorem.

Finite energy paths in the space of Borel probability measures of a compact metric space K, under Wasserstein metric, are related to finite energy K-valued stochastic processes, in L^2 metric. We may realize the former as the latter in an energy preserving manner. This is the content of Theorem 5 below, which is used to prove Theorem 3.

3.1. **Realizing measure valued paths as stochastic processes.** Throughout this section (K, d) denotes a compact metric space and $\mathcal{M}(K)$ denotes the space of Borel probability

measures on K in the Wasserstein metric. A path $\gamma = (\gamma(t); 0 \leq t \leq 1)$ in $\mathcal{M}(K)$ is *realized* by a K-valued stochastic process $X = (X(t); 0 \leq t \leq 1)$ if the following conditions holds.

(1) $X(t) \sim \gamma(t)$ for every t.

(2) X has continuous sample paths almost surely.

The energy of X is as given by (3.1) with respect to the L^2 metric:

$$d_{L^2}(X(t), X(s)) := \mathbb{E}\left[d(X(t), X(s))^2\right]^{1/2}.$$

Theorem 5. *Suppose that γ is a $\mathcal{M}(K)$–valued path with finite energy with respect to the Wasserstein distance. There is a stochastic process X that realizes γ in a energy preserving manner: $\mathcal{E}[X] = \mathcal{E}[\gamma]$.*

We refer to the process X as an optimal coupling *of γ.*

The rest of the section proves Theorem 5. To begin, note the following fact. If $\nu, \nu' \in \mathcal{M}(K)$ are two Borel probability measures then there is a coupling (V, W) of ν with ν' such that $W(\nu, \nu') = \mathbb{E}\left[d(V, W)^2\right]^{1/2}$. This is because K is compact. Using this fact we inductively build up optimal couplings by using the following lemma.

Lemma 3.2. *Let $\gamma_0, \ldots, \gamma_n \in \mathcal{M}(K)$. There exist jointly distributed K–valued random variables $(X_0, \ldots, X_n)$ such that $X_i \sim \gamma_i$ and $\mathbb{E}\left[d(X_{i-1}, X_i)^2\right] = W(\gamma_{i-1}, \gamma_i)^2$ for $1 \leq i \leq n$.*

Proof. We proceed by induction. The case for two measures is mentioned above (see Lemma 7.2). To carry out the induction step we will need the following measure theoretic fact. It is often known as the Disintegration Theorem (see [10, Theorem 5.10]).

Fact: Let $(X, Y) \in K^2$ be jointly distributed random variables. There is a measurable function $g : K \times [0, 1] \to K^2$ such that if $U \sim \text{Uniform}[0, 1]$ and U is independent of (X, Y) then $(X, g(X, U))$ has the same joint distribution as (X, Y).

Suppose the statement of the lemma holds for $\gamma_0, \ldots, \gamma_{n-1}$ with jointly distributed random variables $(X_0, \ldots, X_{n-1})$. Using Lemma 7.2, we find a coupling (X'_{n-1}, X'_n) of γ_{n-1} with γ_n such that $W(\gamma_{n-1}, \gamma_n)^2 = \mathbb{E}\left[d(X'_{n-1}, X'_n)^2\right]$. Let $U \sim \text{Uniform}[0, 1]$ be independent of all the random variables $X_0, \ldots, X_{n-1}, X'_{n-1}$ and X'_n. Let g be as mentioned in the fact above for the pair (X'_{n-1}, X'_n).

Since X_{n-1} has the same distribution as X'_{n-1}, and U is independent of all other random variables, the pair $(X_{n-1}, g(X_{n-1}, U))$ has the same distribution as (X'_{n-1}, X'_n). Let $X_n = g(X_{n-1}, U)$. Thus, $W(\gamma_{n-1}, \gamma_n)^2 = \mathbb{E}\left[d(X'_{n-1}, X'_n)^2\right] = \mathbb{E}\left[d(X_{n-1}, X_n)^2\right]$. The random variables $(X_0, \ldots, X_n)$ provide the desired coupling. $\square$

Now suppose γ is a $\mathcal{M}(K)$ valued path. We may choose a sequence of nested finite partitions $\Pi_0 \subset \Pi_1 \ldots$ such that $\Delta(\Pi_n) \to 0$ and $\mathcal{E}[\gamma, \Pi_n] \nearrow \mathcal{E}[\gamma]$.

For each n, we apply Lemma 3.2 to find coupled random variables $(X_n(t); t \in \Pi_n)$ such that if $\Pi_n = \{0 = t_0 < \ldots < t_k = 1\}$ then

$$\mathbb{E}\left[d(X_n(t_i), X_n(t_{i-1}))^2\right] = W(\gamma(t_i), \gamma(t_{i-1}))^2 \text{ for every } 1 \le i \le k.$$

Fix an $x_0 \in K$. Set $\Pi_\infty = \cup_n \Pi_n$ and extend X_n to Π_∞ by setting $X_n(t) \equiv x_0$ if $t \in \Pi_\infty \setminus \Pi_n$.

The process X_n takes values in K^{Π_∞} for every n. As K^{Π_∞} is compact in the product topology, by applying Prokhorov's Theorem we can find a subsequence $n_i \to \infty$ and a process $(X(t); t \in \Pi_\infty)$ such that $X_{n_i} \to X$ weakly. As the partitions Π_n are nested we may assume w.l.o.g. that $n_i = n$, that is, $X_n \to X$ weakly.

Consider the process $(X(t); t \in \Pi_\infty)$. We must extend X continuously from the dense subset Π_∞ to $[0, 1]$. First, we show that X has finite energy along Π_∞. Let

$$\mathcal{E}[X, \Pi_\infty] := \lim_{n \to \infty} \mathcal{E}[X, \Pi_n],$$

which exists by monotonicity.

