Time Series, Persistent Homology and Chirality

Yuliy Baryshnikov*
Departments of Mathematics and ECE,
University of Illinois at Urbana-Champaign
ymb@illinois.edu

May 9, 2024

Abstract

Interactions of the maxima and minima of the univariate functions appear in combinatorics as Dyck paths, and in topological data analysis as persistent homology. We study these descriptors for Brownian motions with drift, deriving the intensity measure and correlation functions for the persistence diagram point process $\mathbf{PH}_0$, and quantifying the intrinsic asymmetries in the coupling of maxima and minima.

1   Introduction

Motivation for this study comes from a growing body of work addressing Morse theory of random function, in the context of *Topological Data Analysis* (see, e.g. [AT07, BK18]).

Morse theory associates to a smooth function on a manifold a collection of its critical points with some of their local descriptors (such as critical values or indices, for nondegenerate critical points). It became a powerful tool to characterize both the function and the underlying manifold in different contexts, such as spectral theory of differential operators, or diffusions on manifolds.

An application of the Morse descriptors in data analysis emerged over the past decade, popularized under the name of *persistent homology*. The original motivation behind the notion was to provide a scale-independent toolbox for understanding the topology of the unknown underlying model for the data; the properties of the function defining the persistence data was a mere byproduct. With time, however, persistent homology became a powerful tool to sketch the properties of the function itself.

The analysis of the mapping that associates to a function its persistent homology is rather subtle, and it is natural to attempt to understand its output on some "typical functions" proxied, customarily, by realizations of a *random function*. While in general precise characterization of the output of the persistent homology black box is rather implicit (compare, for example, [AT11, KM10, BA11]), there is a class of random functions where that can be described quite precisely: namely, the trajectories of certain Brownian motions. This is the central object of study of this paper. We use the standard properties of Brownian motions to investigate in details 0-dimensional persistent homology, interpreted as a persistence diagram point process. In particular, we find the intensity density and 2-point correlation functions for the standard Brownian motions with drift.

While understanding the structure of persistent homology for random univariate functions was the starting point of this paper, our findings address a somewhat finer descriptor than just persistence diagram. As we discuss below, for univariate functions the zeroth persistent homology is just a sketch of a finer invariant of the function, its *merge tree*. In particular, each bar comes with a

---

right- or left- (**N**- or **И**-, in our nomenclature) orientation, a *chirality*. We find relative frequencies of them.

Main qualitative results of this paper are the formulae for the density of the $\mathbf{PH}_0$ point process and its 2-point correlation functions (Proposition 3.3 and Theorem 5.7) and the results about the expected excess of the number of **N** bars over the number of **И** bars, Theorem 4.4 for Brownian motions with positive drift.

The results of this paper should be seen in the general context of *reparametrization invariant* tools of data analysis, in our situation of analysis of time series. Changing the coordinates of the underlying space corrupts many traditional tools (such as Fourier analysis). By the contrast, persistence diagrams, their various refinements described in this paper, or the characteristics like unimodal category [BG18] survive right compositions with homeomorphism, and therefore describe patterns independent of the semantics attached to the domain coordinates.

The structure of the paper: in Section 2 we recall basics of persistent homology, merge trees; specialize these definitions to the univatiate case, and introduce our notion of chirality. In Section 3 we introduce the random persistence diagrams for trajectories of Brownian motions, and use that in Section 4 to evaluate the average numbers of bars of different chiralities in the Brownian motions with constant drift. In Section 5 we develop the formalism of automata, and use it to compute the 2-point correlation functions for $\mathbf{PH}_0$.

In Appendix we present, inter alia, an efficient algorithm for computing $\mathbf{PH}_0$ for time series, and introduce a monoid structure on merge trees.

2 Persistent Homology and Merge Trees

2.1 Persistent Homology Diagrams

Here we briefly mention some basic notions of the persistent homology, to place our results in the appropriate context.

Consider a continuous function $f : X \to \mathbb{R}$ on a topological space X. (In most applications one can assume that X can be triangulated.) Sublevel sets of f define a filtration of X by the sets $X_s := \{x \in X : f(x) \leqslant s\}$. *Persistent homology* corresponding to such filtration is the collection of morphisms
$$H_k(m_{st}) : H_k(X_s) \to H_k(X_t), k = 0, 1, \ldots$$
induced by the natural embeddings $m_{st} : X_s \hookrightarrow X_t$ defined for any $s < t$. (All homologies here are singular, over a field.)

The persistent homology theory (see, e.g. [EH10, Oud15]) stipulates that for any $k \geqslant 0$, this collection of the homology groups and homomorphisms between them decomposes into a direct sum (possibly, infinite) of elementary pieces, each of which is a trivial bundle of one-dimensional spaces over an interval, with trivialization being the isomorphisms. This means that a collection of homology classes in $H_k(X_b)$ emerges at some $s = b$ and ends at some $s = d$, $b < d$. (One augments this family of isomorphisms by 0-dimensional spaces outside of the $[b, d]$ range and corresponding zero morphisms $H_k(m_{st}) = 0$ whenever $s < b$ or $s > d$.) These chains of isomorphisms are referred to as *bars* $[b, d]$.

A conventional and convenient way to visualize these bars is to place a point charge at $(b, d) \in \mathbb{R}^2$ in the half-space $b < d$ for the bar starting at b and ending at d, the weight being the rank of the corresponding homology group. This results, for each $k \geqslant 0$, in what is known as the k-*dimensional persistent homology diagram* $\mathbf{PH}_k$.

We will identify persistence diagrams with their *counting measures*, i.e. the sum of delta-functions weighted by the ranks of the homologies at (b, d) for each bar. This gives a positive measure (still denoted as $\mathbf{PH}_k$) supported by the half-plane $\{b < d\}$.

For random functions, the resulting random measure is interpreted as a *point process*, a convenient device to describe properties of the random persistence diagrams.

In general, the total mass of the persistence measures $\mathbf{PH}_k$ can be infinite. (In fact, as we prove in [BW16], it *is* infinite for *generic* functions with loose enough moduli of continuity on triangulable

spaces with length metric.)

However, the contents of the displaced quadrants $Q(b_*, d_*) := \{(b, d) : b \leqslant b_*, d \geqslant d_*\}$ are finite for all $b_* < d_*$ in most interesting situations. Filtrations (or functions generating them) for which this holds are called *tame* [Oud15]. Continuous functions on compact metric spaces are tame.

Notice that the points in $Q(b_*, d_*)$ correspond to bars *straddling* the interval $[b_*, d_*]$.

2.2   Merge trees

In this paper we deal exclusively with 0-dimensional persistence homologies $\mathbf{PH}_0$.

It is well known that 0-dimensional persistence diagram $\mathbf{PH}_0$ of the filtration associated with a function f can be defined without invoking any homologies, but rather by simply tracking the connected components of the subgraphs of the function. This procedure is codified by the notion of *merge trees* associated with f (our trees grow downwards).

Let X be a path connected space [BBI22], and f a continuous real-valued function on X.

Definition 2.1. For a pair of points $x, x' \in X$, we say that the point x' *tops* x if there is a path in X from x to x', such that f restricted to this path has a global maximum at its x' end (clearly, this is a transitive relation). A pair of points topping each other are declared equivalent.

The *Merge Tree* T_f of f is the topological space which coincides as a set with the collection of these equivalence classes (we will denote the class of $x \in X$ as $[x]$).

It is clear that the function f is constant over these equivalence classes, thus defining a function on T_f (which we still will be calling f). We will be referring to the (well-defined) value of f at each point of the merge tree as its *height*.

The merge tree is equipped with the roughest topology such that the height is continuous. In this topology, the merge tree is a path metric space.

Quite immediately, the merge tree of the merge tree T_f (viewed as a metric space with the height considered as the defining function) is again T_f.

For compact spaces, the (unique) class of points where f attains its global maximum on X is referred to as the *root*. We note that one can extend the height of the root above the maximum of f on X (by attaching a long whisker sticking up);' sometimes it is useful to extend it to $+\infty$.

The class of a point that tops only points in its equivalence class is a *leaf* of the merge tree (leaves correspond to the local minima of f).

It is easy to see that classes of points topping a given point form a segment in T_f, ending at the root: for any pair of levels $s < t$, and x such that $f(x) = s$, there is a unique point $[x'] \in T_f$ topping $[x]$ with $f([x'] = t$.

