TOPOLOGICAL INFERENCE FOR DEPENDENT STATIONARY RANDOM VARIABLES

SADOK KALLEL AND SANA LOUHICHI

Abstract. In this paper we extend results on reconstruction of probabilistic supports of random i.i.d variables to supports of dependent stationary random variables. All supports are assumed to be compact of positive reach. The main results involve the study of the convergence in the Hausdorff sense of a cloud of stationary correlated points to its support. A novel topological reconstruction result is stated, and a number of illustrative examples are presented. The example of the Möbius Markov chain on the circle is discussed in details with simulations.

1. Introduction

Given a sequence of stationary random variables of unknown common law and unknown compact support $\mathbb{M}$, uncovering topological properties of $\mathbb{M}$ based on the observed data can be very useful in practice. Data analysis in high-dimensional spaces with a probabilistic point of view was initiated in [18] where data was assumed to be drawn from sampling an i.i.d on or near a submanifold $\mathbb{M}$ of Euclidean space. Topological properties of $\mathbb{M}$ (homotopy type and homology) were deduced based on the random samples and the geometrical properties of $\mathbb{M}$. Several papers on probability and topological inference ensued, some taking a persistence homology approach by providing a confidence set for persistence diagrams corresponding to the Hausdorff distance of a sample from a distribution supported on $\mathbb{M}$ [8].

Topology intervenes in Probability through reconstruction results (see [1, 4, 5, 19] and references therein). This research direction is now recognized as part of "manifold learning". Given an n point-cloud $\mathbb{X}_n$ lying in a support $\mathbb{M}$, which is generally assumed to be a compact subspace of $\mathbb{R}^d$ for some $d > 0$, and given a certain probability distribution of these n points across $\mathbb{M}$, one can formulate practical conditions to reconstruct, up to homotopy or up to homology, this support $\mathbb{M}$. The two notions have great overlap. Reconstruction up to homotopy means recovering the homotopy type of $\mathbb{M}$. Reconstruction up to homology means determining, up to certain indeterminacy or up to certain degree, the homology groups of $\mathbb{M}$. Of course knowing homotopy type allows in principle to know the homology groups, but the converse is not in general possible. The advantage of using homology, or persisence homology, lies in the fact that it is combinatorially computable, and so gives concrete invariants of the support.

The goal of our work is to extend work of Nigoyi, Smale and Weinberger [18] to the case of stationary dependent data (not just i.i.d) and for $\mathbb{M}$ a compact space of "positive reach" or PR-set[1] (not just a submanifold). The interest in going beyond independence lies in the fact that many of the observations of everyday life are dependent, and independence is not sufficient to describe these phenomena. The study of the data support topologically and geometrically in this case can be very useful in directional statistics for example, where the observations are often correlated. This can help get information on animal migration paths or wind directions for instance. Modeling by a Markov chain on an unknown compact manifold, with or without boundary, makes it possible to study such models. Other illustrative examples can be found in more applied fields, for instance in cosmology, medicine, imaging, biology, environmental science, etc.

[1]The reach of a set S in a metric space is the supremum τ such that any point within distance less than τ of S has a unique nearest point in S

To get information on an unknown support from stationary dependent data, we need consider, as was done in the i.i.d case, the convergence of the Hausdorff distance d_H of this cloud of data points to its support. More precisely, we make the following definition.

Definition 1.1. *We say that a point-cloud $\mathbb{X}_n$ of n stationary dependent $\mathbb{R}^d$-valued random variables is (ϵ,α)-dense in $\mathbb{M}\subset\mathbb{R}^d$, if given $\epsilon>0$ and $\alpha\in]0,1[$,*

$$\mathbb{P}\left(d_H(\mathbb{X}_n,\mathbb{M})\le\epsilon\right)\ge 1-\alpha.$$

If $\mathbb{X}:=(X_i)_{i\in\mathbb{N}^*}$ is a stationary sequence of $\mathbb{R}^d$-valued random variables, we say that $\mathbb{X}$ is "asymptotically (ϵ,α)-dense" in $\mathbb{M}$ if given $\epsilon>0$ and $0<\alpha<1$, there exists n_0 so $\forall n\ge n_0$, $\mathbb{X}_n$ is (ϵ,α)-dense.

The first undertaking of the paper is to find conditions on $\mathbb{X}$ so that it is asymptotically (ϵ,α)-dense in a compact support. In §3 and §5 we treat explicitly a number of examples and show for all of these that the property of being asymptotically (ϵ,α)-dense holds by means of a key technical Proposition 2.1 which uses blocking techniques to give upper bounds for $\mathbb{P}\left(d_H(\mathbb{X}_n,\mathbb{M})>\epsilon\right)$.

The next step is topological and consists in showing that when the Hausdorff distance between $\mathbb{X}_n$ and the support is sufficiently small, it is possible to reconstruct the support up to homotopy. This is addressed in §4. To use the notation from that section, we write $X\approx Y$ if $Y\subset X$ and X deformation retracts onto Y. Both the probabilistic and topological results can be combined into the following main statement which summarizes the content of the paper.

Main Results:

- *Let $(X_i)_{i\in\mathbb{N}}$ be a stationary sequence of $\mathbb{R}^d$-valued random variables. Suppose that X_1 is with compact support $\mathbb{M}$ having positive reach τ. Let $0<\delta<\frac{2}{5}\tau$ and $\epsilon\in]0,\frac{\tau}{2}-\frac{\delta}{4}[$. Let $\alpha\in]0,1[$ be fixed. Suppose that there exists n_0 such that for any $n\ge n_0$, $d_H(\mathbb{X}_n,\mathbb{M})\le\frac{\epsilon}{2}$ with probability at least $1-\alpha$. Then for any r such that $\epsilon+\delta\le r<\frac{\tau}{2}-\frac{\delta}{4}$,*

$$\mathbb{P}\left(\bigcup_{x\in\mathbb{X}_n}B(x,r)\approx M\right)\ge 1-\alpha,$$

  for each $n\ge n_0$ (Proposition 4.6).
- *Conditions are stated to give explicit expressions of n_0 for the following sequences of random variables: (i) stationary m-dependent sequences (Proposition 3.2), (ii) stationary β-mixing sequences (Proposition 3.3), (iii) stationary weakly dependent sequences (Proposition 3.4), and (iv) stationary Markov chains (Proposition 5.3 and Proposition 5.5).*

Although our main results are mainly of a probabilistic and geometric nature, we can say a little word about its statistical implications. In practice, the point-cloud data are realizations of random variables living in an unknown support $\mathbb{M}\subset\mathbb{R}^d$. We then ask to know if this support is a circle, or a sphere, or a torus or a more complicated object. By taking sufficiently many points $\mathbb{X}_n$, our results tell us that the homology of $\mathbb{M}$ is the same as the homology of the union of balls around the data $\bigcup_{x\in\mathbb{X}_n}B(x,r)$, and this can be computed in general. The uniform radius r depends on the reach of $\mathbb{M}$, which is the only quantity we need to know a priori. Knowing the homology rules out many geometries for $\mathbb{M}$. Note that knowledge of $\mathbb{M}$ is only precise up to homotopy (or deformation). We may want to find ways to distinguish between a support that is a circle and one that is an annulus, however conclusions of the sort are not discussed in this paper.

We now give some more details about the content of the paper and how it is organized. In §2 we state conditions under which $d_H(\mathbb{X}_n,\mathbb{M})\le\epsilon$ with large probability and for n large enough. This is stated in Proposition 2.1 which is the main technical tool of the paper. The strategy here is to use blocking techniques, i.e. to group the n points into k_n blocks, each block with r_n points is considered as a single point in the appropriate Euclidean space of higher dimension. The control of $d_H(\mathbb{X}_n,\mathbb{M})$ is then reduced to the behavior of lower bounds of the concentration function of one block

$$\rho_{r_n}(\epsilon) := \inf_{x \in \mathbb{M}_{dr_n}} \mathbb{P}(\|(X_1, \cdots, X_{r_n})^t - x\| \leq \epsilon), \tag{1}$$

and of

$$\inf_{x \in \mathbb{M}_{dr_n}} \mathbb{P}(\min_{1 \leq i \leq k_n} \|(X_{(i-1)r_n+1}, \cdots, X_{ir_n})^t - x\| \leq \epsilon), \tag{2}$$

where $\mathbb{M}_{dr_n}$ the support of the block $(X_1, \cdots, X_{r_n})^t$ (vector transpose). Clearly, for independent random variables, a lower bound for (1) is directly connected to a lower bound for (2), but this is not the case for dependent random variables, and we need to control (1) and (2) separately. Section 3 gives our main examples of stationary sequences of $\mathbb{R}^d$-valued random variables having good convergence properties, under the Hausdorff metric, to the support. In the case of mixing sequences, the control of $d_H(\mathbb{X}_n, \mathbb{M})$ is based on assumptions on the behavior of some lower bounds for the concentration function $\rho_m(\epsilon)$, for fixed $m \in \mathbb{N} \setminus \{0\}$, in connection with the decay of the mixing dependence coefficients. These lower bounds can be obtained by means of a condition similar to the so-called (a, b)-standard assumption (see for instance [4]) used in the case of i.i.d. sequences (i.e. when $k_n = n$ and $r_n = 1$). Section 4 establishes the reconstruction result we need. Here, the support $\mathbb{M}$ is assumed to be compact of positive reach (or "PR"). Our main result is Proposition 4.6 which is an extension of similar results in [18] and [23]. The trick here is to thicken the PR set to obtain a Riemannian submanifold with boundary, and then apply techniques of [18].

The last sections of the paper give explicit illustrations of our main results and techniques in the case of Markov chains §5. The Möbius Markov chain on the circle is studied in Section 5.2, and an explicit simulation is presented in §6.

2. Hausdorff Distance and the Support

This section states and proves the main technical lemma of this paper. The main result, Proposition 2.1, is general and of independent interest. It is based on blocking techniques and a useful geometrical result, proven in [18], relating the minimal number of a covering of a compact set by open balls to the maximal length of chains of points whose pairwise distances are bounded below.

Let $(X_i)_{i \in \mathbb{N}}$ be a stationary sequence of $\mathbb{R}^d$-valued random variables. Let P be the distribution of X_1. Suppose that P is supported on a compact set $\mathbb{M}$ of $\mathbb{R}^d$, i.e. $\mathbb{M}$ is the smallest closed set carrying the mass of P;

$$\mathbb{M} = \bigcap_{C \subset \mathbb{R}^d, \ P(\overline{C})=1} \overline{C},$$

where $\overline{C}$ means the closure of the set C in Euclidean space. Recall that $\mathbb{X}_n = \{X_1, \cdots, X_n\}$ and this is viewed as a subset of $\mathbb{R}^d$. Throughout, we will be working with the Hausdorff distance d_H, which is a distance between compact subsets, A and B, in $\mathbb{R}^d$ given by

$$d_H(A, B) = \max \left\{ \sup_{x \in A} \inf_{y \in B} \|x - y\|, \sup_{x \in B} \inf_{y \in A} \|x - y\| \right\}, \tag{3}$$

where $\| \cdot \|$ denotes the Eucliean norm. Obviously $d_H(\{x\}, \{y\}) = ||x - y||$ if x, y are points.