Lemma 3.3. *The process $(X(t); t \in \Pi_\infty)$ satisfies $\mathcal{E}[X, \Pi_\infty] \le \mathcal{E}[\gamma]$. Moreover, for every $s < t$ in Π_∞, $\mathbb{E}\left[d(X(t), X(s))^2\right] \le (t - s)\mathcal{E}[\gamma, [s, t]]$.*

Proof. We begin by showing $\mathbb{E}\left[d(X(t), X(s))^2\right] \le (t - s)\mathcal{E}[\gamma, [s, t]]$ for $s < t$ in Π_∞. Suppose that $s < t$ are both in Π_∞. From weak convergence of the X_n and compactness of K we have that $\mathbb{E}\left[d(X(t), X(s))^2\right] = \lim_{n \to \infty} \mathbb{E}\left[d(X_n(t), X_n(s))^2\right]$. We now bound $\mathbb{E}\left[d(X_n(t), X_n(s))^2\right]$. As $s, t \in \Pi_\infty$, there is an N such that $s, t \in \Pi_n$ for $n \ge N$. Suppose that the points of Π_n between s and t are $s = t_{0,n} < t_{1,n} < \ldots < t_{k_n,n} = t$. Using Lemma 3.1 we deduce that for $n \ge N$,

$$\frac{\mathbb{E}\left[d(X_n(t), X_n(s))^2\right]}{t - s} \le \sum_{i=1}^{k_n} \frac{\mathbb{E}\left[d(X_n(t_{n,i}), X_n(t_{n,i-1}))^2\right]}{t_{n,i} - t_{n,i-1}}$$

$$= \sum_{i=1}^{k_n} \frac{W(\gamma(t_{n,i}), \gamma(t_{n,i-1}))^2}{t_{n,i} - t_{n,i-1}}$$

$$\le \mathcal{E}[\gamma, [s, t]].$$

The last inequality follows due to the $t_{n,i}$ forming a partition of $[s, t]$. By letting $n \to \infty$ we conclude from the above estimate that $\mathbb{E}\left[d(X(t), X(s))^2\right] \le (t - s)\mathcal{E}[\gamma, [s, t]]$.

For the partition $\Pi_n = \{0 = t_0 < \ldots < t_n = 1\}$ we deduce from the inequality above that

(3.3) $$\sum_{i=1}^{n} \frac{\mathbb{E}\left[d(X(t_i), X(t_{i-1}))^2\right]}{t_i - t_{i-1}} \le \sum_{i=1}^{n} \mathcal{E}[\gamma, [t_{i-1}, t_i]].$$

From the inequality (3.2) we now deduce that $\sum_{i=1}^{n} \mathcal{E}[\gamma, [t_{i-1}, t_i]] \le \mathcal{E}[\gamma]$. Therefore, $\mathcal{E}[X, \Pi_\infty] \le \mathcal{E}[\gamma]$ as required. $\square$

We now show that X has a continuous extension to a process defined for times $t \in [0, 1]$. Let (Ω, Σ, μ) denote the probability space where $(X(t), t \in \Pi_\infty)$ is jointly defined and let $X_\omega(t)$ denote the outcome of $X(t)$ for $\omega \in \Omega$. The inequality $\mathcal{E}[X, \Pi_\infty] \le \mathcal{E}[\gamma]$ from

Lemma 3.3 implies that for μ-almost every ω the energy of the discrete K-valued path $(X_\omega(t), t \in \Pi_\infty)$ is finite. In particular, for μ-almost every ω there exists a constant C_ω such that

$$d(X_\omega(t), X_\omega(s)) \leq C_\omega \sqrt{|t - s|} \text{ for } s, t \in \Pi_\infty.$$

Since Π_∞ is a dense subset of $[0, 1]$, Lemma 7.3 from the Appendix implies that $(X_\omega(t); t \in \Pi_\infty)$ has a continuous extension to times $t \in [0, 1]$ for μ-almost every ω. We denote this extension by $X = (X(t), 0 \leq t \leq 1)$, which is then a K-valued stochastic process with continuous sample paths.

Now we show that X realizes γ. Certainly, $X(t) \sim \gamma(t)$ for $t \in \Pi_\infty$ because $X_n(t) \to X(t)$ weakly and $X_n(t) \sim \gamma(t)$ for all large n due to the partitions Π_n being nested. Suppose that $t \in [0, 1] \setminus \Pi_\infty$. Choose a sequence $t_n \in \Pi_n$ such that $t_n \to t$. By continuity of X and the bounded convergence theorem we conclude that $\mathbb{E}[d(X(t_n), X(t))] \to 0$. This implies that $X(t_n) \to X(t)$ weakly. The distribution of $X(t_n)$ is $\gamma(t_n)$ and $\gamma(t_n) \to \gamma(t)$ weakly because the path γ is continuous due to having finite energy. Therefore, $X(t) \sim \gamma(t)$ for every t.

Finally we show that $\mathcal{E}[X] = \mathcal{E}[\gamma]$. As X realizes γ, $\mathbb{E}\left[d(X(t), X(s))^2\right] \geq W(\gamma(t), \gamma(s))^2$. Hence, $\mathcal{E}[X] \geq \mathcal{E}[\gamma]$. To get the reverse inequality first recall from Lemma 3.3 that

$$\mathbb{E}\left[d(X(t), X(s))^2\right] \leq (t - s)\,\mathcal{E}[\gamma, [s, t]] \text{ for every } s, t \in \Pi_\infty \text{ with } s < t.$$

Suppose $s < t$ are two arbitrary points in $[0, 1]$. Choose sequences $\{s_n\}$ and $\{t_n\}$ such that $s_n, t_n \in \Pi_n$, $s_n \leq t_n$, $s_n \searrow s$ and $t_n \nearrow t$. From continuity of X and the bounded convergence theorem we have that $\mathbb{E}\left[d(X(t), X(s))^2\right] = \lim_{n\to\infty} \mathbb{E}\left[d(X(t_n), X(s_n))^2\right]$. Since

$$\mathbb{E}\left[d(X(t_n), X(s_n))^2\right] \leq (t_n - s_n)\,\mathcal{E}[\gamma, [s_n, t_n]] \text{ and } \mathcal{E}[\gamma, [s_n, t_n]] \leq \mathcal{E}[\gamma, [s, t]],$$

we conclude that $\mathbb{E}\left[d(X(t), X(s))^2\right] \leq (t - s)\,\mathcal{E}[\gamma, [s, t]]$ for every $s \leq t$. For an arbitrary partition $\Pi = \{0 = t_0 < \ldots < t_n = 1\}$ we use this inequality to deduce that $\mathcal{E}[X, \Pi] \leq \sum_{i=1}^n \mathcal{E}[\gamma, [t_{i-1}, t_i]]$. The inequality (3.2) implies that $\sum_{i=1}^n \mathcal{E}[\gamma, [t_{i-1}, t_i]] \leq \mathcal{E}[\gamma]$. As Π was arbitrary it follows that $\mathcal{E}[X] \leq \mathcal{E}[\gamma]$. This completes the proof.