This correspondence is, of course, the same as the natural mapping $H_0(X_s) \to H_0(X_t)$ used in the definition of the persistent homology: for each (path-)connected component of X_s, there is a unique connected component of X_t containing it.

The term *merge* reflects the obvious from the definition observation that the components of the subgraph can only merge as the level grows, but cannot branch.

In general the merge trees can be quite wild even for Lipschitz functions (can have vertices of infinite degrees etc).

2.2.1   From merge trees to $\mathbf{PH}_0$

Persistence diagrams can be derived from the merge tree by using the following recursive procedure (requiring compactness of sublevel sets) of *tree pruning*, relying on the *Elder Rule* [EH10] :

- Find a global minimum (a leaf of the merge tree with the lowest height), and the unique path to the root of the merge tree from that leaf. The resulting pair of heights, from the lowest to the highest, is a recorded as a unit charge on the persistence diagram.

- We will refer to the path from the selected bottom leaf to the root as the *stem* of the (sub)tree. Removing the stem from the tree leaves a forest (possibly, empty) of merge trees. We will say

that the stems of the each of the resulting subtrees *are attached to*, or *are descendants of* the stem just removed.

- Iterate the procedure on each of the resulting trees recursively. This results in a pile of stems, each corresponding to a bar in the persistence diagram.

From the construction of the stems, it is clear that a stem $[b', d']$ attached to its parent stem $[b, d]$ satisfies $b < b' < d' < d$. In terms of the persistent diagrams, a descendent stem is located South-East off its parent.

The decomposition of a tree into a pile of stems erases a lot of information about the original function: thus, the parentage relationships between stems are lost. This is one source of ambiguity if one attempts to reconstruct the merge tree from its persistence diagram (pile of stems).

Another source of ambiguity (for univariate functions, which are the focus of this paper) is the fact that a stem can be attached to its parent on the right or on the left side.

In combinatorial terms, the merge tree associated to a function on an interval is equipped the structure of a *plane tree*: this means that at each internal vertex one fixes the order in which branches are attached to it.

2.2.2   Univariant Functions and Trees

The merge tree (together with its planar embedding) generates in the standard fashion its *contour* or *height walk*, also known as the *Dyck path* (see, e.g. [LG05]). In the situation when the underlying space is an oriented interval, the correspondence between height walks and plane trees is bijective.
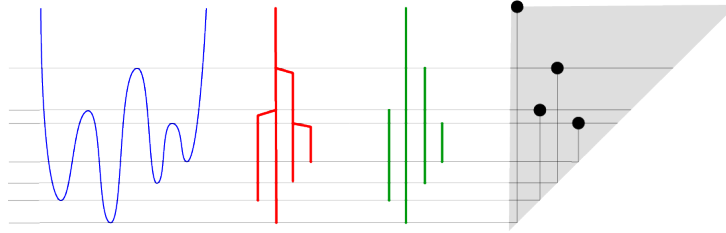

Figure 1: Left to right: graph of a function, its merge tree, the pile of stems (bars) resulting from pruning, and the persistence diagram.

Lemma 2.2. The height walk corresponding to the merge tree T_f is right-equivalent (i.e. equal up to a reparametrization of the argument) to f.

Figure 1 illustrates the relationship between univariate functions, their merge trees and their $\mathbf{PH}_0$.

2.2.3   From $\mathbf{PH}_0$ to merge trees

Constructions above essentially answer the natural question about the space of univariate functions generating a particular collection of bars as its $\mathbf{PH}_o$ diagram. The extra data necessary to rebuild the planar merge tree (and therefore the function generating that merge tree, up to reparametrization) from a (locally finite) collection of bars, are the data a) on their parental attachments, and b) on their planar order.

This result was first published by J. Curry [Cur17]: Let f be a univariant function with the longest bar $-\infty \leqslant b_\infty < d_\infty \leqslant \infty$. Assume that the number of bars straddling each interval (i.e. the $\mathbf{PH}_0$-content of $Q(b, d)$) is finite (i.e. the function is tame).

Proposition 2.3 ([Cur17])**.** The merge tree is uniquely reconstructed by the (arbitrary) attachment relation on the stems (how each bar with exception of the longest one is attached to one of the

straddling bars), and by the planar structure (which we will refer to as chiralities, see section 2.3) of the stems.

We remark that one can generalize Proposition 2.3 to higher dimensions as follows:

Theorem 2.4. Consider the space of Morse functions on $\mathbb{R}^d$ with compact sublevel sets, whose critical points are of indices 0 and 1 only, and increasing to infinity at infinity. The subspace of functions corresponding to a given merge tree (i.e. with a fixed persistence diagram and the stem attachments data) is homotopy equivalent to the product of $(d-1)$-dimensional spheres, one sphere for each bar not ending at $+\infty$.

The proof will appear elsewhere.

2.3   Chirality

A bar (i.e. a point (b, d) in the persistence diagram) corresponds to two events: the birth of a connected component of the sublevel set X. at height b, and its death, i.e. merger of two connected components, at height d. Clearly, this implies that there exists a (local) minimum of f with the critical value b and a (local) maximum with the critical value d.

Definition 2.5. We will be referring to the critical *points* corresponding to a par (b, d) in $\mathbf{PH}_0$ as *coupled*.

Given a coupled pair of local maximum and minimum, we will refer to the right or left orientation of 0-dimensional bar in a persistence diagram of a univariate function, when it is attached to its parent stem, as *chirality*. There are two possible chiralities which we will denote as **И**, and **N**.[1] In terms of the bijection between the Dyck paths and plane trees, **И** corresponds to the branch attached to its parent stem from the left; **N**, from the right.

Formally,

Definition 2.6. A bar $(b, d) = (f(s), f(t))$ (here s, t are the critical points corresponding to the critical values b, d) is an **И**, if the local maximum follows local minimum, i.e. if $s < t$, and an **N**, otherwise, - i.e. if $s > t$.
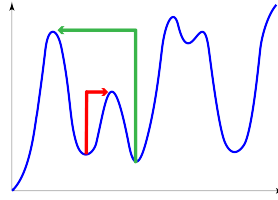

Figure 2: Coupled critical points: **И** in red and **N** in green.

Remark 2.7. Chiralities might be a useful tool to capture potential asymmetry of a time series with respect to the time reversal, a prominent topic in econometric literature. In [LNR$^+$12], an approach somewhat resembling our chiralities was proposed.

Intuitively, if the function is rising, one would expect more of **N**'s than **И**'s. In Section 4 we will compute the average excess of the number of **N**'s over **И**'s in Brownian trajectories with positive drift.

---

[1] **И** is the 11th letter of Ukrainian alphabet, phonetically close to /y/.

2.4 Bars and Windings

In this section we will set up necessary apparatus to analyze bar decomposition of a continuous univariate function. Our key observation is that bars in a persistence diagram for a function on the real line are in direct correspondence with the *windings of the function around an interval*.[2]

In what follows we will be working primarily with the functions on an interval which attain their global minimum on the left end, and their global maximum on the right. One can always relate this particular version with other settings (say, where the function tends to its global supremum on both ends), by stitching on a monotonic function interpolating global maximum and minimum at one of the interval ends: such a transformation alters the persistence diagram at most by one (long) bar.

2.4.1 Winding of a function around an interval

Recall that we say that a point $(b, d), b < d$ in the persistence diagram (or the corresponding bar $[b, d]$) *straddles the interval* $[b', d']$ if $b \leqslant b' < d' \leqslant d$.

Given an interval $[b, d]$, and a continuous function f on $(t_0, t_1) = I \subset \mathbb{R}$, starting below b and ending above d on I (i.e. $\limsup_{t \to t_0} f < b < d < \liminf_{t \to t_1} f$), one can define sequences of alternating b- and d-times $t_0^b < t_1^d < \ldots < t_k^d < t_k^b < \ldots$ by setting, iteratively,

$$t_0^b = t_0; \text{ for } k \geqslant 1, t_k^d = \min\{t : f(t) = d, t \in I, t > t_{k-1}^b\}, t_k^b = \min\{t : f(t) = b, t \in I, t > t_k^d\}. \quad (1)$$

(We set the minimum over an empty subset of I to be t_1).

This sequence is obviously finite, if the interval I is compact.