We wish to give upper bounds for $\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right)$ via a blocking technique. Let k and r be two positive integers such that $kr \leq n$. Define, for $1 \leq i \leq k$, the random vector $Y_{i,r}$ of $\mathbb{R}^{dr}$, by $Y_{i,r} = (X_{(i-1)r+1}, \cdots, X_{ir})^t$. Let

$$\mathbb{Y}_k = \{Y_{1,r}, \cdots, Y_{k,r}\}.$$

Clearly $\mathbb{Y}_k$ is a subset of $\mathbb{R}^{dr}$ of stationary k dependent random vectors. The support $\mathbb{M}_{dr}$ of the vector $Y_{1,r}$ is included in $\mathbb{M} \times \cdots \times \mathbb{M}$ (r times) and since, by definition, $\mathbb{M}_{dr}$ is a closed set, it is necessarily compact in $\mathbb{R}^{dr}$. As we now show, it is possible to reduce the behavior of $d_H(\mathbb{X}_n, \mathbb{M})$ to that of the sequence of vectors $(Y_{i,r})_{1 \leq i \leq k}$ for any k and r for which $kr \leq n$ and under only the assumption of stationarity of $(X_i)_{i \in \mathbb{N}}$.

Proposition 2.1. *With $\epsilon > 0$, k and r any positive integers such that $kr \le n$, it holds that*
$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \le \mathbb{P}\left(d_H(\mathbb{Y}_k, \mathbb{M}_{dr}) > \epsilon\right) \le \frac{\sup_{x\in\mathbb{M}_{dr}} \mathbb{P}\left(\min_{1\le i\le k} \|Y_{i,r} - x\| > \epsilon/2\right)}{1 - \sup_{x\in\mathbb{M}_{dr}} \mathbb{P}\left(\|Y_{1,r} - x\| > \epsilon/4\right)}.$$

Proof. Since $\mathbb{P}(\mathbb{Y}_k \subset \mathbb{M}_{dr}) = 1$, we have almost surely (a.s)
$$d_H(\mathbb{Y}_k, \mathbb{M}_{dr}) = \sup_{x\in\mathbb{M}_{dr}} \min_{1\le j\le k} \|Y_{j,r} - x\|. \tag{4}$$
Since $\mathbb{M}_{dr}$ is compact, there exists a finite set $\mathcal{C}_N = \{c_1, \cdots, c_N\} \subset \mathbb{M}_{dr} \subset \mathbb{R}^{dr}$ of centers of balls, forming a minimal ϵ-covering set for $\mathbb{M}_{dr}$ so that, for a fixed $x \in \mathbb{M}_{dr}$, there exists $c_i \in \mathcal{C}_N \subset \mathbb{M}_{dr}$ such that
$$\|x - c_i\| \le \epsilon.$$
Hence,
$$\|Y_{j,r} - x\| \le \|Y_{j,r} - c_i\| + \|c_i - x\| \le \|Y_{j,r} - c_i\| + \epsilon.$$
Consequently, for any $x \in \mathbb{M}_{dr}$,
$$\min_{1\le j\le k} \|Y_{j,r} - x\| \le \min_{1\le j\le k} \|Y_{j,r} - c_i\| + \epsilon \le \max_{1\le i\le N} \min_{1\le j\le k} \|Y_{j,r} - c_i\| + \epsilon$$
and
$$\sup_{x\in\mathbb{M}_{dr}} \min_{1\le j\le k} \|Y_{j,r} - x\| \le \max_{1\le i\le N} \min_{1\le j\le k} \|Y_{j,r} - c_i\| + \epsilon.$$
Hence,
$$\begin{aligned}
&\mathbb{P}\left(\sup_{x\in\mathbb{M}_{dr}} \min_{1\le j\le k} \|Y_{j,r} - x\| \ge 2\epsilon\right) \le \mathbb{P}\left(\max_{1\le i\le N} \min_{1\le j\le k} \|Y_{j,r} - c_i\| \ge \epsilon\right) \\
&\le N \max_{1\le i\le N} \mathbb{P}\left(\min_{1\le j\le k} \|Y_{j,r} - c_i\| \ge \epsilon\right) \le N \sup_{x\in\mathbb{M}_{dr}} \mathbb{P}\left(\min_{1\le j\le k} \|Y_{j,r} - x\| \ge \epsilon\right). \qquad (5)
\end{aligned}$$
We have now to bound N. For this we use Lemma 5.2 in [18] (as was done in [8]), to get
$$N \le \left(\inf_{x\in\mathbb{M}_{rd}} \mathbb{P}(\|Y_{1,r} - x\| \le \epsilon/2)\right)^{-1} = \left(1 - \sup_{x\in\mathbb{M}_{rd}} \mathbb{P}(\|Y_{1,r} - x\| > \epsilon/2)\right)^{-1}. \tag{6}$$
Hence, by (4) together with (5) and (6),
$$\begin{aligned}
&\mathbb{P}\left(d_H(\mathbb{Y}_k, \mathbb{M}_{dr}) > 2\epsilon\right) \qquad (7)\\
&\le \left(1 - \sup_{x\in\mathbb{M}_{rd}} \mathbb{P}(\|Y_{1,r} - x\| > \epsilon/2)\right)^{-1} \sup_{x\in\mathbb{M}_{rd}} \mathbb{P}\left(\min_{1\le j\le k} \|Y_{j,r} - x\| \ge \epsilon\right).
\end{aligned}$$
Thanks to (7), the proof of this proposition is complete if we prove that,
$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \le \mathbb{P}\left(d_H(\mathbb{Y}_k, \mathbb{M}_{dr}) > \epsilon\right). \tag{8}$$
Recall that $\mathbb{P}(\mathbb{X}_n \subset \mathbb{M}) = 1$, so that $d_H(\mathbb{X}_n, \mathbb{M}) = \sup_{x\in\mathbb{M}} \min_{1\le j\le n} \|X_j - x\|$, and, since $kr \le n$,
$$d_H(\mathbb{X}_n, \mathbb{M}) = \sup_{x\in\mathbb{M}} \min_{1\le j\le n} \|X_j - x\| \le \sup_{x\in\mathbb{M}} \min_{1\le j\le kr} \|X_j - x\| = d_H(\mathbb{X}_{kr}, \mathbb{M}).$$
From this we deduce that,
$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \le \mathbb{P}\left(d_H(\mathbb{X}_{kr}, \mathbb{M}) > \epsilon\right). \tag{9}$$
Now, we have to prove that
$$\mathbb{P}\left(d_H(\mathbb{X}_{kr}, \mathbb{M}) > \epsilon\right) \le \mathbb{P}\left(d_H(\mathbb{Y}_k, \mathbb{M}_{dr}) > \epsilon\right). \tag{10}$$
For this, let $X_j \in \mathbb{X}_{kr}$ and $x \in \mathbb{M}$. Then there exist l and i such that X_j is the l-th component of the vector $Y_{i,r}$. We claim also that there exists $\tilde{x} \in \mathbb{M}_{dr}$ such that x is the l-th component of the vector $\tilde{x}$. In fact suppose that, for any $x_1, \cdots, x_{l-1}, x_{l+1}, \cdots, x_r \in \mathbb{M}$, the vector $\tilde{x} = (x_1, \cdots, x_{l-1}, x, x_{l+1}, \cdots, x_r)^t \notin \mathbb{M}_{dr}$, i.e., $\tilde{x}$ cannot be a realisation of the vector $Y_{i,r}$ while x is

a realisation of the vector X_j. Suppose without loss of generality that all the evoked random vectors have density (we denote by f_Z the density of Z), then $f_{X_j}(x) > 0$ and $f_{Y_{i,r}}(\tilde{x}) = 0$. Now,

$$f_{X_j}(x) = \int \cdots \int f_{Y_{i,r}}(\tilde{x}) dx_1, \cdots, dx_{l-1}, dx_{l+1}, \cdots, dx_r = 0,$$

which is in contradiction with the fact that x belongs to $\mathbb{M}$, the support of X_j, i.e. $f_{X_j}(x) > 0$.
From this, we deduce that, for any $X_j \in \mathbb{X}_{kr}$ and $x \in \mathbb{M}$, there exist $1 \leq i \leq k$ and $\tilde{x} \in \mathbb{M}_{dr}$ such that,

$$\|X_j - x\| \leq \|Y_{i,r} - \tilde{x}\|.$$

Hence,

$$\inf_{X_j \in \mathbb{X}_{kr}} \|X_j - x\| \leq \inf_{Y_{i,r} \in \mathbb{Y}_k} \|Y_{i,r} - \tilde{x}\| \leq d_H(\mathbb{Y}_k, \mathbb{M}_{dr}).$$

Consequently, (recall that $\mathbb{P}(\mathbb{X}_{kr} \subset \mathbb{M}) = 1$),

$$d_H(\mathbb{X}_{kr}, \mathbb{M}) = \sup_{x \in \mathbb{M}} \inf_{X_j \in \mathbb{X}_{kr}} \|X_j - x\| \leq d_H(\mathbb{Y}_k, \mathbb{M}_{dr}).$$

From this we get (10). Now (10) together with (9) prove (8). The proof of this proposition is then complete. □

3. (ϵ, α)-dense sequences of random variables

As indicated in the introduction, our main goal is to find conditions on P, the distribution of X_1, under which a sequence $\mathbb{X}$ is asymptotically (ϵ, α)-dense. In this section, we give several examples of dependent random variables for which this is the case. This property is established every time by means of Proposition 2.1 applied with suitable choices of k and r, and for all these examples, it holds that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(d_H(\mathbb{Y}_k, \mathbb{M}_{dr}) > \epsilon\right) = \lim_{n\to\infty} \mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) = 0.$$

All proofs of the following Propositions are relegated to §7.

3.0.1. *Stationary m-dependent sequence on a compact set.* Recall that $(X_i)_{i\in\mathbb{N}}$ is m-dependent for some $m \geq 0$ if for any $i \geq 1$ the two σ-fields $\sigma(X_1, \cdots, X_i)$ and $\sigma(X_{i+j}, X_{i+j+1}, \cdots)$ are independent whenever $j > m$.

Example 3.1. *(m-dependent sequence).*
Let $(T_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with values in $\mathbb{R}^d$. Let h be a real-valued function defined on $\mathbb{R}^{dm}$. The stationary sequence $(X_n)_{n\in\mathbb{N}}$ defined by $X_n = h(T_n, T_{n+1}, \cdots, T_{n+m})$ is a stationary sequence of m-dependent random variables.

Define, for $m \in \mathbb{N}\backslash\{0\}$ and for $\epsilon > 0$, as in the introduction, $Y_{1,m} = (X_1, \cdots, X_m)^t$, the concentration function of the vector $Y_{1,m}$,

$$\rho_m(\epsilon) = \inf_{x \in \mathbb{M}_{dm}} \mathbb{P}\left(\|Y_{1,m} - x\| \leq \epsilon\right). \tag{11}$$

The following proposition gives conditions on $\rho_m(\epsilon)$ under which (1.1) is satisfied.