4. MINIMAL ENERGY PERMUTON PROCESSES FROM IDENTITY TO REVERSE

In this section we prove Theorem 2 and Corollary 2.1. The proof of Theorem 2 uses the following lemma about minimal energy paths on a Hilbert sphere. A proof is provided in the Appendix.

Lemma 4.1. *Let γ be a path on the unit sphere of a Hilbert space between two antipodal points $\gamma(0)$ and $-\gamma(0) = \gamma(1)$. Then $\mathcal{E}[\gamma] \geq \pi^2$ with equality if and only if*

$$\gamma(t) = \cos(\pi t)\gamma(0) + \sin(\pi t)\gamma(1/2).$$

Proof of Theorem 2. Suppose $X = (X(t); 0 \leq t \leq 1)$ is a permuton process with $X(0) = -X(1)$. Since $\mathbb{E}\left[X(t)^2\right] = 1/3$, the process X is a path between two antipodal points on the sphere of radius $1/\sqrt{3}$ in the Hilbert space $L^2(\Omega, \Sigma, P)$, where (Ω, Σ, P) is the probability space over which the process X is defined. From Lemma 4.1 we see that $\mathcal{E}[X] \geq \pi^2/3$ with equality if and only only if $X(t) = \cos(\pi t)X(0) + \sin(\pi t)X(1/2)$. In case of equality,

since $X(t) \sim \text{Uniform}[-1,1]$ for every t, this equation for X implies that the projection of $(X(0), X(1/2))$ onto any line through the origin has the $\text{Uniform}[-1,1]$ distribution. Thus, $(X(0), X(1/2))$ is distributed according to the Archimedean measure and X is the Archimedean process. $\qquad \square$

Proof of Corollary 2.1. Theorem 2 implies that all limit points of X^n are supported on permuton processes with energy at least $\pi^2/3$. Due to the uniqueness of energy minimizers, X^n will converge to the Archimedean process if the expected energy of any limit point X of X^n is at most $\pi^2/3$. By bounded convergence theorem and assumption (1.8) we have

$$\mathbb{E}\left[(X(t) - X(0))^2\right] \leq \frac{(\pi t)^2}{3}.$$

The ϵ_n-stationarity of X^n implies ϵ-stationarity of X for all $\epsilon > 0$, so

$$\mathbb{E}\left[(X(t) - X(s))^2\right] \leq \frac{(\pi(s-t))^2}{3},$$

and therefore for every partition $\Pi = \{t_0 = 0, \ldots, t_k = 1\}$ we have

$$\sum_{i=1}^{k} \frac{\mathbb{E}\left[(X(t_i) - X(t_{i-1}))^2\right]}{t_i - t_{i-1}} \leq \frac{\pi^2}{3}. \qquad \square$$

Note that conversely, the Archimedean process limit and the bounded convergence theorem would imply that for every t, $\mathbb{E}\left[(X^n(t) - X^n(0))^2\right] \to \frac{2}{3}(1 - \cos(\pi t))$.

5. Minimal energy permuton paths from identity to reverse

In this section, we prove Theorem 3. Using Theorem 5 we transfer the study of paths in $\mathcal{P}$ to $[-1,1]^2$–valued stochastic processes. Then we solve the corresponding energy minimization problem for stochastic processes. We verify that there is an unique energy minimizer and its fixed time marginals agree with the Archimedean path.

1-dimensional energy minimization. In proving Theorem 3 we will reduce the 2-dimensional energy minimization problem to a pair of 1-dimensional energy minimization problems. Here we solve that 1-dimensional problem.

For continuous $f : [0,1] \to \mathbb{R}$, let $\mathcal{E}[f]$ and $\mathcal{E}[f, \Pi]$ denote energy w.r.t. the Euclidean metric on $\mathbb{R}$.

Lemma 5.1. *Let $X = (X(t); 0 \leq t \leq 1)$ be a continuous $\mathbb{R}$-valued stochastic process such that $\mathbb{E}[X(t)] = 0$ and $\mathbb{E}[X(t)^2] < \infty$ for every t. Set $\sigma(t) = \mathbb{E}[X(t)^2]^{1/2}$. Then $\mathcal{E}[X] \geq \mathcal{E}[\sigma]$. Here the energy of X is w.r.t. the L^2-metric and the energy of σ is w.r.t. the Euclidean metric.*

Moreover, suppose that $\mathcal{E}[X] = \mathcal{E}[\sigma]$, $\mathcal{E}[\sigma] < \infty$ and $\sigma(t) > 0$ for $t > 0$. Then the following holds almost surely,

$$X(t) = \frac{\sigma(t)}{\sigma(1)} X(1) \text{ for every } 0 \leq t \leq 1.$$

Proof. We have that $\mathbb{E}\left[|X(t) - X(s)|^2\right] = \sigma(t)^2 - 2\mathbb{E}\left[X(t)X(s)\right] + \sigma(s)^2$. The Cauchy-Schwarz inequality implies $\mathbb{E}\left[X(t)X(s)\right] \leq \sigma(t)\sigma(s)$, and hence, $\mathbb{E}\left[|X(t) - X(s)|^2\right] \geq (\sigma(t) - \sigma(s))^2$. From this inequality it is immediate that $\mathcal{E}[X] \geq \mathcal{E}[\sigma]$.