Definition 2.8. A *winding of f around the interval* $[b, d]$ is a pair $\{t_k^b, t_k^d, k \geqslant 1\}$ in the sequence thus generated.

(See Fig. 4, right display, for an example of a function with two windings around an interval.) We remark, that the continuity of f implies that the set of windings in any compact subset of I is finite.

One has the following

Lemma 2.9. The number of the bars in $\mathbf{PH}_0(f)$ straddling the interval $[b, d]$, or, equivalently, the $\mathbf{PH}_0$ content of the quadrant $Q(b, d)$, is equal to the number of windings of f around $[b, d]$, plus one.

Proof. On one hand, to each winding (t_k^b, t_k^d) one can associate the unique connected component of $I_b := \{f \leqslant b\}$ having $t_k^b > t_0$ as its left end. On the other hand, to any component of $[t_-, t_+] \subset I_b$ outside the one having t_0 in its closure, one can associate the winding with the left end t_-. This is manifestly a bijection between the windings and all but the leftmost components of I_b. $\square$

This definition of windings is, of course, identical to the well-known construction of "upcrossings" used in the standard proof of Doob's martingale convergence theorem.

3 Persistence Diagram for Brownian Walks

Our focus in this note on the properties of persistent diagrams for Brownian motions, the simplest Markov processes with continuous trajectories.

Specifically, we consider Brownian motion with a drift, that is

$$f(t) = \mathbf{B}_o(t) + mt, \quad (2)$$

where $\mathbf{B}$ is the standard Brownian motion starting at 0, on the ray $(0, \infty)$. In this and following sections we will be concerned with the persistence diagrams of a typical trajectory.

---

[2]The term is clarified on Figure 4.

We will be dealing with the intervals $0 < b < d$. For general intervals (and Brownian motions with drift starting at 0) we can condition on the process to ever reach the level b; restarting the process at this stopping time leads to immediate generalizations to arbitrary b.

3.1 Persistence Diagrams as Point Processes

Almost surely, there is unique bar $(b_\infty, d_\infty) = (\inf_{t>0} f(t), \infty)$.

Other bars are finite, forming a random persistence diagram $\mathbf{PH}_0$. Our interpretation of this persistence diagram is to view it as a *point process*, a random measure consisting of sum of $\delta_{(b,d)}$ over all bars (b, d).

Remark 3.1. Almost sure tameness of the this point process, that is the fact that the content of any NW-quadrant with its apex above the diagonal is almost surely finite, will be proven below. However, this can be seen from the mentioned above general result in [BW16], that the persistence diagram of any Hölder function on a compact metric space is tame, and the well-known fact that the trajectories of Brownian motions (with any smooth drift) are almost surely $1/2$-Hölder.

3.1.1 Process decomposition

Following the definitions of Section 2.4.1, we consider the alternating sequences of b- and d-times.

These times are, clearly, stopping times with respect to the filtration $\{\mathcal{F}_\tau\}_{\tau \geqslant 0}$ generated by the Brownian motion. Hence, the intervals $t_k^b - t_{k-1}^b, k = 1, 2, \ldots$ between the winding events are independent (thanks to the standard fact that for any stopping time τ, the process $\mathbf{B}_{t-\tau}^m - \mathbf{B}^m(\tau)$ is independent of $\mathcal{F}_\tau$).

As an immediate consequence, we obtain

Proposition 3.2. Consider the standard Brownian motion with constant drift $m > 0$. Then the total number of bars in $\mathbf{PH}_0$-persistence diagram straddling an interval $[b, d], b > 0$ (or, equivalently, the $\mathbf{PH}_0$-content of the quadrant $Q(b, d) = \{(x, y) : x < b < d < y\}$) is geometrically distributed with parameter $p = p_m(\Delta) = 1 - \exp(-2m\Delta)$, where $\Delta = d - b$.

Proof. Indeed, the total number of windings around $[x, y]$ is the total number of transitions from d to b. The probability that the trajectory of the standard Brownian motion with drift m starting at y reaches x before escaping to infinity is $\exp(-2m(y - x))$. Strong Markov property implies that the escapes to infinity after each hitting of y are independent, proving the conclusion. □

3.1.2 Invariance properties

Let us state explicitly two invariances.

First, the standard rescaling properties and the invariance of the persistence diagrams under the reparameterization of the argument imply that the same result - and all results on the distributional properties of the persistence diagrams and chiralities for the standard Brownian motion and drift m remain valid for the Brownian motion with quadratic variation σ^2 and drift $m\sigma^2$.

Further, for $b \geqslant 0$ the distribution of the $\mathbf{PH}_0$ content of $Q(b, d)$ depends only on the length Δ of the bar $[b, d]$. This is again an immediate property of the Markov property of the process. Similarly, the joint distributions of the windings around several intervals remains invariant under the simultaneous shifts of these intervals as long they all remain within the positive half-line.

3.1.3 Other Brownian motions

We should notice that reflection principle allows one to derive quite easily the expected number of bars straddling a particular interval for other Brownian motions. For example, for the standard Brownian bridge on the unit interval, and an interval of $[b, d], 0 < b < d$, it is equal to

$$\mathbf{PH}_0(Q(b, d)) = \sum_{n \geqslant 1} \exp(-2(n\Delta + b)^2),$$

7

(wit $\Delta = d - b$). For the intervals containing the origin, the expression is a bit more cumbersome, but still easy to derive.

3.1.4 Intensity measure for $\mathbf{PH}_0$ for Brownian motion with drift

Corollary 3.2 immediately leads to an explicit expression for the intensity measure $\mathbb{E}\mathbf{PH}_0$ for the point process of zero-dimensional persistence from standard Brownian motion with drift:

Proposition 3.3. The intensity measure of $\mathbf{PH}_0$ on $0 < b < d$ is given by the density

$$\mu = \frac{4m^2 e^{-2m\Delta}(1 + e^{-2m\Delta})}{(1 - e^{-2m\Delta})^3}, \tag{3}$$

where $\Delta = d - b$.

In particular, near the diagonal the density explodes as $1/m(d-b)^3$.

Proof. By 3.2, the expected $\mathbf{PH}_0$ content of the quadrants $Q_{b,d}$ is

$$q_{b,d} := \mathbb{E}\mathbf{PH}_0(Q_{b,d}) = p_m(\Delta)/q_m(\Delta) = \frac{1}{e^{2m\Delta} - 1}.$$

The density of the intensity measure is, clearly, given by

$$\mu(b, d) = -\frac{\partial^2 q_{b,d}}{\partial b \partial d},$$

which gives the stated result. $\qquad\square$

The fact that the density grows as $(m\Delta)^{-3}$ as $\Delta \to 0$ is consistent with the scaling laws of [BW16], and holds for other Brownian motions (for Brownian bridges this follows from the formula in section 3.1.3), see also [Per20].

4 Chiralities in Brownian trajectories

Fix an interval $B = [b, d]$. Corollary 3.2 describes the distribution of the total number of the bars $\mathbf{PH}_0(Q(b, d))$ of the Brownian motion with drift $\mathbf{B}^m$ straddling B. In this section, we address the question: *how many of them will be* ⋁*, and how many* ⋀*?*

4.1 Deconstructing into the bars

Define, as in (1) the stopping times $t_l^b, t_l^d, l = 1, \ldots$. A trajectory straddles the interval exactly k times if $(k+1)$ is the first index for which $t_{k+1}^b = \infty$. In this case there are k pairs of times, $(t_l^d < t_l^b)$ forming an increasing sequence collectively: $t_0^b < t_1^d < t_1^b < t_2^d < t_2^b < \ldots < t_k^d < t_k^b < t_{k+1}^d$ (it is convenient to add the boundary points to the sequence).

We will use these times to chop the trajectory $\mathbf{B}^m$ into $2k$ pieces, $f_l^\downarrow, f_l^\uparrow$ by setting

$$f_l^\downarrow(t) := \mathbf{B}^m(t_l^d + t), 0 \leqslant t \leqslant t_l^b - t_l^d, 1 \leqslant l \leqslant k$$

being the fragments of the trajectory traveling from d to b, and

$$f_l^\uparrow(t) := \mathbf{B}^m(t_l^b + t), 0 \leqslant t \leqslant t_{l+1}^d - t_l^b, 1 \leqslant l \leqslant k$$

their upward counterparts.