Proposition 3.2. *Let $(X_i)_{i\in\mathbb{N}}$ be a stationary sequence of m-dependent, $\mathbb{R}^d$-valued random vectors. Suppose that X_1 is with compact support $\mathbb{M}$. Suppose that for any $\epsilon > 0$, there exists a strictly positive constant κ_ϵ such that,*

$$\rho_m(\epsilon) \geq \kappa_\epsilon,$$

then it holds for any $\epsilon > 0$ and any $n \geq m$,

$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \leq \frac{(1 - \kappa_{\frac{\epsilon}{2}})^{[\frac{1}{2}[\frac{n}{m}]]}}{\kappa_{\frac{\epsilon}{4}}},$$

where $[\cdot]$ *denotes the integer part.*
Consequently, for any $\alpha \in]0,1[$ *and any* $n \geq \frac{2m}{\kappa_{\frac{\epsilon}{2}}}\left(\log\left(\frac{1}{\alpha}\right)+\log\left(\frac{1}{\kappa_{\frac{\epsilon}{4}}}\right)\right)+3m$*,* $d_H(\mathbb{X}_n,\mathbb{M}) \leq \epsilon$ *with probability at least* $1-\alpha$*.*

3.0.2. *Stationary* β*-mixing sequence on a compact set.* Recall that the stationary sequence $(X_n)_{n\in\mathbb{N}}$ is β-mixing if β_n tends to 0 when n tends to infinity where the coefficients β_n are defined by, (see [3]),

$$\beta_n = \sup_{l\geq 1}\mathbb{E}\left\{\sup\left|\mathbb{P}\left(B|\sigma(X_1,\cdots,X_l)\right)-\mathbb{P}(B)\right|,\ B\in\sigma(\sigma_i,\ i\geq l+n)\right\}.$$

The following corollary gives conditions on the behaviors of the two sequences $(\rho_n(\epsilon))_n$ and $(\beta_n)_n$ under which (1.1) is satisfied.

Proposition 3.3. *Let* $(X_n)_{n\geq 0}$ *be a stationary* β*-mixing sequence. Suppose that* X_1 *is supported on a compact set* $\mathbb{M}$*. Then it holds, for any* $\epsilon>0$ *and any sequences* k_n *and* r_n *such that* $k_n r_n \leq n$*,*

$$\mathbb{P}\left(d_H(\mathbb{X}_n,\mathbb{M})>\epsilon\right) \leq \frac{k_n^2\beta_{r_n}+k_n\exp\left(-[\frac{k_n}{2}]\rho_{r_n}(\epsilon/2)\right)}{k_n\rho_{r_n}(\epsilon/4)}.$$

Suppose moreover that for some $\beta>1$*, and any* $\epsilon>0$*,*

$$\lim_{m\to\infty}\rho_m(\epsilon)\frac{e^{m^\beta}}{m^{1+\beta}}=\infty,\ \ \textit{and}\ \ \lim_{m\to\infty}\frac{e^{2m^\beta}}{m^2}\beta_m=0.$$

Then for any $\alpha\in]0,1[$ *there exists a positive integer* n_0 *such that for any* $n\geq n_0$*,* $d_H(\mathbb{X}_n,\mathbb{M})\leq\epsilon$ *with probability at least* $1-\alpha$*.*

3.0.3. *Stationary weakly dependent sequence on a compact set.* We suppose here that $(X_i)_i$ is a stationary sequence such that X_1 takes values in a compact support $\mathbb{M}$. We suppose also that this sequence is weakly dependent in the sense of [6]. More precisely, we suppose that there exists a non-increasing function Ψ such that $\lim_{r\to\infty}\Psi(r)=0$, that for any measurable functions f and g bounded by 1 and for any $1\leq i_1\leq\cdots\leq i_k<i_k+r\leq i_{k+1}\leq\cdots\leq i_n$ one has

$$\left|\mathrm{Cov}(h(X_{i_1},\cdots,X_{i_k}),g(X_{i_{k+1}},\cdots,X_{i_n})\right|\leq\Psi(r). \tag{12}$$

This dependence condition is weaker than the Rosenblatt strong mixing dependence as was introduced in [21].

Proposition 3.4. *Let* $(X_n)_{n\geq 0}$ *be a sequence of stationary, weakly dependent in the sense of (12). Suppose that* X_1 *is supported on a compact set* $\mathbb{M}$*. Then it holds, for any* $\epsilon>0$ *and any sequences* k_n *and* r_n *such that* $k_n r_n\leq n$*,*

$$\mathbb{P}\left(d_H(\mathbb{X}_n,\mathbb{M})>\epsilon\right)\leq\frac{k_n^2\Psi(r_n)+k_n\exp\left(-[\frac{k_n}{2}]\rho_{r_n}(\epsilon/2)\right)}{k_n\rho_{r_n}(\epsilon/4)}.$$

Suppose moreover that, for some $\beta>1$*, and any* $\epsilon>0$*,*

$$\lim_{m\to\infty}\rho_m(\epsilon)\frac{e^{m^\beta}}{m^{1+\beta}}=\infty,\ \ \textit{and that}\ \ \lim_{m\to\infty}\frac{e^{2m^\beta}}{m^2}\Psi(m)=0.$$

Then for any $\alpha\in]0,1[$ *there exists a positive integer* n_0 *such that for any* $n\geq n_0$*,* $d_H(\mathbb{X}_n,\mathbb{M})\leq\epsilon$ *with probability at least* $1-\alpha$

Remark 3.5. The condition of Proposition 3.2 and the first conditions of Proposition 3.3 and 3.4 are all satisfied, in particular, if $\inf_{m\geq 1}\rho_m(\epsilon)$ has a strictly positive lower bound.

4. A reconstruction result

Given a point-cloud $\mathbb{X}_n$ on a support $\mathbb{M}$, a standard problem is to reconstruct this support from the given distribution of points as n gets large. Various reconstruction processes in the literature are based on the Nerve theorem. This basic but foundational result can be found in introductory books in algebraic topology ([12], chapter 4). It is based on the existence of a *good* cover of X, meaning a cover by open contractible subspaces U_i such that the intersection of any number of U_i's is either empty or contractible. The *nerve* of this cover is now the abstract simplicial complex whose $n-1$ dimensional simplices correspond to any non-empty intersection of n-open sets of the cover, and the "Nerve Theorem" states that the realization of the nerve of a good cover of a subspace $M \subset \mathbb{R}^d$ has the same homotopy type as M. We will use the notation $|S|$ to refer to the underlying geometric complex which is the realization of an abstract simplicial complex S.

Starting with a point-cloud in M, the nerve complex associated to this cloud appears as either a Vietoris complex or a Cech complex construction. These constructions are very closely related and we refer to [1, 17] for general discussions and results. In our case, we will always consider $\mathbb{X}_n = \{x_1, \dots, x_n\} \subset M$ with M a compact metric space. For a given $r > 0$, take open balls $U_i = B(x_i, r)$ around each x_i in $\mathbb{X}_n$. The nerve complex associated to this collection $\mathcal{U} = \{U_i\}$ is called the Cech complex and is written $Cech(\mathcal{U}, r)$ [19]. We have the homotopy equivalence

$$|Cech(\mathcal{U}, r)| \simeq \bigcup_{x_i \in \mathbb{X}} B(x_i, r)$$

as a direct consequence of the Nerve theorem since the intersection of any number of balls is always a convex set thus contractible. The reconstruction results we are interested in consists in finding suitable r and suitable conditions on $\mathbb{X}_n$ so that $\bigcup \simeq M$, or even better, such that $|Cech(\mathcal{U}, r)|$ is a deformation retract

One of the earliest reconstruction attempts along the lines indicated above seems to be [13] which shows essentially that reconstructions by finite sets is indeed possible. Assume M is a closed Riemannian manifold. Then there exists $\epsilon_0 > 0$ such that for every $0 < r < \epsilon_0$, there exists a $\delta > 0$ such that for any cloud of points $\mathbb{X}$ that has Gromov-Hausdorff distance less than δ from M, the geometric realization $|\mathbb{X}_r|$ is homotopy equivalent to M. This theorem is clearly not constructive in nature.

Note that in the case of Riemannian manifolds M, there is an appealing method for reconstruction using "geodesic balls". Let ρ_c (or $\rho_c(M)$) be the *convexity radius.* Around each $p \in M$, there is a "geodesic ball" $B_g(p, \rho_c)$ which is convex, meaning that any two points in this neighborhood are joined by a unique geodesic in that neighborhood. These geodesic balls are contractible. If $\mathbb{X} = \{x_1, \dots, x_n\}$ is a pointcloud such that $\mathcal{U}_c := \{B_c(x_i, \rho_c)\}$ is a cover of M, and since the intersection of geodesic balls is contactible, then

$$|Cech(\mathcal{U}_c, \rho_c)| \simeq M$$

For some submanifolds $M \in \mathbb{R}^d$, it might happen that $\rho_c > \tau$, where τ is the reach of M. This is the case of the unit circle $\mathcal{C}$ for example since $\rho_c(\mathcal{C}) = \pi/2$ and $\tau = 1$.

The fist main constructive reconstruction result of use in the literature seems to be Proposition 2.1 of [18]. As in the case of [13], one needs restrict to Riemannian manifolds. Let M be a compact Riemannian manifold and let τ be defined as the largest number having the property that the open normal bundle about M of radius r is embedded in $\mathbb{R}^d$ for every $r < \tau$. This number is precisely the reach of M. Let $\mathbb{X}$ be any finite collection of points in $\mathbb{R}^d$ that is $\frac{r}{2}$ dense in M, meaning any point in $\mathbb{X}$ is at most this much distance from a point in M. Assume $r < \sqrt{\frac{3}{5}}\tau$. Then $M \subset |Cech(\mathcal{U}, r)| = \bigcup_{x \in \mathbb{X}} B(x, r)$ is a deformation retract so that M and $|Cech(\mathcal{U}, r)|$ have the same homotopy type. The bulk of the proof is to construct a map $|Cech(\mathcal{U}, r)| \to M$ which is proper with contractible fibers.

An extension of Proposition 3.1 of [18] from closed Riemannian manifolds to manifolds with boundary is given in [23], while various other versions for general compact metric spaces are in [1, 17]. More particularly, and in ways closer to the spirit of this paper, [19] proves a general homotopy reconstruction result for compact sets with positive reach (or PR sets, see [7]), which is a collection of spaces that subsumes Riemannian manifolds. See Figure 2 for an example of such space. The statement of the result is in terms of "subspace balls". Take a sufficiently dense collection $\mathbb{X}$ in M and a radius r so that $M \subset \bigcup_i B(x_i, r)$. Then $\mathcal{U}_M = \{B(x_i, r) \cap M\}$ is an open cover of M by "subspace balls". If this cover is good, then as we know by the nerve theorem, $|\mathcal{U}_M|$ is of the homotopy type of M. The main result of [19] asserts that if M is any subset of $\mathbb{R}^d$ of positive reach $\tau > 0$, and $\mathcal{U} = \{B(x_i, r)\}$ is a finite collection of balls that cover M with $r < \tau$, then $\mathcal{U}_M$ is necessarily good and $|Cech(\mathcal{U}_M, r)| \simeq M$. The bulk of the proof of [19] is to show that all possible non-empty intersections of subspace balls $B(x_i, r) \cap M$ are contractible.

Note that there is an inclusion of simplicial complexes $Cech(\mathcal{U}_M, r) \hookrightarrow Cech(\mathcal{U}, r)$ but in general, both spaces need not have the same homotopy type. The complex $|Cech(\mathcal{U}, r)|$ will be detecting the homotopy type of the so-called *offset*

$$M^{\oplus r} := \{p \in \mathbb{R}^d \mid d(p, M) := \inf_{x \in M} ||x - p|| \leq r\}$$

This is not surprising since $|Cech(\mathcal{U}, r)| \simeq \mathbb{X}^{\oplus r} = \bigcup_{x \in \mathbb{X}} B(x, r)$. Many of the existing theorems in homotopic and homological inference are about offsets. The next section explains how $\mathbb{X}^{\oplus r}$ can recover the homotopy type of M for compact spaces with positive reach. Our Proposition 4.6 seems to give a more streamlined result than what is existing in the literature.