Now suppose that $\mathcal{E}[X] = \mathcal{E}[\sigma]$, $\mathcal{E}[\sigma]$ is finite and $\sigma(t) > 0$ for every $t > 0$. If we show that $X(t)/\sigma(t)$ is almost surely constant on the interval $[\epsilon, 1]$, for any $\epsilon > 0$, then the continuity of X implies that $X(t)/\sigma(t)$ is almost surely constant on $[0, 1]$. Therefore, we may assume that $\sigma(t) > 0$ for $t \in [0, 1]$.

Set $\delta = \inf_{t \in [0,1]} \sigma(t)$. Then $\delta > 0$ since $\sigma(t)$ is continuous and positive on $[0, 1]$. Set $Y(t) = \frac{X(t)}{\sigma(t)}$. For $0 \leq s \leq t \leq 1$,

$$\mathbb{E}\left[|Y(t) - Y(s)|^2\right] = \frac{\mathbb{E}\left[|X(t) - X(s)|^2\right] - |\sigma(t) - \sigma(s)|^2}{\sigma(s)\sigma(t)}$$

$$\leq \frac{\mathbb{E}\left[|X(t) - X(s)|^2\right] - |\sigma(t) - \sigma(s)|^2}{\delta^2}.$$

The estimate above implies that for any finite partition Π of $[0, 1]$,

$$\mathcal{E}[Y, \Pi] \leq \delta^{-2}\big(\mathcal{E}[X, \Pi] - \mathcal{E}[\sigma, \Pi]\big) \leq \delta^{-2}\big(\mathcal{E}[X] - \mathcal{E}[\sigma, \Pi]\big).$$

Choose a sequence of nested partitions $\Pi_0 \subset \Pi_1 \subset \cdots$ such that $\mathcal{E}[\sigma, \Pi_n] \to \mathcal{E}[\sigma]$. We deduce from the above that $\mathcal{E}[Y, \Pi_n] \to 0$ due to $\mathcal{E}[X] = \mathcal{E}[\sigma]$. Since $\mathcal{E}[Y, \Pi_n]$ is monotone increasing we conclude that $\mathcal{E}[Y, \Pi_n] = 0$ for every n. Set $\Pi = \cup_n \Pi_n$. Then for every $s, t \in \Pi$,

$$\mathbb{E}\left[|Y(t) - Y(s)|^2\right] \leq |t - s| \cdot \left(\sup_n \mathcal{E}[Y, \Pi_n]\right) = 0.$$

Fix an arbitrary $p \in \Pi$. We deduce from the above that for every $q \in \Pi$, $\mathbb{P}\left[Y(q) = Y(p)\right] = 1$. Taking the countable intersection of these events over all $q \in \Pi$ we conclude that

$$\mathbb{P}\left[Y(q) = Y(p) \text{ for every } q \in \Pi\right] = 1.$$

The continuity of Y and the fact that Π is a dense subset of $[0, 1]$ imply that almost surely, $Y(t) \equiv Y(p)$ for every $t \in [0, 1]$. In other words, $X(t) = \frac{\sigma(t)}{\sigma(1)}X(1)$ for every t, almost surely, as required. $\qquad\square$

Proof of Theorem 3. Let $\mathbf{P} = (\mathbf{P}(t); 0 \leq t \leq 1)$ be a path in $\mathcal{P}$ from **id** to **rev** such that $\mathcal{E}[\mathbf{P}] < \infty$ in the Wasserstein metric. Using Theorem 5 we may realize $\mathbf{P}$ as a $[-1, 1]^2$-valued continuous stochastic process X such that $\mathcal{E}[\mathbf{P}] = \mathcal{E}[X]$.

Write $X(t) = (x(t), y(t))$. Then $x(t)$ and $y(t)$ are distributed as Uniform$[-1, 1]$ since $\mathbf{P}(t)$ is a permuton. Also, $x(0) = y(0)$ and $x(1) = -y(1)$ due to $\mathbf{P}(0) = \mathbf{id}$ and $\mathbf{P}(1) = \mathbf{rev}$. Set

$$u(t) = \frac{x(t) - y(t)}{\sqrt{2}},$$

$$v(t) = \frac{x(t) + y(t)}{\sqrt{2}}.$$

Then $\mathbb{E}\left[u(t)\right] = \mathbb{E}\left[v(t)\right] = 0$ for every t. For the boundary conditions we have $u(0) = 0$ and $u(1) = \sqrt{2}x(1)$, while $v(0) = \sqrt{2}x(0)$ and $v(1) = 0$. Set

$$\sigma_u^2(t) = \mathbb{E}\left[u(t)^2\right] \quad \text{and} \quad \sigma_v^2(t) = \mathbb{E}\left[v(t)^2\right].$$

Since $u(t)^2 + v(t)^2 = x(t)^2 + y(t)^2$, we see that $\sigma_u^2(t) + \sigma_v^2(t) = \mathbb{E}\left[x(t)^2 + y(t)^2\right] = 2/3$ due to $x(t)$ and $y(t)$ being distributed as Uniform$[-1, 1]$.

The map $t \to (\sigma_u(t), \sigma_v(t))$ is a path on the circle of radius $\sqrt{2/3}$ that begins at $(0, \sqrt{2/3})$ and ends at $(\sqrt{2/3}, 0)$. It is well known that there is a unique path of minimal energy on such a circle from $(0, \sqrt{2/3})$ to $(\sqrt{2/3}, 0)$. This is the minor arc going from $(0, \sqrt{2/3})$ to $(\sqrt{2/3}, 0)$, and uniquely parametrized by $t \to \sqrt{2/3}\,(\sin(\frac{\pi}{2}t), \cos(\frac{\pi}{2}t))$. The energy of this path is

$$\frac{\pi^2}{6} \int_0^1 \cos\left(\frac{\pi}{2}t\right)^2 + \sin\left(\frac{\pi}{2}t\right)^2 \; dt = \frac{\pi^2}{6}.$$

Consequently, $\mathcal{E}[(\sigma_u, \sigma_v), \Pi]^2 \geq \pi^2/6$. In case of equality we must have

(5.1) $$\sigma_u(t) = \sqrt{2/3}\,\sin\left(\frac{\pi}{2}t\right) \quad \text{and} \quad \sigma_v(t) = \sqrt{2/3}\,\cos\left(\frac{\pi}{2}t\right).$$