Proposition 4.1. Conditioned on the event that the trajectory straddles $[b, d]$ $m \geqslant k$ times, the random processes $f_l^\uparrow, f_l^\downarrow, l = 1, \ldots, k$ are independent. Moreover the processes $f_l^\uparrow, l = 1, \ldots, k$ are identically distributed, as the processes $f_l^\downarrow, l = 1, \ldots, k$ are.

Proof. This follows, again, from the strong Markov property of the Brownian motion with drift. □

We remark that the distributions of the downward segments $f^\downarrow$ are those of the standard (drift-less) Brownian motion starting at d and stopped once it reaches b, while the distributions of $f^\uparrow$ are those of the Brownian motion with drift m started at b and stopped when it reaches d.

The *set* of the maximal values of the processes $f_l^\downarrow, l = 1, \ldots, k$ coincides with the set of the right ends of the bars straddling the interval $[b, d]$, while the minimal values of the processes $f_l^\uparrow, l = 1, \ldots, k$ form the set of their left ends. Of course, d_l is not necessarily coupled to b_l (although it could be).

4.2 Symmetries of the process

Almost surely, these local maximal and minimal values are distinct. When this is the case, we form two *rank* processes, $\sigma_+(l)$ and $\sigma_-(l)$, where $1 \leqslant l \leqslant k$, σ_+ is the rank of d_l among all $\{d_1, \ldots, d_k\}$ and $\sigma_-(l)$ is the rank of b_l in $\{b_1, \ldots, b_k\}$. The sequences σ_+ and σ_- are permutations of $\{1, \ldots, k\}$.

The Proposition 4.1 implies

Proposition 4.2. Conditioned on the $\mathbf{PH}_o(Q_{a,b}) = k$ the permutations σ_+ and σ_- are independent, uniformly distributed on the symmetric group $\mathfrak{S}_k$.

Proof. Both statements follow immediately from the mutual independences of $f_l^\uparrow, f_l^\downarrow$, $l = 1, \ldots, k$. □

Which of the local minima and maxima $b_\cdot, d_\cdot$ are coupled, depends only on the permutations σ_+, σ_-. The left display of Fig. 3 shows a configuration of maxima and minima of the processes $f_\cdot^\uparrow, f_\cdot^\downarrow$.

We will augment the sequence σ_+ with $(k + 1)$ on its right, and the sequence σ_- with 0 on its left.

For a maximum of rank $r, 1 \leqslant r \leqslant k$ at position l we call the rightmost index $m < l$ for which $\sigma_+(m) > \sigma_+(k)$ its *left wall*. Note that the left wall does not exist if (and only if) $\sigma_+(l)$ is a *record* in the sequences σ_+ (that is it beats all the earlier elements). Thus $\sigma_+(1)$ is always a record.

Similarly we define the *right wall* of $\sigma_+(l) = r$, as the leftmost index $n > l$ for which $\sigma_+(n) > r$. Note that thanks to our convention that $\sigma_+(k + 1) = k + 1$, the right wall always exists.

We say that the maximum of rank r is walled by ranks $s, t > r$ if s and t are ranks of the left and right walls of the rank r maximum (in particular, implying that r is not a record in the permutation σ_+). In this situation, we denote the positions of the elements s, r, t as m, l, n.

Summarizing, for a two-side walled (i.e., non-record) element r,

$$\sigma_+(m) = s > \sigma_+(l) = r < \sigma_+(n) = t, m < l < n \text{ and } \sigma_+(u) < r \text{ for } m < u < l \text{ or } l < u < n. \quad (4)$$

If element r at position l is walled by s, t at positions m, n we will form its *left* and *right* wells, defined as

$$\mathtt{LV}(r) := \min_{m \leqslant u < l} \sigma_-(u); \mathtt{RV}(r) := \min_{l \leqslant v < n} \sigma_-(v). \quad (5)$$

Lemma 4.3. If element r at position l is walled by s, t at positions m, n then it is coupled with one of its wells: with the left well if $\mathtt{LV}(r) > \mathtt{RV}(r)$ and the right well otherwise.

Proof. The components merging at r are born, obviously, at its wells. The choice of the well to couple with follows from the Elder Rule. □

Thus, if the left well is deeper, the bar corresponding to r couples to the right (i.e. is an **N**); if the right one is deeper, it is an **И**.

Conforming with the intuition, if the element is a record in σ_+, its left well at 0 is the deeper one, and the element is coupled to an index on its right, that is, it is a **N**.
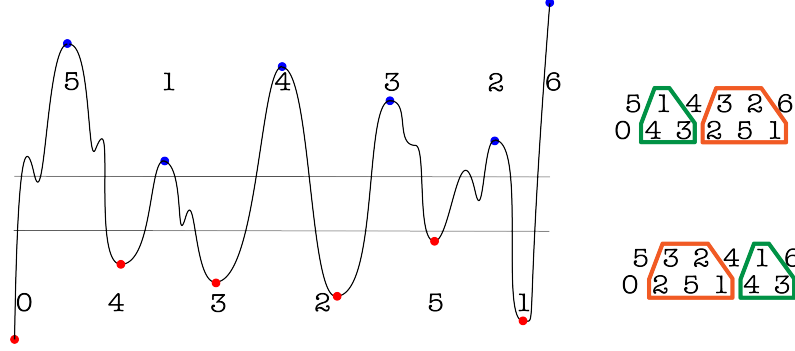
Figure 3: *Left display:* There are 5 bars straddling I in this picture. The permutations of maxima and minima are, correspondingly, $\sigma_+ = (514326)$ and $\sigma_- = (043251)$. The walls for $r = \sigma_+(3) = 4$ are $s = 5$ and $t = 6$, at the positions 1 and 6 respectively, the left well is 3, the right well is 1 and $r = 4$ is coupled with the $\sigma_-(2) = 3$, thus a И. *Right display:* swapping the segments.

4.3 More N's than И's

Finally we can deduce the excess of the expected numbers of N's over И's in a Brownian motion with drift.

Theorem 4.4. Conditioned on the number k of bars straddling an interval I, in a trajectory of $\mathbf{B}^m$, $m > 0$, the expected excess of the numbers of N's over И's is equal to the expected number of records in a random k-permutation, that is k-th harmonic sum,

$$\mathbb{E}(|\mathsf{N}| - |\mathsf{И}|) = H_k = 1 + 1/2 + \ldots + 1/k. \tag{6}$$

Proof. The couplings of the elements of σ_+ to the elements of σ_- depend only on those permutations, so we can focus on them.

Fix a rank r. Partition the set of pairs $\Sigma := (\sigma_+, \sigma_-)$ into Σ_*, corresponding to permutations σ_+ for which r is a record, and

$$\amalg_{s,t>r,s\neq t}\Sigma_{s,t}$$

of permutations for which s, t are respectively the left and the right walls for r.

One can define an involution on each of the collections $\Sigma_{s,t}$. If the position of s, r, t are, respectively, $m < l < n$, then swap the segments

$$(\sigma_+(m+1), \ldots \sigma_+(l-1)) \text{ and } (\sigma_+(l+1), \ldots \sigma_+(n-1))$$

(keeping $\sigma_+(l) = r$ between them), and swap the segments

$$(\sigma_-(m), \ldots \sigma_-(l-1)) \text{ and } (\sigma_-(l), \ldots \sigma_+(n-1))$$

(right display on Fig.3 shows this swap for a configuration in $\Sigma_{5,6}$ for $r = 4$).

Under such a swap, the chirality of the coupling changes sign (as the wells swap). This shows that the contributions from the partitions $\Sigma_{s,t}$ to the difference between the numbers of N's and И's coupled to r over these subsets vanishes.

Thus only the records contribute. The probability that r is a record is $1/(k - r + 1)$ (or, see, e.g. [Pit06]), and the result follows. □

Corollary 4.5. For Brownian motion with drift m, the expected excess $\mathbb{E}(|\mathsf{N}| - |\mathsf{И}|)$ in the bars straddling an interval of length Δ is

$$-\log(1 - \exp(-2m\Delta)).$$

Proof. The sum (6) gives the excess conditioned on the number of straddling bars. Summing these harmonic sums with weights pq^k, where $q = \exp(-2m\Delta)$, the probability given by Corollary 3.2, gives the desired result, after switching order of summation in the (obviously) converging series. $\square$

In particular, the expected excess of N over $и$ grows logarithmically (as $|\log(2m\Delta)|$) for small $m\Delta$; consequently the fraction of the excess among all bars straddling a short interval, disappears, as the length of the interval decreases to 0.