4.1. **Manifolds with boundary.** A measure of distance between two closed subspaces in a metric space is the Hausdorff distance d_H (3). This is a "coarse" metric in the sense that two closed spaces A and B can be very different topologically and yet be arbitrarily close in Hausdorff distance. In particular, this metric is very sensitive to outliers (See [20]). If $\mathbb{X}$ is a point cloud inside M, then d_H can be viewed as a measure of density. Indeed in this case, we can write $d_H(\mathbb{X}, M) = \sup_{y \in M} \inf_{x \in \mathbb{X}} d(x, y)$, where d is Euclidean distance. Saying that $d_H(\mathbb{X}, M) < \epsilon$ means that for all $y \in M$, $\inf_{x \in \mathbb{X}} d(x, y) = \min_{x \in \mathbb{X}} d(x, y) < \epsilon$, and so there is a point in $\mathbb{X}$ within distance ϵ from y.

Definition 4.1. *We say that a subset $\mathbb{X}$ is ϵ-dense in M if $d_H(\mathbb{X}, M) \leq \epsilon$, or alternatively if $B(p, \epsilon) \cap \mathbb{X} \neq \emptyset$ for each $p \in M$.*

With this interpretation of d_H, the reconstruction result of [19] alluded to earlier takes the following useful form. First we indicate that if M is smooth submanifold in $\mathbb{R}^d$, then the condition number τ of [18] about embedding tubular neighborhoods of M coincides with the reach of M.

Proposition 4.2. *([18], Proposition 3.1) Let M be a compact Riemannian submanifold of $\mathbb{R}^d$ with positive reach τ, and $\mathbb{X} \subset M$ an $\frac{\epsilon}{2}$-dense finite subset. Then for any $\epsilon \leq r < \sqrt{\frac{3}{5}}\tau$, $\bigcup_{x \in \mathbb{X}} B(x, r) \simeq M$.*

An extension of this proposition to Riemannian manifolds with boundary is given in [23]. In this case one obtains the analogue of Proposition 4.2 where now M is a compact manifold with boundary, and the bound $\sqrt{\frac{3}{5}}\tau$ is replaced by $\frac{\delta}{2}$, where $\delta = \min(\tau(M), \tau(\partial M))$. Note that in the case of [18], the submanifold must have codimension at least 1 since it is closed. The codimension 0 case means that M is necessarily a (compact) manifold with boundary and this is the case we need. The reach of ∂M (manifold boundary) and M are not comparable in general. Indeed, take M to be the closed upper hemisphere of the unit sphere in $\mathbb{R}^3$. Then $\tau(M) < \tau(\partial M)$. Take now a closed disk M in $\mathbb{R}^2$. Then $\tau(\partial M) < \tau(M) = \infty$. By inspecting both proofs of [18, 23], we can ignore $\tau(\partial M)$ and a bound of $\tau/2$ on r is sufficient.

Proposition 4.3. *Let M be a compact Riemannian submanifold of $\mathbb{R}^d$ with boundary, having codimension 0 and positive reach τ, and let $\mathbb{X} \subset M$ be an $\frac{\epsilon}{2}$-dense finite subset. Then for any r such that $\epsilon \leq r < \frac{\tau}{2}$, $\bigcup_{x \in \mathbb{X}} B(x, r) \simeq M$.*

This Proposition is a straightforward adaptation of results in §4 of [18]. It is an improvement on [23] since we do not need to demand that $\epsilon < \frac{1}{2}\min(\tau(M), \tau(\partial M))$ which can be fairly restrictive in applications (i.e. consider a very flat ellipse in the plane. This is convex, so ϵ can be chosen as large as needed, but $\tau(\partial M)$ will impose on ϵ to be very small).

Proof. In codimension 0, the boundary is an oriented hypersurface, and divides Euclidean space into two regions. Let τ^+ denote the reach of ∂M in the unbounded region, and τ^- its reach within the bounded region. Then $\tau(M) = \tau^+$, while evidently $\tau(\partial M) = \min\{\tau^+, \tau^-\}$. The key point in the proof of Proposition 3.1 of [18], asserting that $\bigcup_{x\in\mathbb{X}}(B(x,r) \simeq N$, for N closed submanifold in Euclidean space, $\mathbb{X} \subset N$, is to show that any point $v \in T_p(N)^\perp$, "not far away" from p, which is in a ball $B(q,\epsilon)$ with $q \in \mathbb{X}$ but $q \notin B(p,\epsilon)$, must be in another ball $B(x,\epsilon)$, with $x \in B(p,\epsilon) \cap \mathbb{X}$. In our case, we will apply this computation to $N = \partial M$. To measure this "not far away" quantity, we look at extreme cases where q is on a tangent circle to $T_p(\partial M)$ of curvature $\frac{1}{\tau^+} = \frac{1}{\tau}$ (the closed manifold ∂M cannot bend further according to the relationship between curvature and reach as described in §6 of [18]). Lemma 4.1 shows that for such a configuration v, p, q to exist (i.e. $|q-p| > \epsilon$, $|v-p| < \tau$ and $|v-q| = \epsilon$) one must have $|v-p| < \frac{\epsilon^2}{\tau}$. This is all illustrated in our Figure 1 which is the analog of Fig.2 of [18]. So if $\frac{\epsilon^2}{\tau} < \frac{\epsilon}{2}$, that is if $\epsilon < \frac{\tau}{2}$, then $|v-p| < \frac{\epsilon}{2}$, and since there is an $x \in \mathbb{X}$ that is within distance $\frac{\epsilon}{2}$ from p, necessarily $|x-v| < \epsilon$. Remaining arguments as can be found in §4 of [18] prove the Proposition. Note that in the case of Proposition 3.1 of [18], that x couldn't be "anywhere" possibly around p, but had to lie on $N = \partial M$, which was the Riemannian manifold under investigation, thus the authors get a different bound on ϵ. □

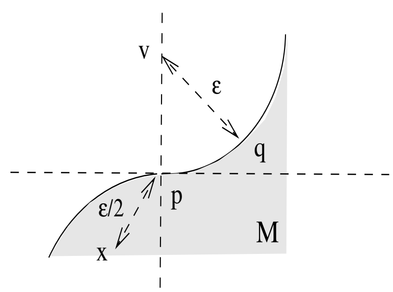


FIGURE 1. An extreme disposition of points, $x, q \in \mathbb{X}$ and $p, q \in \partial M$. The points q, p are on a circle tangent to $T_p(M)$, of radius τ and center on the vertical dashed line representing the normal direction $T_p(M)^\perp$, while y is on a circle of radius τ^- with center on the normal.

We next extend Proposition 4.3 to M compact in $\mathbb{R}^d$ whose reach is strictly positive. Figure 2 gives an example of such a space. We will always denote by $\tau(M)$, or just τ, the reach of M. The quintessential property of PR-sets (i.e. positive reach sets) is the existence, for $r < \tau$, of the "unique closest point" projection

$$\pi_M : M^{\oplus r} \longrightarrow M \quad , \quad ||y - \pi_M(y)|| = d_H(y, M) \tag{13}$$

sending y to its nearest unique point $x \in M$.

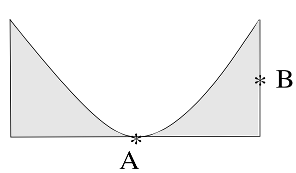


FIGURE 2. This space has positive reach τ in $\mathbb{R}^2$ but is not a submanifold.

PR-sets are necessarily closed, thus compact if bounded. If M is a compact PR-set, then so is $M^{\oplus r}$, for $r<\tau$, where τ is the reach of M. We write $B(x,r)$ the open ball around x of radius r, and $\overline{B}(x,r)$ its closure. The following observation is a consequence of work of Federer [10].

Proposition 4.4. *Let $M\subset\mathbb{R}^d$ be connected and compact of positive reach. For $0<r<\tau$, $M^{\oplus r}$ is a compact manifold with boundary, of dimension n, homotopy equivalent to X.*

Proof. Let d_M be the distance function to M. Elementary pointset topology shows that the interior of the r-offset of M is $int(M^{\oplus r})=\bigcup_{x\in M}B(x,r)=d_M^{-1}[0,r)$, $M^{\oplus r}=\bigcup_{x\in M}\overline{B}(x,r)$ (this is a consequence of compactness of M) and the topological boundary is $d_M^{-1}(r)$ (this is a consequence of the continuity of the Hausdorff metric).

The interior of $M^{\oplus r}$ is open in $\mathbb{R}^d$ so is necessarily a smooth submanifold of codimension 0. It is bounded with non-trivial topological boundary $\partial M^{\oplus r}$. By work of Federer (see [7], Lemma 1.10), $\partial M^{\oplus r}$ is a $C^{1,1}$-hypersurface in $\mathbb{R}^d$, in particular it is C^1, oriented. To deduce from this that $M^{\oplus r}$ is a submanifold with boundary, we need verify that there exist open Euclidean neighborhoods U of $y\in\partial M^{\oplus r}$, and V and a diffeomorphism $g:U\to V$ such that $g(U\cap\partial M^{\oplus r})\cong V\cap H^n$, where H^n is half-space. Since the boundary is a manifold of codimension 1, we know there are such U and V, and a g so that $g(U\cap M^{\oplus r})\cong V\cap(\mathbb{R}^{n-1}\times 0)$. If suffices now to see if a point in the interior part $int(M^{\oplus r})\cap U$ goes to $V\cap H^n$, then all points in that interior part must map to $V\cap H^n$. Suppose this is not the case and two points p,q in $int(M^{\oplus r})\cap U$ map under g to two different hemispheres $V\cap H^n_+$ and $V\cap H^n_-$, then every path linking p and q must cross the boundary, since its image by a diffeomorphism g must cross $\mathbb{R}^{n-1}\times\{0\}$ and this cannot be true in general.

Finally, we show that $M^{\oplus r}\simeq M$. To this end, we use the map (13) and claim that it is continuous and proper (this is clear) and that it has contractible preimages. This would imply that both spaces are homotopy equivalent. So for every fixed $x\in M$, we must show that $\pi^{-1}(x)\subset M^{\oplus r}$ is contractible. Pick y in this preimage. We claim that the interval $[x,y]$ in $\mathbb{R}^d$ is entirely in $\pi_M^{-1}(x)$. Let $z\in[x,y]$ and suppose there is $x'\neq x$, $x'\in M$, so that $\pi_M(z)=x'$, that is $||z-x'||<||z-x||$. Then

$$||y-x'||\le||y-z||+||z-x'||<||y-z||+||z-x||=||y-x||$$

which contradicts the fact that $x=\pi_M(y)$. So $\pi(z)=x$ for all $z\in[x,y]$ and $[x,y]\subset\pi_M^{-1}(x)$. The preimage is starshaped so must be contractible (by the contraction that collapses line segments to x). The proof is complete. ☐

We next need to compare the reach of M to that of $M^{\oplus r}$, $r<$reach$(M)=\tau$. Note that the reach is not always well-behaved for nested compact sets. By this we mean that if (K_2,K_1) is a pair of nested compact sets in $\mathbb{R}^d$, $K_1\subset K_2$, then both cases $\tau_1<\tau_2$ or $\tau_2<\tau_1$ can occur, where τ_i is the reach of K_i. For example and in the former case, take K_1 to be the circle and K_2 to be the closed disk, while for the latter case, take K_1 to be a point in a finite reach K_2. The case of $(K_2,K_1)=(M^{\oplus r},M)$ is special.