Note that $\mathbb{E}\left[|X(t) - X(s)|^2\right] = |u(t) - u(s)|^2 + |v(t) - v(s)|^2$, where the distance for X is in the Euclidean metric of $\mathbb{R}^2$. This implies that $\mathcal{E}[X] = \mathcal{E}[u] + \mathcal{E}[v]$. Lemma 5.1 then implies $\mathcal{E}[u] + \mathcal{E}[v] \geq \mathcal{E}[\sigma_u] + \mathcal{E}[\sigma_v]$. Therefore,

$$\mathcal{E}[\gamma] = \mathcal{E}[X] = \mathcal{E}[u] + \mathcal{E}[v] \geq \mathcal{E}[\sigma_u] + \mathcal{E}[\sigma_v].$$

However, $\mathcal{E}[\sigma_u] + \mathcal{E}[\sigma_v] = \mathcal{E}[(\sigma_u, \sigma_v)] \geq \pi^2/6$.

We have deduced that $\mathcal{E}[\gamma] \geq \pi^2/6$ for any path γ in $\mathcal{P}$. As $\mathcal{E}[\mathbf{A}] = \pi^2/6$, we deduce that $(\mathbf{A}(t); 0 \leq t \leq 1)$ has minimal energy among all paths from $\mathbf{id}$ to $\mathbf{rev}$ in $\mathcal{P}$. If $\mathcal{E}[\gamma] = \pi^2/6$ then the functions σ_u and σ_v must equal the functions from (5.1). In this case we may apply the case of equality from Lemma 5.1 to the processes $u(t)$ and $v(t)$ to conclude that $u(t) = \sqrt{2}\sin(\frac{\pi}{2}t)\,x(1)$ and $v(t) = \sqrt{2}\cos(\frac{\pi}{2}t)\,x(0)$ for every t, almost surely. In terms of X we get that almost surely, for all $0 \leq t \leq 1$,

(5.2) $$x(t) = \cos\left(\frac{\pi}{2}t\right)x(0) + \sin\left(\frac{\pi}{2}t\right)x(1)$$
$$y(t) = \cos\left(\frac{\pi}{2}t\right)x(0) - \sin\left(\frac{\pi}{2}t\right)x(1).$$

We claim that (5.2) implies $(x(0), x(1))$ is distributed as the Archimedean measure. If this holds then we have $\gamma(t) = \mathbf{A}(t)$ because $X(t) \sim \gamma(t)$ and (5.2) implies that $X(t) \sim \mathbf{A}(t)$. The latter holds because the formula above implies the density function of $X(t)$ agrees with that of $\mathbf{A}(t)$ if $(x(0), x(1))$ is distributed as the Archimedean measure.

To see that $(x(0), x(1))$ is distributed as the Archimedean measure observe that $x(t)$ is the projection of $(x(0), x(1))$ onto the line through the origin with angle $\frac{\pi}{2}t$. Also, $y(t)$ is the projection of $(x(0), x(1))$ on the line through the origin with angle $-\frac{\pi}{2}t$. As $x(t)$ and $y(t)$ are distributed as Uniform$[-1, 1]$ for every t, it follows that the distribution of the projection of $(x(0), x(1))$ onto any line through the origin is Uniform$[-1, 1]$. This property determines the Archimedean measure. This completes the proof of Theorem 3.

6. Permutons in Wasserstein metric

In the final section of the paper we establish a formula for the Wasserstein distance from the identity to any permuton and use it to prove Theorem 4.

Lemma 6.1. *Let σ and τ be permutations in $\mathfrak{S}_n$. Then,*
$$W(\mu_\sigma, \mu_\tau)^2 = \frac{4}{n^3} \inf_{\pi \in \mathfrak{S}_n} \sum_i (i - \pi(i))^2 + (\sigma(i) - \tau(\pi(i)))^2.$$

The proof is in the Appendix.

Theorem 6. *The Wasserstein distance of the identity permuton* **id** *from any permuton* $\mathbf{P} = (X, Y)$ *is as follows. Let* (X', Y') *denote an independent copy of* (X, Y). *Then,*
$$W(\mathbf{id}, \mathbf{P})^2 = \frac{4}{3} - 2\mathbb{E}\left[\max\left\{X + Y, X' + Y'\right\}\right].$$

Proof. We first derive the analogue of the above formula for permutations and then take limits to get the final result. For a permutation $\sigma \in \mathfrak{S}_n$, consider
$$\sum_i (i - \pi(i))^2 + (i - \sigma\pi(i))^2 = 4\sum_i i^2 - 2\sum_i i \cdot (\pi(i) + \sigma\pi(i)).$$
Reindexing the latter sum by setting $i := \pi^{-1}(i)$ and replacing π by π^{-1} we get that
$$\inf_{\pi \in \mathfrak{S}_n} \sum_i (i - \pi(i))^2 + (i - \sigma\pi(i))^2 = 4\sum_i i^2 - 2\sup_\pi \sum_i \pi(i)(i + \sigma(i)).$$

The sum $\sum_i \pi(i) \cdot (i + \sigma(i))$ is maximized by choosing $\pi(i)$ to be the rank of $i + \sigma(i)$ in the sequence $1 + \sigma(1), \ldots, n + \sigma(n)$. We can write the maximizing permutation π as $\pi(i) = \sum_j \mathbf{1}_{\{j+\sigma(j) \leq i+\sigma(i)\}}$, whence,
$$\sup_\pi \sum_i \pi(i) \cdot (i + \sigma(i)) = \sum_i \sum_j \mathbf{1}_{\{j+\sigma(j) \leq i+\sigma(i)\}}(i + \sigma(i))$$
$$= \frac{1}{2} \sum_{i,j} \max\left\{i + \sigma(i), j + \sigma(j)\right\}.$$
Therefore,
$$(6.1) \qquad \inf_{\pi \in \mathfrak{S}_n} \sum_i (i - \pi(i))^2 + (i - \sigma\pi(i))^2 = 4\sum_i i^2 - \sum_{i,j} \max\left\{i + \sigma(i), j + \sigma(j)\right\}.$$