5 Automata and Correlation Functions for PH_0.

In this section we will deploy somewhat more involved techniques to understand the correlations between the windings around different intervals (which are essential to understanding the correlation functions of the PH_0 point process). This relies on the constructions of finite state automata translating continuous functions into symbolic trajectories. For continuous trajectories given by strong Markov processes, thus generated symbolic trajectories become Markov processes, and the questions about the joint distributions of windings around different intervals become standard questions about numbers of specific transitions in realizations of these Markov chains. We will use these ideas to derive some second-order characteristics of the PH_0 point processes.

Recall that any point process is characterized by its *factorial moments*, which in our tame setting can be characterized by their densities, i.e. *correlation functions* [Rue99].

The goal of this section is to outline a formalism to deduce those correlation function, and to compute a second order correlation function μ_2 in some situations.

5.1 Windings and Symbolic Trajectories

To study windings of a function around an interval, or more generally, the interactions of the function with one or more intervals, it is convenient to encode them with a scheme using finite automata. We will use an intermediate construction that will turn a continuous function of time into a symbolic trajectory.

5.1.1 Coverings and Lifts

Consider an open interval U in the real line, perhaps semi-infinite, and its finite covering by open intervals (or half-rays)

$$U_{\alpha} = (h^-(\alpha), h^+(\alpha)), \alpha \in A.$$

We will denote by $\mathbf{BA}$ the collection of *of finite* ends of these intervals:

$$\mathbf{BA} := \{\alpha^{\pm} : -\infty < h^{\pm}(\alpha) < \infty\}.$$

Fix a map τ from this set $\mathbf{BA}$ of boundaries to the intervals $\mathbf{A}$, subject to the condition that each finite end is mapped into a (necessarily, different) interval which contains it:

$$h^{\pm}(\alpha) \in U_{\tau(\alpha^{\pm})}.$$

Given these conditions, a continuous function $f : [t_s, t_f) \to U$, can be lifted unambiguously to the covering $\{U_{\alpha}\}_{\alpha \in A}$, if one assigns the initial covering interval α_s (such that $f(t_s) \in U_{\alpha_s}$). Such a lifting is a countable partition of the interval $[t_s, t_f) = \amalg_k[t_k, t_{k+1})$, and a sequence of intervals α_k defined inductively as follows:

- One initializes with $t_0 = t_s, \alpha_0 = \alpha_s$;
- Throughout, for $t_k \leqslant t < t_{k+1}$, one has $f(t) \in \alpha_k$;
- The switch times t_{k+1} are defined as $\inf\{t > t_k | f(t) \notin U_{\alpha_k}\}$, and

- The next interval is, naturally, given by the transition map τ: $\alpha_{k+1} = s(\alpha_k^\pm)$ ($\pm$ depending on which end of α_k the trajectory hits at t_{k+1}).

In other words, the k-sojourn lasts while the trajectory stays within the corresponding interval; as soon as it hits a boundary, it instantaneously jumps to whichever interval the switch map s sends it, where it continues to track the function f.

Remark 5.1. This construction of lifting a continuous trajectory can be easily generalized to, say, coverings of manifolds by the open patches with well-behaving boundaries: one needs to partition those boundaries into disjoint subsets such that each of them is contained in a patch; the switch map would then send the trajectory hitting the boundary at a component to one of the patches containing that component. This construction is an alternative way to define the diffusions on cellular complexes, a notoriously messy problem (compare [Nye20]).

5.1.2 Finite State Automata

Given the data described in Section 5.1.1 we can form finite state automata whose states will correspond to the finite height endpoints of intervals U_α. A bit counterintuitively, the state corresponding to $\alpha^\pm$ will be located in the interval $s(\alpha^\pm)$: at the exit point of the switch. Once in that interval, the process exits through one of the ends to the next interval and so on.

In terms of the (directed) graph description of the automaton, each state will have at most two potential out-arrows.

We will fix an interval (h^α, h^ω) containing all the intervals $U_\alpha, \alpha \in A$, referring to it as the *span*.

We add the absorbing state ω and the initial state, α corresponding to the endpoints of the span. The initial state α has no in-, and the absorbing state ω, no out-arrows. esponding automaton).

5.1.3 Symbolic trajectories

Fix a finite automaton described in 5.1.2 with the span (h^α, h^ω).

Consider a continuous function f *crossing the span of the automaton*, and such that on its interval of the definition $I = (t_0, t_1)$ contains the lifespan: such a function starts below the span of the automaton, and ends above it (i.e. $f(t_0+) < h^\alpha; f(t_1-) > h^\omega$).

Each function crossing the span defines a finite *symbolic trajectory*: a sequence of states, starting with α, ending with ω and transitions between finite states in between. These transition track the exits of the lift of f to the intervals U_α, as described in Section 5.1.1.
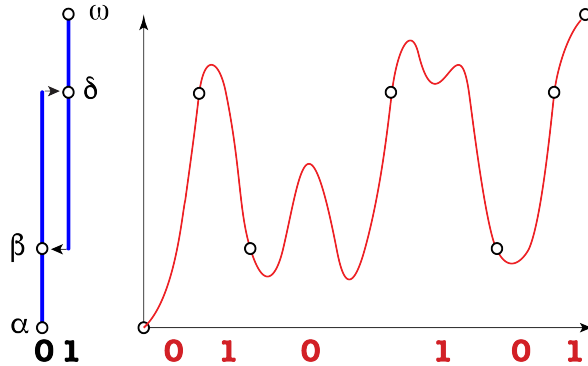

Figure 4: The finite automaton justifies the term *windings*: the number times a function straddles an interval is equal to the number of times its lift cycles around. The symbolic trajectory corresponding to the function shown above is $\alpha \to \delta \to \beta \to \delta \to \beta \to \delta \to \omega$. There are two transitions $\delta \to \beta$, and two windings around the interval $[b, d]$.

The resulting sequence of states and arrows is called *symbolic trajectory* corresponding to the function f.

The following is immediate:

Lemma 5.2. The symbolic trajectory of a continuous function crossing the span of a finite automaton is finite.

$\square$

5.1.4 Automata and windings around an interval

The simplest useful finite automaton of this kind described above (shown on the left display of Fig. 4) has the starting state α at level h^α; two finite states β and δ with levels $b := h^\beta$ and $d := h^\delta$, and the absorbing state ω with level h^ω. The transitions are $\alpha \to \delta, \delta \to \beta, \delta \to \omega$, and $\beta \to \delta$.

Again, the following is immediate:

Proposition 5.3. For a continuous function f crossing the span of this automaton, the number of $\delta \to \beta$ transitions in the corresponding symbolic trajectory is one less than to the number of bars straddling the interval (b, d) in the persistence diagram $\mathbf{PH}_0$ of f.

(Remark: The one extra bar is the "infinite" one, spanning the $\inf f$ and $\sup f$.)

5.2 Lifts of Brownian Paths

The formalism of lifts of trajectories to the automata described in Section 5.1.2 is well suited to address the windings of Brownian motions around intervals. Namely, in our model situation, - where the lifted function is a continuous trajectory of strong Markov process, - the symbolic trajectory becomes a Markov process itself.

5.2.1 Transition Probabilities for Brownian Motion with Drift

It is easy to compute the transition probabilities in the situation of Brownian motion with drift.

Let's introduce the character defined by

$$\mathbf{e}(s) = \exp(-2ms). \tag{7}$$

Evaluated on the Brownian motion with drift m it becomes a martingale: hence its ubiquitous presence in the formulae for various exit probabilities.

In particular, if the Brownian motion with drift $\mathbf{B}^m$ starts at s, the probability that it exits the interval $[s - \Delta_l, s + \Delta_r]$ at the right (left) end is, respectively,

$$p(\Delta_l, \Delta_r) = \frac{1 - \mathbf{e}(\Delta_l)}{1 - \mathbf{e}(\Delta_l + \Delta_r)}; q(\Delta_l, \Delta_r) = \frac{\mathbf{e}(\Delta_l) - \mathbf{e}(\Delta_l + \Delta_r)}{1 - \mathbf{e}(\Delta_l + \Delta_r)}. \tag{8}$$

5.2.2 Automata and Sums over Paths

Consider a general finite state automaton, with edges marked by weights, - monomials in some collection of formal variables $z_1, z_2, \ldots, z_m$ with real coefficients. We will denote the weight corresponding to the edge (β, β') as $w(\beta, \beta')$. For a finite path π in the automaton, we define the weight $w(\pi)$ of the path to be the product of all the weights of the transitions.