Lemma 4.5. *Let τ_r be the reach of $M^{\oplus r}$, $r<\tau$. Then $(M^{\oplus r})^{\oplus\tau_r}=M^{\oplus\tau}$. In particular, if M is convex, then so is $M^{\oplus r}$, and if τ is finite, $\tau_r=\tau-r$.*

Proof. Pick y not in $M^{\oplus r}$, and let x be one of its closest point in M. We claim that $y_r:=[y,x]\cap\partial M^{\oplus r}=\pi_{M^{\oplus r}}(y)$ is one of the closest points to y in the offset $M^{\oplus r}$. This is clear because if say there is z in the boundary that is closest to y, and $w=\pi_M(z)$, then

$$d(y,w)\le d(y,z)+d(z,w)=d(y,z)+r<d(y,y_r)+d(y_r,M)=d(y,x)$$

and this is a contradiction. The last equality follows from the fact already shown in the proof of Proposition 4.4 which is that if $y\in M^{\oplus r}$, then $\pi_M(z)=\pi_M(y)$ for all $z\in[y,\pi_M(y)]$.

Start now with y having two distinct projections x_1,x_2 onto M (i.e. $y\notin M^{\oplus\tau}$). Then $[y,x_1]$ and $[y,x_2]$ intersect $\partial M^{\oplus r}$ at z_1,z_2 with $d(z_1,x)=d(z_2,x)=r$. Moreover these are closest to y in $M^{\oplus r}$ by the preceding paragraph. Since $z_1\neq z_2$, then $y\notin(M^{\oplus r})^{\oplus\tau_r}$. This shows that $(M^{\oplus r})^{\oplus\tau_r}\subset M^{\oplus\tau}$. A slight rephrase to this same argument shows that $M^{\oplus\tau}\subset(M^{\oplus r})^{\oplus\tau_r}$, and so both subspaces are equal.

To summarize, if $y \in M^{\oplus\tau} \setminus M^{\oplus r}$ and $y_r = [y, \pi_M(y)] \cap \partial M^{\oplus r}$, then $y_r = \pi_{M^{\oplus r}}(y)$ and

$$d(y, M) = d(y, M^{\oplus r}) + r$$

Since $y \in (M^{\oplus r})^{\oplus \tau_r}$, then $d(y, M) \leq \tau_r + r$. Since this is true for any $y \in M^{\oplus\tau}$, so $\tau \leq \tau_r + r$. A similar argument gives that $\tau_r + r \leq \tau$, so here too equality holds. □

We are now ready to prove the main result of this section.

Proposition 4.6. *Let M be a compact space in $\mathbb{R}^d$ with positive reach τ and let $0 < \delta < \frac{2}{5}\tau$. Let ϵ be such that $0 < \epsilon < \frac{\tau}{2} - \frac{5\delta}{4}$. Assume that $\mathbb{X}$ is $\frac{\epsilon}{2}$ dense in M. Then for any r such that $\epsilon + \delta \leq r < \frac{\tau}{2} - \frac{\delta}{4}$, $\bigcup_{x \in \mathbb{X}} B(x, r) \simeq M$.*

Proof. If $\mathbb{X}$ is an $\frac{\epsilon' - \delta}{2}$-dense sample in M, $\mathbb{X} \subset M$, then it is an $\frac{\epsilon'}{2}$-dense sample in $M^{\oplus\delta/2}$. This offset is a codimension 0 manifold with boundary containing M and reach $\tau' = \tau - \frac{\delta}{2}$. Proposition 4.3 implies then that for all $\epsilon' \leq r < \frac{\tau - \delta/2}{2}$, $\bigcup_{x \in \mathbb{X}} B(x, r) \simeq M^{\oplus\delta/2}$. By setting $\epsilon = \epsilon' - \delta$, we get the hypotheses of the Theorem. But $\frac{\delta}{2} < \tau$, and so according to Proposition 4.4, $M^{\oplus\delta/2} \simeq M$ and the proof is complete. □

5. Application to stationary Markov chains on a compact state space

This section gives conditions on stationary Markov chains on a compact state space so that they are asymptotically (ϵ, α)-dense. Those conditions can be checked by studying the β-mixing properties of these Markov chains and by applying Proposition 3.3. We choose however in this section to be even more precise by adopting specific models and carrying out explicit calculations.

Let $(X_n)_{n\geq 0}$ be an homogeneous Markov chain satisfying the following two assumptions.

- $(\mathcal{A}_1)$ This Markov chain has an invariant measure μ with compact support $\mathbb{M}$ (and then the chain is stationary).
- $(\mathcal{A}_2)$ The transition probability kernel K, defined for $x \in \mathbb{M}$, by

$$K(x, \cdot) = \mathbb{P}(X_1 \in \cdot | X_0 = x)$$

is absolutely continuous with respect to some measure ν on $\mathbb{M}$, i.e. there exists a positive measure ν and a positive function k such that for any $x \in \mathbb{M}$, $K(x, \nu(dy)) = k(x, y)\nu(dy)$. Suppose that, for some $b > 0$ and $\epsilon_0 > 0$,

$$V_d := \inf_{x \in \mathbb{M}} \inf_{0<\epsilon<\epsilon_0} \left(\frac{1}{\epsilon^b} \int_{B(x,\epsilon)\cap\mathbb{M}} \nu(dx_1) \right) > 0$$

and that there exists a positive constant κ such that $\inf_{x\in\mathbb{M},\, y\in\mathbb{M}} k(x, y) \geq \kappa > 0.$.

Recall that $\mathbb{P}_x$ (resp. $\mathbb{P}_\mu$) denotes the probability when the initial condition $X_0 = x$ (resp. X_0 is distributed as the stationary measure μ). We need the following two lemmas in order to check the conditions of Proposition 2.1 (with $r_n = 1$).

Lemma 5.1. *Let $(X_n)_{n\geq 0}$ be a Markov chain satisfying Assumptions $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$. Then, it holds, for any $0 < \epsilon < \epsilon_0$ and any $x_0 \in \mathbb{M}$,*

$$\inf_{x \in \mathbb{M}} \mathbb{P}_{x_0}(\|X_1 - x\| \leq \epsilon) \geq \kappa\epsilon^b V_d, \quad \inf_{x \in \mathbb{M}} \mathbb{P}_{\mu}(\|X_1 - x\| \leq \epsilon) \geq \kappa\epsilon^b V_d.$$

Proof. We have, using Assumption $(\mathcal{A}_2)$,

$$\mathbb{P}_{x_0}\left(\|X_1 - x\| \le \epsilon\right) = \mathbb{P}_{x_0}\left(X_1 \in B(x,\epsilon)\cap \mathbb{M}\right) = \int_{B(x,\epsilon)\cap\mathbb{M}} K(x_0, \nu(dx_1))$$
$$= \int_{B(x,\epsilon)\cap\mathbb{M}} k(x_0,x_1)\nu(dx_1)$$
$$\ge \kappa \int_{B(x,\epsilon)\cap\mathbb{M}} \nu(dx_1) \ge \kappa\epsilon^b \inf_{0<\epsilon<\epsilon_0}\left(\frac{1}{\epsilon^b}\int_{B(x,\epsilon)\cap\mathbb{M}} \nu(dx_1)\right) \ge \kappa\epsilon^b V_d.$$

The proof of Lemma 5.1 is complete since $\mathbb{P}_\mu\left(\|X_1 - x\| \le \epsilon\right) = \int \mathbb{P}_{x_0}\left(\|X_1 - x\| \le \epsilon\right) d\mu(x_0)$. □

Lemma 5.2. *Let $(X_n)_{n\ge 0}$ be a Markov chain satisfying Assumptions $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$. Then, it holds, for any $0<\epsilon<\epsilon_0$ and $k\in\mathbb{N}\setminus\{0\}$,*

$$\sup_{x\in\mathbb{M}} \mathbb{P}_\mu\left(\min_{1\le i\le k}\|X_i - x\| > \epsilon\right) \le \left(1-\kappa\epsilon^b V_d\right)^k.$$

Proof. [2] Set $\mathcal{F}_n = \sigma(X_0,\dots,X_n)$. By Markov property and Lemma 5.1

$$\mathbb{P}_\mu\left(\min_{1\le i\le k}\|X_i - x\| > \epsilon\right) = \mathbb{P}_\mu(\forall\, 1\le i\le k,\ X_i\notin B(x,\epsilon))$$
$$= \mathbb{E}_\mu\left(\prod_{i=1}^{k-1} 1_{\{X_i\notin B(x,\epsilon)\}}\mathbb{E}(\mathbb{1}_{\{X_k\notin B(x,\epsilon)\}}|\mathcal{F}_{k-1})\right)$$
$$= \mathbb{E}_\mu\left(\prod_{i=1}^{k-1} \mathbb{1}_{\{X_i\notin B(x,\epsilon)\}}\mathbb{E}_{X_{k-1}}(\mathbb{1}_{\{X_k\notin B(x,\epsilon)\}})\right)$$
$$\le (1-\kappa\epsilon^b V_d)\mathbb{E}_\mu\left(\prod_{i=1}^{k-1} \mathbb{1}_{\{X_i\notin B(x,\epsilon)\}}\right)$$
$$\le (1-\kappa\epsilon^b V_d)\mathbb{P}_\mu(\forall\, 1\le i\le k-1,\ X_i\notin B(x,\epsilon)).$$

Lemma 5.2 is proved using the last bound together with an induction reasoning on k. □

Proposition 5.3. *Suppose that Assumptions $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$ are satisfied for some Markov chain $(X_n)_{n\ge1}$. Then the bounds of Proposition 2.1 are satisfied for $r_n = r = 1$, $k_n = n$ and*

$$\mathbb{P}\left(d_H(\mathbb{X}_n,\mathbb{M}) > \epsilon\right) \le \frac{4^b(1-\kappa\epsilon^b V_d/2^b)^n}{\kappa\epsilon^b V_d}.$$

Consequently, for any $\alpha\in]0,1[$ and any $n \ge \frac{2^b}{\kappa\epsilon^b V_d}\left(\ln\left(\frac{4^b}{\kappa\epsilon^b V_d}\right)+\ln\left(\frac{1}{\alpha}\right)\right)$,

$$d_H(\mathbb{X}_n,\mathbb{M}) \le \epsilon,$$

with probability at least $1-\alpha$.

Proof. We have using Lemmas 5.1 and 5.2,

$$\sup_{x\in\mathbb{M}} \mathbb{P}_\mu\left(\min_{1\le i\le n}\|X_i - x\| > \epsilon\right) \le (1-\kappa\epsilon^b V_d)^n \le \exp(-n\kappa\epsilon^b V_d),$$
$$1 - \sup_{x\in\mathbb{M}} \mathbb{P}\left(\|X_1 - x\| > \epsilon\right) \ge \kappa\epsilon^b V_d > 0.$$

Consequently the conclusion of Proposition 2.1 is satisfied with $r_n = 1$, $k_n = n$. More precisely, it holds

$$\mathbb{P}\left(d_H(\mathbb{X}_n,\mathbb{M}) > \epsilon\right) \le \frac{4^b\exp(-n\kappa\epsilon^b V_d/2^b)}{\kappa\epsilon^b V_d}$$

[2] We are grateful to Sophie Lemaire for the present form of the proof of Lemma 5.2.