Let μ_σ be the empirical distribution associated to σ. If (X, Y) and (X', Y') are two independent random variables with distribution μ_σ then from (6.1) and elementary simplifications we get

$$\mathbb{E}\left[\max\left\{X + Y, X' + Y'\right\}\right] = \frac{2}{n^2} \sum_{i,j} \max\left\{\frac{i + \sigma(i)}{n} - 1, \frac{j + \sigma(j)}{n} - 1\right\}$$

$$= \frac{2}{n^3}\left[4\sum_i i^2 - \inf_\pi \sum_i (i - \pi(i))^2 + (i - \sigma(\pi(i)))^2\right] - 2$$

$$= \frac{2}{n^3}\left[4\sum_i i^2 - \frac{n^3}{4} W(\mu^{\mathrm{id}^n}, \mu^\sigma)^2\right] - 2,$$

where the last equality follows from Lemma 6.1. Since $\sum_i i^2 = n^3/3 + O(n^2)$, the above implies

$$W(\mu^{\mathrm{id}^n}, \mu^\sigma)^2 = \frac{4}{3} - 2\mathbb{E}\left[\max\left\{X + Y, X' + Y'\right\}\right] + O(1/n).$$

There exits permutations $\sigma^n \in \mathfrak{S}_n$ such that μ^{σ^n} converges to $\mathbf{P}$ in the Wasserstein distance by [9, Theorem 1.6] and Lemma 7.1. Therefore, $W(\mathbf{id}, \mathbf{P})^2 = \lim_n \mathbb{E}\left[W(\mu^{\mathrm{id}^n}, \mu^{\sigma^n})^2\right]$. The formula for $W(\mathbf{id}, \mathbf{P})^2$ follows from this convergence upon taking the large n limit of the formula above. $\qquad\square$

Observe that for a permuton $\mathbf{P} = (X, Y)$ we have $W((X, Y), \mathbf{rev}) = W((X, -Y), \mathbf{id})$. From Theorem 6 we conclude that for any permuton $\mathbf{P} = (X, Y)$,
(6.2)
$$W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2 = \frac{8}{3} - 2\mathbb{E}\left[\max\left\{X + Y, X' + Y'\right\} + \max\left\{X - Y, X' - Y'\right\}\right].$$

Proof of Theorem 4. Given $\mathbf{P} = (X, Y)$, let $W = \frac{X - Y}{\sqrt{2}}$, $V = \frac{X + Y}{\sqrt{2}}$. Define (W', V') analogously for the pair (X', Y'). From (6.2) we have

$$W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2 = \frac{8}{3} - 2\sqrt{2}\,\mathbb{E}\left[\max\left\{W, W'\right\} + \max\left\{V, V'\right\}\right].$$

Suppose Z is an integrable random variable and Z' is an independent copy of Z. Then

$$\mathbb{E}\left[\max\{Z, Z'\}\right] = 2\mathbb{E}\left[Z\,\mathbb{P}_{Z'}[Z' < Z]\right] + \mathbb{E}\left[Z\,\mathbb{P}_{Z'}[Z' = Z]\right].$$

Define $F(z, u) = \mathbb{P}[Z < z] + u\mathbb{P}[Z = z]$ for $z \in \mathbb{R}$ and $0 < u < 1$. The function F is called the "distributional transform" of Z. If U is independent of Z and distributed as Uniform$[0, 1]$ then $F(Z, U)$ is distributed as Uniform$[0, 1]$ as well [14, Proposition 2.1]. From the definition of $F(z, u)$ we obtain

$$\mathbb{E}\left[Z\,\mathbb{P}\left[Z' < Z\right]\right] = \mathbb{E}\left[ZF(Z, U)\right] - \frac{1}{2}\mathbb{E}\left[Z\,\mathbb{P}_{Z'}[Z = Z']\right].$$

In particular, $\mathbb{E}\left[\max\{Z, Z'\}\right] = 2\mathbb{E}\left[ZF(Z, U)\right]$. Hence,

$$\mathbb{E}\left[\max\left\{W, W'\right\} + \max\left\{V, V'\right\}\right] = 2\mathbb{E}\left[WF(W, U) + VG(V, U)\right],$$

where F and G are the distributional transforms of W and V.

Set $\hat{F}(x, u) = 2F(x, u) - 1$ and $\hat{G}(x, u) = 2G(x, u) - 1$. Observe that

$$WF(W, U) + VG(V, U) = \frac{W\hat{F}(W, U)}{2} + \frac{W}{2} + \frac{V\hat{G}_U(V)}{2} + \frac{V}{2}.$$

We take expectations of this equation and use that $\mathbb{E}[W] = \mathbb{E}[V] = 0$. Then in order to bound the expectation of the r.h.s. we use the inequality $ab \le \frac{a^2 + b^2}{2}$. This gives that $W\hat{F}(W, U) \le (W^2 + \hat{F}(W, U)^2)/2$ and $V\hat{G}(V, U) \le (V^2 + \hat{G}(V, U)^2)/2$. Since $\mathbb{E}[W^2 + V^2] = \mathbb{E}[X^2 + Y^2] = 2/3$, we conclude that

$$\mathbb{E}[WF(W, U) + VG(V, U)] \le \frac{1}{4}\mathbb{E}\left[W^2 + \hat{F}(W, U)^2 + V^2 + \hat{G}(V, U)^2\right] = \frac{1}{3}.$$

Furthermore, there is equality if and only if $W = \hat{F}(W, U)$ and $V = \hat{G}(V, U)$, which is equivalent to (W, V) being a permuton. As a result,

$$W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2 \ge \frac{8 - 4\sqrt{2}}{3},$$

with equality if and only if $\left(\frac{X-Y}{\sqrt{2}}, \frac{X+Y}{\sqrt{2}}\right)$ is a permuton. $\square$

ACKNOWLEDGEMENTS

7. APPENDIX

Proof of Lemma 2.1.