For a pair of states β, β' in the automaton, consider the formal power series (with nonnegative real coefficients) formed by a sum over all paths going from β to β':

$$F_{\beta, \beta'}(z) = \sum_{\beta \xrightarrow{\pi} \beta'} w(\pi), \tag{9}$$

The usual rules apply, for example, the the most important being the *last edge decomposition*,

$$F_{\beta,\beta''} = \delta_{\beta,\beta''} + \sum_{\beta \to \beta'} F_{\beta,\beta'} w(\beta', \beta''). \qquad (10)$$

This leads, as usual, to rational generating functions for the enumerations of weights over trajectories in the automaton.

5.2.3 Distributions of the Windings around Intervals

In the situation where the edge weights are transition probabilities multiplied by monomials in formal variables, the sums over paths (9) become partition functions of the various descriptors of the trajectories. If the trajectory spends only finite time in the span of the automaton, then the sum (9) has well-defined terms, whose nonnegative coefficients sum up to the unity (and converging in the unit polydisk $\{|z_l| \leqslant 1, l = 1, \ldots, m\}$), and the standard computations render each of $F_{\beta,\beta'}(z)$ a rational function of the variables $z, \ldots, z_m$.

As an example, if one associates to the automaton shown on Fig. 4 transition probabilities

$$p_{\delta,\beta} = \mathbf{e}(\Delta); p_{\delta,\omega} = 1 - \mathbf{e}(\Delta), \text{ where } \Delta = d - b; \text{ all others } = 1,$$

and sets the weight on (δ, β) equal to $p_{\delta,\beta} z$, then

$$F_{\alpha,\omega} = \sum_k p_k z^k,$$

where p_k is the probability that a trajectory of the Brownian motion with drift has k finite bars spanning $[b, d]$. Applying (10) we obtain

$$F_{\alpha,\alpha} = 1; F_{\alpha,\beta} = F_{\alpha,\alpha} + z\mathbf{e}(\Delta)F_{\alpha,\delta}; F_{\alpha,\delta} = F_{\alpha,\beta}; F_{\alpha,\omega} = (1 - \mathbf{e}(\Delta))F_{\alpha,\delta},$$

resolving which leads, unsurprisingly, to

$$F_{\alpha,\omega} = \frac{1 - \mathbf{e}(\Delta)}{1 - z\mathbf{e}(\Delta)}.$$

5.3 Higher Windings

Understanding joint windings around several intervals requires construction of somewhat more complicated automata.

One useful construction is that of the *product* of two automata.

5.3.1 Product of Automata

Consider a pair of automata of the kind described in Section 5.1.2: two collections of intervals, $\mathbf{A}_1, \mathbf{A}_2$, and the corresponding switch maps s_1, s_2 defined on the endpoints $\mathbf{BA}_1, \mathbf{BA}_2$. Assume, for simplicity, that the heights of the finite ends of the intervals in different collections are all disjoint.

Then one can form a new automaton, *the product of the original ones*. The intervals of the product automaton would be the formed by the pairs $U_{\alpha_1,\alpha_2} = U_{\alpha_1} \cap U_{\alpha_2}$ for which this overlap is nonempty. Note that heights of different intervals of this kind can coincide: those intervals are enumerated not by their heights, but by their indices (α_1, α_2).

Each end of the interval U_{α_1,α_2} comes unambiguously from one if the collections, $\mathbf{A}_1$ or $\mathbf{A}_2$. This would allow to define the switch map on the product of the original automata. If, say, the upper end of the interval $(\alpha_1, \alpha_2)^+$ stems from the first automaton (that is the upper height of α_1 is less than that of α_2), then the switch map s of the sends $(\alpha_1, \alpha_2)^+$ to $(\tau_1(\alpha_1^+), \alpha_2)$, and so on.

The Fig. 5 illustrates the products of two elementary automata for differently overlapping intervals.

We remark that in the case when the intervals are not nested, one of the branches is a *"garden of Eden"*, - a trajectory starting outside it will never enter it. Hence one can ignore it; this reduces the effective number of states to six.

5.3.2 Windings around a Pair of Intervals

Let $I_1 = [b_1, d_1]$, $I_2 = [b_2, d_2]$ are two intervals. We will be looking into the joint distribution of the numbers b_1, b_2 of bars spanning these intervals.

Assume that the intervals are not nested, so that $b_1 < b_2$; $d_1 < d_2$. In this case, the relevant automaton is shown on the left display of Fig. 5. Any continuous function that starts below b_1 and ends above d_2 generates a symbolic trajectory. The following is immediate:

Lemma 5.4. The total number of windings around the interval I_1 is counted by the number of transitions into the state β_1, (i.e. either $\delta_1 \to \beta_1$ or $\beta_2 \to \beta_1$), while the number of windings around I_2 is given by the number of transitions $\delta_2 \to \beta_2$.

Proof. Indeed, lifts of the trajectory to either of the factors coincides with the projection to that factor of the lift of the trajectory into the product, and the transitions in either lift happen at the same positions. □
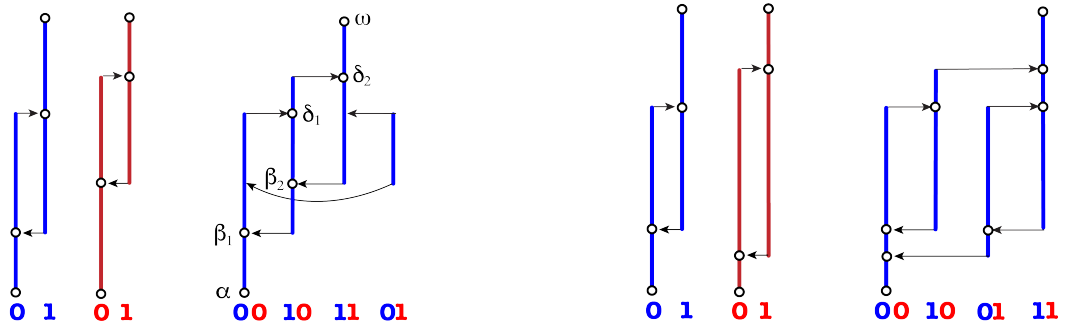

Figure 5: Windings around a pair of intervals.

Let us introduce the generating functions for the numbers of windings around the intervals I_1, I_2 for the paths starting at α and ending at a state v:

$$H_v(x, y) := F_{\alpha, v}(x, y) = \sum_{\alpha \xrightarrow{\pi} v} w(\pi) = \sum_{\alpha \xrightarrow{\pi} v} x^{k(\pi)} y^{l(\pi)} \mathbb{P}_\pi : \qquad (11)$$

here $k(\pi), l(\pi)$ are numbers of windings of the path π around the intervals I_1, I_2, respectively, $\mathbb{P}(\pi)$ is the probability of the realization of the symbolic path π, and the summation is over all paths starting in α and ending in v.

We will conduct the computations only for the non-nested intervals, depicted on the left display of Fig. 5. The case of nested intervals, as well as more correlation functions of higher orders will be addressed elsewhere.

Using the last edge decomposition (10) for the chain on the Figure 5 we arrive at

$$
\begin{aligned}
H_\alpha &= 1 \\
H_{\beta_1} &= H_\alpha + x(q_{\delta_1} H_{\delta_1} + q_{\beta_2} H_{\beta_2}) \\
H_{\beta_2} &= y q_{\delta_2} H_{\delta_2} \\
H_{\delta_1} &= H_{\beta_1} \\
H_{\delta_2} &= p_{\delta_1} H_{\delta_1} + p_{\beta_2} H_{\beta_2} \\
H_\omega &= p_{\delta_2} H_{\delta_2}.
\end{aligned}
\qquad (12)
$$

Here we use the shortcuts p_v, q_v to denote the probabilities to exit from the top or bottom end, respectively, from the interval containing the state v, conditioned on the Brownian trajectory starting at v.