The proof of this proposition is complete since $\alpha \geq \frac{4^b \exp(-n\kappa\epsilon^b V_d/2^b)}{\kappa\epsilon^b V_d}$ is equivalent to

$$n \geq \frac{2^b}{\kappa\epsilon^b V_d} \ln\left(\frac{4^b}{\alpha\kappa\epsilon^b V_d}\right).$$

□

We next give examples of Markov chains satisfying the requirements of Proposition 5.3. Those examples concern stationary Markov chains on the balls and stationary Markov chains on the circles.

5.1. Stationary Markov chains on a ball of $\mathbb{R}^d$.

5.1.1. *Random difference equations.* Let $(X_n)_{n\geq 0}$ be defined by,

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \;\; n \geq 0 \tag{14}$$

where A_{n+1} is a $(d \times d)$-matrix, $X_n \in \mathbb{R}^d,\ B_n \in \mathbb{R}^d$, $(A_n, B_n)_{n\geq 0}$ is an i.i.d. sequence independent of X_0. Recall that for a matrix M, $\|M\|$ is the operator norm defined by $\|M\| = \sup_{x\in\mathbb{R}^d,\, \|x\|=1} \|Mx\|$. It is well known see for instance [16] that, for any $n \geq 1$, X_n is distributed as $\sum_{k=1}^n A_1 \cdots A_{k-1}B_k + A_1 \cdots A_n X_0$, that the following conditions (see [15], or [11])

$$\mathbb{E}(\ln^+ \|A_1\|) < \infty,\ \mathbb{E}(\ln^+ \|B_1\|) < \infty,\ \lim_{n\to\infty} \frac{1}{n} \ln \|A_1 \cdots A_n\| < 0, \tag{15}$$

ensure the existence of a stationary solution to (14) and that $\|A_1 \cdots A_n\|$ tends to 0 exponentially fast. If $\mathbb{E}\|B_1\|^\beta < \infty$ for some $\beta > 0$ then the series $R := \sum_{i=1}^\infty A_1 \cdots A_{i-1}B_i$ converges a.s. and the distribution of X_n converges to that of R, independently of X_0. The distribution of R is then that of the stationary measure of the chain.

Compact state space. If $\|B_1\| \leq c < \infty$ for some fixed c, then this stationary Markov chain is $\mathbb{M}$-compactly supported. In particular if $\|A_1\| \leq \rho < 1$ for some fixed ρ, then $\mathbb{M}$ is included in the ball $B_d(0, \frac{c}{1-\rho})$ of $\mathbb{R}^d$.

Transition kernel. Suppose that, for any $x \in \mathbb{M}$, the random vector A_1x+B_1 has a density $f_{A_1x+B_1}$ with respect to the Lebesgue measure (here ν is the Lebesgue measure) satisfying $\inf_{x,\, y\in\mathbb{M}} f_{A_1x+B_1}(y) \geq \kappa$, then $k(x,y) = f_{A_1x+B_1}(y) \geq \kappa > 0$.

We collect all the above results in the following corollary.

Corollary 5.4. *Suppose that in the model (14), Conditions (15) are satisfied with moreover $\|B_1\| \leq c < \infty$. If $f_{A_1x+B_1}$, the density of $A_1x + B_1$, satisfies $\inf_{x,\, y\in\mathbb{M}} f_{A_1x+B_1}(y) \geq \kappa > 0$ for some positive κ, then Assumptions $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$ are satisfied with $b = d$ and ν is the Lebesgue measure on $\mathbb{R}^d$.*

Example: The AR(1) process. We consider a particular case of the Markov chain as defined in (14) where, for each n, $A_n = \rho$ with $|\rho| < 1$. We obtain the standard first order linear Auto-Regressive process, that is

$$X_{n+1} = \rho X_n + B_{n+1},$$

we suppose that

- B_1 has a density function f_B supported on $[-c, c]$ for some $c > 0$ with $\kappa := \inf_{x\in[-c,c]} f_B(x) > 0$
- $X_0 \in [\frac{-c}{1-|\rho|}, \frac{c}{1-|\rho|}]$

This Markov chain evolves in a compact state which is a subset of $[\frac{-c}{1-|\rho|}, \frac{c}{1-|\rho|}]$. Thanks to Corollary 5.4, $(X_n)_n$ admits a stationary measure μ. We have, moreover,

$$k(x,y) = f_{B_1}(y - \rho x) \geq \kappa,\ \forall\, x \in \mathbb{M},\ \forall\, y \in \mathbb{M}.$$

Assumptions $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$ are then satisfied with $b = 1$ and ν is the Lebesgue measure on $\mathbb{R}$.

Example: The AR(k) process. The AR(k) is defined by,

$$Y_n = \alpha_1 Y_{n-1} + \alpha_2 Y_{n-2} + \cdots + \alpha_k Y_{n-k} + \epsilon_n,$$

where $\alpha_1, \cdots, \alpha_k \in \mathbb{R}$. Since this model can be written in the form of (14) with,

$$X_n = (Y_n, Y_{n-1}, \cdots, Y_{n-k+1})^t, \quad B_n = (\epsilon_n, 0, \cdots, 0)^t, \quad A_n = \begin{pmatrix} \alpha_1 & \cdots & \alpha_k \\ I_{k-1} & & 0 \end{pmatrix}$$

all the above results, for random difference equations, apply under the corresponding assumptions. In particular the process AR(2) is stationary as soon as $|\alpha_2| < 1$ and $\alpha_2 + |\alpha_1| < 1$.

5.2. **The Möbius Markov chain on the circle.** Our purpose is to give an example of Markov chain on the unit circle, known as Möbius Markov chain, satisfying the requirements of Proposition 5.3. This Markov chain $(X_n)_{n\in\mathbb{N}}$ is introduced in [14] and is defined as follows.

- X_0 is a random variable which takes values on the unit circle.
- For $n \geq 1$,
$$X_n = \frac{X_{n-1} + \beta}{\beta X_{n-1} + 1}\epsilon_n,$$
where $\beta \in]-1, 1[$ and $(\epsilon_n)_{n\geq 1}$ is a sequence of i.i.d. random variables which are independent of X_0 and distributed as the wrapped Cauchy distribution with a common density with respect to the arc length measure ν on the unit circle $\partial B(0,1)$,
$$f_\varphi(z) = \frac{1}{2\pi}\frac{1-\varphi^2}{|z-\varphi|^2}, \ \varphi \in [0, 1[, \ \ z \in \partial B(0,1).$$

The following proposition holds.

Proposition 5.5. *Let $(X_n)_{n\geq 0}$ be the Möbius Markov chain on the unit circle as defined above. Then this Markov chain admits a unique invariant distribution, denoted by μ. If X_0 is distributed as μ then the set $\mathbb{X}_n = \{X_1, \cdots, X_n\}$ converges in probability, as n tends to infinity, in the Hausdorff distance to the unit circle $\partial B(0,1)$, more precisely, for any $\alpha \in]0, 1[$ and any $n \geq \frac{2}{\kappa v \epsilon}\left(\ln(\frac{1}{\alpha}) + \ln(\frac{4}{\epsilon\kappa v})\right)$*

$$d_H(\mathbb{X}_n, \partial B(0,1)) \leq \epsilon,$$

with probability at least $1-\alpha$. Here v is a finite positive constant and $\kappa = \frac{1}{2\pi}\frac{1-\varphi}{1+\varphi}$.

Proof. We have to prove that all the requirements of Proposition 5.3 are satisfied. Our main reference for this proof is [14]. It is proved there that this Markov chain has a unique invariant measure μ on the unit circle. Assumption $(\mathcal{A}_1)$ is then satisfied with $\mathbb{M} = \partial B(0,1)$. The task now is to check Assumption $(\mathcal{A}_2)$. We have also, for $x \in \partial B(0,1)$,

$$K(x, \nu(dz)) = \mathbb{P}(X_1 \in \nu(dz) | X_0 = x) = k(x,z)\nu(dz), \tag{16}$$

where ν is the arc length measure on the unit circle and for $x, z \in \partial B(0,1)$,

$$k(x,z) = \frac{1}{2\pi}\frac{1-|\phi_1(x)|^2}{|z-\phi_1(x)|^2},$$

with

$$\phi_1(x) = \frac{\varphi x + \beta\varphi}{\beta x + 1}.$$

We obtain, since $\frac{x+\beta}{\beta x+1} \in \partial B(0,1)$ whenever $x \in \partial B(0,1)$,

$$|\phi_1(x)|^2 = \varphi^2.$$

Now, for $x, z \in \partial B(0,1)$,

$$|z - \phi_1(x)| \leq |z| + |\phi_1(x)| \leq 1 + \varphi.$$

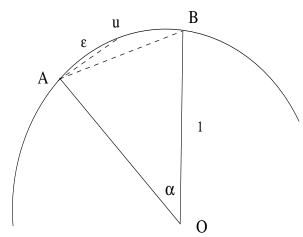
Hence,

$$k(x,z) \geq \frac{1}{2\pi}\frac{1-\varphi^2}{(1+\varphi)^2} = \frac{1}{2\pi}\frac{1-\varphi}{1+\varphi} > 0. \tag{17}$$

We have now, to check that, for some $\epsilon_0 > 0$

$$v := \inf_{u \in \partial B(0,1)} \inf_{0<\epsilon<\epsilon_0} \left(\epsilon^{-1} \int_{\partial B(0,1)\cap B(u,\epsilon)} \nu(dx_1) \right) > 0. \tag{18}$$

For this let $u \in \partial B(0,1)$, define $\overset{\frown}{AB} = \displaystyle\int_{\partial B(0,1)\cap B(u,\epsilon)} \nu(dx_1)$.



We have $|uA| = |uB| = \epsilon$. Let $\alpha = \widehat{AOB}$, then on the one hand $\overset{\frown}{AB} = \alpha$. On the other hand, since the triangle OAu is isosceles, with an angle of $\alpha/2$ in O, then $\epsilon = 2\sin(\alpha/4)$. We thus obtain,

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon}\overset{\frown}{AB} = \lim_{\epsilon\to 0} \frac{\alpha}{\epsilon} = \lim_{\alpha\to 0} \frac{\alpha}{2\sin(\alpha/4)} = 2,$$

from this (18) is satisfied.
Assumption $(\mathcal{A}_2)$ is satisfied thanks to (16), (17) and (18). The proof of Proposition 5.5 is complete by using Proposition 5.3. □

6. Simulations

The purpose of this section is to simulate a Möbius Markov process on the unit circle (as defined in Subsection 5.2) and to illustrate the results of Proposition 5.5 and of Theorem **??**. More precisely, we simulate

- X_0 is a random variable with the uniform law on the unit circle $\partial B(0,1)$, that is X_0 has the density,

$$f(z) = \frac{1}{2\pi} \ \forall\ z \in \partial B(0,1)$$

- For $n \geq 1$,

$$X_n = X_{n-1}\epsilon_n,$$

  where $(\epsilon_n)_{n\geq 1}$ is a sequence of i.i.d. random variables which are independent of X_0 and distributed as the wrapped Cauchy distribution with a common density with respect to the arc length measure ν on the unit circle $\partial B(0,1)$,

$$f_\varphi(z) = \frac{1}{2\pi}\frac{1-\varphi^2}{|z-\varphi|^2},\ \varphi \in [0,1[,\ \ z \in \partial B(0,1).$$

It is proved in [14] in this case, that this Markov chain is stationary with stationary measure the uniform law on the unit circle.