Proof. Note that $m^\delta(Y)$ is non increasing in δ and converges to zero as $\delta \to 0$ almost surely since $t \to Y(t)$ is uniformly continuous. Also, $m^\delta(Y) \le 2$, and thus, $\mathbb{E}[m^\delta(Y)] \to 0$ as $\delta \to 0$ by the bounded convergence Theorem. For the second claim observe that $\left||Y(t) - Y(s)| - |\hat{Y}(t) - \hat{Y}(s)|\right| \le 2||Y - \hat{Y}||_\infty$ by the triangle inequality. Therefore, $m^\delta(Y) \le m^\delta(\hat{Y}) + 2||Y - \hat{Y}||_\infty$ and vice-versa, which implies the claim. Finally, for the third claim notice that

$$|\text{Lin}(n, Y)(t) - Y(t)| = \sum_{i=1}^{n} |Y(\frac{i}{n}) - Y(t) + n(Y(\frac{i-1}{n}) - Y(\frac{i}{n}))(t - \frac{i}{n})| \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n}]}(t) \le 2m^{1/n}(Y).$$

The claim now follows from the assertion of the first claim. $\square$

Proof of Lemma 3.1.

Proof. Suppose that $\Pi = \{0 = t_0 < t_1 \ldots < t_n = 1\}$. As Π' is a refinement of Π it contains points between the t_i. Suppose the points of Π' are indexed as $t_{i,j}$ for $0 \le i \le n$ and $0 \le j \le k_i$ such that $t_i = t_{i,0} < t_{i,1} < \ldots < t_{i,k_i} = t_{i+1,0} = t_{i+1}$. From the triangle

inequality, $d(\gamma(t_i), \gamma(t_{i-1})) \leq \sum_{j=1}^{k_{i-1}} d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))$. So we deduce from the Cauchy-Schwarz inequality that

$$d(\gamma(t_i), \gamma(t_{i-1})) \leq \sum_{j=1}^{k_{i-1}} \sqrt{t_{i-1,j} - t_{i-1,j-1}} \quad \frac{d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))}{\sqrt{t_{i-1,j} - t_{i-1,j-1}}}$$

$$\leq \sqrt{t_i - t_{i-1}} \left[\sum_{j=1}^{k_{i-1}} \frac{d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))^2}{t_{i-1,j} - t_{i-1,j-1}} \right]^{1/2}.$$

We conclude that

$$\frac{d(\gamma(t_i), \gamma(t_{i-1}))^2}{t_i - t_{i-1}} \leq \sum_{j=1}^{k_{i-1}} \frac{d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))^2}{t_{i-1,j} - t_{i-1,j-1}}.$$

Summing the inequality above over i implies that $\mathcal{E}[\gamma, \Pi] \leq \mathcal{E}[\gamma, \Pi']$. $\qquad\square$

Proof of Lemma 4.1.

Proof. Suppose that γ is a path on the unit sphere of a Hilbert space H from the vector $\gamma(0)$ to its antipode $-\gamma(0)$. For unit vectors a and b it is easily seen that $||a - b||^2 = 4\sin(\theta/2)^2$ where $\theta = \arccos(\langle a, b \rangle)$ is the unique angle between a and b in the interval $[0, \pi]$. Let $\theta(s, t)$ denote the angle between $\gamma(s)$ and $\gamma(t)$. Using the inequality $x - (x^3/6) \leq \sin(x) \leq x$ for $0 \leq x \leq \pi$, we see that for any partition $\Pi = \{0 = t_0 < \ldots < t_n = 1\}$,

$$(1 - \delta(\Pi))^2 \sum_{i=1}^{n} \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i} \leq \mathcal{E}[\gamma, \Pi] \leq \sum_{i=1}^{n} \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i},$$

where $\delta(\Pi) = \max_i \{\theta(t_{i-1}, t_i)^2\}/24$. Note that $\theta(s, t)$ is continuous as γ is continuous, and in particular, $\delta(\Pi) \to 0$ as the mesh size $\Delta(\Pi) \to 0$. This implies that $\mathcal{E}[\gamma] = \sup_{\Pi} \sum_i \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i}$.

If a, b and c are unit vectors in H then the corresponding angles between them satisfy $\theta(a, b) + \theta(b, c) \geq \theta(a, c)$. This elementary fact can be deduced by considering unit vectors in $\mathbb{R}^3$ since a, b and c lie is a 3-dimensional subspace. In particular,

$$\sum_i \theta(t_{i-1}, t_i) \geq \theta(0, 1) = \pi$$

because $\gamma(0)$ and $\gamma(1)$ are antipodal. From the Cauchy-Schwarz inequality we conclude that

$$\pi^2 \leq \sum_i (\sqrt{t_{i-1} - t_i})^2 \cdot \sum_i \frac{\theta(t_{i-1}, t_i)^2}{(\sqrt{t_{i-1} - t_i})^2} = \sum_i \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i},$$

with equality only if $\theta(t_{i-1}, t_i) = \pi(t_{i-1} - t_i)$. We may now conclude that $\mathcal{E}[\gamma] \geq \pi^2$, and moreover, if there is equality then $\theta(s, t) = \pi|s - t|$.

Now suppose that $\mathcal{E}[\gamma] = \pi^2$. We show that $\gamma(t) = \cos(\pi t)\gamma(0) + \sin(\pi t)\gamma(1/2)$. From the fact that $\theta(s, t) = \pi|t - s|$ we see that $\theta(0, t) = \pi t$ and we may write

$$\gamma(t) = \cos(\pi t)\,\gamma(0) + \sin(\pi t)\,x(t),$$

where $x(t)$ is a unit vector that is orthogonal to $\gamma(0)$. Noting that $x(1/2) = \gamma(1/2)$ we have that $\langle \gamma(t), \gamma(1/2) \rangle = \sin(\pi t) \langle x(t), x(1/2) \rangle$. Now $\langle \gamma(t), \gamma(1/2) \rangle = \cos(\theta(t, 1/2))$, which equals $\sin(\pi t)$ because $\theta(t, 1/2) = \pi|t - \frac{1}{2}|$. Therefore, $\langle x(t), x(1/2) \rangle = 1$ for $0 < t < 1$, which implies that $x(t) = x(1/2)$ as both of these are unit vectors. Consequently, $\gamma(t) = \cos(\pi t)\gamma(0) + \sin(\pi t)\gamma(1/2)$. $\qquad \square$

Proof of Lemma 6.1.