15

Now we are ready to compute the joint distributions of the numbers of windings around the intervals I_1, I_2.

The formal power series (11) is, clearly, the generating function for the numbers of windings around the intervals I_1, I_2.

Proposition 5.5. The generating function $H := H_\omega$ is ginven by

$$H(x, y) = -\frac{(\mathbf{b}_1 - \mathbf{d}_1)(\mathbf{b}_2 - \mathbf{d}_2)}{\mathbf{d}_1(\mathbf{b}_1 - \mathbf{d}_2) + \mathbf{b}_1(\mathbf{d}_2 - \mathbf{d}_1)x + \mathbf{d}_1(\mathbf{b}_2 - \mathbf{b}_1)y + \mathbf{b}_1(\mathbf{d}_1 - \mathbf{b}_2)xy}. \tag{13}$$

where we denote $\mathbf{b}_k = \mathbf{e}(-b_k); \mathbf{d}_k = \mathbf{e}(-b_k), k = 1, 2$.

Proof. The work is shown in the Appendix B. $\qquad\square$

5.4 Correlation density

Using equation (13), we can easily find the covariance of the numbers $\mathbf{w}_1, \mathbf{w}_2$ of windings around the intervals I_1, I_2: indeed,

$$\text{Cov}(I_1, I_2) = \mathbb{E}\mathbf{w}_1\mathbf{w}_2 - \mathbb{E}\mathbf{w}_1\mathbb{E}\mathbf{w}_2 = \left. \frac{\partial^2 \log H}{\partial x \partial y} \right|_{x=y=1}. \tag{14}$$

Performing the calculations results in

$$\text{Cov}(I_1, I_2) = \frac{\exp(-2m(d_2 - b_1))}{(\exp(-2m(d_1 - b_1)) - 1)(\exp(-2m(d_2 - b_2)) - 1)}. \tag{15}$$

Recall that any point process is characterized by its *factorial moments*, which in our tame setting can be characterized by their densities, i.e. *correlation functions*.

Proposition 5.6. The second moment density function of $\mathbf{PH}_0$ for the non-nested intervals (i.e. for $b_1 < b_2, d_1 < d_2$) is given by

$$\mu_2(b_1, d_1, b_2, d_2) = \frac{\partial^4 \text{Cov}(b_1, d_1, b_2, d_2)}{\partial b_1 \partial d_1 \partial b_2 \partial d_2}. \tag{16}$$

Proof. The covariance (15) can be interpreted as a function of the endpoints (b_1, d_1, b_2, d_2) as follows:

$$\text{Cov}(b_1, d_1, b_2, d_2) = \mathbb{E}\left(\int \phi_{b_1, d_1} d\mathbf{PH}_0 \int \phi_{b_2, d_2} d\mathbf{PH}_0 \right) - \mathbb{E}\left(\int \phi_{b_1, d_1} d\mathbf{PH}_0 \right) \mathbb{E}\left(\int \phi_{b_2, d_2} d\mathbf{PH}_0 \right), \tag{17}$$

where $\phi_{b,d}$ is the indicator function of the NW coordinate quadrant with the apex at (b, d).

The bilinearity of (17) with respect to ϕ's implies that the covariance of the $\mathbf{PH}_0$-contents of the small coordinate aligned squares centered at (b_1, d_1) and (b_2, d_2) can be obtained by taking first order differences with respect to four variables (b_1, d_1, b_2, d_2) of (15).

Dividing by the areas of these squares and passing to the limit, we obtain (17) $\qquad\square$

Lengthy but elementary computations yield

Theorem 5.7. For non-nested intervals $b_1 < b_2; d_1 < d_2$, the 2-point correlation function for the point process $\mathbf{PH}_0$ of 0-dimensional persistence for the standard Brownian motion with drift m is given by

$$\mu_2(b_1, d_1; b_2, d_2) = \frac{64m^4 \exp(2b_1 + b_2 + 2d_1 + d_2)}{(\exp(2mb_1) - \exp(2md_1))^3 (\exp(2mb_2) - \exp(2md_2))^3}. \tag{18}$$

The expression (18) can be rendered as

$$\mu_2(b_1, d_1; b_2, d_2) = \frac{64m^4 e(\Delta_1) e(\Delta_2) e(\Delta)}{(1 - e(\Delta_1))^3 (1 - e(\Delta_2))^3} = \mu(\Delta_1)\mu(\Delta_2)\frac{4e(\Delta)}{(1 + e(\Delta_1))(1 + e(\Delta_2))};$$

here μ is the intensity density of the point process $\mathbf{PH}_0$ (see the Proposition 3.3), Δ_1, Δ_2 are the lengths of the intervals I_1, I_2 and $\Delta = d_2 - b_1$ the span of the union of the intervals.

Remark 5.8. The Theorem 5.7 implies, in particular, that the *pair correlation function* is positive, at least for the non-nested intervals.

Another implication is the exponential decay of correlation function. By standard arguments, this implies that the expected number and the variance of the number of bars in the region

$$Q_c(L) := \{0 < b < d < L, d - b > c > 0\}$$

grow asymptotically linearly with L, and that the number itself is asymptotically normal, as $L \to \infty$.

6 Concluding remarks

- As we mentioned above, it would be of interest to understand the overall structure of the point process, in particular to derive its higher factorial moments. The techniques of the products of elementary automata makes the problem computable.

- Similarly, the question of the *joint distribution* of the numbers of bars of either type, N and $И$, is of interest.

- We focused in this note on the special class of Brownian motions. One might expect some universality near the diagonal.

- An alternative approach to description of the persistence diagrams of trajectories of rather general $\mathbb{R}$-valued Markov random processes was addressed by J. Picard in [Pic08]: he generated a sequence of stopping times marking emergence of all bars of length above some threshold, and used its properties to derive quite general results on the resulting point process. This line of research was also pursued by D. Perez, see e.g. [Per20].

- As we mention in Appendix A, the behavior of the stack sizes in the algorithm for finding the barcodes of a time series seems nontrivial and interesting, with random *updown permutation* [Arn92] being the natural model for the algorithm input.

- Generalizing to higher-dimensional domains seems difficult but not entirely out of reach, at least for some classes of Gaussian processes, using the techniques on the bar decompositions of the products of random merge trees. This line of research will be pursued elsewhere.

Acknowledgements

7 Appendices

A. Finding the Bars

Finding the bars of a time series can be done quite efficiently.

Assume that a function is given as a time series, i.e. a list of its values $f(k), k = 1, \ldots, N$ at consecutive points $t_1 < t_2 < \ldots < t_N$ between which the function is assumed to be monotonic (say, linearly interpolating). Further we will assume that the function is generic.

We augment the series by setting $f(0)$ and $f(N+1)$ to respectively global minimum and maximum values: $f(0) < f(k) < f(N+1), k = 1, \ldots, N$.

The output is a list with entries having the structure $(d, t_d; b, t_b)$, where $(d < b)$ is a bar, and (t_d, t_b) are the locations of the corresponding (local) maxima and minima.

The algorithm maintains at all times two stacks, one with local minima, one with local maxima, ordered by critical value within each stack. Removals from the stacks happen in pairs and produce bars.

Algorithm 1 One pass bar algorithm for time series.

`Maxima, Minima ←` empty stacks   ▷ Initializing
`Direction ← +1`   ▷ Initially, the function increases
for $t = 1, \ldots, N$ **do**
    if $(f(t) - f(t-1)) \times$ `Direction` < 0 **then**   ▷ Direction changes, so either
        if `Direction` $= +1$ **then**
            `Maxima.Push(`$(t-1, f(t-1))$`)`   ▷ add the local maximum
        else
            `Minima.Push(`$(t-1, f(t-1))$`)`   ▷ ...or minimum to respective stack,
        end if
        `Direction = −Direction`   ▷ and record change of the direction.
    else
        if $($`Direction` $= +1 \&\& f(t) >$ `Maxima(0)`$)||($`Direction` $= −1 \&\& f(t) <$ `Minima(0)`$)$ **then**
            `Output(Maxima.Pop, Minima.Pop)`   ▷ If hitting a wall, output a bar.
        end if
    end if
end for

While the time execution of the algorithm is clearly linear in the length of the time series, an interesting question arises on the depth of stacks (memory) required for its execution. In the worst case (when the local maxima are descending and the local minima ascending) the depth required is linear as well. What about the average case?