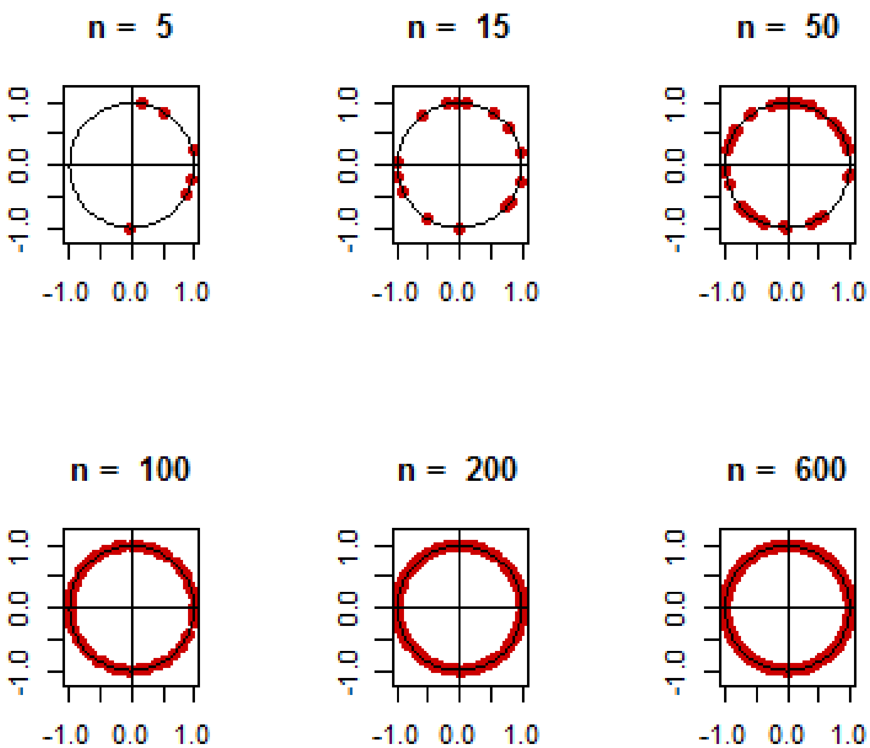


FIGURE 3. Illustrations of the set $\{x_1, \cdots, x_n\}$ which is a realisation of the stationary points $\mathbb{X}_n$ for different values of n with $\varphi = 0$.

7. Proofs for the examples

7.1. **Proof of Proposition 3.2.** Recall that, for any $1 \le i \le k_n = [n/m]$, the vectors

$$Y_{i,m} = (X_{(i-1)m+1}, \cdots, X_{im})^t$$

are dependent but, thanks to the m-dependence property, the two families $\{Y_{1,m}, Y_{3,m}, Y_{5,m}, \cdots\}$ and $\{Y_{2,m}, Y_{4,m}, Y_{6,m}, \cdots\}$ are each of i.i.d. random vectors. Hence, for any $\epsilon > 0$, (recall that $\rho_m(\epsilon) \ge \kappa_\epsilon$)

$$\sup_{x\in \mathbb{M}_{dm}} \mathbb{P}\left(\min_{1\le i\le k_n} \|Y_{i,m} - x\| > \frac{\epsilon}{2}\right) \le \sup_{x\in \mathbb{M}_{dm}} \mathbb{P}\left(\min_{1\le 2i\le k_n} \|Y_{2i,m} - x\| > \frac{\epsilon}{2}\right)$$
$$\le \sup_{x\in \mathbb{M}_{dm}} \mathbb{P}^{[k_n/2]}\left(\|Y_{1,m} - x\| > \frac{\epsilon}{2}\right) \le \left(1 - \rho_m(\frac{\epsilon}{2})\right)^{[k_n/2]} \le \left(1 - \kappa_{\frac{\epsilon}{2}}\right)^{[k_n/2]},$$

and

$$1 - \sup_{x\in \mathbb{M}_{dr}} \mathbb{P}\left(\|Y_{1,m} - x\| > \frac{\epsilon}{4}\right) \ge \kappa_{\frac{\epsilon}{4}}.$$

The bounds of Proposition 2.1 are then satisfied with $r = m$. This gives that, for any $\epsilon > 0$,

$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \le \mathbb{P}\left(d_H(\{Y_{1,m}, \cdots, Y_{k_n,m}\}, \mathbb{M}_{dm}) > \epsilon\right) \le \frac{\left(1 - \kappa_{\frac{\epsilon}{2}}\right)^{[k_n/2]}}{\kappa_{\frac{\epsilon}{4}}} \le \frac{\exp\left(-\kappa_{\frac{\epsilon}{2}}[k_n/2]\right)}{\kappa_{\frac{\epsilon}{4}}}.$$

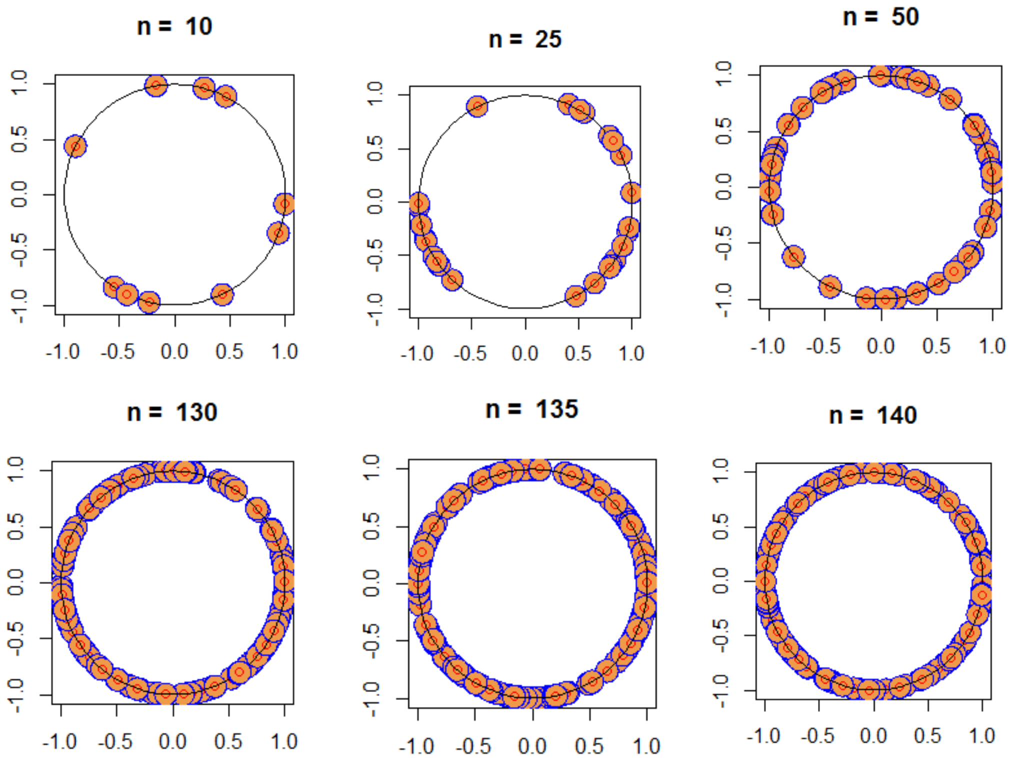


FIGURE 4. In the above graphics, the points of $\mathbb{X}_n$ are in red. Each of these points is the center of the circle with radius $r = 0.1$. This is an illustration of the reconstruction result, $\bigcup_{x\in\mathbb{X}_n} B(x,r) \simeq M$, with different values of n and with $r = 0.1$.

Let $\alpha \in]0,1[$ be such that

$$\frac{\exp\left(-\kappa_{\frac{\epsilon}{2}}[k_n/2]\right)}{\kappa_{\frac{\epsilon}{4}}} \leq \alpha,$$

which is equivalent to,

$$[k_n/2] \geq \frac{1}{\kappa_{\frac{\epsilon}{2}}} \log\left(\frac{1}{\alpha\kappa_{\frac{\epsilon}{4}}}\right).$$

Consequently, for any $n \geq \frac{2m}{\kappa_{\frac{\epsilon}{2}}} \log\left(\frac{1}{\alpha\kappa_{\frac{\epsilon}{4}}}\right) + 3m$,

$$[k_n/2] \geq k_n/2 - 1 \geq \frac{n}{2m} - 3/2 \geq \frac{1}{\kappa_{\frac{\epsilon}{2}}} \log\left(\frac{1}{\alpha\kappa_{\frac{\epsilon}{4}}}\right).$$

and then,

$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \leq \alpha.$$

The proof of Proposition 3.2 is complete. $\square$

7.2. **Proof of Proposition 3.3.** We use the blocking method of [24] to transform the dependent β-mixing sequence $(X_n)_{n\in\mathbb{N}}$ into a sequence of nearly independent blocks. Let $Z_{2i,r_n} = (\xi_j,\ j \in \{(2i-1)r_n+1,\cdots,2ir_n\})^t$ be a sequence of i.i.d. random vectors independent of the sequence $(X_i)_i$

such that, for any i, Z_{2i,r_n} is distributed as Y_{2i,r_n} (which is distributed as Y_{1,r_n}). Lemma 4.1 of [24] proves that the two vectors $(Z_{2i,r_n})_i$ and $(Y_{2i,r_n})_i$ are related thanks to the following relation,

$$|\mathbb{E}(h(Z_{2i,r_n}, 1 \le 2i \le k_n)) - \mathbb{E}(h(Y_{2i,r_n}, 1 \le 2i \le k_n))| \le k_n \beta_{r_n},$$

which is true for any measurable function bounded by 1. We then have, using the last bound,

$$\begin{aligned}
& k_n \sup_{x \in \mathbb{M}_{dr_n}} \mathbb{P}\left(\min_{1 \le i \le k_n} \|Y_{i,r_n} - x\| > \epsilon\right) \le k_n \sup_{x \in \mathbb{M}_{dr_n}} \mathbb{P}\left(\min_{1 \le 2i \le k_n} \|Y_{2i,r_n} - x\| > \epsilon\right) \\
& \le k_n \sup_{x \in \mathbb{M}_{dr_n}} \left|\mathbb{P}\left(\min_{1 \le 2i \le k_n} \|Y_{2i,r_n} - x\| > \epsilon\right) - \mathbb{P}\left(\min_{1 \le 2i \le k_n} \|Z_{2i,r_n} - x\| > \epsilon\right)\right| \\
& + k_n \sup_{x \in \mathbb{M}_{dr_n}} \mathbb{P}\left(\min_{1 \le 2i \le k_n} \|Z_{2i,r_n} - x\| > \epsilon\right) \\
& \le k_n^2 \beta_{r_n} + k_n \sup_{x \in \mathbb{M}_{dr_n}} \mathbb{P}\left(\min_{1 \le 2i \le k_n} \|Z_{2i,r_n} - x\| > \epsilon\right) \\
& \le k_n^2 \beta_{r_n} + k_n \sup_{x \in \mathbb{M}_{dr_n}} \mathbb{P}^{[k_n/2]}\left(\|Y_{1,r_n} - x\| > \epsilon\right) \\
& \le k_n^2 \beta_{r_n} + k_n (1 - \rho_{r_n}(\epsilon))^{[k_n/2]} \\
& \le k_n^2 \beta_{r_n} + k_n \exp\left(-[\frac{k_n}{2}] \rho_{r_n}(\epsilon)\right).
\end{aligned}$$