Proof. Couplings between μ_σ and μ_τ are supported on the points $\left(\frac{2i}{n} - 1, \frac{2\sigma(i)}{n} - 1, \frac{2j}{n} - 1, \frac{2\tau(j)}{n} - 1 \right)$ for $1 \leq i, j \leq n$. Thus, a coupling (V, W) between μ_σ and μ_τ is always described by the array of numbers $[\alpha_{i,j}]_{1 \leq i,j \leq n}$ such that

$$(7.1) \qquad \alpha_{i,j} = \mathbb{P}\left[V = \left(\frac{2i}{n} - 1, \frac{2\sigma(i)}{n} - 1 \right), W = \left(\frac{2j}{n} - 1, \frac{2\tau(j)}{n} - 1 \right) \right].$$

The constraints $V \sim \mu_\sigma$ and $W \sim \mu_\tau$ is equivalent to the matrix $M = [n\alpha_{i,j}]$ being doubly stochastic. Denoting $M = [m_{i,j}]$, we get that

$$(7.2) \qquad \mathbb{E}\left[||V - W||^2 \right] = \frac{4}{n^3} \sum_{i,j} m_{i,j}[(i - j)^2 + (\sigma(i) - \tau(j))^2].$$

Let $\mathcal{B}_n$ be the set of all $n \times n$ doubly stochastic matrices. The map taking $M = [m_{ij}] \in \mathcal{B}_n$ to the r.h.s. of (7.2) is linear, and hence minimized at one of the extreme points of the convex set $\mathcal{B}_n$. These are the permutation matrices P_π for $\pi \in \mathfrak{S}_n$. For a permutation matrix P_π, we have that $\sum_{i,j} m_{i,j}[(i - j)^2 + (\sigma(i) - \tau(j))^2] = \sum_i (i - \pi(i))^2 + (\sigma(i) - \tau(\pi(i)))^2$. As a result, we conclude from (7.2) that

$$W(\mu_\sigma, \mu_\tau)^2 = \frac{4}{n^3} \left[\inf_{\pi \in \mathfrak{S}_n} \sum_i (i - \pi(i))^2 + (\sigma(i) - \tau(\pi(i)))^2 \right].$$

$\qquad \square$

The proofs of the following lemmas use standard arguments. We omit them for brevity.

Lemma 7.1. *Let (K, d) be a compact metric space. Let ν_n be a sequence of Borel probability measures on K. Then ν_n converges weakly to a measure ν if and only if $W(\nu_n, \nu) \to 0$.*

Lemma 7.2. *Let ν, ν' be Borel probability measures on a compact metric space (K, d). There exists a coupling (V, W) of ν with ν' such that $W(\nu, \nu') = \mathbb{E}\left[d(V, W)^2 \right]^{1/2}$.*

Lemma 7.3. *Let K be a complete metric space and $S \subset [0, 1]$ a countable dense set. Suppose $f : S \to K$ has modulus of continuity m on S, i.e., $d(f(t), f(s)) \leq m(|t - s|)$ for $s, t \in S$. Then f has an extension to $[0, 1]$ with modulus of continuity m.*

References

[1] O. Angel, D. Dauvergne, A. E. Holroyd, and B. Virág. The local limit of random sorting networks. To appear in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques.* arXiv:1702.08368.

[2] O. Angel, V. Gorin, and A. E. Holroyd. A pattern theorem for random sorting networks. *Electronic Journal of Probability*, 17(99):1–16, 2012.

[3] O. Angel, A. E. Holroyd, D. Romik, and B. Virág. Random sorting networks. *Advances in Mathematics*, 215(2):839–868, 2007.

[4] D. Dauvergne. The Archimedean limit of random sorting networks. *Preprint*, 2018. arXiv:1802.08934.

[5] D. Dauvergne and B. Virág. Circular support in random sorting networks. *Preprint*, 2018. arXiv:1802.08933.

[6] P. Edelman and C. Greene. Balanced tableaux. *Advances in Mathematics*, 63(1):42–99, 1987.

[7] R. Glebov, A. Grzesik, T. Klimošová, and D. Král. Finitely forcible graphons and permutons. *Journal of Combinatorial Theory, Series B.*, 110:112–135, 2015.

[8] V. Gorin and M. Rahman. Random sorting networks: Local statistics via random matrix laws. *Preprint*, 2017. arXiv:1702.07895.

[9] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Rath, and R. M. Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory, Series B.*, 102(1):93–113, 2013.

[10] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag New York, 2nd edition, 2002.

[11] R. Kenyon, D. Král, C. Radin, and P. Winkler. Permutations with fixed pattern densities. *Preprint*, 2015, arXiv:1506.02340.

[12] M. Kotowski. Limits of random permuton processes and large deviations of the interchange process. *PhD Thesis*, University of Toronto, 2016.

[13] M. Rahman and B. Virág. Brownian motion as limit of the interchange process. *Preprint*, 2016. arXiv:1609.07745.

[14] L. Rüschendorf. On the distributional transform, Sklar's theorem and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009.

[15] R. P. Stanley. On the number of reduced decompositions of elements of Coxeter groups. *European Journal of Combinatorics*, 5(4):359–372, 1984.

[16] S. Starr. Thermodynamic limit for the Mallows model on S_n. *Journal of Mathematical Physics*, 50, 2009.

DEPARTMENT OF MATHEMATICS, MIT, CAMBRIDGE, MA, USA.

E-mail address: `mustazee@gmail.com`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, TORONTO, ON, CANADA AND MTA ALFRED RÉNYI INSTITUTE OF MATHEMATICS, BUDAPEST, HUNGARY.

E-mail address: `balint@math.toronto.edu`

MTA ALFRED RÉNYI INSTITUTE OF MATHEMATICS, BUDAPEST, HUNGARY.

E-mail address: `vizermate@gmail.com`