We remark that one can adapt this algorithm to register arbitrarily complicated snake-like patterns in time series.

B. Merge Tree Monoid

Consider two path-connected spaces with continuous functions on them: $f : X \to \mathbb{R}, g : Y \to \mathbb{R}$. We define the Thom-Sebastiani sum of the functions f and g as the function $f \oplus g : X \times Y \to \mathbb{R}$ taking $f \oplus g : (x, y) \mapsto f(x) + g(y)$ (motivated by [ST71]).

One can prove

Proposition 7.1. The merge tree $T_{f \oplus g}$ of the Thom-Sabatini sum depends only on the merge trees T_f, T_g.

In particular, one can consider $\oplus$ as an operation on the merge trees (with roots placed at the infinite height).

Definition 7.2. The operation $\oplus$ turns the collection of merge trees bounded from below into a *commutative monoid*, with the unity given by the merge tree T_0 whose sole leaf is at height 0 (the merge tree of any positive definite quadratic function, analogously to Thom-Sabatini setup).

We note that the merge trees corresponding to $x^2 + r$, i.e. tree with a single leaf at the height r form an embedding of $\mathbb{R}$ into the merge trees monoid, and addition with them realizes the shift action of $\mathbb{R}$ on the monoid.

To recover $T_{f \oplus g}$ from the merge trees T_f, T_g, up to chiralities, it is enough to identify all bars (i.e. stems resulting from the recursive pruning of the tree, see Section 2.2.1) and their attachments. The collection of bars of $T_{f \oplus g}$ is easy to recover: they are in one-to-one correspondence to pairs of bars $[b_f, d_f], [b_g, d_g]$ of f and g, and have birth at $b_f + b_g$, and have the lifespan equal to the smaller of the lifespans of $d_f - b_f, d_g - b_g$.

Consider two merge trees shown on Fig. 6 together with their product; the right display shows some of the level sets of $f(x) + g(y)$, where f, g are the univariate functions whose merge trees correspond to the factors on the left.
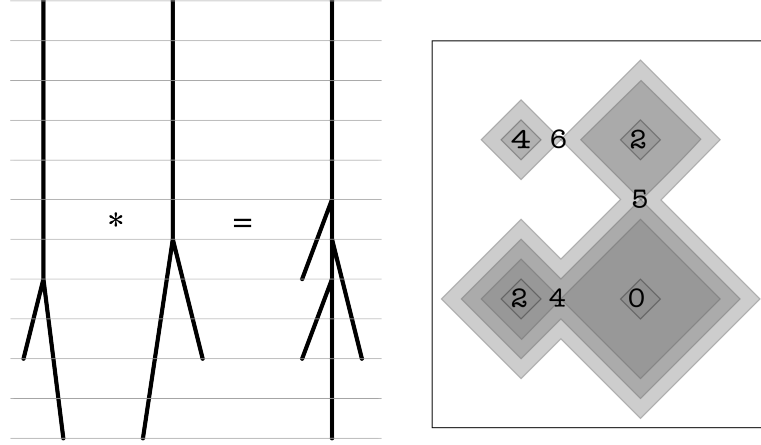

Figure 6: Product of merge trees.

The rules of the attachments, details of the construction, as well as applications to the study of persistence diagrams of (tilted) Brownian sheets will be presented in a forthcoming publication [Le22].

Another application of merge tree monoid structure could be the model of *corner percolation*, see [Pet08]. There one studies the structure of the level sets of functions $f(x) - g(y)$ where f, g are (univariate) random functions representing the trajectories of random excursions of equal height.

Appendix C: Generating functions for windings around two intervals

Solving for F_ω is routine, the result is

$$H(x, y) = pp_1/(1 - xq_1 - yqp_2 + xyq(q_1p_2 - p_1q_2)) \tag{19}$$

where we use shorthands (relying on (8))

$$p = p(d_2 - b_2, \infty); q = q(d_2 - b_2, \infty);$$
$$p_1 = p(d_1 - b_1, d_2 - d_1); q_1 = q(d_1 - b_1, d_2 - d_1);$$
$$p_2 = p(b_2 - b_1, d_2 - b_2); q_2 = q(b_2 - b_1, d_2 - b_2).$$

Using the expressions (8) for p's and q's we find that $H(x, y)$ is given by

$$-\frac{(\mathbf{e}(b_1) - \mathbf{e}(d_1))(\mathbf{e}(b_2) - \mathbf{e}(d_2))}{\mathbf{e}(d_1)(\mathbf{e}(b_1) - \mathbf{e}(d_2)) + \mathbf{e}(b_1)(\mathbf{e}(d_2) - \mathbf{e}(d_1))x + \mathbf{e}(d_1)(\mathbf{e}(b_2) - \mathbf{e}(b_1))y + \mathbf{e}(b_1)(\mathbf{e}(d_1) - \mathbf{e}(b_2))xy},$$

which simplifies to (13).

8 Declarations

8.1 Ethical Approval

Not applicable.

8.2 Funding

8.3 Availability of data and materials

Not applicable

References

[Arn92]   Vladimir I. Arnol'd. The calculus of snakes and the combinatorics of Bernoulli, Euler and Springer numbers of Coxeter groups. *Russian Mathematical Surveys*, 47(1):1, 1992.

[AT07]   Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York, 2007.

[AT11]   Robert J. Adler and Jonathan E. Taylor. *Topological complexity of smooth random functions: École d'Été de Probabilités de Saint-Flour XXXIX - 2009*. Number 2019 in Lecture notes in mathematics. Springer, Heidelberg ; New York, 2011. OCLC: ocn733988174.

[BA11]   Omer Bobrowski and Robert J. Adler. Distance Functions, Critical Points, and the Topology of Random Čech Complexes. *arXiv:1107.4775 [math]*, July 2011. arXiv: 1107.4775.

[BBI22]   Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, 2022.

[BG18]   Yuliy Baryshnikov and Robert Ghrist. Minimal Unimodal Decompositions on Trees. *arXiv:1806.09673 [math, stat]*, June 2018. arXiv: 1806.09673.

[BK18]   Omer Bobrowski and Matthew Kahle. Topology of random geometric complexes: a survey. *J. Appl. Comput. Topol.*, 1(3-4):331–364, 2018.

[BW16]   Yuliy Baryshnikov and Shmuel. Weinberger. Persistence jitter. *Preprint*, pages 1–14, 2016.

[Cur17]   J. Curry. The Fiber of the Persistence Map. *ArXiv e-prints*, June 2017.

[EH10]   Herbert Edelsbrunner and John L. Harer. *Computational topology*. American Mathematical Society, Providence, RI, 2010. An introduction.

[KM10]   Matthew Kahle and Elizabeth Meckes. Limit theorems for Betti numbers of random simplicial complexes. *arXiv:1009.4130 [math]*, September 2010. arXiv: 1009.4130.

[Le22]   Phuong Le. Merge trees and brownian motions. *preprint*, 2022.

[LG05]   Jean-François Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.

[LNR+12] Lucas Lacasa, Ángel M. Núñez, Édgar Roldán, Juan M. R. Parrondo, and Bartolo Luque. Time series irreversibility: a visibility graph approach. *The European Physical Journal B*, 85(6), June 2012. arXiv: 1108.1691.

[Nye20]   Tom M. W. Nye. Random walks and Brownian motion on cubical complexes. *Stochastic Processes and their Applications*, 130(4):2185–2199, April 2020.

[Oud15]   Steve Y. Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[Per20]   Daniel Perez. On c^0-persistent homology and trees. *arXiv preprint arXiv:2012.02634*, 2020.

[Pet08]   Gábor Pete. Corner percolation on $\mathbb{Z}^2$ and the square root of 17. *The Annals of Probability*, 36(5):1711–1747, September 2008.

[Pic08]   Jean Picard. A tree approach to p-variation and to integration. *The Annals of Probability*, 36(6), November 2008.

[Pit06]   Jim Pitman. *Combinatorial Stochastic Processes: Ecole d'Eté de Probabilités de Saint-Flour XXXII - 2002*. Springer, July 2006.

[Rue99]   David Ruelle. *Statistical mechanics: Rigorous results*. World Scientific, 1999.

[ST71]   M. Sebastiani and R. Thom. Un résultat sur la monodromie. *Inventiones mathematicae*, 13(1):90–96, March 1971.