Consequently,

$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \le \frac{k_n^2 \beta_{r_n} + k_n \exp\left(-[\frac{k_n}{2}] \rho_{r_n}(\epsilon/2)\right)}{k_n \rho_{r_n}(\epsilon/4)}. \tag{19}$$

We have now to construct two sequences k_n and r_n such that $k_n r_n \le n$ and that

$$\lim_{n\to\infty} k_n^2 \beta_{r_n} = 0, \ \lim_{n\to\infty} k_n \rho_{r_n}(\epsilon) = \infty, \ \lim_{n\to\infty} k_n \exp\left(-\frac{k_n}{2} \rho_{r_n}(\epsilon)\right) = 0. \tag{20}$$

We have supposed that $\lim_{m\to\infty} \rho_m(\epsilon) \frac{e^{m^\beta}}{m^{1+\beta}} = \infty$ for some $\beta > 1$. Define $\alpha = 1/\beta \in]0,1[$ and

$$k_n = [\frac{n}{(\ln n)^\alpha}], \ \ r_n = [(\ln n)^\alpha].$$

We have then, (letting $m = r_n = [(\ln n)^\alpha]$), $\lim_{n\to\infty} k_n \frac{\rho_{r_n}(\epsilon)}{\ln n} = \infty$ and then (since $k_n \le n$),

$$\lim_{n\to\infty} k_n \frac{\rho_{r_n}(\epsilon)}{\ln(k_n)} = \infty.$$

The last limit gives that $\lim_{n\to\infty} k_n \rho_{r_n}(\epsilon) = \infty$ and for n large enough and for some $C > 2$, $k_n \frac{\rho_{r_n}(\epsilon/2)}{\ln(k_n)} \ge C$, so that,

$$k_n \exp\left(-\frac{k_n}{2} \rho_{r_n}(\epsilon)\right) \le k_n^{1-C/2}.$$

Consequently, $\lim_{n\to\infty} k_n \exp\left(-\frac{k_n}{2} \rho_{r_n}(\epsilon)\right) = 0$. Now, we deduce from $\lim_{m\to\infty} \frac{e^{2m^\beta}}{m^2} \beta_m = 0$ that (letting $m = r_n = [(\ln n)^\alpha]$)

$$\lim_{n\to\infty} k_n^2 \beta_{r_n} = 0.$$

The two sequences k_n and r_n, so constructed, satisfy (20) and then it holds for those sequences

$$\lim_{n\to\infty} \frac{k_n^2 \beta_{r_n} + k_n \exp\left(-\frac{k_n}{2} \rho_{r_n}(\epsilon/2)\right)}{k_n \rho_{r_n}(\epsilon/4)} = 0,$$

hence for any $\alpha \in]0,1[$, there exists an integer n_0 such that for any $n \geq n_0$,

$$\frac{k_n^2 \beta_{r_n} + k_n \exp\left(-\frac{k_n}{2}\rho_{r_n}(\epsilon/2)\right)}{k_n \rho_{r_n}(\epsilon/4)} \leq \alpha.$$

The proof of Proposition 3.3 is complete, combining the last bound together with (19). □

7.3. **Proof of Proposition 3.4.** We have,

$$\begin{aligned} k_n \mathbb{P}\left(\min_{1\leq i\leq k_n} \|Y_{i,r_n} - x\| > \epsilon\right) &\leq k_n \mathbb{P}\left(\min_{1\leq 2i\leq k_n} \|Y_{2i,r_n} - x\| > \epsilon\right) \\ &\leq k_n \left| \mathbb{P}\left(\min_{1\leq 2i\leq k_n} \|Y_{2i,r_n} - x\| > \epsilon\right) - \prod_{i:\,1\leq 2i\leq k_n} \mathbb{P}\left(\|Y_{2i,r_n} - x\| > \epsilon\right)\right| \\ &+ k_n \prod_{i:\,1\leq 2i\leq k_n} \mathbb{P}\left(\|Y_{2i,r_n} - x\| > \epsilon\right). \end{aligned} \tag{21}$$

We have, for s events $A_1, \cdots, A_s$, (with the convention that, $\prod_{j=1}^{0} \mathbb{P}(A_j) = 1$)

$$\mathbb{P}(A_1 \cap \cdots \cap A_s) - \prod_{i=1}^{s} \mathbb{P}(A_i) = \sum_{i=1}^{s-1} \mathbb{P}(A_1)\cdots\mathbb{P}(A_{i-1})\mathrm{Cov}(\mathbb{1}_{A_i}, \mathbb{1}_{A_{i+1}\cap\cdots\cap A_s}).$$

Hence,

$$\left|\mathbb{P}(A_1 \cap \cdots \cap A_s) - \prod_{i=1}^{s} \mathbb{P}(A_i)\right| \leq \sum_{i=1}^{s-1} \left|\mathrm{Cov}(\mathbb{1}_{A_i}, \mathbb{1}_{A_{i+1}\cap\cdots\cap A_s})\right|.$$

We apply the last bound with $A_i = (\|Y_{2i,r_n} - x\| > \epsilon)$ and we use (12), we get

$$\left|\mathrm{Cov}(\mathbb{1}_{A_i}, \mathbb{1}_{A_{i+1}\cap\cdots\cap A_s})\right| \leq \Psi(r_n),$$

and

$$\left| \mathbb{P}\left(\min_{1\leq 2i\leq k_n} \|Y_{2i,r_n} - x\| > \epsilon\right) - \prod_{i:\,1\leq 2i\leq k_n} \mathbb{P}\left(\|Y_{2i,r_n} - x\| > \epsilon\right)\right| \leq k_n \Psi(r_n). \tag{22}$$

We deduce, combining (21) and (22),

$$\begin{aligned} k_n \mathbb{P}\left(\min_{1\leq i\leq k_n} \|Y_{i,r_n} - x\| > \epsilon\right) &\leq k_n^2 \Psi(r_n) + k_n(1 - \rho_{r_n}(\epsilon))^{[k_n/2]} \\ &\leq k_n^2 \Psi(r_n) + k_n \exp(-[k_n/2]\rho_{r_n}(\epsilon)). \end{aligned}$$

Consequently,

$$\mathbb{P}\left(d_H(\mathbb{X}_n, \mathbb{M}) > \epsilon\right) \leq \frac{k_n^2 \Psi(r_n) + k_n \exp\left(-[\frac{k_n}{2}]\rho_{r_n}(\epsilon/2)\right)}{k_n \rho_{r_n}(\epsilon/4)}.$$

We have now to construct two sequences r_n and k_n such that

$$\lim_{n\to\infty} k_n \exp(-k_n \rho_{r_n}(\epsilon)/2) = 0, \quad \lim_{n\to\infty} k_n^2 \Psi(r_n) = 0, \quad \lim_{n\to\infty} k_n \rho_{r_n}(\epsilon) = \infty.$$

The last construction is possible (we argue as the end of the proof of Proposition 3.3). □

Acknowledgement: The first author is supported by an SFRG-grant from the American University of Sharjah (AUS, UAE). He also wishes to thank the LGK (laboratoire Jean Kuntzmann, Université de Grenoble) for its invitation to work on the project.

[1] D. Attali, A. Lieutier, and D. Salinas, *Vietoris-Rips complexes also provide topologically correct reconstructions of sampled shapes*, Computational Geometry: Theory and Applications, 46(4) 2013, 448–465.
[2] R. C. Bradley, *Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions*, Probab. Surveys, **2** (2005), 107–144.
[3] R. C. Bradley, *Absolute regularity and functions of Markov chains. Stochastic Process*, Appl. 14-1 (1983), 67–77.
[4] F. Chazal, M. Glisse, C. Labruère, M. Michel, *Optimal rates of convergence for persistence diagrams in Topological Data Analysis*, Journal of Machine Learning Research **16** (2015), 3603–3635.
[5] F. Chazal and S. Oudot, *Towards persistence-based reconstruction in Euclidean spaces*, Proc. 24th Ann.Sympos. Comput. Geom. (2008), 232–241.
[6] P. Doukhan, S. Louhichi, *A new weak dependence condition and applications to moment inequalities*, Stochastic Process. Appl., **84** no.2 (1999), 313–342.
[7] J.C. Ellis, *On the Geometry of Sets of Positive Reach*, thesis, University of Georgia (2012).
[8] B.T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh, *Confidence sets for persistence diagrams*, Ann. Stat., 42-6 (2014), 2301–2339.
[9] B.T. Fasy, R. Komendarzyk, S. Majhy, C. Wenk, *Topological and geometric reconstruction of metric graphs in* $\mathbb{R}^d$, https://arxiv.org/abs/1912.03134
[10] H. Federer, *Curvature measures*, Trans. Amer. Math. Soc. **93** (1959), 418–491.
[11] C.M. Goldie, R.A Maller, *Stability of perpetuities*, Ann. Probab. 28-3 (2000), 1195-1218.
[12] A. Hatcher, *Algebraic Topology*, Oxford University Press.
[13] J. Latschev, *Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold*, Arch. Math. **77** (2001) 522–528.
[14] S. Kato, A Markov, *Process for Circular Data*, J. R. Statist. Soc. B. 72-5 (2010), 655–672.
[15] H. Kesten, *Renewal Theory for Functionals of a Markov Chain with General State Space*, Ann. Probab. 2-3 (1974), 355–386.
[16] H. Kesten, *Random difference equations and Renewal theory for products of random matrices*, Acta Math. **131** (1973), 207–248.
[17] J. Kim, J. Shin, F. Chazal, A. Rinaldo, L. Wasserman, *Homotopy reconstruction via the Cech complex and the Vietoris-Rips complex*,
[18] P. Niyogi, S. Smale, S. Weinberger, *Finding the homology of submanifolds with high confidence from random samples*, Discrete Comput. Geom. **39** (2008), 419–441.
[19] J. Kim, J. Shin, A. Rinalo, L. Wassermann, *Nerve theorem on a positive reach set*,
[20] R. Moreno, S. Koppal, E. de Muinck, *Robust estimation of distance between sets of points*, Pattern Recognition Letters **34**, Issue 16 (2013), 2192–2198.
[21] M. Rosenblatt, *A central limit theorem and a strong mixing condition*, Proc. Natl. Acad. Sci. USA **42**, (1956), 43-47.
[22] E. Rio, *Inequalities and limit theorems for weakly dependent sequences* 3rd cycle. pp.170 (2013). cel-00867106v1.
[23] Y. Wang, B. Wang *Topological inference of manifolds with boundary*, Computational Geometry 88 (2020) 101606.
[24] Yu, B. (1994). Rates of Convergence for Empirical Processes of Stationary Mixing Sequences. Ann. Probab. 22, 94-116.

Sadok Kallel: American University of Sharjah, UAE
Email address: `skallel@aus.edu`

Sana Louhichi: Laboratoire Jean Kuntzmann (CNRS 5224), Université Grenoble Alpes, 700 Avenue Centrale, 38401 Saint-Martin-d'Hères, France.
Email address: `sana.louhichi@univ-grenoble-alpes.fr`